# nQuery Advanced

## MANUAL

Statsols

# nQuery Advisor® Sample Size Calculator

**v. 9.1.1.0 User Manual**

Statsols

**One International Place,**
**100 Oliver Street, Suite 1400,**
**Boston MA, 02110, U.S.A.**

**Statistical Solutions Ltd.**
**Building 4500, Airport Business Park,**
**Cork, Ireland,**

**Registered in Ireland Reg No. 233638**

**US: +1 (617) 830 1836 EU/ROW: +353 21 48 39 100**
Web: http://www.statsols.com Email: sales@statsols.com

©2022 Statistical Solutions Ltd.

Printed in the Republic of Ireland. No part of this manual may be reproduced, stored in a retrieval system, transmitted, translated into any other language, or distributed in any form by any means without prior permission of Statistical Solutions Ltd.

# Acknowledgements

We would like to sincerely thank all those who made the production of nQuery Advisor from Statsols possible. The Statsols R&D Team: Christine Carey, David Lemoine, Steven Keady, Cathal O'Gorman, Anita Coffey, Sean Tynan, Ronan Fitzpatrick, Brian Fox, Paul Murphy, Brian Ronayne, Denis Desmond and the Statsols Team: Brendan Nyhan, David Cleary, Liam Burke and Marjo Curran.

Special Thanks to: Mary Byrne, Kevin Sievewright, Diana Scriven, Caroline Costello, Helen Murphy, Eimear Keyes, Adrian O'Leary, Aodhan O'Leary, Dr. Janet Elashoff, Kieran Logan, Tricia Baalfe, Andrew Grannell, Brian Sullivan, Eoghan Murphy, Laura O'Donovan, Luke Rothwell, Sorcha Conaghan, Aisling Leen, Paul Kenneally, Louise Byrne, Layla Tobin, Garry McCarthy, Dan Andrei Chirac, Andrei Banc, Claudiu Balanean and Cristina Pop from Lasting Software, and Dr. Niall Fitzgerald.

IMPORTANT: READ BEFORE INSTALLING SOFTWARE

**nQuery Advisor License Agreement**

This document sets forth the terms and conditions of the license and the limited warranty for nQuery Advisor, hereinafter referred to as "nQuery". Proceeding with this installation constitutes your acceptance of this license agreement with respect to all accompanying nQuery software received by you. If you do not accept this agreement, you may cease the installation within 10 days of receipt of installation key and your money will be refunded.

THIS LICENSE AGREEMENT is made and entered into by and between Statistical Solutions Limited, an Irish registered company, hereinafter referred to as "LICENSOR", and you, hereinafter referred to as "LICENSEE".

WHEREAS, LICENSOR has ownership rights of nQuery ("SOFTWARE") and LICENSEE desires to obtain from LICENSOR, and LICENSOR desires to grant to LICENSEE, a license to use the aforementioned nQuery.

NOW, THEREFORE, in consideration of the mutual covenants, conditions and terms hereinafter set forth, and for other good and valuable consideration, the parties hereto hereby agree as follows:

I. GRANT

A. Subject to the provisions contained herein, LICENSOR grants to LICENSEE a non-exclusive, non-transferable license for the use of SOFTWARE ("LICENSE") which shall not and will not be copied by the LICENSEE and will not be passed by the LICENSEE to any other person or party, or will the whole or any part of it be disclosed by the LICENSEE to any other or party.

B. Unless otherwise specifically agreed in writing by the LICENSOR, SOFTWARE may be used solely by LICENSEE.

C. No title or ownership rights to SOFTWARE are transferred to LICENSEE by this Agreement.

II. MAINTENANCE

A. LICENSOR agrees to provide maintenance for SOFTWARE for the duration of this Agreement.

B. MAINTENANCE includes the following services and materials.

1. Updates and improvements to SOFTWARE, as these become available.
2. Reasonable written or telephone consultations.

C. Maintenance will be provided only for the release of SOFTWARE which is current at the time maintenance is required.

III. REGISTERED USE

A. LICENSEE agrees that it will maintain SOFTWARE (including any future forms of SOFTWARE provided to LICENSEE) in confidence.

B. LICENSEE agrees not to knowingly provide, distribute or otherwise make available, in whole or in part to those outside of LICENSEE'S site, SOFTWARE, or any derivative thereof, including without limitation program listings and code.

IV. ACCEPTANCE AND PAYMENTS

A. All terms and conditions of this Agreement shall be binding upon purchase of LICENSE by LICENSEE.

B. LICENSEE, for and in consideration and as a condition of purchasing this LICENSE, hereby agrees to pay over to LICENSOR an annual fee. The yearly fee in effect will become due unless LICENSEE has properly exercised its right to terminate this Agreement pursuant to paragraph V.B. below.

V. TERM

A. This Agreement shall become effective upon purchase of LICENSE by LICENSEE and shall remain in effect for an initial period of one (1) year, and thereafter shall be automatically renewed for successive periods of one (1) year, unless terminated as provided in this Agreement.

B. LICENSEE or LICENSOR shall have the right to terminate this Agreement, at the end of any licensing period upon written notice, which shall be given to the other party at least sixty (60) days prior to the expiration of the yearly licensing period. In the event that notice of termination is not given by either party at least 60 days prior to the expiration of the license period the Agreement shall be deemed effective for another one year.

C. Either party shall have the right to terminate this Agreement at any time upon violation by the other party of any of the provisions of this Agreement.

D. Upon termination, of this Agreement, LICENSEE shall destroy all installations of SOFTWARE, LICENSEE shall promptly forward written notice to LICENSOR that all programs and materials containing SOFTWARE have been destroyed and are no longer in use or usable by LICENSEE.

VI. WARRANTIES

A. LICENSOR warrants that it has the right to grant this License.

B. LIMITED WARRANTY AND REMEDY. Both parties hereto agreeing that this is not a consumer contract, LICENSOR warrants that for sixty (60) days after purchase of LICENSE by the LICENSEE, SOFTWARE shall reasonably perform in accordance with the accompanying documentation (manuals and computer aided instructions). LICENSOR specifically DOES NOT warrant that SOFTWARE will operate uninterrupted or error free.

LICENSEE'S SOLE AND EXCLUSIVE REMEDY for a failure of SOFTWARE to perform in accordance with the accompanying documentation is for the LICENSEE to return SOFTWARE to LICENSOR in writing of the non-performance within sixty (60) days of purchase of the media. LICENSOR'S sole obligation within a reasonable time after receiving notification of non-performance shall be, at LICENSOR'S option to:

1. correct demonstrable and significant program errors.
2. provide the LICENSEE with functionally equivalent software, or
3. provide or authorise a refund.

THE LICENSOR'S LIMITED WARRANTY SET FORTH ABOVE IS IN LIEU OF ANY OTHER WARRANTIES. LICENSOR DISCLAIMS ANY IMPLIED WARRANTY INCLUDING, BUT NOT LIMITED TO ANY IMPLIED WARRANTY OF FITNESS FOR A PARTICULAR PURPOSE OR MERCHANTABILITY. LICENSOR DISCLAIMS ALL OTHER WARRANTIES INCLUDING, BUT NOT LIMITED TO, THOSE MADE BY DISTRIBUTORS OF LICENSOR'S SOFTWARE, BUT DOES ATTEMPT TO DISCLAIM OR EXCLUDE LIABILITY FOR PERSONAL INJURIES WHERE THE LICENSOR HAS AN OBLIGATION FOR SAME UNDER LAW.

A. In no event shall LICENSOR become liable to LICENSEE, or any other party, for any loss or damages, consequential or otherwise, including but not limited to time, money or goodwill arising from the use, operation or modification of SOFTWARE by LICENSEE. LICENSEE further agrees to indemnify, defend and hold harmless, its agents or employees from and LICENSOR against all loss or expense resulting from liability of any nature arising or resulting from the acts or omissions of LICENSEE, its employees or agents, in its use, possession or operation of SOFTWARE.

B. If any of the provisions, or portion thereof, of this Agreement is invalid under any applicable statute of law, they are to that extent to be deemed omitted. This Agreement shall constitute the entire agreement between the parties hereto with respect to the subject matter hereof, and shall supersede all oral, promotional or advertising presentations. This Agreement may not be modified except in writing signed by an authorised representative of LICENSOR. The prevailing party in the legal action brought

to construe the terms of or enforce this Agreement shall be entitled to reasonable legal fees and costs. Under no circumstances shall this Agreement or any of the rights granted to LICENSEE hereunder be sold, assigned or sub-licensed, voluntarily or by operation of law, to any other person or entity, and any such purported sale, assigned or sublicenses shall be void.

VII. This agreement shall be governed by the laws of Ireland and the parties agree to the exclusive jurisdiction of the Irish Court.

VIII. DATA PROTECTION

Each of the parties shall observe all of their respective obligations under all legislation and regulations relating to the protection of personal data (including Regulation (EU) 2016/679) that arise in connection with this Agreement. The parties acknowledge that LICENSEE will not be disclosing or transferring personal data to LICENSOR and that LICENSOR will not be processing personal data on behalf of LICENSEE.

# Contents


**1 nQuery Basics** **1**
1.1 System Requirements . . . 1
1.2 Help and Support . . . 1
1.3 Starting nQuery . . . 1
1.4 nQuery Start-up and Layout . . . 2
1.5 Opening a Design Table . . . 9
1.6 Design Table Layout . . . 13
1.7 Using nQuery Design Tables . . . 15
1.8 Design Table Tools . . . 23
1.9 Editing Plots . . . 32

**2 Advanced nQuery Table Features** **34**
2.1 Manual Solver Activation . . . 34
2.2 Errors and Warnings . . . 38
2.3 Options Menu . . . 41
2.4 Print and Report Options . . . 47
2.5 Other Topics . . . 49

**3 Advanced nQuery Assistants** **54**
3.1 Design Table Side-Tables . . . 54
3.2 Compute Standard Deviation Assistants . . . 62
3.3 Data Entry Assistant . . . 63
3.4 Cumulative Distribution Function Assistants . . . 65
3.5 Survival Parameter Conversion Assistant . . . 66
3.6 Posterior Error Rate Calculator Assistant . . . 67

**4 Group Sequential Designs, Interim Monitoring & Unblinded Sample Size Re-estimation** **69**
4.1 Introduction . . . 69
4.2 Group Sequential Test Design . . . 69
4.3 Interim Monitoring and Unblinded Sample Size Re-estimation . . . 80

**5 Multiple Comparisons Procedure - Modelling (MCP-Mod)** **102**
5.1 Multiple Comparison Procedures and Modelling (MCP-Mod) Design . 102
5.2 MCP-Mod Demonstration . . . 108

**6 nQuery Predict** **117**
6.1 Trial Milestone Prediction . . . . . 117
6.2 nQuery Predict Demonstration . . . . . 128

**7 nQuery Qualification Tools** **366**
7.1 Installation Qualification (IQ) Tool . . . . . 366
7.2 Operational Qualification (OQ) Tool . . . . . 367

**8 nQuery Updates and Licensing** **372**
8.1 Renewing a License . . . . . 372
8.2 Activating an Add-on Module . . . . . 373
8.3 Checking for Software Updates . . . . . 374

**Bibliography** **375**

# 1 nQuery Basics

This chapter is a quick guide for users to familiarize themselves with how to access nQuery Advisor (referred to as nQuery from hereon) and conduct a basic sample size analysis. Many individual elements of using nQuery are described in more detail in subsequent chapters.

Note that installation is not covered in this tutorial as installation guides will have been provided to you with the application and are also available online in your nQuery account.

## 1.1 System Requirements

Minimum system requirements for nQuery 9.1.1.0

- Processor: 2 core 1 GHz processor
- Memory: 3 GB of available hard disk space
- RAM: 2GB
- OS: Windows 7 Service Pack 1 (Home Premium) or higher
- Other: .NET Framework 4.6.2

## 1.2 Help and Support

The easiest way to get statistical or technical support is to contact Statsols using their nQuery Community User account (see subsection 1.4.1) or to use the Help Centre button found in the bottom right of the nQuery application Home tab (see subsection 1.4.5). The support centre can also be found at the following URL:

- info.statsols.com/help-center

## 1.3 Starting nQuery

There are two main ways to open nQuery. Firstly, nQuery can be opened by double-clicking on the desktop icon (if selected during installation) and then nQuery will be

automatically launched. Secondly, it can be found in the Windows Start menu by clicking on the Windows icon in the bottom-left. In each OS, the easiest way to find the application is to enter “nQuery” into the search dialog and select the “nQuery” application when it appears in the results. Alternatively, to find the application manually, use the following guide for your respective OS:

Windows 11: Click the Start menu in the lower-left. Click the “All apps” button on the upper-right. Scroll down to the “N” section in the alphabetised list of apps. Click on the “nQuery” application in that section

Windows 10: Click the Start menu in the bottom-left. In the “All Apps” menu on the left-hand side of the Start menu, go to the “N” section in the alphabetised list, select the “nQuery” folder and select the “nQuery” application.

Windows 8/8.1: Click the Start menu in the bottom-left. Select the “All Apps” menu using the arrow at the bottom of the screen. Scroll to the nQuery application or enter “nQuery” in the search window in the top-right and select the “nQuery” application.

Windows 7: Click the Start menu in the bottom-left. Then select “All Programs” and you can locate nQuery in the alphabetised list. Click on this folder and select the “nQuery” application.

## 1.4 nQuery Start-up and Layout

Once the user has launched nQuery, the application will appear as illustrated in Figure 1.1. The five major elements in the application on start-up are the Home tab (1), the toolbar (2), the menu bar (3), the tab menu (4) and the information bar (5). These are summarised in the following section.

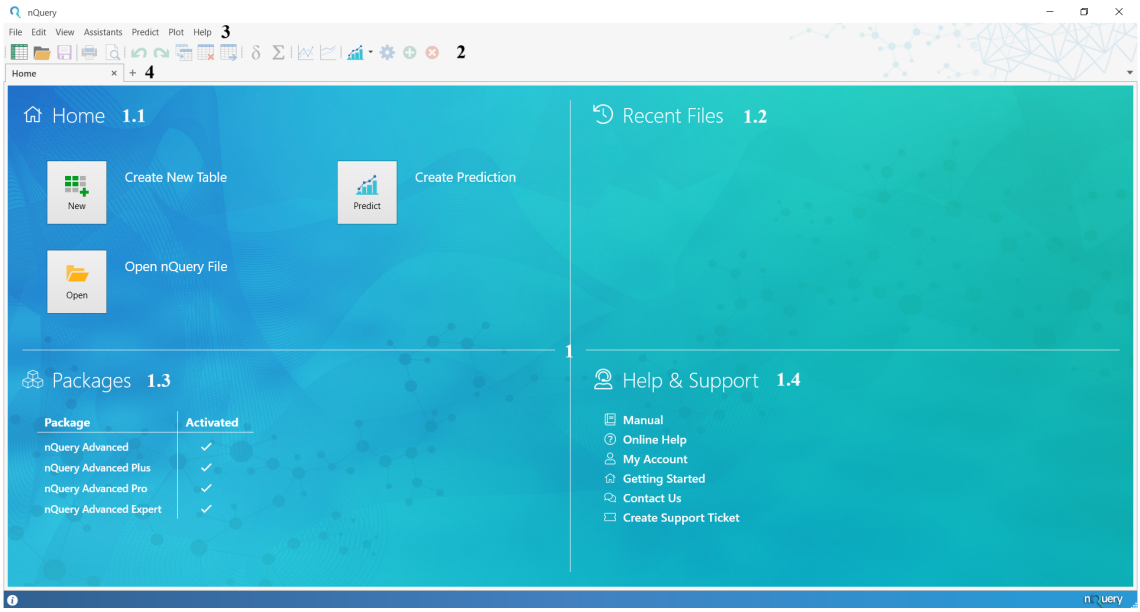



**Figure 1.1:** nQuery on Start-up

## 1.4.1 Home Tab (1)

When you open nQuery, the software will open the "Home" tab. This window contains options to open new tables or pre-existing nQuery files, open recent nQuery files, see which nQuery modules are currently active and support options if you need further help in using the software. The four quadrants (1.1 to 1.4 from top-left to bottom-right) in the Home tab are described in more detail below.

### 1.4.1.1 Home (1.1)

The Home quadrant (top-left) contains two buttons: "Create a New Table" and "Open an Existing Table".

Selecting the "Create a New Table" button will open the "Select Test" window where you can open a new table. This window is described in section 1.5

Selecting the "Open nQuery File" button will open the "Open" window where you can open any pre-existing nQuery (.nqt, .nqp, .nqd) file. This is a standard Windows "Open File" dialog.

Selecting the "Create Prediction" button will open the a new nQuery Predict workspace. nQuery Predict is described in chapter 6. This option will not be shown unless a nQuery Advanced Expert license is active. Active packages are displayed in the Packages section. To purchase additional packages, see `www.statsols.com`.

#### 1.4.1.2 Recent Files (1.2)

The Recent File quadrant (top-right) will contain a list of all save files recently opened in nQuery. To open a file, select the required file from this list and the save file will open automatically.

#### 1.4.1.3 Packages (1.3)

The Packages quadrant (bottom-left) will contain a list of the nQuery packages which are currently active under your current nQuery license. Packages which are active will have a tick to the right of the package name. For details on nQuery packages and activating them, see section 8.2. To purchase additional packages, see `www.statsols.com`.

The current packages in nQuery Advanced (largest areas covered in brackets) are:

- nQuery Advanced (Fixed Term Trial Sample Size)
- nQuery Advanced Plus (Bayesian Sample Size)
- nQuery Advanced Pro (Adaptive Design)
- nQuery Advanced Expert (Milestone Prediction)

#### 1.4.1.4 Help & Support (1.4)

The Help & Support quadrant (bottom-right) contains a range of help, support and troubleshooting options for nQuery. These options are as follows:

1. Manual: This will open the User Manual (this document)
2. Online Help: This will open the nQuery Knowledge Base website. This website provides a comprehensive set of instructions for installation, using the software and common troubleshooting issues and solutions.
3. My Account: This will open the nQuery Account Administrator Portal. Your nQuery account website will contain access to the relevant installers, activation keys and installer guides.
4. Getting Started: This will open the "How to Use nQuery" website. This website contains videos and instructions on all the basic elements of using nQuery for sample size determination.
5. Contact Us: This will open the "nQuery FAQ" website. This website will contain answers to frequently asked questions (FAQs) about nQuery and provides an "Ask a Question" facility where you can send a custom query to us online by filling out the relevant form.
6. Create a Support Ticket: This will bring you the "Log a Support Ticket" website. Select the "Log a Ticket" option to start this process. Note an nQuery account is required to use this facility.

### 1.4.2 Toolbar (2)

The toolbar provides easy access to the most commonly used functions in nQuery. The toolbar options can be split into five main categories: file options, print options, table options, side-table options, plotting options and nQuery Predict options. Note that the nQuery Predict options (15 - 18) will be described further in chapter 6 but will not be displayed unless an nQuery Advanced Expert license is active. Active packages are shown in the Packages section of the Home tab. The toolbar is shown in detail in Figure 1.2.

**Figure 1.2:** Toolbar

There are fourteen options in the menu bar. These are as follows:

1. New Table (New): Open the Select Table screen
2. Open Existing Table (Open): Open nQuery table save files
3. Save Table (Save): Save the currently open nQuery table
4. Print: Print the currently open nQuery table
5. Print Preview: Open Print Preview of currently open nQuery table
6. Undo Action (Undo): Undo last action within currently open nQuery table
7. Redo Action (Redo): Redo last undone action with currently open nQuery table
8. Copy Table: Copy all contents of currently open nQuery table
9. Clear Table: Clear all entries from currently open nQuery table
10. Fill Right: Fill all entries to the right with the same selected cell value
11. Compute Effect Size: Open the Effect Size side-table (if available)
12. Compute Covariance Matrix: Open the Covariance Matrix side-table (if available)
13. Plot Power vs Sample Size: Plot Power vs Sample plot for selected column(s)
14. Plot User Selected Rows: Open custom plot dialog for selected column(s)
15. Create New Prediction Workspace (New Prediction): Open new nQuery Predict workspace. When selected a drop down two options will be displayed:
    a) Prediction with Subject-level Data: Open “New Prediction - Import Datasets” dialog for uploading interim data

b) Prediction with Fixed Parameters: Open new nQuery Predict Workspace.

16. Specify the Simulation Parameters (Configure): Return to the Setup stage in the current nQuery Predict workspace
17. Add a New Dataset to the Workspace (Add Dataset): Open Import Datasets dialog to add new dataset to current nQuery Predict workspace
18. Remove the Selected Dataset from the Workspace (Remove Dataset): Remove the selected Dataset in Data section from current nQuery Predict workspace.

These options will be explored in further detail later in this manual.

### 1.4.3 Menu Bar

There are six options on the menu bar: File, Edit, View, Assistants, Plot and Help. These are highlighted in Figure 1.3.

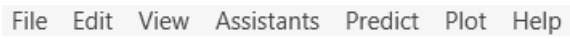



**Figure 1.3:** Menu Bar

The **File** menu allows the user to open a new or previously saved design table, as well as enabling the user to save a design and allowing the user to exit nQuery whenever they wish. Design tables can be saved as .nqt format, which is the Statsols file format for nQuery. It also includes the options for printing, the option menu and to close nQuery.

The **Edit** menu enables the user to undo and redo actions within the current nQuery table (shortcuts are provided in text). The user has options to copy and clear a table or fill a table using the Fill Right option. The fill right option is where the user, when defining multiple columns, enters certain information into a column and can copy this information across the remaining empty columns.

The **View** menu options are initially unavailable (greyed out) until the user opens a design table except for the "Home" option. The "Home" option will open or close the initial "Home" tab in nQuery. Once a table has been opened, options appear enabling the user to open or close the table elements of the output statement, the specify multiple factors tool, the help panel and the notes tool. These table elements are described in section 1.8.

The **Assistants** menu gives access to several useful tools that can help with a user's calculation. These include access to the "Calculate Effect Size" and "Specify Covariance Matrix" side-tables, access to utilities for deriving the standard deviation, access to cumulative distribution functions for multiple statistical distributions, a data entry tool, a survival parameter converter tool, a Bayesian Posterior Error calculator, the Report feature and a shortcut to the windows calculator.

The **Predict** menu gives access to options to create a new nQuery Predict workspace and add or remove datasets in the current nQuery Predict workspace. "New Prediction" will open a menu with two options: "Prediction with Subject-level Data" and "Prediction with Fixed Parameters". "Prediction with Subject-level Data" opens the "New Prediction - Import Datasets" dialog for uploading interim data. "Prediction with Fixed Parameters" opens a new blank nQuery Predict Workspace. "Add Dataset" opens the Import Datasets dialog to upload additional datasets. "Remove Dataset" removes a dataset selected from the Data field in the current nQuery Predict workspace. Note that the Predict menu will not be displayed unless an nQuery Advanced Expert license is active

The **Plot** menu is initially unavailable until the user opens a design table. Once a table has been opened and filled appropriately, the user can use this menu to create Power vs. Sample Size and Plot User Selected rows plots. There are also table specific options for Boundary and Inverse Boundary plots (group sequential design tables e.g. Two Means GST - MTT12) and the Survival vs Time plot (simulation survival tables - STT3/STT3u).

The **Help** menu gives access to the nQuery user manual (this document), the "About" page containing the nQuery version information and license agreement, the Installation (IQ) and Operational Qualification (OQ) tools and utilities for activating/renewing a license, enabling add-on modules and for checking for the latest nQuery updates.

Below is a complete list of menu options:

**File:** New, Open, Save, Save As, Print, Print Preview, Recent (horizontal menu), Options, Close

**Edit:** Undo, Redo, Fill Right, Copy Table, Clear Table

**View:** Home, Specify Multiple Factors, Output, Help, Notes

**Assistants:** Compute Effect Size, Compute Covariance Matrix, Standard Deviation, Data Entry, Distribution Functions, Survival Parameter Converter, Posterior Error Rate Calculator, Report, Windows Calculator

**Predict:** New Prediction (Prediction with Subject-level Data, Prediction with Fixed Parameters), Add Dataset, Remove Dataset - nQuery Advanced Expert Only

**Plot:** Power vs. Sample Size, User-Selected Rows, Survival vs Time plot (STT3/STT3u only), Multiple Boundary (Group Sequential Design only)

**Help:** nQuery Manual, Activate/Renew License, Enable Modules, Check for Updates, Installation Qualification, Operational Qualification, About

### 1.4.4 Tab Menu

The tab menu provides an easy way to navigate between different tables within nQuery and quickly create fresh copies of open tables. It is shown in Figure 1.4 (note

that we have opened some additional tables in nQuery for illustration purposes)

**Figure 1.4:** Tab Menu

There are three main elements to the tab menu: the tabs, the fresh table shortcut and the table navigation drop-down.

1. Tabs: Individual tabs for each table opened in the session and the "Home" tab. The tab also has an "x" button on its right-hand side to close an individual tab. The current open tab will be highlighted compared to the other tabs.
2. Fresh Tab: If the Home tab is selected or no tabs are available, will open the "Select Test" menu. If a nQuery table is selected, it will open a clean version of that table in a new tab. If a nQuery Predict workspace is opened, it will open a copy of the current workspace state in a new tab.
3. Tab Navigation Menu: If selected, will open a drop-down menu of all tabs currently open in nQuery. When a tab is selected, this table is displayed in nQuery.

### 1.4.5 Information Bar

The information bar provides information on the currently opened table, the currently selected cell and access to the nQuery help centre. It is shown in Figure 1.5 (note that we have selected a cell in a table for illustration purposes)

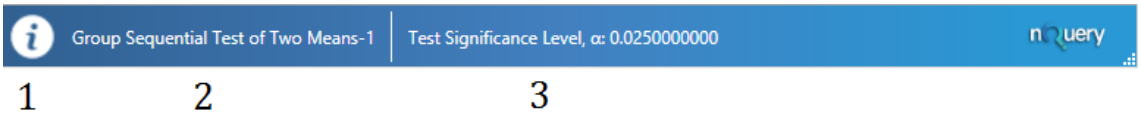



**Figure 1.5:** Information Bar

There are three main elements to the information bar: the help centre, the table name and cell information

1. Help Centre: Selecting the information symbol will open the online help centre in a browser. It provides links to guides, troubleshooting and create support ticket
2. Table Name: Provides the name of the currently open table
3. Cell Information: Provides the name and the exact value (to 10 decimal places) in the currently selected table cell

## 1.5 Opening a Design Table

The next aspect of the interface that will be reviewed is the opening of a new design table. There are three ways in which the user can open a new design table in nQuery:

1. Selecting New from the File menu
2. Selecting the New option from the menu bar
3. Selecting the "+" symbol in the tab menu when the Home tab is open or no tables are open in nQuery.

When these options are selected, the "Select Table" window will open. This is shown in Figure 1.6.

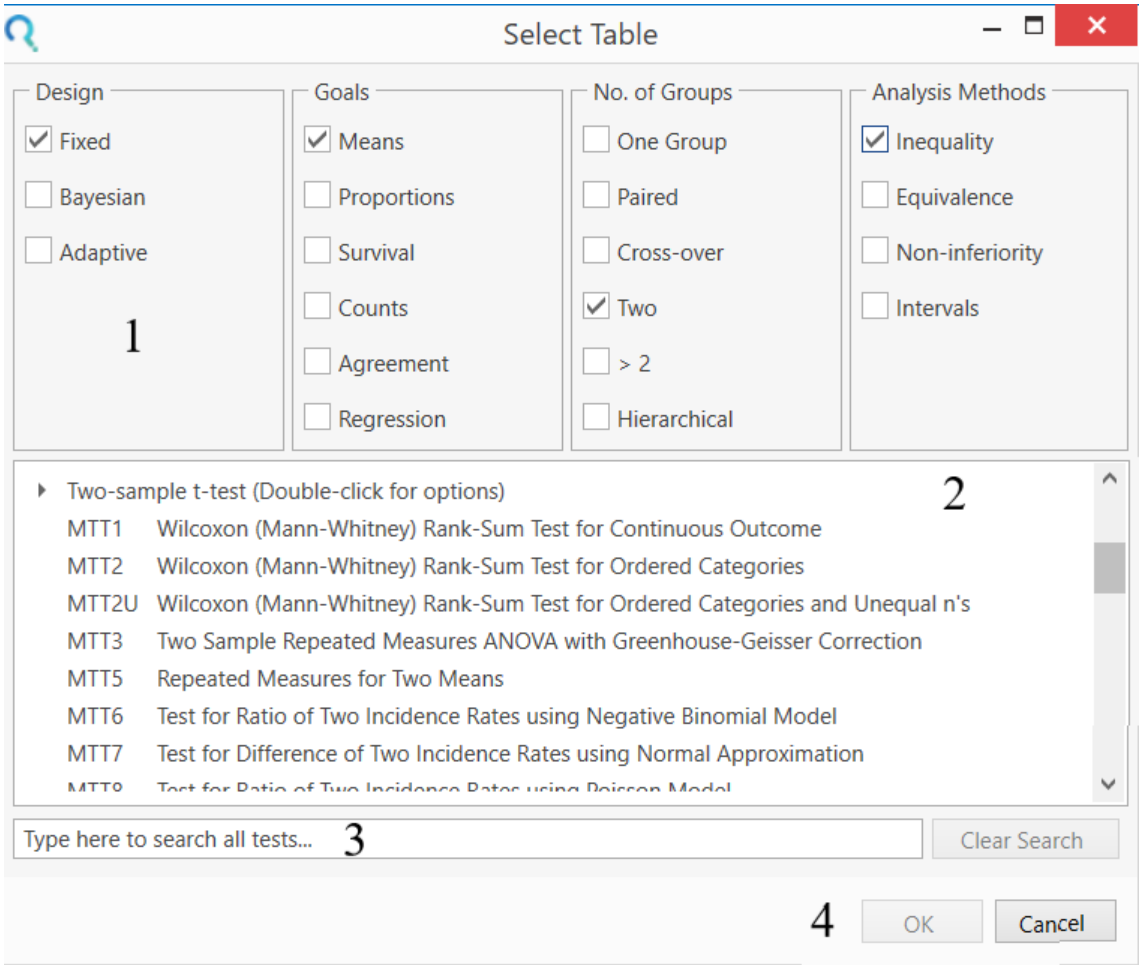



**Figure 1.6:** Select Table Window

The Select Table window consists of four main elements: the table selection menus (1), the table selection window (2), the table search bar (3) and the OK/Cancel buttons (4).

In nQuery, there are two main ways to find a specific design table: using the table selection menus or using the table search bar.

**Table Selection Menus:** The table selection menu consists of four radio button columns. These are for the Design, the Goal, the No. of Groups and the Analysis Methods. A brief description of these follows:

1. Design: Specify whether the study analysis will use a Fixed, Bayesian or Adaptive design.
2. Goal: Specify the type of data which will be analysed in the study. Choose between Means, Proportions, Survival, Counts, Agreement or Regression.
3. No. of Groups: Specify the number of groups which will be compared in the study. Choose between One Group, Paired, Cross-over, Two, >2 (Greater than Two) or Hierarchical
4. Analysis Type: Specify the type of analysis and hypothesis type which will be used in the study. Choose between Inequality, Equivalence, Non-inferiority, Intervals

By default no values are selected from the Table Selection Menus and the Table Selection Window (2, see below) will display a list of every table ordered alphabetically by table code to the left of each design table name. As options are selected from the Table Selection Menus, the list is filtered to include only those tables which fulfil that radio button condition. Note that some tables will be shown for multiple options within a given radio button column. The tables filtered for with each radio button are as follows:

In the **Design** menu, Fixed will select all tables which assume a fixed term analysis i.e. where analysis occurs once all subject data is available. Bayesian will select all tables which either use Bayesian inference to improve power calculations or find the sample size for Bayesian analysis, test or interval. Adaptive will select all tables which use an adaptive design i.e. where decisions or changes are made to the trial while it is on-going based on interim data.

In the **Goal** menu, Means will select all tables related to continuous, ordinal, incidence rate or variances data analysis. Proportions will select tables related to categorical and ordinal data analysis. Survival will select tables related to time-to-event data analysis. Agreement will select tables related to agreement, correlation or diagnostic screening data analysis or testing. Regression will select tables related to regression analysis or testing for continuous, binary, survival and incidence rate data. Cluster Randomized will select tables related to the analysis or testing of cluster randomized data.

In the **No. of Groups** menu, One Group will select all tables where a single group will be analyzed, often against a reference value. Paired will select all tables which use a paired analysis for related measurements. Cross-over will select all tables which use a cross-over design. Two will select all tables where two independent

groups are compared. >2 (Greater than Two) will select all tables where more than two independent tables are compared. Hierarchical will select all tables which use a hierarchical design including repeated measures, cluster randomized ,multi-center or mixed designs.

In the **Analysis Methods** menu, Inequality will select tables where the power for an inequality (also commonly referred to as superiority) hypothesis test is of interest. Equivalence will select tables where the power of an equivalence hypothesis test is of interest. Non-inferiority will select tables where the power of an non-inferiority hypothesis test is of interest. Note that the Non-inferiority option will also include Supersuperiority (also known as Superiority by a Margin) tables. Intervals will select tables where a specific width of the confidence interval is targeted or where another interval type is of interest. At present the other interval types available in nQuery are prediction, tolerance and credible intervals.

**Table Selection Window:** For a given set of table selection menu options or a specific table search query, the table selection window will display all of the tables which are consistent with the menu options or search query. For example, in Figure 1.6 the options selected from the Table Selection Menu are Fixed > Means > Two Groups > Inequality which corresponds to tables which use a fixed term analysis for the power of a test for a two sample superiority hypothesis. Relevant options for this combination of menu options include common tests such as Two Sample t-tests, Wilcoxon/Mann-Whitney Rank-Sum (U) tests and Two Group Repeated Measure ANOVA tests.

To open a table, select the desired table from this Window and select OK in the bottom-right of the window.

If the number of available tables for a given selection criteria is high, you can use the scroll bar on the right-hand side to find tables further down the list for the given query.

Note that some types of statistical test have a large number of variants. These variants are then contained within a sub-menu to reduce the number of options shown by default for easier searchability. These sub-menus are indicated by a "▶" symbol and the test prompt "(Double-click for options)". To open the sub-menu select the ▶ symbol or double-click the table name and the options will appear automatically below the sub-menu title. For example, in Figure 1.6 the option of "Two-Sample t-test" is a sub-menu option. If we select the ▶ drop-down menu then the options will appear below the "Two-Sample t-test" option. This is shown in Figure 1.7.

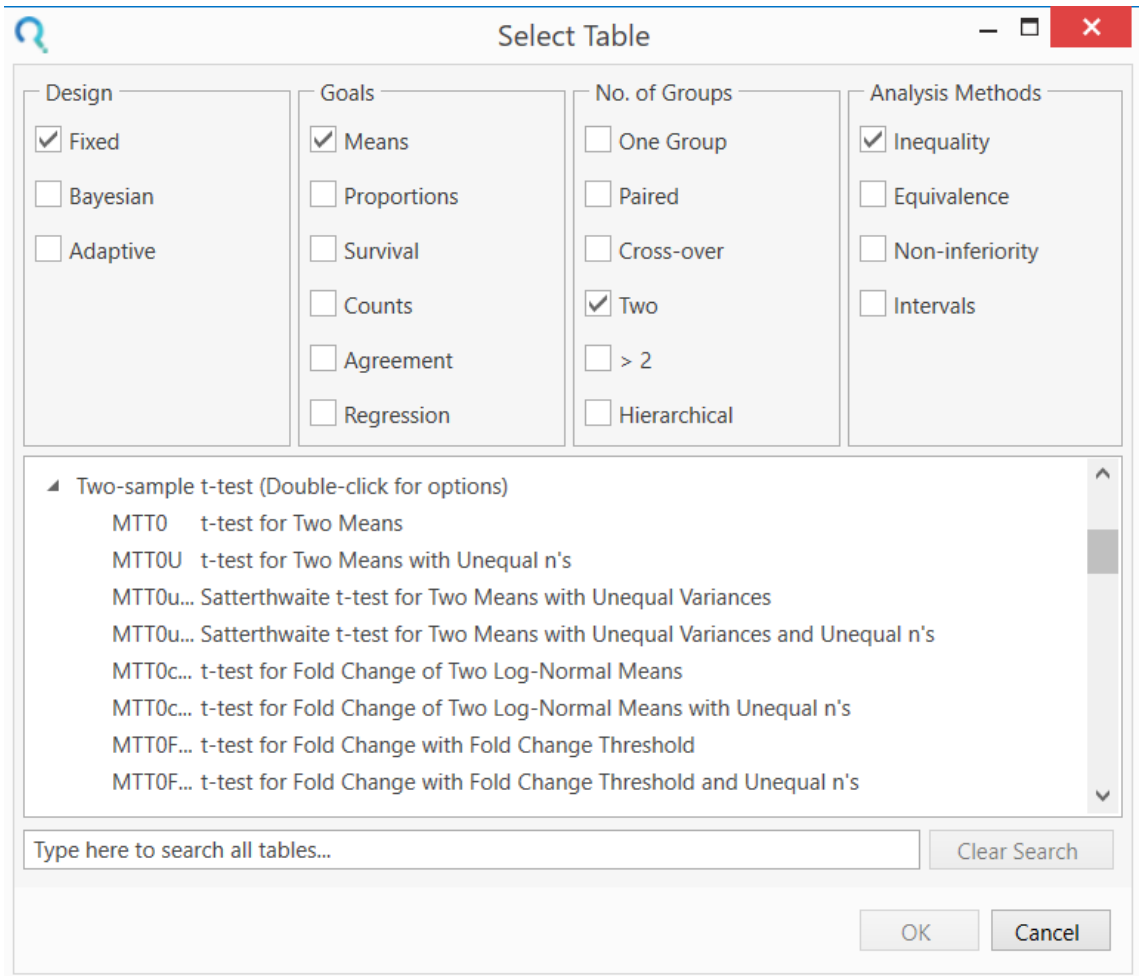



**Figure 1.7:** Select Table Drop-down

**Table Search Bar:** The table search bar allows the user to search for tables using a given text search query. For example if a t-test analysis was being considered, entering "t-test" into this search bar field would show all of the potential t-test tables in nQuery. The tables which match the query will be shown in the Table Selection Window. To open the table, select the table in Table Selection Window and select "OK". An example of this is shown in Figure 1.8.

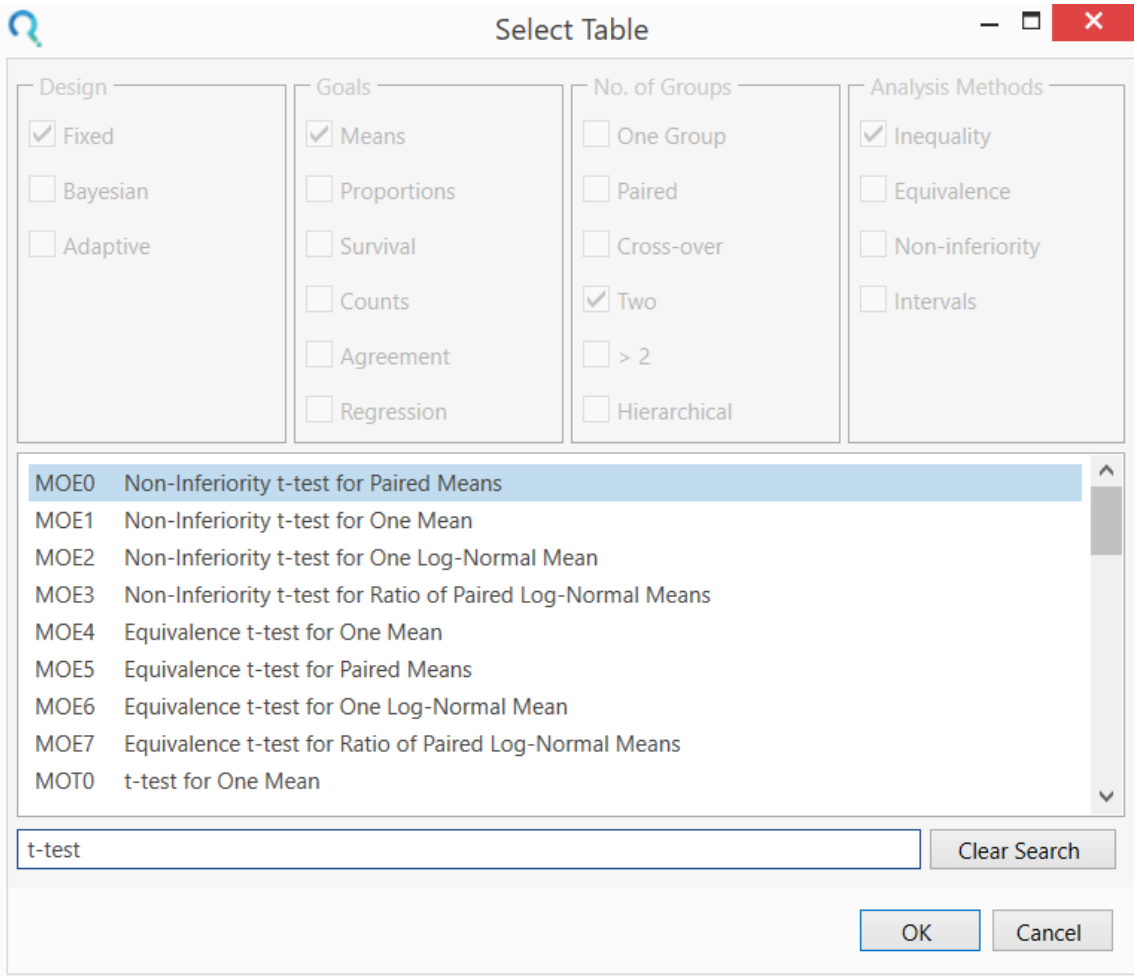



**Figure 1.8:** Table Search Query Example

Note that while a search query is active the Table Select Menu will be disabled (greyed out). To re-activate the Table Select Menu, click the "Clear Search" button.

**OK/Cancel Buttons:** These buttons complete the table selection process. When the desired table from the Table Selection Window is selected, click "OK" and that table will be opened in nQuery. To close the Select Test Design and Goal window without opening a table, select the "Cancel" button. Note that selecting the close window option in the top-right is equivalent to selecting "Cancel".

## 1.6 Design Table Layout

nQuery design tables have a large number of shared design elements. This section will outline the default layout of an nQuery design table and a brief description of these elements and their usage. Many of these elements will be covered in more

detail in future chapters. The default layout of the One Way Analysis of Variance (ANOVA) table is shown in Figure 1.9.

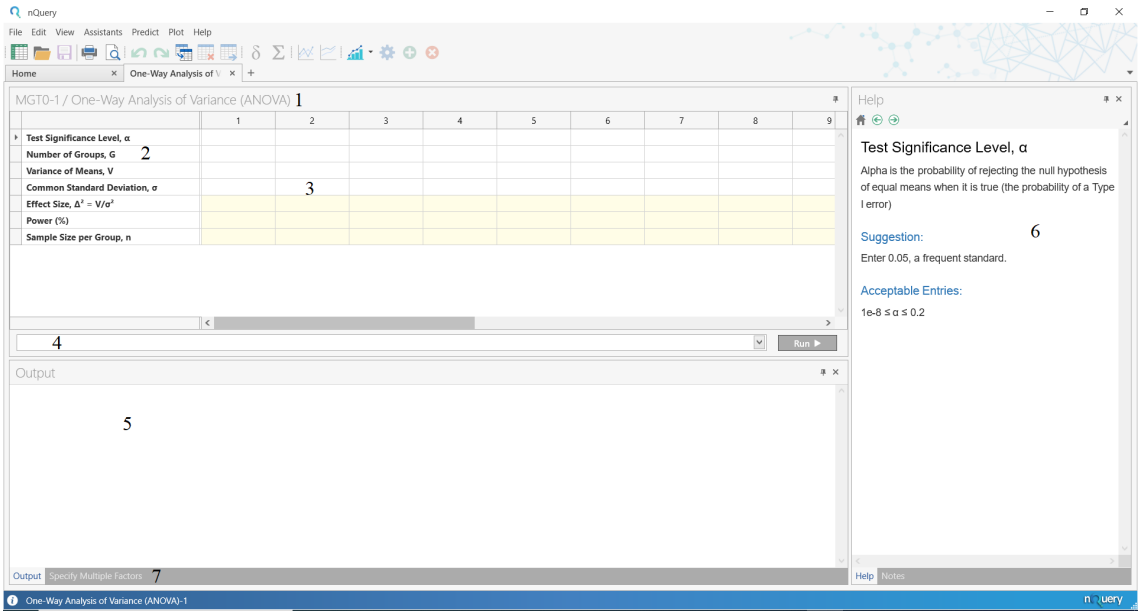



**Figure 1.9:** One Way Analysis of Variance Design Table

When a design table is opened, there are seven major elements of interest: the table title, the design parameter row-names, the main design table, the solver selection menu, the Output window, the Help window and the window selection tabs.

1. **Table Title:** This is the full name of the design table for the currently open table
2. **Design Parameter Row-Names:** The names of the design parameters required for the calculations in this design table.
3. **Main Design Table:** The table where the values of the design parameters will be entered for a specific calculation.
4. **Solver Drop-down Menu:** The menu containing the names of the solvers in this design table and the "Run" button to manually activate solvers.
5. **Output Window:** The window containing the output statement summary for a solver calculation.
6. **Help Window:** The window containing the help card for each parameter row.
7. **Window Selection Tabs:** Tabs to navigate between a window's options. In the output window, we can select the Specify Multiple Factors Tool and in the Help window, the Notes tool.

The usage of these elements will be explored in the next section using a tutorial example for nQuery.

# 1.7 Using nQuery Design Tables

## 1.7.1 Introduction

In this section, a basic example of using an nQuery design table is provided. In addition, the usage of the additional tools and options to assist in conducting a sample size or power calculation is illustrated. Some of these elements will be explored in detail later in this manual.

For this example, the One-Way ANOVA table used in section 1.5 and section 1.6 will continue to be used. This table is shown in Figure 1.10.

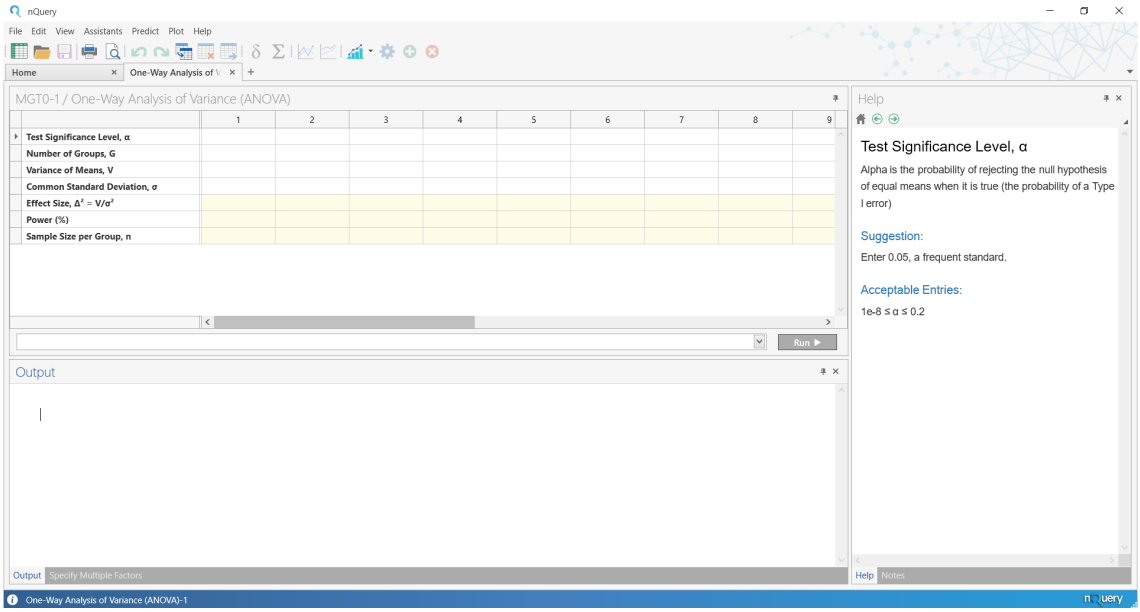



**Figure 1.10:** Clean One-Way ANOVA Table Example

In the main design table, each row corresponds to a design parameter which is set by the user and each column corresponds to an individual design specification. nQuery allows users to easily calculate the parameters of interest using multiple design specifications using its intuitive spreadsheet format.

Main table inputs will consist either numeric inputs (within the allowable range, see subsection 1.7.2 and subsection 2.2.1) or drop down menus.

Numeric inputs will either be floating point values (e.g. Variance of Means, V), percentages (e.g. Power) or integers (e.g. Sample Size). Integer rows will prevent decimal values being inputted.

Dropdown menus will consist of a list of options for that row. For example, a list of possible tests or variance calculations. The list of options is viewable by selecting the downwards arrow button on the left of the selected cell to open a drop down

menu of all available options for that row. The user can then select the required drop down option for that column. In cases where the drop down options are all numeric (e.g. 1 or 2 Sided?), the user can use the drop down menu or enter the numeric input directly.

For each row in the left-most column, the definition of that row parameter is given. In this example table, the first row corresponds to the test significance level of the proposed study design. For example, this could be set to 0.05, an industry standard.

To conduct a calculation the full set of mandatory design parameters must be filled in a column except for the desired value of interest (e.g. sample size or power). In this table, the design parameters are the test significance level, the number of groups, the variance of means, the common standard deviation, the effect size, the power and the sample size per group.

### 1.7.2 Table Help

To find additional information or guidance for a design parameter, use the help cards provided for that parameter. If a specific table row is selected, the help window on the right-hand side will dynamically update to relevant help for that design parameter. These help cards have four main potential sections: Main Text, Suggestion, Acceptable Entries and Aid. The **Main Text** section provides a definition of the parameter for the design. The **Suggestion** section provides advice on common values for a parameter or advice on how to derive the value from other known parameters. The **Acceptable Entries** section provides the range of values which are enter-able in a given cell with values outside this range not being usable in the table solvers. The **Aid** section provides information on table assistants which can be used to derive this cell value from other known information (e.g. Effect Size side table, Data Entry tool)

In Figure 1.10, the help card for the test significance level is shown on the right-hand side. This shows the definition of the significance level as the Type I error, the suggested value of 0.05 and the acceptable entries of 1E-8 to 0.20. In Figure 1.11, the help card for the Variance of Means row is shown. In this help card, there is an example of an Aid section which refers to the usage of an Effect Size side-table to derive the value for this parameter.

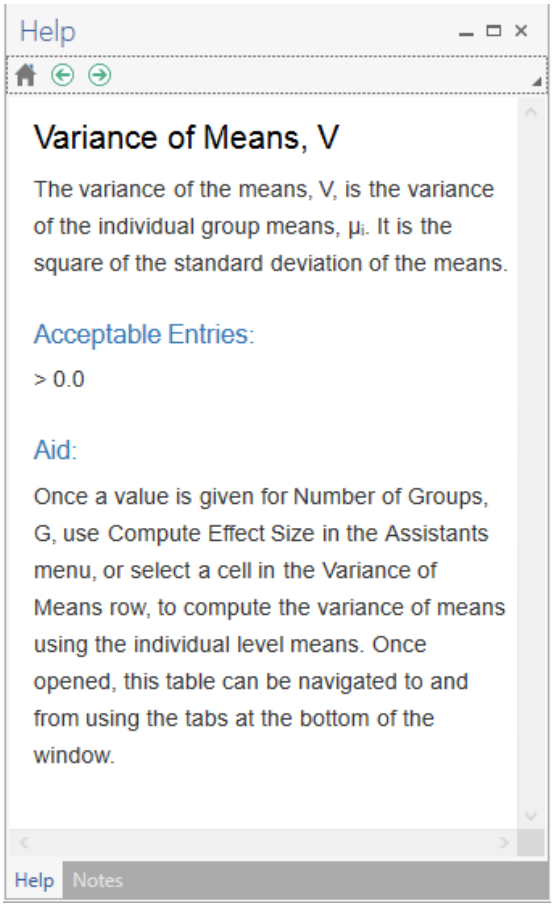



**Figure 1.11:** Variance of Means Help Card Example

If you want an overall summary for the table, select the Home icon at the top of the Help window. This will provide a brief summary on how to use the table and provide the reference materials used for the design table solver methods. The Home Card for this table is shown in Figure 1.12.

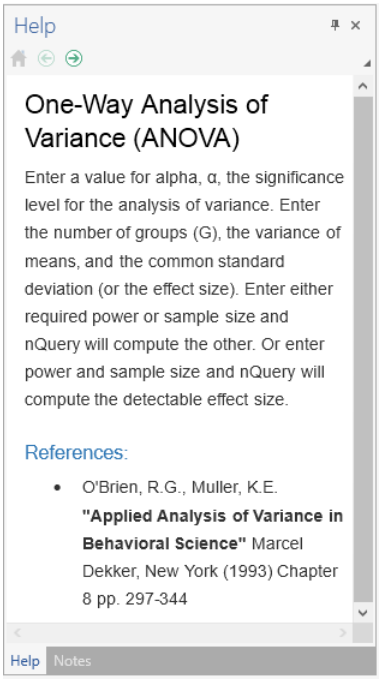



**Figure 1.12:** Home Card Example

### 1.7.3 Activating a Table Solver

To conduct a calculation of interest, we require the usage of a table solver. Rows which are highlighted in yellow correspond to a table solver i.e. rows for which there is a solver available. In this example, there are solvers available for the effect size, the power and the sample size.

The easiest way to activate a specific solver is to fill in all of the other mandatory design parameters in a table column except for the parameter of interest. The other methods for selecting and activating solvers will be discussed in later chapters.

After the values for the design parameters which will be used in this study design is decided, enter these into the relevant column. In this example, assume a sample size calculation is requested and that values given in Table 1.1 were used for the mandatory design parameters. This design corresponds to a study which will use a one-way analysis of variance to test the null hypothesis that 3 independent group means are equal at the 0.05 significance level with 80% power, assuming a "medium" effect size of 0.5 [Cohen, 1988]. Note that in this table, the variance of means and common standard deviation are optional parameters and are thus not required for a solver to activate. These rows are provided for user convenience to derive the effect size.

**Table 1.1:** Example Parameter Values

| Parameter | Value |
|---|---|
| Test Significance Level | 0.05 |
| Number of Groups | 3 |
| Effect Size | 0.5 |
| Power | 80 |

When these values are entered into column one of the design table, the sample size is automatically calculated due to being a solver row and all other mandatory parameters being specified. This is illustrated in Figure 1.13

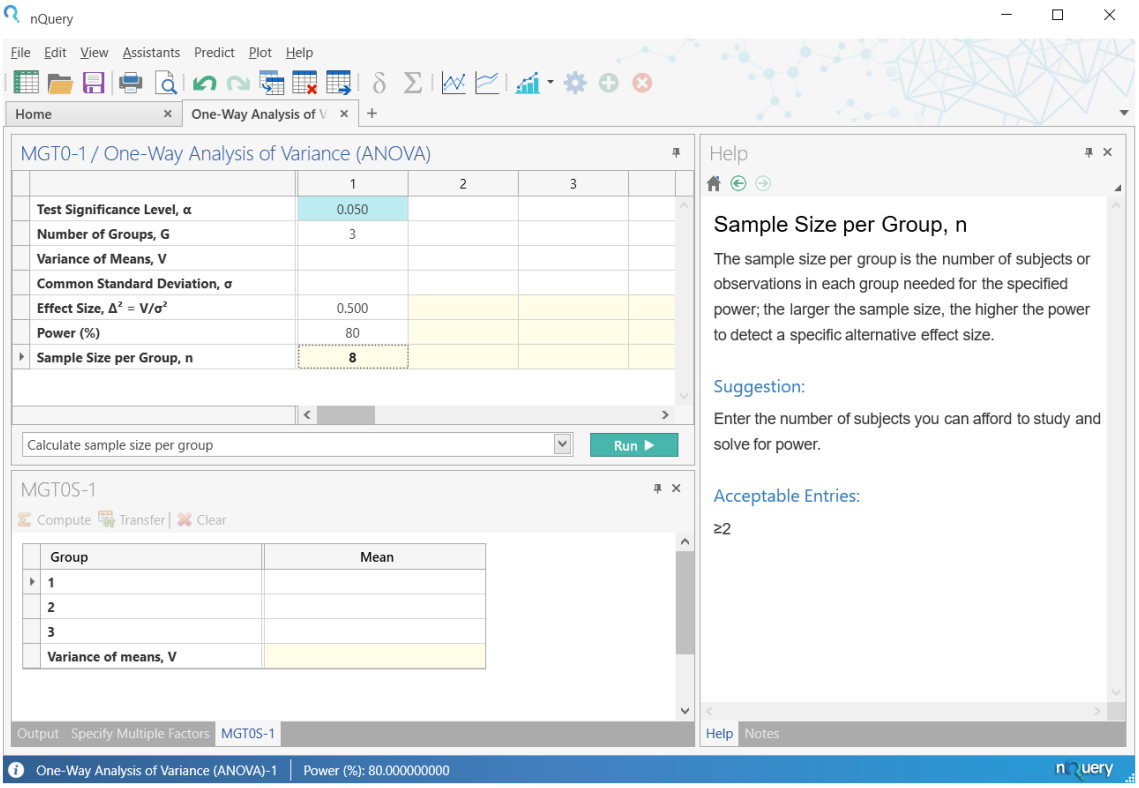



**Figure 1.13:** Example Solver Answer

In this example, the sample size per group required is 8. This corresponds to a total sample size of 24 (8 x 3). When a solver is activated the solver output (in this case the sample size per group in column 1) will be the only cell still highlighted in yellow and that the text of the solver answer will be highlighted in bold. Note that some tables will automatically update the power to its exact value after a non-power solver is used, in these cases the solver will highlight the row of interest and the power row containing the exact power.

The other solvers for power and effect size can be activated in a similar fashion. Note that the "Calculate sample size per group" solver is now activated in this column unless this column is subsequently fully cleared using the "Clear Table" option or "Clear" from the right-click context menu while the column cells are highlighted. For more detail on how to manually select and change solvers, see section 2.1.

This very basic example illustrates the ability of nQuery to easily and quickly find the appropriate sample size for your study.

### 1.7.4 Effect Size Side-table and Auto-Calculations

Let's expand upon this example by using the optional elements and effect size side table to derive the sample size for a study design where the user can take advantage of the additional information they have available to find better estimates for the required design parameters.

For this example, assume that the common (within-group) standard deviation is assumed to be equal to 1.1. Assume that the expected group means are also known and are equal to the values in Table 1.2.

**Table 1.2:** Tutorial Group Means Example

| Group # | Mean Value |
|---|---|
| 1 | 1.1 |
| 2 | 2.3 |
| 3 | 3.0 |

To use these group mean values to derive the Variance in Means, use the Effect Size side-table. To open this side-table, the number of groups must be specified. Once the number of groups is specified in a column, the "Calculate Effect Size" options will become active in the Assistants menu and menu bar when that column is selected. However, the easiest way to open the side-table in a column is to select the derived row in that column and the side-table will automatically open below the main design table (i.e. in the same window as "Output").

Once the side-table is open, enter the group means into the relevant rows. This will activate the "Calculate" button in the top-left of the side-table window. Select "Calculate" and this will show the calculated value of approximately 0.616 in the Variance of Means row in the side-table. This activates the Transfer button. Selecting "Transfer" will transfer the value for the Variance of Means to the relevant row for the current side-table's column of the main table. Note that some tables will only have a "Transfer" button as the required outputs will be calculated automatically for the user in the side-table and in other tables there will be neither a "Calculate" or "Transfer" button, in these cases the side-table will automatically calculate and

transfer the required outputs to the main table when the side-table is filled. This is shown in column two in Figure 1.14.

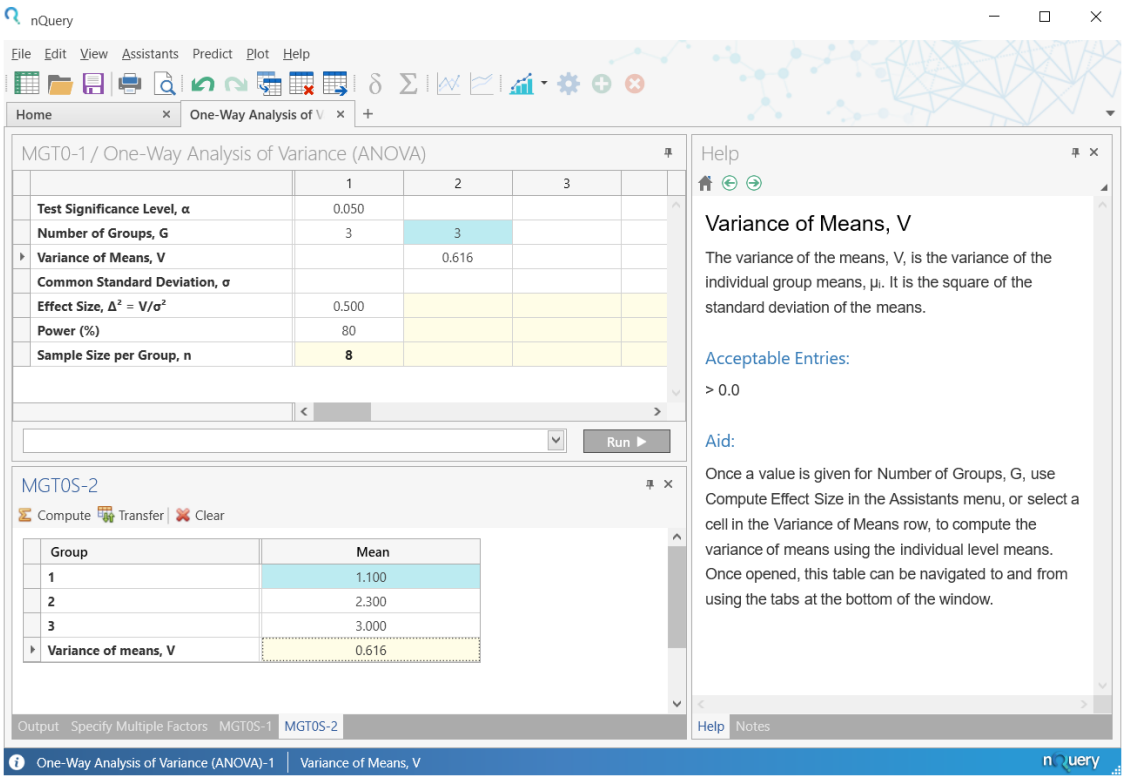



**Figure 1.14:** Side Table Example

Now enter the common standard deviation value of 1.1. In this table, the effect size can be derived from the variance in means and the common standard deviation and thus when these are both specified the effect size will automatically update to derived value for the values set for those two parameters. In this example, this equals approximately 0.509. This is shown in Figure 1.15.

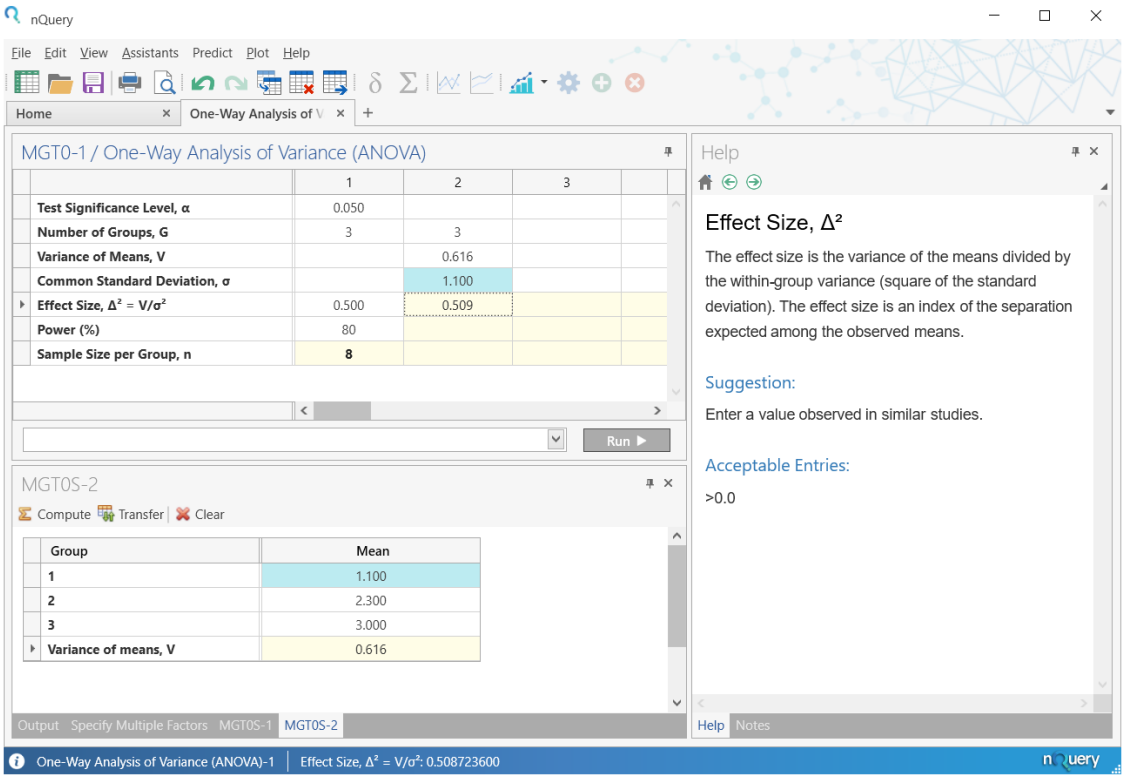



**Figure 1.15:** Effect Size Auto-Calculation Example

Note that entering two out of three of these parameters (variance, standard deviation, effect size) will always return the third value and that changing one after all three are specified will cause one of the other parameters to change to ensure consistency (e.g. updating the variance of means or common standard deviation would cause the effect size row to update).

To complete the example, enter the same significance level (0.05) and power (80) as for the previous example. This will give the same sample size per group of 8.

## 1.7.5 Saving and Opening nQuery Files

nQuery comes with a save file format named .nqt. This file format allows users to save their work in a design table and use it again in future sessions. The save file will retain the design table, side-table, output statements, plots and notes from the time of saving.

To save an nQuery file, the user can use the Save or Save as... options from the File menu or by using the Save option icon in the tool bar. If the currently open design table has been saved before then selecting the save option will automatically overwrite that file with the current state of the design table. If the design table has

not been saved before or the user selects the Save as... option then the Save menu will appear. Select the desired folder and edit the default save name in the File Name field if desired. Select Save to save the file.

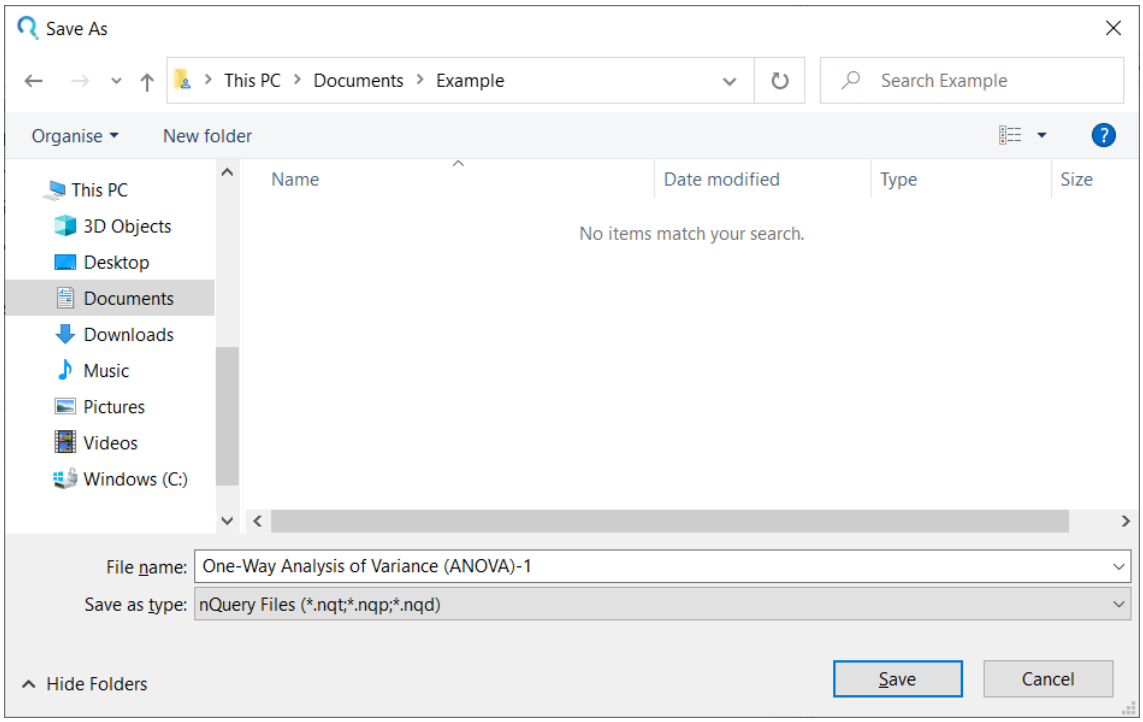



**Figure 1.16:** Save Window

To open an nQuery file, the user can select the Open option from the File menu or by using the Open option icon in the tool bar. This will open the Open menu. Select the folder which contains the previously saved .nqt file and select the file. Select Open to open the file in nQuery. You can also open an nQuery (.nqt) file by double-clicking the file within Windows explorer. This will open the save file in a currently open instance of nQuery or open an instance of nQuery and open the file in the new instance.

## 1.8 Design Table Tools

nQuery features a number of additional tools to help users understand and summarise their calculations. The tools which will be covered in this section are the Output statement, the Notes tool, the Specify Multiple Factors tool, the plotting tools and saving and opening an nQuery table.

### 1.8.1 Output Statement

The output statement provides a verbal summary of the results given in a column after a solver has been activated. This statement can be used as a template for the sample size justification given in a study protocol, academic paper or similar document. The output statement is shown by default in the window below the main design table in nQuery. In this example, due to the usage of the effect size side-table, that window currently shows the effect size side table for column two of the main table. To return to the Output statement, select the "Output" tab on the left of the tab bar at the bottom of the lower window. The output statement for the column two calculation will be shown below the main table in Figure 1.17.

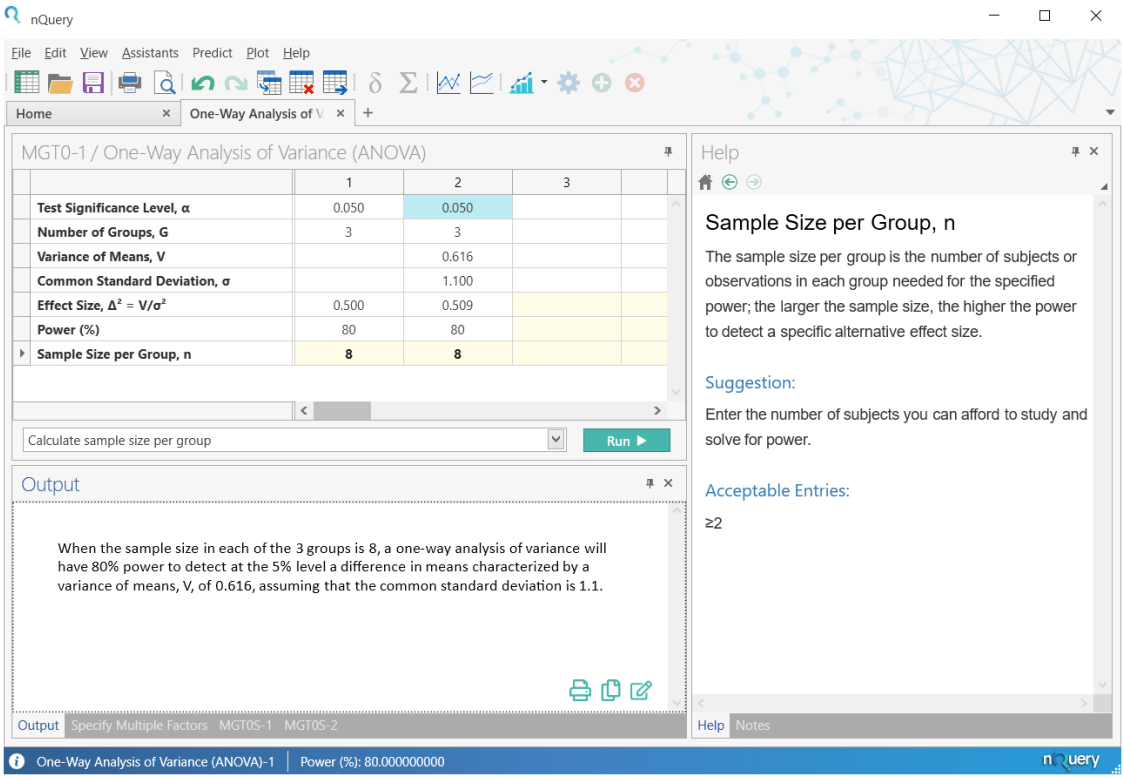



**Figure 1.17:** Output Statement Tutorial Example

To print the output statement select the Print button (1st icon) in the bottom-right of the output statement window or use the Print or Print Preview options from the right-click context menu. The Print icon will open a Print Preview window.

To copy the output statement to the clipboard, either use the Copy All button (2nd icon) in the bottom-right of the output statement window or use the Copy option in the right-click context menu. The Copy option will default to Copy All if no text is selected or will copy the selected text if part of output statement text is highlighted.

To edit the output statement, select the Edit button (3rd icon) in the bottom-right of the output statement window or use the Edit option in the right-click context menu. When finished editing, select the Stop Editing button (same position as Edit icon).

To transfer the current statement into the Notes tool (see subsection 1.8.2), select Add to Notes from the right-click context menu.

Note that the output statement will change depending on whether the optional elements have been fully specified in the selected column. For example, the output statement in column 1 will remove the references to optional elements of the variance of means and the common standard deviations and be replaced with the effect size. This is shown in Figure 1.18

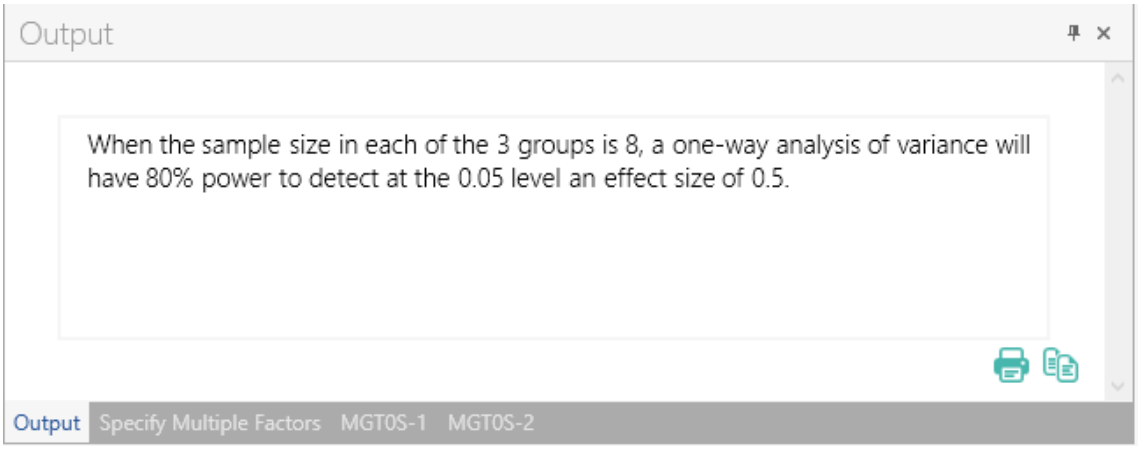



**Figure 1.18:** No Optional Elements Output Example

## 1.8.2 Notes Tool

The notes tool is a word processor facility integrated into nQuery which allows users to write and save notes about their calculations directly into their nQuery save file. By default, it is found in the window to the right of the main table. To open the Notes tool, select the "Notes" option from the bottom of the Help window in the tab bar. This Notes Tool window can be seen on the right-hand side of Figure 1.19.

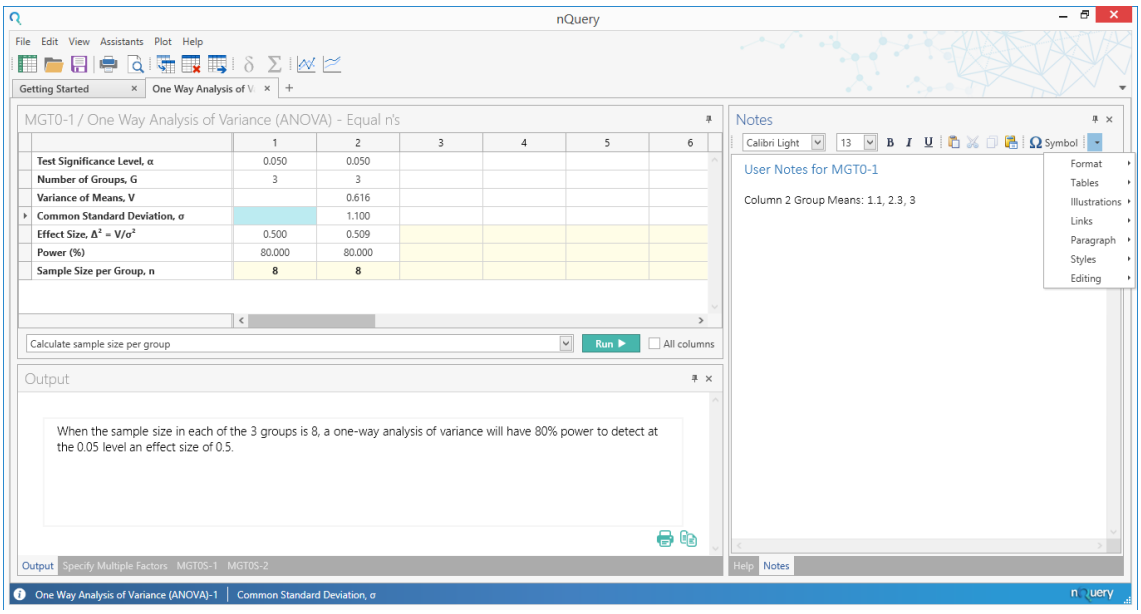



**Figure 1.19:** Notes Tool Example

At the top of the Notes Tool, the toolbar will contain a drop-down for the font and text size and buttons for the style (bold, italics, underline), copy, cut, paste, paste special, print and a Symbols menu. The arrow on the right-hand side of the toolbar gives access to a large variety of advanced tools for formatting, tables, images and other options. Right-clicking within the editor will open a context menu containing the Edit tools (copy, cut, paste), indent options, font and paragraph menus, bookmark and hyperlink options.

subsection 1.8.1 shows that Output Statements can be transferred to the Notes tool using the "Add to Notes" option from the right-click context menu within the Output tab. This will be automatically placed below any material in the Notes tab at that point in time.

### 1.8.3 Specify Multiple Factors Tool

The Specify Multiple Factors tool provides a method to quickly enter multiple values for a design parameter and to generate all combinations if more than one design parameter is varied simultaneously. The Specify Multiple Factors tool is found by default in the window below the main table. To open the Specify Multiple Factors tool, select the "Specify Multiple Factors" option from the bottom of the Output/Side-table window in the tab bar. This Specify Multiple Factors Tool window can be seen on the bottom of Figure 1.20.

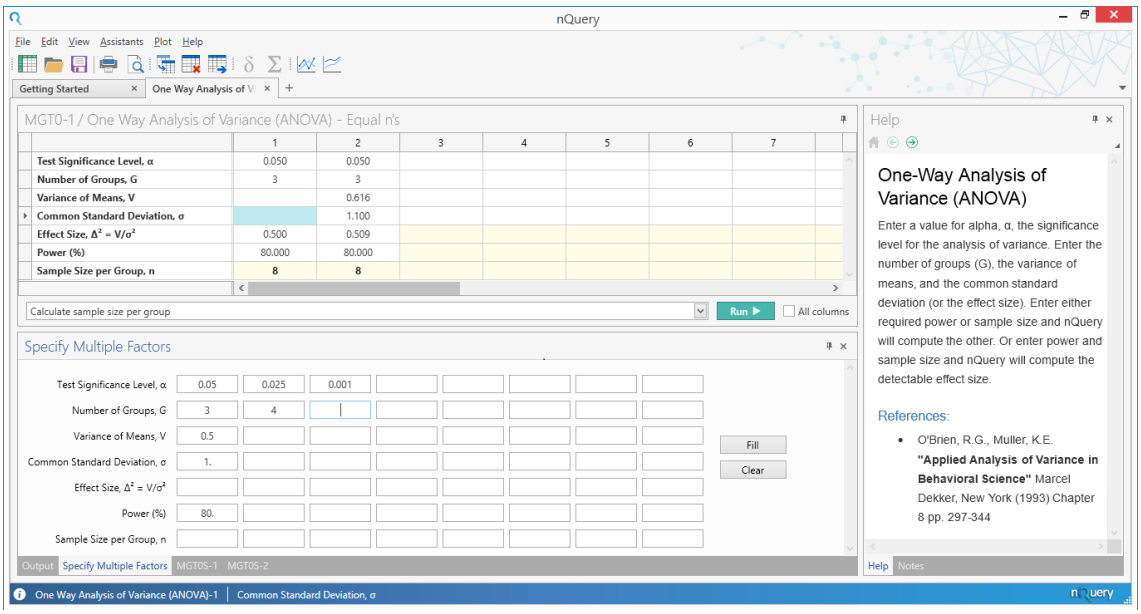



**Figure 1.20:** Specify Multiple Factors Tool Example

The Specify Multiple Factors tool allows a maximum of eight values per design parameter.

In this example, the significance level will be evaluated at 0.05, 0.025 and 0.001 and the number of groups will be evaluated at 3 and 4. The variance in means, common standard deviation and power will be fixed at 0.5, 1 and 80 respectively.

Note that the effect size row values are left empty since it will be auto-calculated from the variance of means and common standard deviation. The sample size is left empty as this will be solver targetted for this example.

To transfer these inputs into the main table by selecting the "Fill" button on the right-hand side of the window. The results for this set of inputs is shown in Figure 1.21.

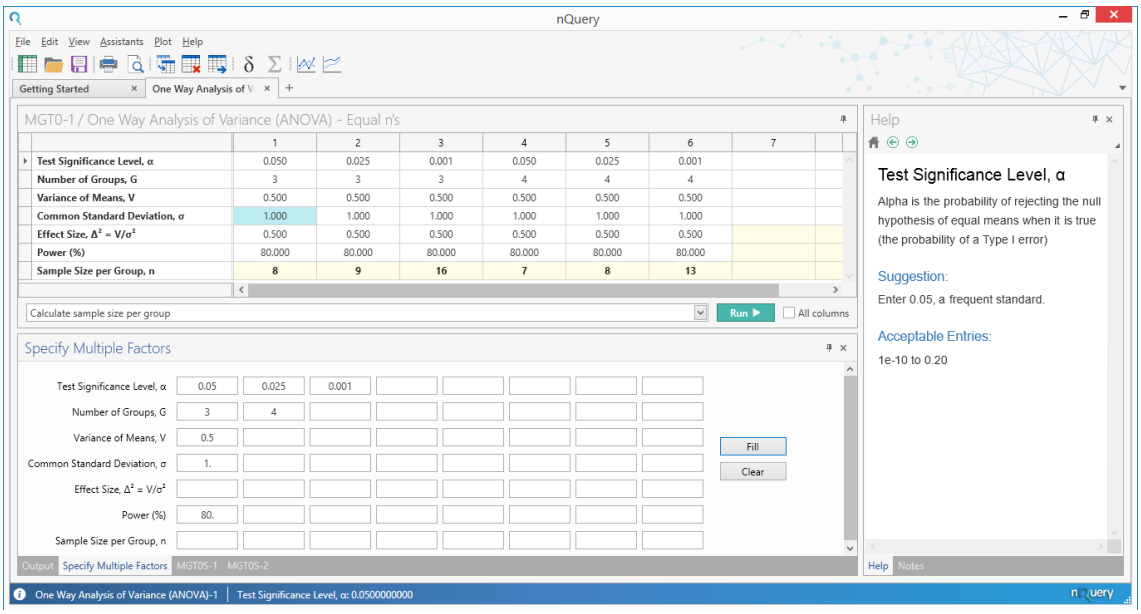



**Figure 1.21:** Specify Multiple Factors Output Example

Note that the number of columns filled by the Specify Multiple Factors is equal to product of the number of values entered for each row. In this case this equals 3 x 2 x 1 x 1 x 1 = 6.

To clear the inputs from the Specify Multiple Factors tool and reset the tool, select the "Clear" button on the right of the window.

If the number of column combinations exceeds the number of columns open in the table, the table will automatically create additional columns up to the maximum number of columns

For rows which contain a drop down menu (e.g. 1 or 2 Sided Test?), a drop down menu will exist for all cells for that row in the Specify Multiple Factors tool. By default, all cells will have the same drop down option selected but the user can manually select other drop down options as per the main table. If there are duplicate cells in that row of the Specify Multiple Factors tool, the duplicates will be ignored.

If the number of column combinations exceeds the allowed maximum number of columns then a warning will be given and only the first number of combinations equal to the maximum number of columns will be shown. The maximum number of columns is configurable in Options menu.

### 1.8.4 Plot Power vs Sample Size

The Power vs Sample Size plot provides a quick way to assess the relationship between power and sample size over a wider range while fixing the other elements

in calculation. The Power vs. Sample Size plot can be activated in two ways: using the Plot menu drop-down option "Plot Power vs Sample Size" or using the tool bar "Plot Power vs Sample Size" button . Both of these options will be inactive until a column is selected which has all of the mandatory elements except for the power and sample size.

After a column is filled appropriately, select the Plot menu or menu bar options and the plot will appear in the application foreground. If you want to compare up to 6 columns, select cells in the desired columns and open the plot and the results for all columns will be displayed simultaneously. Note that holding Ctrl (to select individual columns) or Shift (to select a range of columns) while clicking columns may assist in selecting multiple columns.

An example of an individual and multiple column versions of the Power vs Sample Size plot are shown in Figure 1.22. In this example, the plots are for the six columns generated from the Specify Multiple Factors example above, with the individual plot being from column 1.

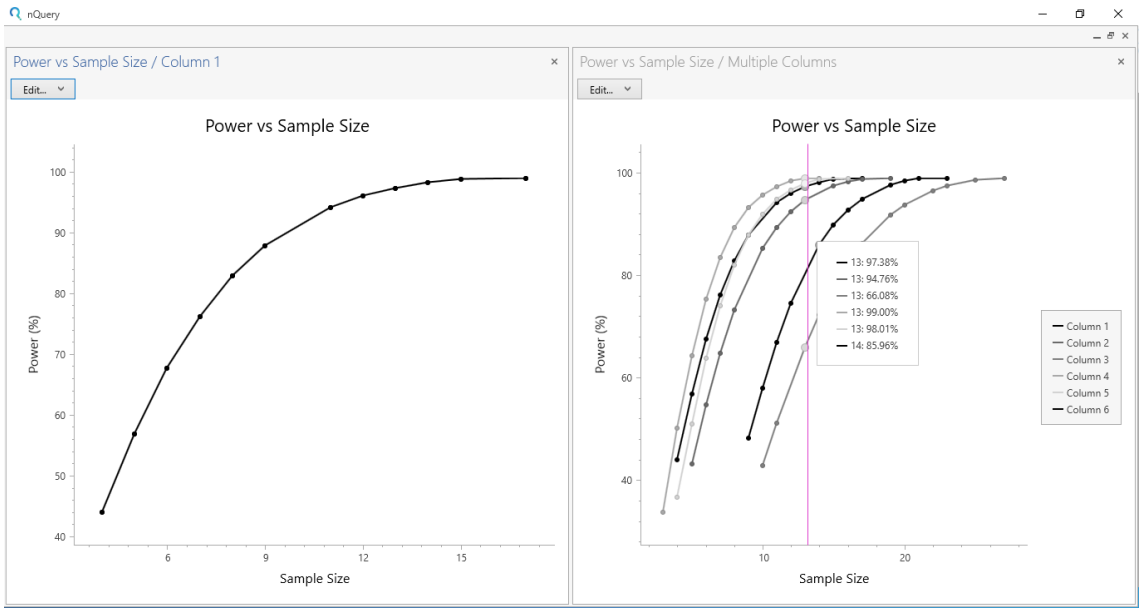



**Figure 1.22:** Power vs Sample Size Example

Highlighting a point in the plot will display the exact values for the X-axis and Y-axis value selected for each column.

Right-clicking on the plot will open a context menu which provides options to Print, open a Print Preview, Save the plot as an image file and options to edit the plot (seesection 1.9). The plot can also be edited using the Edit... menu in the top-left of the plot window.

### 1.8.5 Plot User Selected Rows

The Plot User Selected Rows plot provides a visual and customised way to assess the relationship between any of the continuous or integer design parameters and any of the solvers given all other unselected design parameters are fixed. The Plot User Selected Rows menu can be activated in two ways: using the Plot menu drop-down option "Plot User Selected Rows" or using the tool bar "Plot User Selected Row" button . Both of these options will be inactive until a column is filled sufficiently for at least one valid custom plot to be available.

After a column is filled appropriately, select the Plot menu or menu bar option and this will open the "Select X-axis, Y-axis" window. This menu is shown in Figure 1.23.

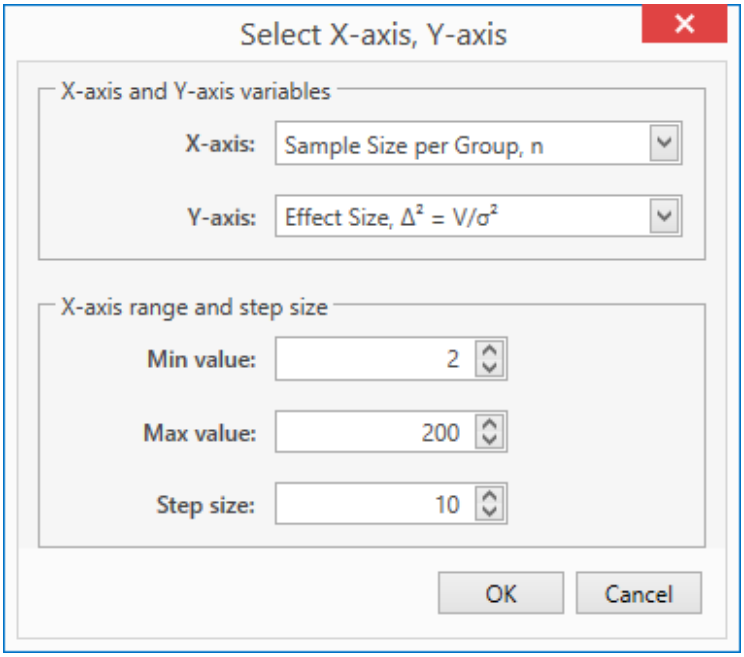



**Figure 1.23:** "Select X-axis, Y-axis" Plot User Selected Rows Menu Example

In this menu, the X-axis drop-down will contain the full list of design parameters which are of either a continuous or integer type. The Y-axis drop-down will contain the list of the all the solvers available in the table. This table allows us to create a combination of any of these two lists to create a custom user plot.

When an X-axis option is selected, the user can specify a range over which that design parameter will be varied in the plot by specifying a minimum value and maximum value for this range. Note that the OK button will be disabled and an error displayed if the "Min value" is below the lower limit of the acceptable entries for the selected design parameter or is greater than the current "Max value" or if

the “Max value” is above the upper limit of the acceptable entries for the selected design parameter or is lower than the current “Min value”.

After a range is specified, the user can select the “Step Size”. This equals the increments that the X-axis parameter will be increased in when moving from the specified Min and Max values. The number of steps can be found by dividing the range by the step size (round down if answer is a non-integer). Note that the Step Size must be positive and must give a number of steps which is between 4 and 1000. If this condition is not met then an error will be displayed and the OK button greyed out.

Default values will be given for the Max value, Min value and Step size for each X-axis option in a given table.

If you want to compare up to 6 columns for given X-axis and Y-axis option selection, select cells in the desired columns and open the plot and the results for all columns will be displayed simultaneously. Note that holding Ctrl (to select individual columns) or Shift (to select a range of columns) while clicking columns may assist in selecting multiple columns.

An example of an individual and multiple column versions of the Plot User Selected Rows plot are shown in Figure 1.24. In this example, the plots are for the six columns generated from the Specify Multiple Factors example above. The individual column plot is for column 1 where the X-axis option was the Test Significance Level, using the default 0.01 to 0.2 by 0.0095 X-axis range, and Y-axis option was the Effect Size. The multiple columns plot is for the six column where the X-axis option was the Sample Size per Group, with a custom range of 5 to 30 by 1 X-axis range, and Y-axis option was Effect Size.

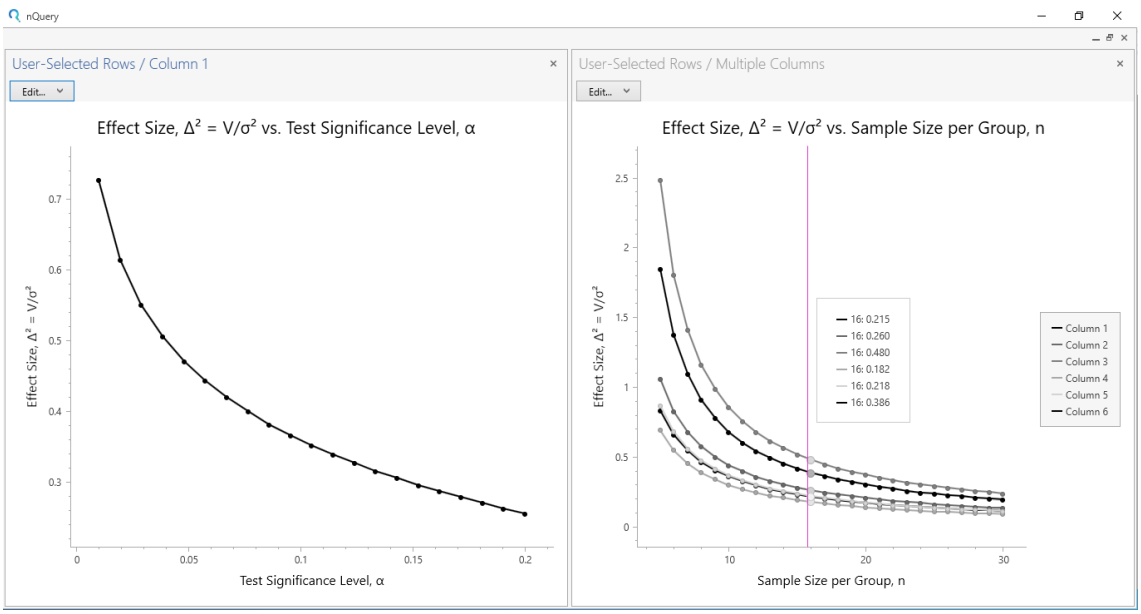



**Figure 1.24:** Plot User Selected Rows Example

Highlighting a point in the plot will display the exact values for the X-axis and Y-axis value selected for each column.

Right-clicking on the plot will open a context menu which provides options to Print, open a Print Preview, Save the plot as an image file and options to edit the plot (seesection 1.9). The plot can also be edited using the Edit... menu in the top-left of the plot window.

## 1.9 Editing Plots

nQuery provides options to edit the default Power vs N and User Selected rows plots. To edit these plots, select the "Edit..." button from the top-left corner of the plot or use the context menu by right-clicking on the plot area.

In the edit menu, there will be options for Print, Print Preview and Save As which allow the printing and saving of the current plot. There are the following options for editing a plot

1. Titles (Edit Table and Axis Titles)
2. Series (Edit Series names and colors)
3. Gridlines (Add and Edit Gridlines)
4. Legend (Add and Edit Legend)

### 1.9.1 Titles Options

The Titles menu contains options for the three main table titles: the Chart Title (above plot), the X-axis Title (below plot), the Y-axis Title (left of plot). Each of these titles has the same three options for editing:

1. Hide/Show Title: Hide or show selected title in plot
2. Rename: Rename title in "Rename Title" dialog window
3. Font: Edit Font, Size, Style (Bold, Italics) and Text Color ("More Colors" for custom colors) in "Update Title Font" dialog window

### 1.9.2 Series Options

The Series menu contains options for updating the x-axis range and editing the color and name of each series (column) in the plot. If a single series is in the current plot, the following three options will be available:

1. Update X-axis range: Edit range of x-axis displayed in plot
2. Rename: Rename series (column) in "Rename Series" dialog window

3. Color: Change the color used for series (column) in "Select Series Color" dialog window. Custom colors can be created by selecting the "More colors" option

If more than two series (columns) are selected in the current plot, this menu will have the "Update X-axis range" option and the names of all the series in the current plot. If you select a specific series, the Rename and Color options will be available for that series. These options work as above.

### 1.9.3 Gridlines Options

The Gridlines menu contains options to add and remove major gridlines for the X and Y axes. This menu contains the following options:

1. Hide/Show X-axis gridlines: Hide or show X-axis gridlines depending if shown at present
2. Hide/Show Y-axis gridlines: Hide or show Y-axis gridlines depending if shown at present

These gridlines will be in-line with the tick marks for the relevant axis.

### 1.9.4 Legend Options

The Legend menu contains options to add and remove the legend and to edit the horizontal and vertical position of the legend. This menu contains the following options:

1. Hide/Show Legend: Hide or show legend depending if shown at present
2. Horizontal Position: Selecting this will open a sub-menu for legend locations. This menu gives five options to place the legend: Left inside, Left outside, Center, Right inside, Right outside. Select the desired option to change the position
3. Vertical Position: Selecting this will open a sub-menu for legend locations. This menu gives five options to place the legend: Top inside, Top outside, Center, Bottom inside, Bottom outside. Select the desired option to change the position

"Inside" and "Outside" refer to whether the legend should be included in the plot area itself (inside) or place outside the plot area (outside). Note that changing the line colors and legend series names is done using the Series options described above. Series names can also be edited using the column name feature by double-clicking the column title in the main table before creating a plot.

# 2 Advanced nQuery Table Features

## 2.1 Manual Solver Activation

In nQuery, the easiest way to generate a result is to use the auto-activation of a solver. An example of auto-activation is given in subsection 1.7.3. In short, a solver will automatically activate if all the mandatory elements in a design table are filled except for a solver row (solver rows are orange in the design table).

However, in certain cases, a user may wish to fix a column to use a specific solver regardless of user inputs or change the solver being used in an already complete column. Both of these are covered in the following section and will make use of the **Solver Drop-down Menu and Run button** referenced in Figure 1.9.

### 2.1.1 Fixing a Column's Solver

To fix the solver which will be used in a given column, open a design table and select the column or columns which will have their solver fixed. To do this, select a cell or cells in the column(s) of interest or select the full column(s) by selecting the column title(s). This can be achieved by clicking and dragging for multiple cells or by holding down Ctrl (individual cells/columns) or Shift (range of cells/columns) while selecting cells or columns.

After the cell(s) or column(s) are selected, the solver can be fixed by selecting a specific option from the Solver Drop-down menu. These options are illustrated for the One Sample Chi-Square Test in Figure 2.1.

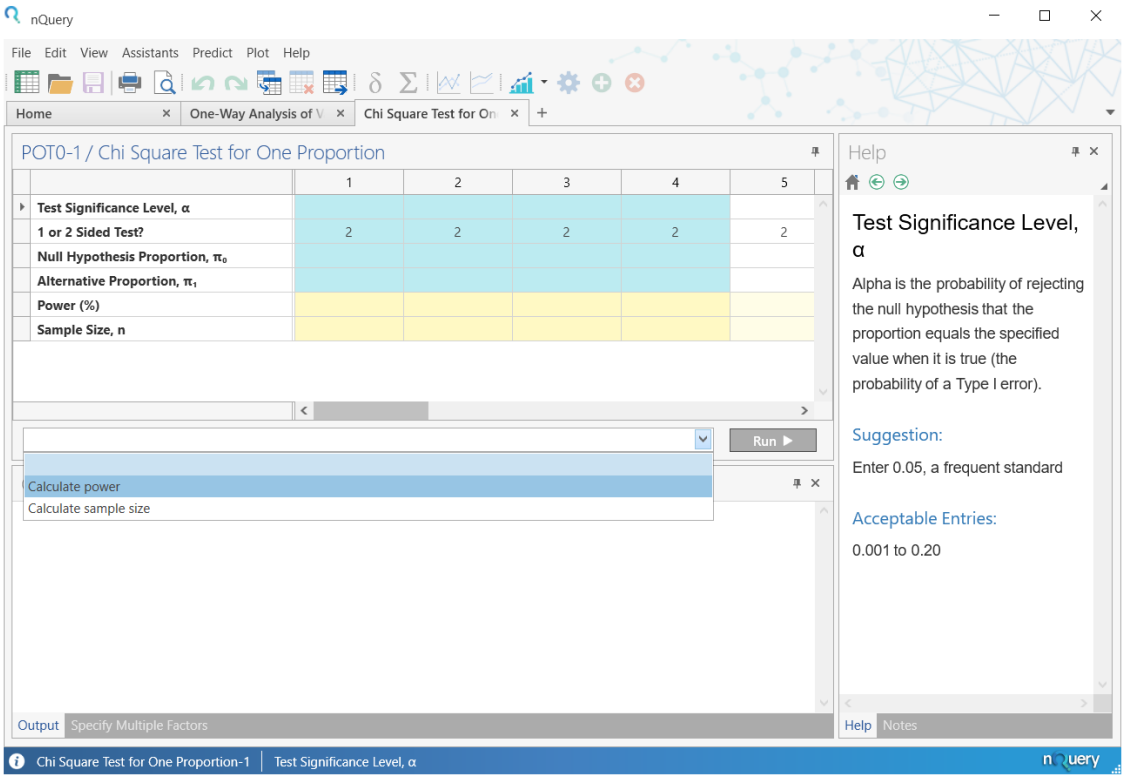



**Figure 2.1:** Solver Drop-down Menu Example

The One Sample Chi-Square test has two solvers available: "Calculate power" and "Calculate sample size". In this example, the power solver will be fixed for the first three columns of the table by selecting the "Calculate power" option from the solver drop-down while the first three columns are highlighted as in Figure 2.1.

As the solver is fixed for these three columns, entering the design parameters in a column which would activate the power solver will calculate the appropriate power for those design parameters. However, entering parameter values which would activate the sample size solver does not generate any solver output. This is illustrated in Figure 2.2.

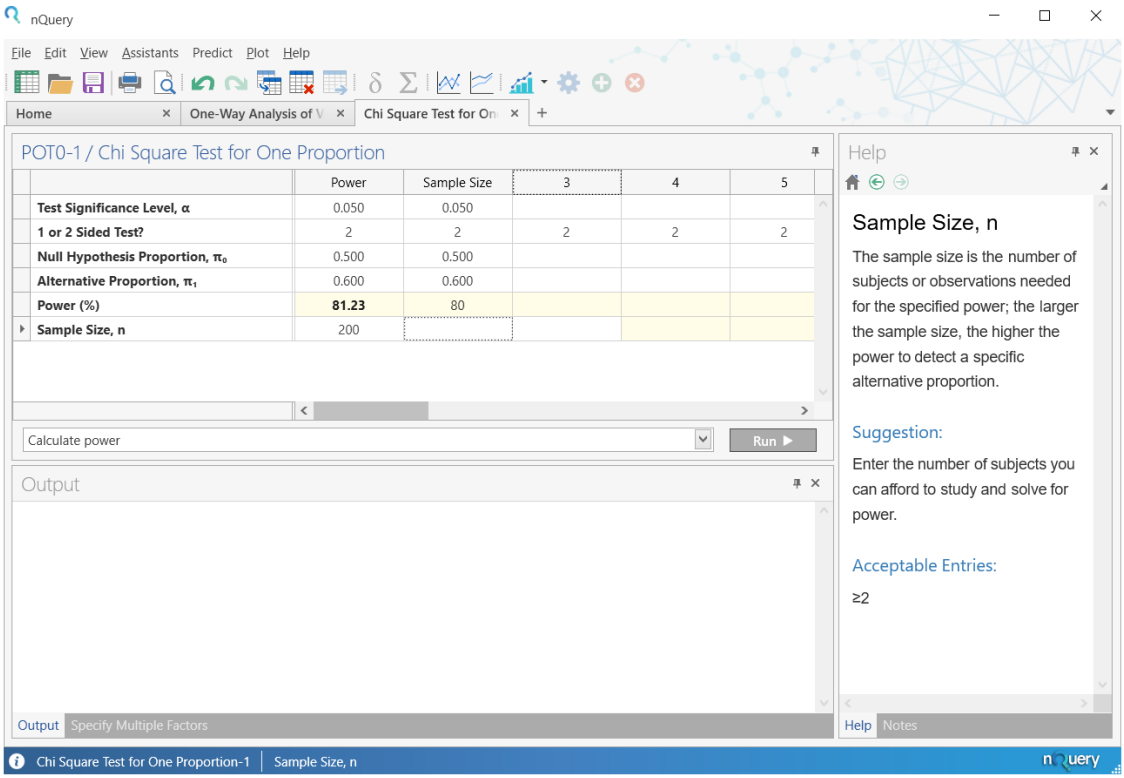



**Figure 2.2:** Solver Manual Activation Example

In this example, the first column has successfully generated the power but in the second column the values generate no result even though this column would generate the sample size if it was in its default state. The other solver(s) would be selected and would work in the same fashion.

In every table the first option in the solver drop-down is a "blank" option. If this option is selected, the selected column(s) will reset to auto-selecting a solver based on the current column inputs.

## 2.1.2 Changing a Filled Column's Solver

After a solver is activated and generates a solver output result (either automatically or manually) in a column, the user may be interested in changing the solver being used in that column. There are two main methods to achieve this: Reset the column or Select an alternative solver

### 2.1.2.1 Reset Column

To reset a column, there are three main methods: Clear Table, Clear Column, Blank Solver Option

1. Clear Table: Selecting Clear Table from the Edit menu or the menu bar will reset all columns to their default state where any solver can be automatically or manually activated.
2. Clear Column: Highlighting a column(s) by selecting the column title or all the cells in a column and then selecting "Clear" from the right-click context menu will return the column to its default state
3. Blank Solver Option: Selecting a column and selecting the first "blank" option from the Solver Drop-down menu will reset the column to its default state

Unlike the Clear options, the Blank Solver option can be used on a partially filled or a fully filled column. If a column if fully filled and the blank option is selected, a new calculation can be activated by changing a value in the affected column(s) or by selecting the "Run" button to the right of the solver drop-down menu. The first calculation in the solver drop down list will be selected if the column was full.

#### 2.1.2.2 Select Alternative Solver

To select an alternative solver select the column(s) or interest, open the Solver Drop-down menu and select the solver which is desired. This works the same as the blank solver option above. After an alternative solver is selected, it will be activated if a cell is changed in the affected column(s) or by using the "Run" button to the right of the solver drop-down menu. Selecting the sample size solver and using the "Run" button is illustrated in Figure 2.3.

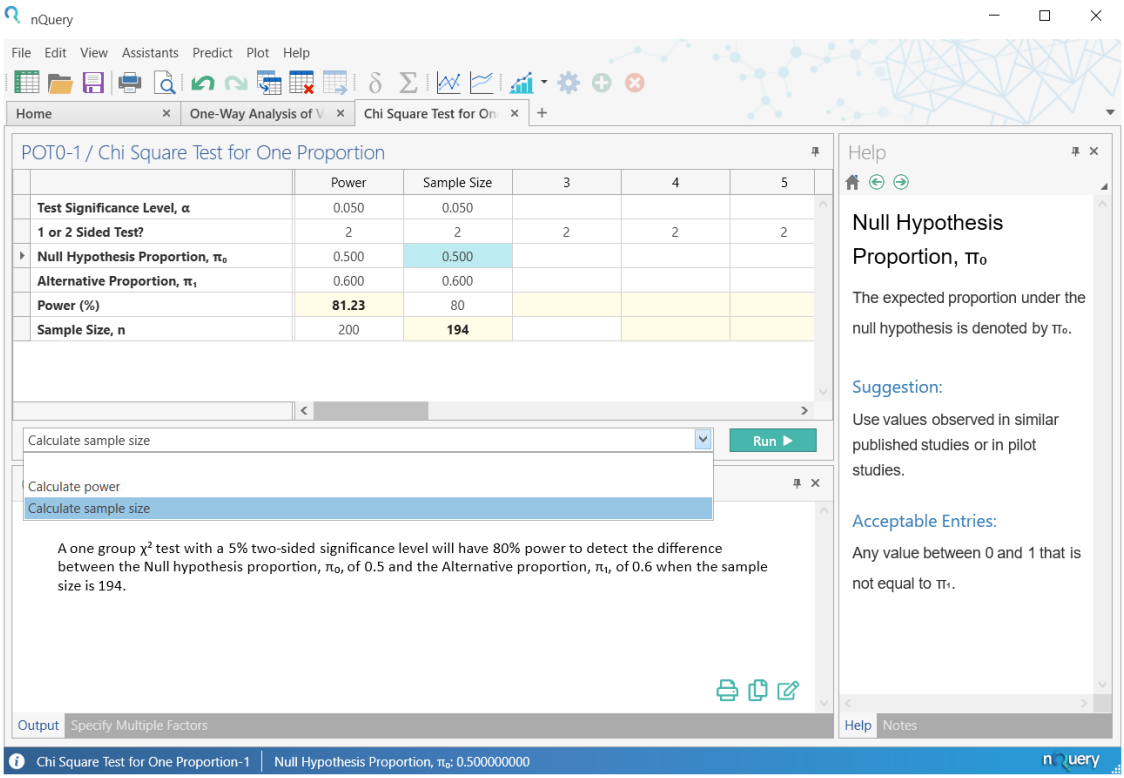



**Figure 2.3:** Manual Solver Change Example

## 2.2 Errors and Warnings

nQuery provides a number of errors and warnings to prevent incorrect values being entered or when a solver cannot find a legitimate answer. There are three primary types of errors and warnings in nQuery: Out of Range errors, Solver errors, Solver Warnings.

### 2.2.1 Out of Range Errors

Out of range errors occur when a value is entered into a design table or side-table cell which is outside the acceptable entries for that cell. The acceptable entries for a given cell can be found in the Help card for that cell (see subsection 1.7.2).

Note that out of range errors can occur due to table auto-calculations. For example, the auto-calculation for the Effect Size row in One Sample t-test can be outside the acceptable entries of 0.1 to 40 when derived from the optional element values of the null hypothesis mean, alternative hypothesis mean and standard deviation. This is also common in side-tables where there are read-only auto-calculated cells

which contain the sum of the writeable elements of the side-tables (e.g. POT2 - Chi-Square Test for Proportions in Multiple Categories).

When an out or range error occurs, a red ⊗ symbol will appear on the left-hand side of the affected cell. Placing the mouse over the cell will provide a tooltip which gives a brief summary of the allowable values in the cell. An example of this is shown in Figure 2.4.

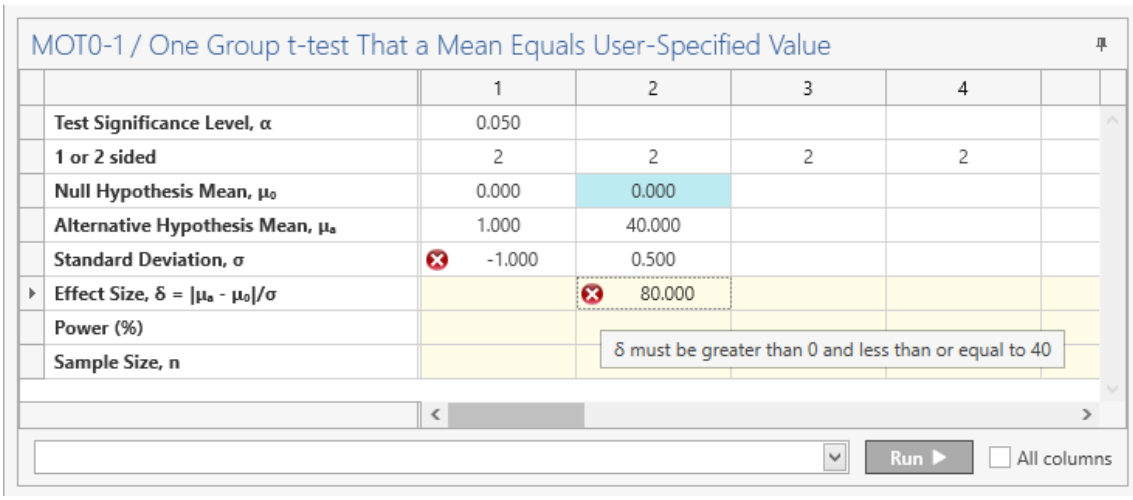



**Figure 2.4:** Out of Range Error Example

While an out of range error is present in a design table column, no solvers will activate in that column. In a side-table, an out of range error will prevent the Compute/Transfer buttons from activating.

### 2.2.2 Solver Errors

Solver errors occur when a solver cannot find a solution for the given inputs. This is most common when the design parameters used are extreme or when the solver needs the use of array search or memory intensive methods. These errors indicate that either there is no correct result for the design parameters given or that the correct result was too extreme to be found by the solver algorithm.

When a solver error occurs, the affected cell will be highlighted in red. If you hover the mouse over the affected cell, a brief description of the error will be given. An example is given in Figure 2.5 where the population size is too low causing there to be no legitimate output for the Lgamma function due to there being no value for the adjusted sample size which would achieve 80% power.

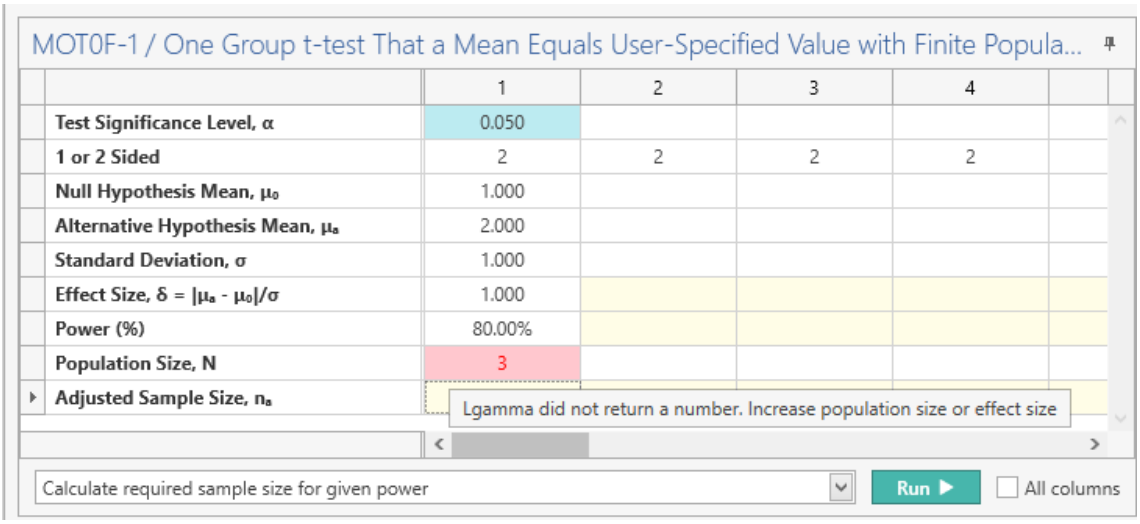



**Figure 2.5:** Solver Error Example

### 2.2.3 Solver Warnings

Solver warnings occur when a solver result is given but there are issues which would cause concern over the veracity of the result given. The two major categories of solver warnings are rounding warnings and assumption warnings.

1. Rounding Warnings: These occur if the solver output was rounded due to being too extreme originally.
2. Assumption Warnings: These occur if an underlying assumption of study design is not met by the current design parameters.

Examples of rounding warnings include rounding sample size up to two when it was below two and rounding the power down to 99% when it was greater than 99%. Examples of assumption warnings include the minimum cell count being too low for chi-square tests and the normal approximation not holding when the population size and sample size are similar when the finite population adjustment is being used.

When a solver warning occurs, the background is clear and the solver answer value is red. This is illustrated in Figure 2.6 where column one has a "Calculate power" solver warning for the minimum cell count being low and the column two has a "Calculate power" solver warning that the power was rounded down to 99% due to being "unrealistically" high. As per the errors above, the specific warning is viewable by placing your cursor over the affected cell(s).

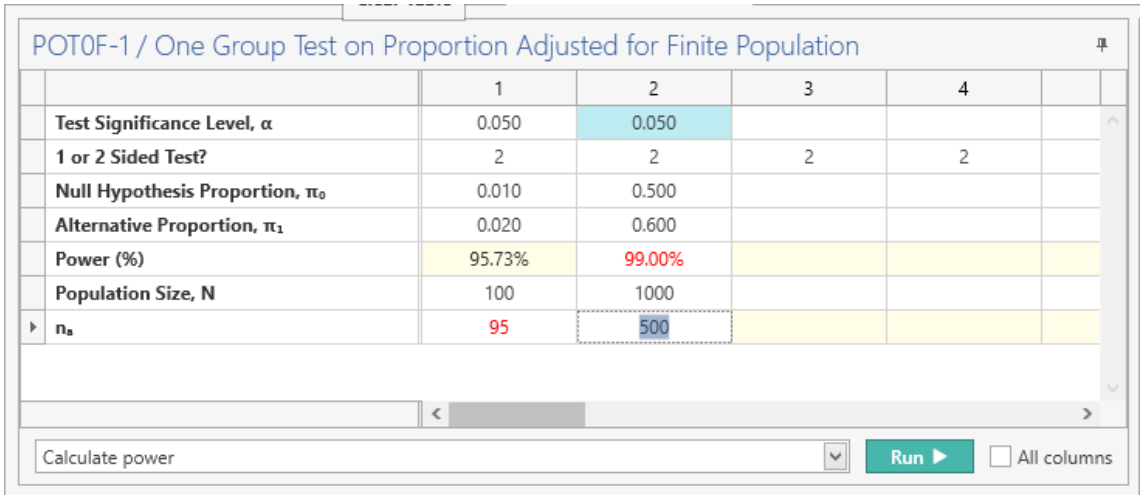



**Figure 2.6:** Solver Warning Example

### 2.2.4 Application Logging

When nQuery opens a table or encounters an error, these are logged automatically by nQuery. If you want to track the errors in the application or wish to provide additional support to Statsols Technical Support then these logs will be useful. An example is given in Figure 2.7.

By default, these logs are saved in "C:/Users/<Username>/AppData/Roaming/nQuery/Logs". The save location can be changed in the Options menu (see section 2.3).

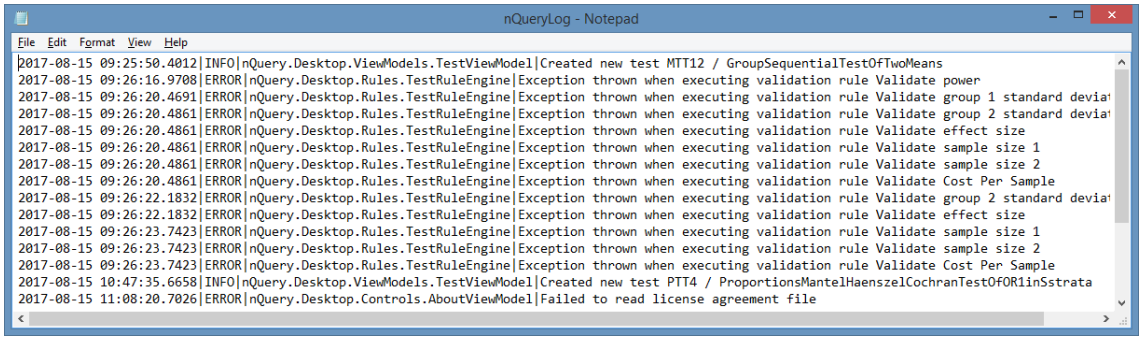



**Figure 2.7:** nQuery Log Example

## 2.3 Options Menu

The options menu gives the user the ability to change a number of nQuery features to reflect a users preferences.

### 2.3.1 Application-Wide Options

The application-wide options are shown in Figure 2.8 with their default settings.

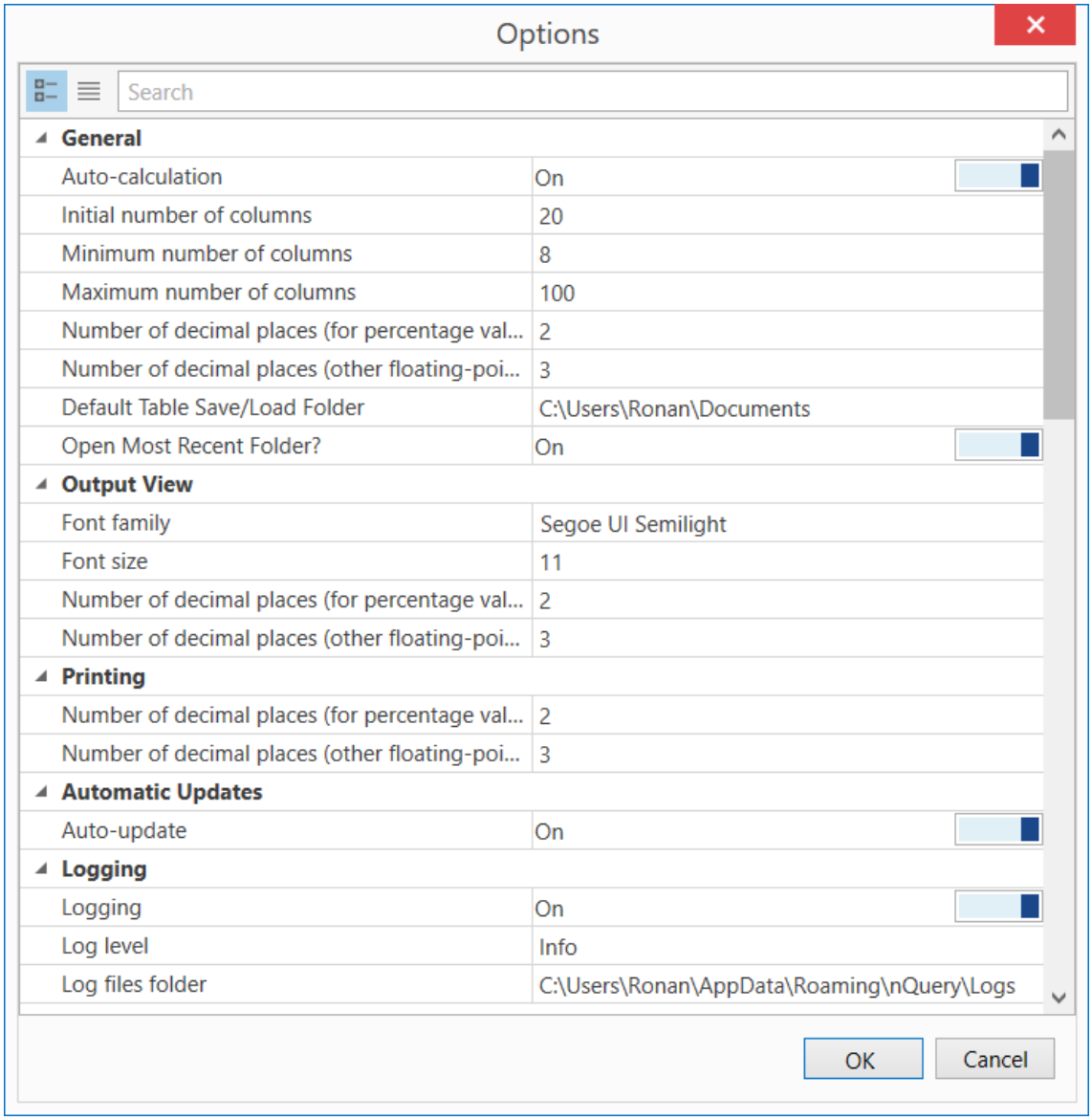



**Figure 2.8:** Application Options Menu

There are three main categories of application-wide settings. These are named and described below:

1. General Settings: Auto-calculation, Initial number of columns, Minimum number of columns, Maximum number of columns, Number of decimal places (percentages, other values)

   a) Auto-calculation: Toggle whether solvers are activated automatically if column is filled appropriately. If disabled, solvers are only activated via the Run button.
   b) Initial number of columns: Set the number of columns included in a design table when opened initially
   c) Minimum number of columns: Set the minimum number of columns in a design table. Resets to this value if the number of Specify Multiple Factors combinations is less than this value.
   d) Maximum number of columns: Set the maximum number of columns in a design table. Resets to this value if the number of Specify Multiple Factors combinations is more than this value.
   e) Number of decimal places (for percentages): Set number of decimals displayed for percentage rows in tables e.g. power
   f) Number of decimal places (other floating-point values): Set number of decimals displayed for all non-percentage rows in tables e.g. means, proportions
   g) Default Table Save/Load Folder: Default folder shown in Save/Open windows in nQuery
   h) Open Most Recent Folder?: Toggle for whether to show folder most recently used to save or open an nQuery table file (.nqt)
2. Output View: Font family, Font size, Number of decimal places (percentages, other values)
   a) Font Family: Select the font family used in the Output window text
   b) Font Size: Select the font size used in the Output window text
   c) Number of decimal places (for percentages): Set number of decimals displayed for percentage rows in output statement e.g. power
   d) Number of decimal places (other floating-point values): Set number of decimals displayed for all non-percentage rows in output statement e.g. means, proportions
3. Printing : Font family, Font size, Number of decimal places (percentages, other values)
   a) Number of decimal places (for percentages): Set number of decimals displayed for percentage rows when printing e.g. power
   b) Number of decimal places (other floating-point values): Set number of decimals displayed for all non-percentage rows when printing e.g. means, proportions
4. Automatic Updates: Auto-update

   a) Auto-update: Select if software should attempt to automatically download and install latest version if available on start-up

5. Logging: Logging, Log level, Log files folder.
   a) Logging: Toggle whether logging occurs
   b) Log Level: Change the level of event logged by nQuery. Options are Info, Debug, Error, Warning
   c) Log Files Folder: Change location where log file is saved

### 2.3.2 Chart (Plotting) Options

This section contains options to edit all default options for plots. The plot options are discussed in section 1.9. These options are shown in Figure 2.9 with their default settings.

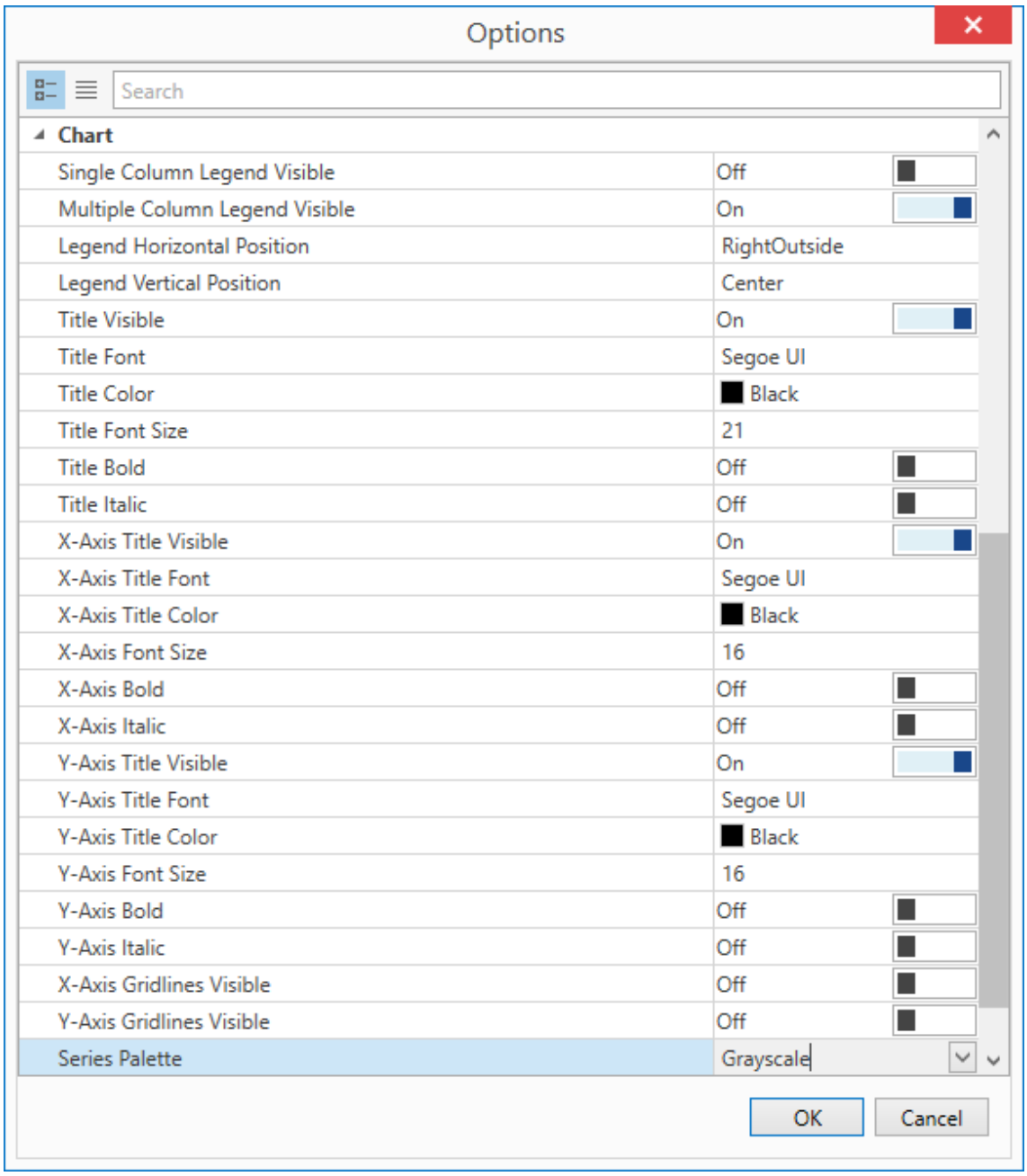



**Figure 2.9:** Plot Options

### 2.3.3 nQuery Predict Options (nQuery Advanced Expert only)

This section contains the options for nQuery Predict. These options are shown in Figure 2.10 with their default settings. More detail on nQuery Predict can be found in chapter 6.

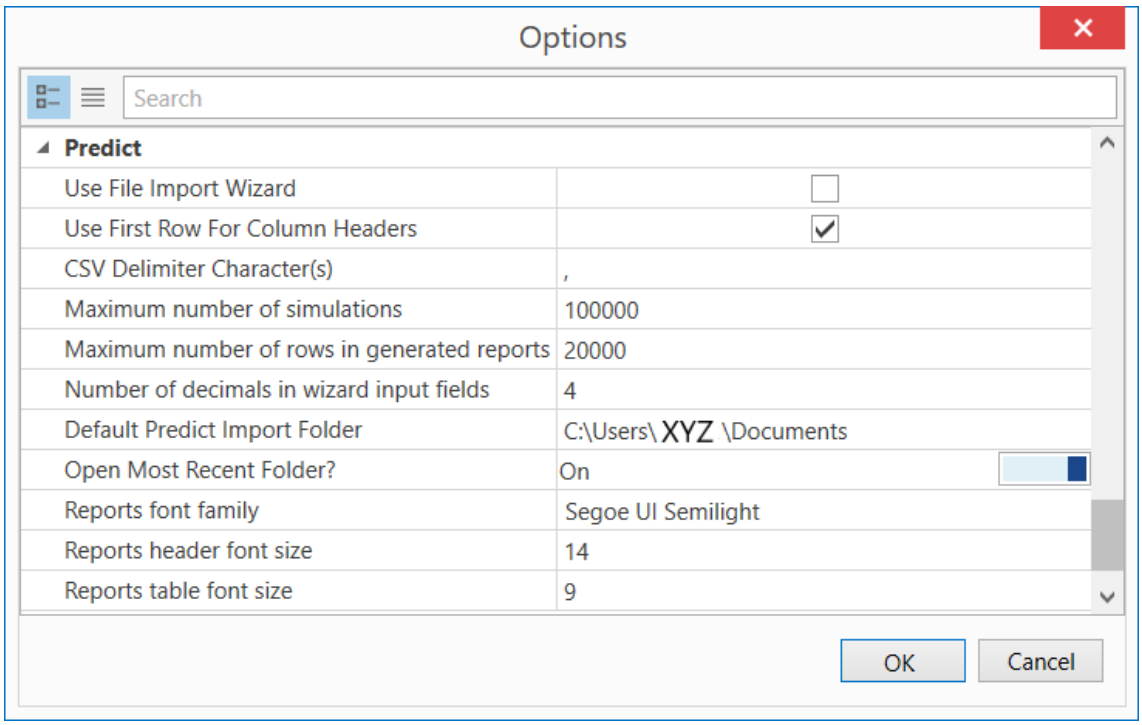



**Figure 2.10:** nQuery Predict Options

The options for nQuery Predict are as follows:

1. Use File Import Wizard: Toggle if Advanced File Import Wizard is needed
2. Use First Row for Column Headers: Toggle if first row of imported datasets should be assigned as column headers
3. CSV Delimiter Character(s): Enter delimiter between entries in imported CSV files
4. Maximum Number of Simulations: Maximum number of simulations allowed in nQuery Predict workspace
5. Maximum Number of Rows in Generated Reports: Maximum number of rows displayed in generated spreadsheet reports
6. Number of decimals in wizard input fields: Number of decimal places displayed in nQuery Predict numeric input fields and Report output fields
7. Default Predict Import Folder: Default folder displayed when importing datasets
8. Open Most Recent Folder?: Open most recently used folder in nQuery Predict workspace when importing datasets
9. Reports Font Family: Font used in nQuery Predict Reports
10. Reports Header Font Size: Font size of headers in nQuery Predict Reports

11. Reports Table Font Size: Font size of values in tables contained in nQuery Predict Reports

### 2.3.4 Per-Table Options

The General Settings options can also be applied on a per-table basis. To achieve this, select the arrow to the left of "Tables" under the "Per-Table Configuration Settings" field found at the end of the Options menu window. Then select the arrow to the right of table code name of interest. The same settings and defaults are assumed as for the Application-wide setting but these can then be edited for that specific table. An example of this is shown in Figure 2.11.

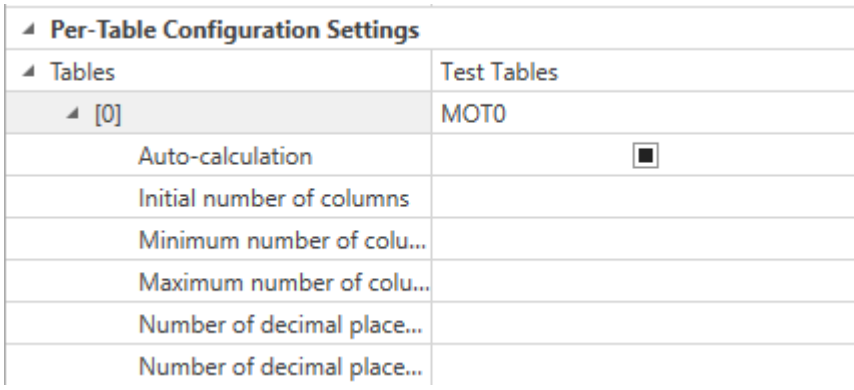



**Figure 2.11:** Per-Table Options Menu Example

## 2.4 Print and Report Options

nQuery provides printing options for the design tables and other table outputs.

### 2.4.1 Print Table

To Print the contents of the table, select either the "Print" option under the File menu or from the tool bar. This will display the default Windows print screen where you can edit the printer used, the printer preferences, the pages printed and number of copies printed.

#### 2.4.1.1 Print Preview

To have nQuery generate a Print Preview, select either "Print Preview" option under the File menu or from the tool bar. This will display a Print Preview screen with an example shown in Figure 2.12.

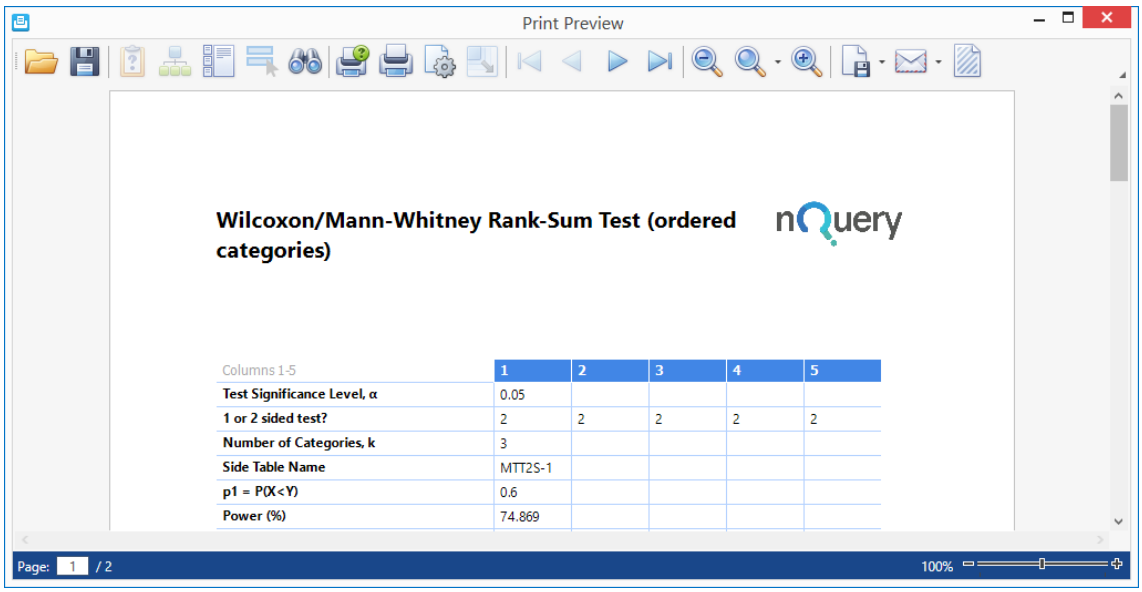



**Figure 2.12:** Print Preview Example

The toolbar at the top of the Print Preview screen gives options to open, save, search, print, quick print, adjust settings, page up, page down, first page, last page, zoom, zoom in, zoom out, export, send or watermark. A hand tool is also available from the right-click context menu.

### 2.4.2 Other Print Options

Print options are available for the following table outputs: Plots and the Output Statement.

To print plots, use the Print option in the right-click context menu used within a plot.

To print the output statement, use the Print button in the bottom-right of the Output window or in the right-click context menu.

### 2.4.3 Report

nQuery provides a report function which allows users to summarise the results for specific design table column in a print-ready format. To open a report for a specific column, select that column by clicking the column name bar above the column. Then right-click the column and select the "Report" option at the bottom of the context menu or select "Report" from the Assistants file menu. This will open a column report which will include a summary of the design parameter values in that column and the output statement for that column if a solver is active. The report

will be opened in Print Preview window and has the same options and functionality described in subsubsection 2.4.1.1.

An example of report is shown in Figure 2.13.

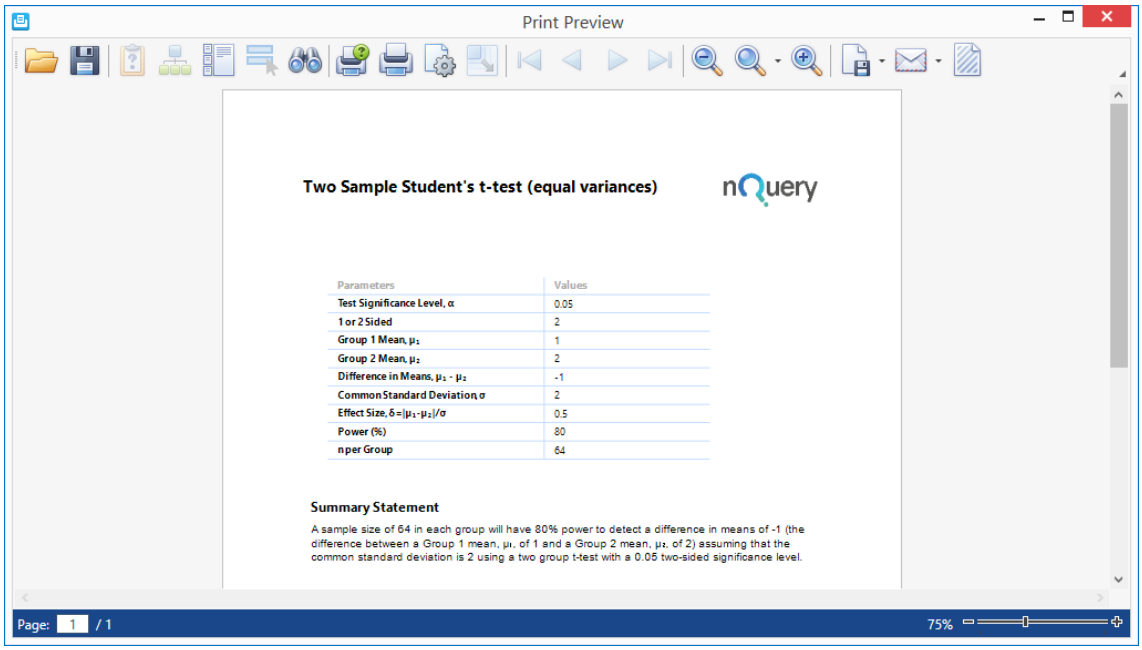



**Figure 2.13:** Report Example

## 2.5 Other Topics

### 2.5.1 Design Table Context Menu and Shortcuts

Within an nQuery design table when a cell or cells is selected, right-clicking will open a context menu for those cells. In nQuery when a cell or multiple cells are selected, the context menu will contain options for Copy, Cut, Paste, Select All, Fill Right, Clear and Copy Table.

Many shortcuts for standard Windows commands (e.g. Ctrl+C for Copy, Ctrl+V for Paste) are used in nQuery. These shortcuts are displayed to the right of the name within the file menus and context menus where appropriate.

If a full column is selected either by dragging over all the cells in a column or by clicking the column title bar, selecting Clear will reset the solver in the column. It also opens a Report option which is described in subsection 2.4.3.

To select multiple entries in a design table, select a cell and then hold down the Shift key (to select all cells between the first and second selected cells) or Ctrl key (to select multiple individual cells) and select another cell.

### 2.5.2 Design Table Drop-downs

In some nQuery design tables, a row is a drop-down menu rather than a numeric input. To select a value from this drop-down, select the relevant cell and click the downwards arrow on the right-hand side of the cell. This will display the full set of options available for that row. The help card will provide a description of the solver drop-down options, their meaning and usually a Suggestion for this option.

Where all options for a drop down are numeric, the numeric input can be inputted directly. One example of this class of drop-down menu is for "1 or 2 Sided Test?" or "1 or 2 Sided Interval?" rows. In these rows, you have the option to either enter 1 or 2 directly into the cell or use the drop-down menu as above.

These rows are treated differently in the Specify Multiple Factors tool and the Plot User Selected rows tool to other rows. In the Specify Multiple Factors tool, these rows will have a drop-down menu similar to design table drop-down. In Plot User Selected rows, these rows will not be available as an option in the X-axis drop-down. To plot the effect of changing the options in these cells, fill multiple columns with all other values fixed and then apply the relevant plotting method to all the relevant columns in the same plot.

### 2.5.3 Customising Table Layout

#### 2.5.3.1 Changing Window Size and Docking

In nQuery, the default size and positioning of each window and element is editable by user to create the layout which best suits their needs. There are two main elements which can be edited: windows and tabs.

**Manipulating nQuery Windows** Windows are the primary user interface unit of an nQuery design table. By default the three windows displayed are the design table window (top-left), the Output/Specify Multiple Factors window (bottom-left) and the Help/Notes window (right). If a plot or plots are created, the Plot window can be manipulated in the same fashion.

There are two main actions that can be applied to a window: docking/undocking and resizing.

- Docking/Undocking

To move a window, click and drag the window bar at the top of the window of interest. This bar will contain the name of the tab currently open in that window (e.g. Help, Output etc.). While the window is being dragged, two options are given for that window: dock the window or undock the window.

To dock the window, drag the cursor while the window is selected to a "docking square" and nQuery will highlight in blue where the window will be docked. Four

edge docking squares will be available on the centre left, right, top and bottom of the nQuery window which will place the current element to the left/right/above/below all other nQuery windows. In addition, the "current" nQuery window over which the cursor is placed will contain the relational docking squares. The centre square will combine the current window and the docking window into a single window with all of their tabs combined. The other squares will place the docking window to the left/right/above/below the current window .

To undock the window, drag the cursor away from the "docking squares" and to the undocked placement of choice.

In both cases, when the desired placement is achieved let go of the left mouse button (or other "clicking" method) and the window will appear in the desired place. An example of the docking process "in action" for the Output window where it is being placed between the MTT0 Design Table and Help windows is given in Figure 2.14. The effect of this placement (after resizing) is shown in Figure 2.15.

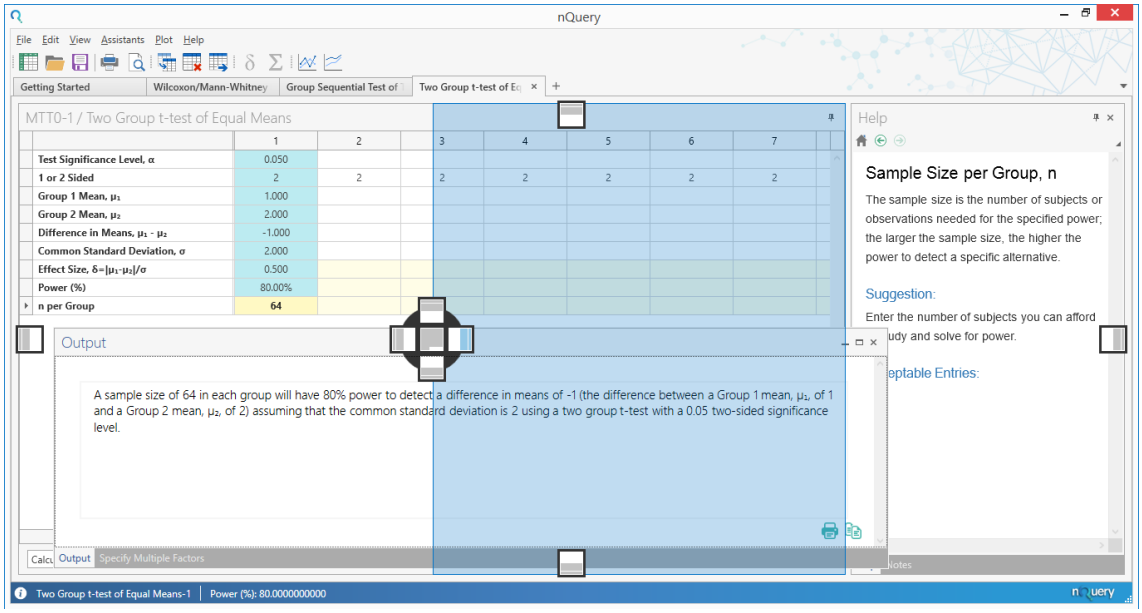



**Figure 2.14:** nQuery Window Docking Example

- Resizing

To resize a window, place the cursor at the edge of the nQuery window of interest. The mouse cursor will change to a ↔ symbol. Then hold down the left mouse button and drag the cursor left/right (for a vertical edge) or up/down (for a horizontal edge) until the window is the desired size. Note that elements in the table will dynamically update to show how they will look for a given window size. An example of the output statement being resized after the docking example above is shown in Figure 2.15.

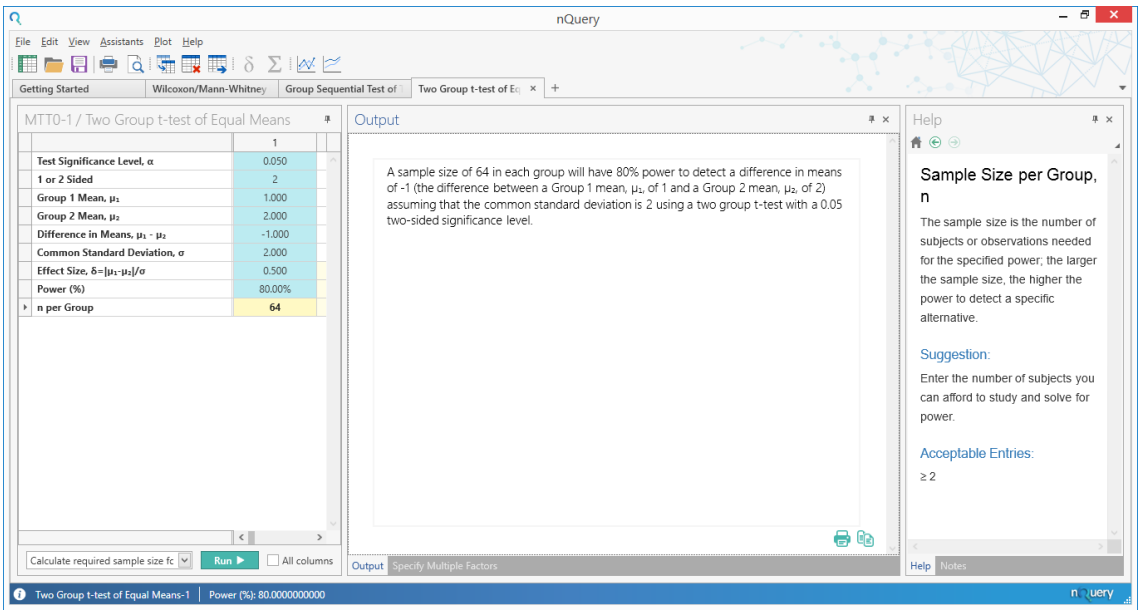



**Figure 2.15:** nQuery Window Resizing Example

**Manipulating Window Tabs** Window tabs have a single applicable operation: docking/undocking. This process works similarly to undocking/docking a window. For tabs, the window element which is clicked and dragged is the tab bar name rather than the window name. The tab names are found in the grey tab bar at the bottom of an nQuery window. When a tab is being dragged, the same options are available as for a window. The three main actions are to undock the tab, place the tab relative to the other windows or dock the tab into another window.

To undock, drag the select tab away from any of the docking squares and leave. When a tab is undocked, it will become an nQuery window.

To place relative to the other windows, select on of the left/right/up/down docking squares and place in the desired position which will be previewed in blue by hovering over a docking square while dragging.

To dock the tab into another window, drag the cursor over the window of interest and select the centre docking square in the middle of the current window.

An example where the Output tab is currently being moved and where the Notes tab has been moved into the main design table window is shown in Figure 2.16.

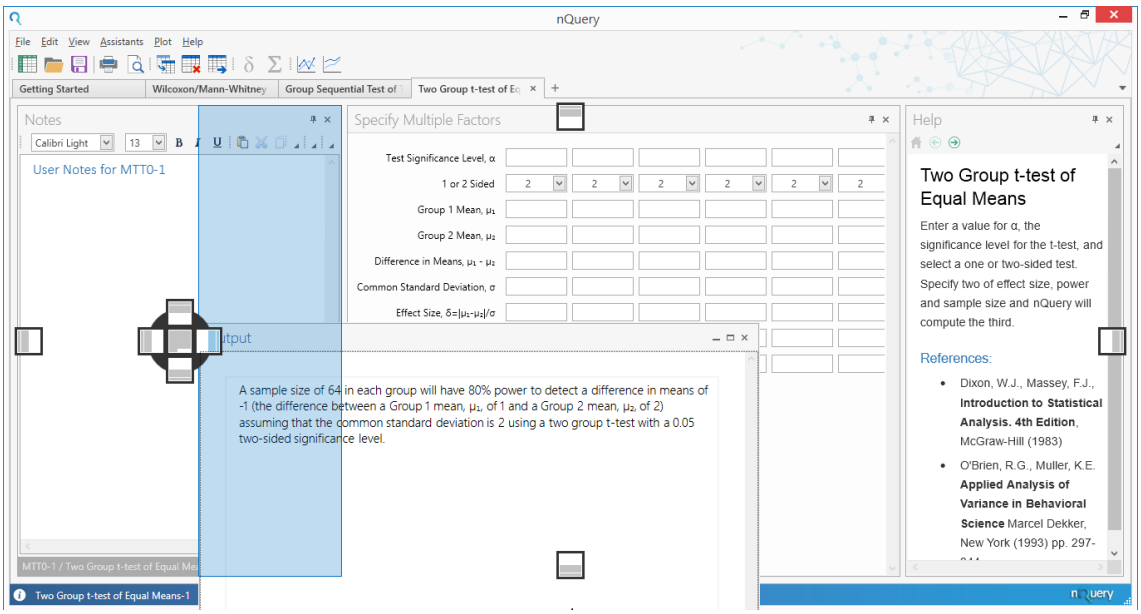



**Figure 2.16:** nQuery Tab Docking Example

Note that you can also drag tab names within the tab bar to place it relative to the other tab names in the same click and drag fashion.

The user can also close or open table elements by selecting these from the View file menu.

#### 2.5.3.2 Editing Column Names

nQuery provides the user the ability to change a column name. To change a column name, double-click the number in the column title bar above the column cells. The number will appear in a white square and will now be editable. Replace the number with the desired column name and this name will then be displayed in the design table. An example of a design table with changed column names was shown in Figure 2.2.

Note that these column names will be reflected automatically in the Legend of plots involving that include the renamed columns.

# 3 Advanced nQuery Assistants

## 3.1 Design Table Side-Tables

Side-tables provide table specific tools which allow users to calculate design parameters based on additional information which is commonly known or used in a particular design type. This section will go over the three primary types of side-tables: Effect Size Side-Tables, Covariance Matrix Side-Tables and Mandatory Side-Tables

### 3.1.1 Compute Effect Size Side-Table

For an introduction in how to open and use effect size side-tables, see subsection 1.7.4. In this section, we will briefly summarise the process of opening the side-table, using the side-table and cover some additional information not included in subsection 1.7.4.

#### 3.1.1.1 Opening an Effect Size Side-Table

There are three main routes to open an effect size side-table: using the "Compute Effect Size" option in the Assistants menu, using the Compute Effect Size button in the menu bar or by selecting the cell(s) in design table which are transferred from that side-table. Column cells which are the transfer target for an Effect Size side-table will be indicated in the cell's Help Card by its "Aid" section.

When a side-table is opened for a column, it will appear below the main analysis table in the same window as the Output and Specify Multiple Factors tools.

For each column in which a side-table is opened, a unique side-table will open associated with that selected column. The names of these side-tables are found in the tab bar at the bottom of the side-table window.

The most common template for a side-table name is [Table Code]S-[Column Number]. For example, opening a side-table in MGT0 (One Way Analysis of Variance) for column one will open a side-table named "MGT0S-1". Note there are some exceptions e.g. Looks-[Column Number] for Group Sequential Design.

Note that users can edit and transfer from other column side-tables into that column by selecting them in the tab menu at the bottom of the side-table window.

#### 3.1.1.2 Using an Effect Size Side-Table

For a specific example of an effect size being used, see subsection 1.7.4. The format for an effect size side-table will be unique depending on the design table being used. For guidance on using a specific side-table refer to the help materials, in particular the Home Card and Help Card "Aid" sections of relevant cells.

In general, the effect size side-table works by the user fully specifying the required cells in the table.

When the minimum amount of information required in a side-table is completed, the "Compute" button in the upper-left of the side-table window will become active (ungreyed). Selecting the Compute button will calculate the design parameter(s) which will be transferred to the main table and other parameters of interest in certain design tables.

After the Compute button is used, the "Transfer" button (right of the Compute button) will become active and when selected this will transfer the relevant design parameters into the associated column in the main design table.

Note that some tables will not contain a "Compute" and/or "Transfer" button. Where the Compute button is missing, nQuery will automatically calculate the required outputs in the side-table. Where the Transfer button is missing, nQuery will automatically transfer the required side-table outputs to main table when the side-table is completed.

To return the side-table to its original state, use the "Clear" button (to the right of the Transfer button) to remove all user entries, read-only entries and computed values from the side-table.

Using the right-click context menu gives access to same edit tools and shortcuts as the main table (see subsection 2.5.1).

An example of an effect size side-table is shown in Figure 3.1.

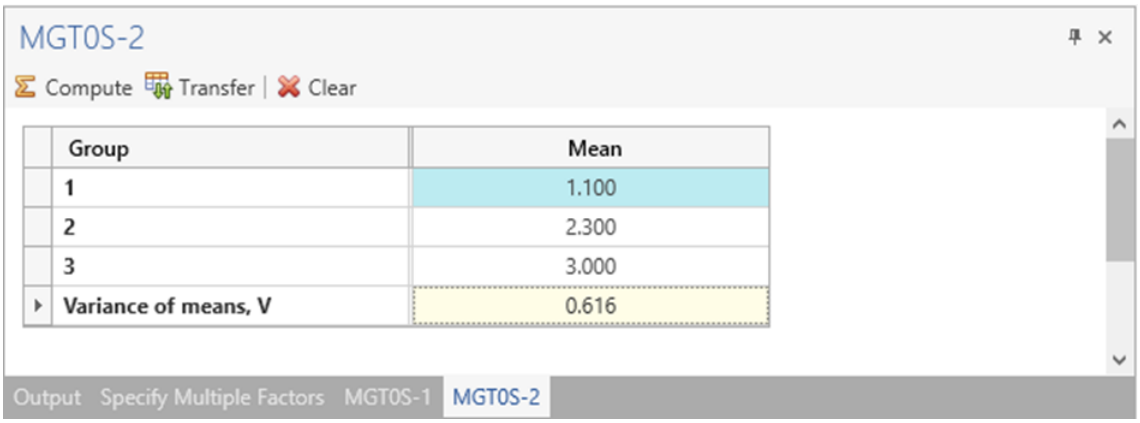



**Figure 3.1:** Effect Size Side-Table Example

#### 3.1.1.3 Closing an Effect Size Side-Table

To close a specific effect size side-table, use the "x" in the top-right of the effect size side-table window. You can also hide the side-table using the pin symbol in the top-right or by right-clicking on the tab menu name.

### 3.1.2 Specify Covariance Matrix Side-Table

The Specify Covariance Matrix side-table provides a convenient method to calculate the error and Greenhouse-Geisser correction terms for the Repeated Measures ANOVA tables which use the Greenhouse-Geisser approximation in nQuery. These are MOT4 and MTT3.

Note that a small number of other tables use this option as generic way to open a second side-table where required.

#### 3.1.2.1 Opening an Effect Size Side-Table

There are three main routes to open an Specify Covariance Matrix side-table: using the "Specify Covariance Matrix" option in the Assistants menu, using the Specify Covariance Matrix button in the menu bar or by selecting the cell(s) in design table which are transferred from the side-table. These are the "error-term", sphericity and bias term rows. This will be indicated in the relevant cell's Help Card by its "Aid" section.

As per the effect size side-tables, when a covariance matrix side-table is opened for a column, it will appear below the main analysis table in the same window as the Output and Specify Multiple Factors tools.

For each column in which a side-table is opened, a unique covariance matrix side-table will open associated with that selected column. The names of these side-tables are found in the tab bar at the bottom of the side-table window. The template for a side-table name is [Table Code]C-[Column Number]. For example, opening a covariance matrix side-table in MOT4 (One Way Repeated Measures ANOVA (Greenhouse Geisser Approximation) for column one will open a side-table named "MOT4C-1".

Note that users can edit and transfer from other column side-tables (effect size or covariance matrix) into that column by selecting them in the tab menu at the bottom of the side-table window.

#### 3.1.2.2 Using a Specify Covariance Matrix Side-Table

There are two modes for filling the Specify Covariance Matrix side-table: the Specify Standard Deviations and Correlations mode and the Specify Full Covariance Matrix

mode. To switch between the modes, use arrow on the right of the mode drop-down to the right of the "Transfer" button at the top of the side-table window. This drop-down is shown in Figure 3.2.

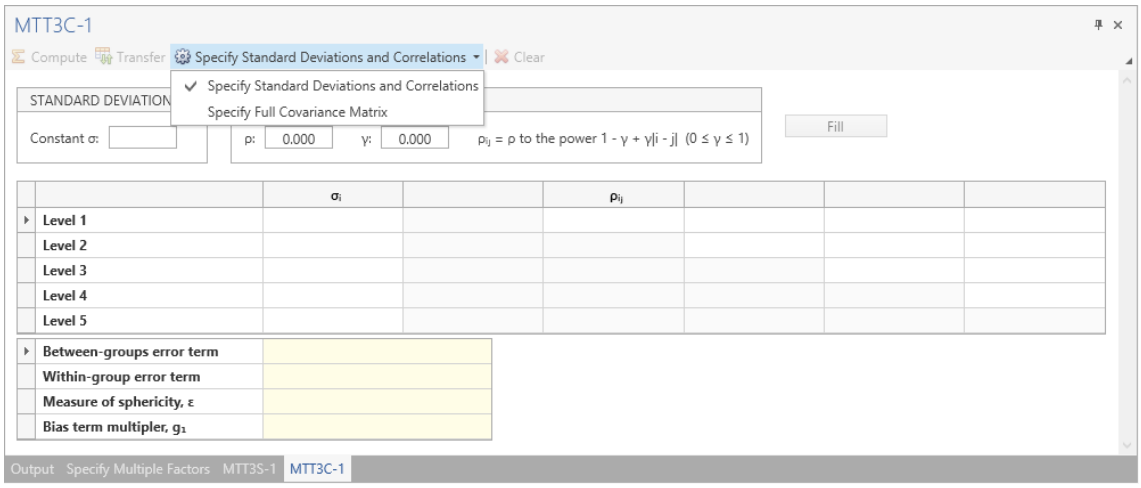



**Figure 3.2:** Specify Covariance Matrix Mode Drop-down Example

**Specify Standard Deviations and Correlations Mode** The Specify Standard Deviations and Correlations mode has two primary methods to activate the side-table: using the Constant Standard Deviations and Correlation Pattern shortcuts or directly filling the $\sigma_i$ and $\rho_{ij}$ table entries.

Note that changes made in one mode will affect the table values given in the other mode.

### 1) Constant Standard Deviations and Correlation Pattern

To use the Constant Standard Deviations and Correlation Patterns shortcuts you need to fill three inputs: the common standard deviation $\sigma$, the correlation $\rho$ and the pattern term $\gamma$.

The common standard deviation value will fill $\sigma_i$ in the table below with the same common standard deviation values. This value must be greater than zero.

The correlation will be the value included in the first diagonal of the $\rho_{ij}$ table below and will adjusted downwards by the pattern term as you rightwards in the $\rho_{ij}$ table. This value must be between 0 and 1

The pattern term defines the adjustment applied to the correlation as you move right-wards within the $\rho_{ij}$ table. The formula for this adjustment is given the table and is $1 - \gamma + \gamma|i - j|$ where $|i - j|$ is the absolute distance between the first cell in the row in the $\rho_{ij}$ and the cell of interest. This value must be between 0 and 1.

The common standard deviation will assign a common standard for the measurements at each measurement level and the correlation and pattern terms will represent

the correlation between measurements taken on a subject at a given "measurement distance". For example, the 2nd column cell in row 1 of the $\rho_{ij}$ table represents the average correlation between the 1st and 3rd measurement for a subject.

To transfer the relevant $\sigma_i$ and $\rho_{ij}$ values for the specified values given in the Shortcuts menu, select the "Fill" button to the right of the shortcuts menus. This will activate the Compute button which will calculate the relevant error-terms, bias and sphericity and activate the Transfer button. Selecting Transfer will place these values in the associated main design table column.

To use a compound symmetry (CS) model, set the constant correlation to the desired value and set the pattern term to zero. In the CS matrix, the correlation between measurements is assumed to be constant regardless of the distance between measurements.

To use an autoregressive(1) (AR(1)) model, set the correlation at the desired value (this will be the "distance 1" correlation) and set the pattern term to one. In the AR(1) matrix, the correlation between measurements is assumed to decay via an AR(1) process as the distance between measurement increases.

To achieve a pattern where the correlation does decay but slower than the AR(1) model, set the correlation at the desired value (this will be the "distance 1" correlation) and set the pattern term to between zero and one.

An example of using the AR(1) model computed using the shortcuts method is shown in Figure 3.3.

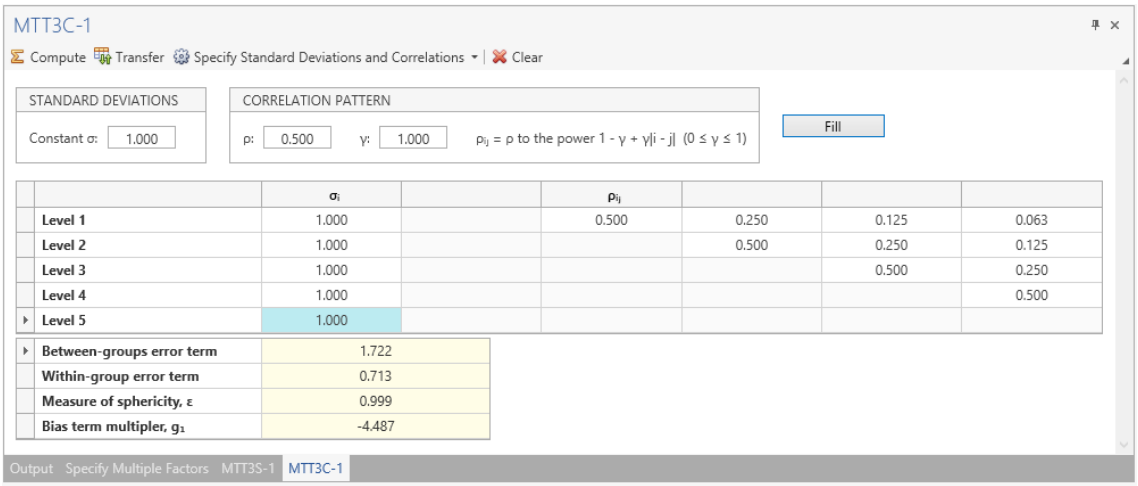



**Figure 3.3:** Specify Covariance Matrix SD + Correlation Example

### 2) Directly Filling Table

In addition to using the shortcuts method, you can fill the $\sigma_i$ and $\rho_{ij}$ in the table directly. This can be done with a blank table or the values can be edited from

a shortcuts generated template. This gives the user a large amount of flexibility to set the table to their specifications. Once all $\sigma_i$ and $\rho_{ij}$ cells are filled, this will activate the Compute button which will calculate the relevant error-terms, bias and sphericity and activate the Transfer button. Selecting Transfer will place these values in the associated main design table column.

An example where we edited from the Figure 3.3 template to have the individual level standard deviations vary and the distance 4 (i.e. between measurement 1 and 5) correlation equal to zero is shown in Figure 3.4.

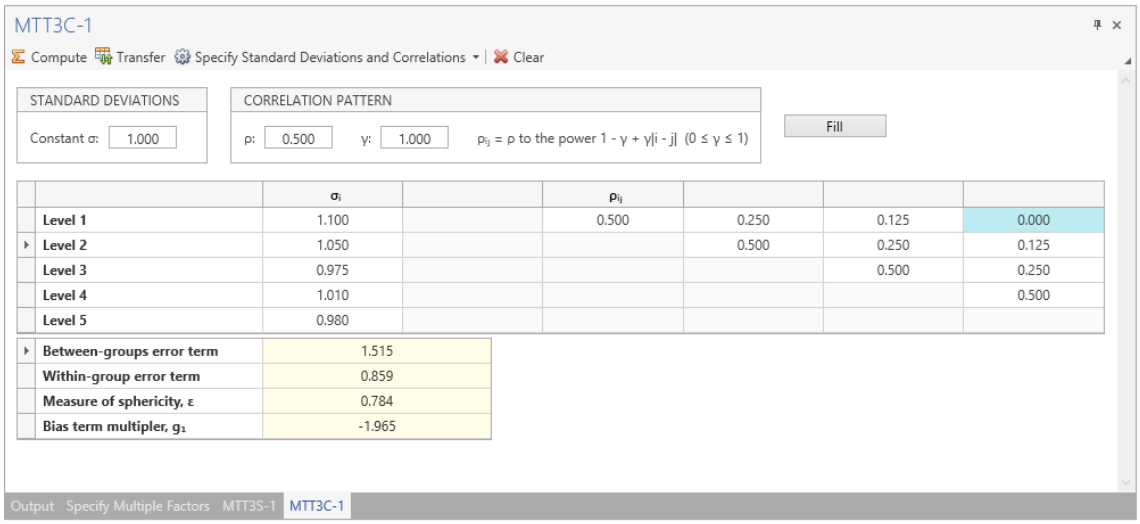



**Figure 3.4:** Specify Covariance Matrix Manual SD + Correlation Example

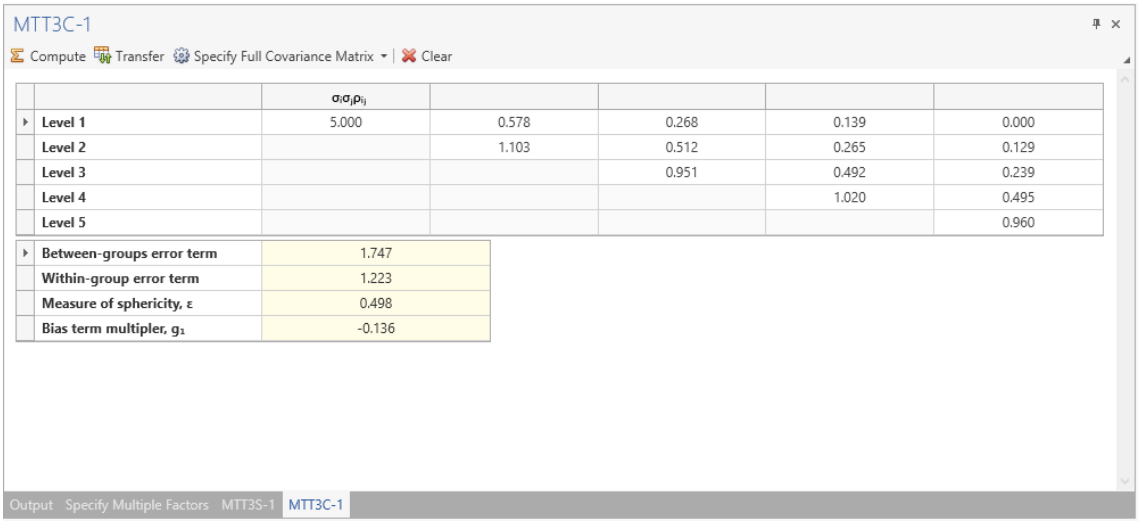
**Specify Full Covariance Matrix mode** The Specify Full Covariance mode works similarly to directly filling the standard deviations and correlations table. However, the covariance matrix is entered directly in this case. In the covariance matrix, the main diagonal elements are equal to within-measurement level variance and off-diagonal elements represent the covariance between main diagonal element level and the measurement at that distance (e.g. the third column entry in row one equals the average covariance between the first level subject measurement and the third level subject measurement).

There are two main methods to fill the covariance matrix: automatically fill using Constant Standard Deviations and Correlation Pattern mode or manually fill.

As noted at the start of this section, if the standard deviations and correlations table has been filled (either using the shortcuts menu or manually) in the other mode then the Covariance Matrix will be filled in this mode. Alternatively, you can fill or edit the covariance matrix manually to the study's specifications. Once all elements are filled in the covariance matrix, this will activate the Compute button which will calculate the relevant error-terms, bias and sphericity and activate the

Transfer button. Selecting Transfer will place these values in the associated main design table column.

An example of an Specify Covariance Matrix side-table computed using the Specify Full Covariance Matrix mode is shown in Figure 3.5.




| | $\sigma_i\sigma_j\rho_{ij}$ | | | | |
|---|---|---|---|---|---|
| Level 1 | 5.000 | 0.578 | 0.268 | 0.139 | 0.000 |
| Level 2 | | 1.103 | 0.512 | 0.265 | 0.129 |
| Level 3 | | | 0.951 | 0.492 | 0.239 |
| Level 4 | | | | 1.020 | 0.495 |
| Level 5 | | | | | 0.960 |

| | |
|---|---|
| Between-groups error term | 1.747 |
| Within-group error term | 1.223 |
| Measure of sphericity, ε | 0.498 |
| Bias term multipler, $g_1$ | -0.136 |

**Figure 3.5:** Specify Covariance Matrix Matrix Example

#### 3.1.2.3 Closing a Covariance Matrix Side-Table

To close a specific covariance matrix side-table, use the "x" in the top-right of the effect size side-table window. You can also hide the side-table using the pin symbol in the top-right or by right-clicking on the tab menu name.

### 3.1.3 Mandatory Side-Tables

In the majority of nQuery design tables which use side-tables, the side-table is optional. In these cases, the role of the side-table is to provide a method for users to derive required design parameters based on other information they may have access to.

In a minority of nQuery design tables, the side-table is mandatory. In these cases, the inputs in the side-table(s) are used directly by the solvers of that table. The following is a list of examples where the table has a mandatory side-table(s) (Table Codes in Brackets):

- Group Sequential Tests (Means: MOT26, MTT12, MTE32; Incidence Rates/Counts: MTT40, MTT42; Proportions: POT8, POT13, PTT12; Survival: STT12, STT15, STT23, STT25) (Discussed in chapter 4)

- Multiple Comparison Procedure - Modeling (MCP-Mod) (MGT5, MGT7, MGT8, MGT9, PGT3) (Discussed in chapter 5)
- Multi-Arm Multi-Stage (MAMS) Group Sequential Tests (MGT6, PGT4)
- Bayesian Assurance using Custom Prior (MOT25, MTT21, POT7, PTT13, STT14)
- Wilcoxon/Mann-Whitney Rank-Sum Test (Ordered Categories) (MTT2/MTT2U)
- Multivariate Analysis of Variance (MANOVA) (MGT3)
- Mantel-Haenszel (Cochran) Test of OR=1 in S Strata (PTT4/PTT4CC)
- Log-Rank Test, User Specified Survival Rates, Accrual, Dropouts (STT22/STT24)
- Log-Rank Test, User Specified Survival Rates, Accrual, Dropouts (Simulation) (STT3/STT3U)
- Maximum Combination Linear Rank Tests (STT33)
- Two Sample Stratified Log-Rank Test (STT31)
- Stepped-Wedge Cluster Randomized Designs (CRT16, CRT17, CRT18)
- Poisson Regression *(Rate Ratio solvers only)* (ROT7)

From this list, all except MGT3 and the Group Sequential (see chapter 4) tables, use the same instructions as for the Effect Size side-tables above to open the side-table. In each of these design tables except MGT3 and the group sequential tables, there is a row named "Side Table Name" which will also open the side-table for a column automatically (assuming all side-table opening conditions are met).

In MGT3, there are three separate side-tables per column and these are opened by selecting the Factor Level Table, Means Matrix and Covariance Matrix read-only cells in a column. In the group sequential tables, selecting any cell in a column will open the Looks side-table.

The side-tables are the same between MTT2/MTT2U, PTT4/PTT4cc, STT3/STT3U, the Bayesian Assurance custom prior tables (MOT25/POT7 etc.) and the group sequential tables (MTT12/PTT12 etc.). These side-tables operate the same as optional side-tables except they do not have Compute/Transfer buttons. For mandatory side-tables when the table is filled sufficiently to be used in a solver, the name of the side-table in "Side Table Name" row (or MGT3 equivalent) will go from grey to black. The table otherwise will operate as per any other design table.

Note that Looks side-table for the group sequential tables does not require any editing to be active and is by default read-only. It is only required to be edited if the Information Times or Spending Function rows are set to "User Specified". See the help for the group sequential details for further details.

If a side-table has been completed in any column, you can automatically use that side-table in another column by selecting the relevant side-table name from the drop-down options in that column's "Side Table Name" row.

An example of a mandatory side-table being used in MTT2 is shown in Figure 3.6.

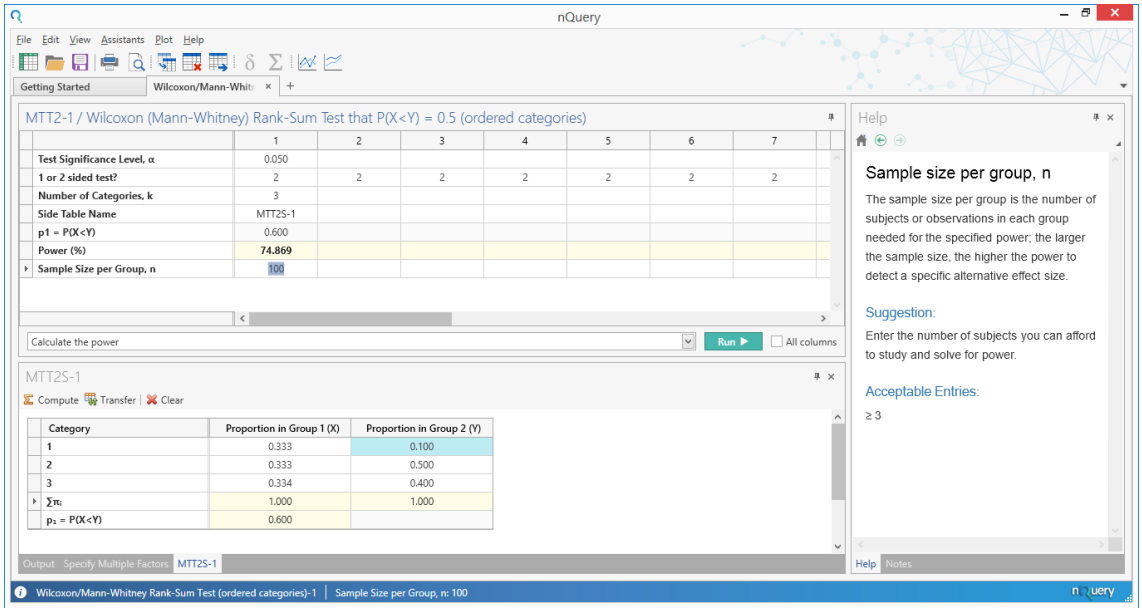



**Figure 3.6:** MTT2 Mandatory Side-Table Example

## 3.2 Compute Standard Deviation Assistants

nQuery provides a number of assistants which allow a user to derive the standard deviation for a study based on other parameters or data which the user has.

To access these assistants, select the "Standard Deviation" option from the Assistants file menu. This will open a radio button list menu containing the standard deviation assistants. Select an assistant radio button and select OK to open the relevant assistant.

There are 10 standard deviations (SD) assistants available in nQuery. These are (with a brief description):

1. From Standard Error: Derive SD from sample size (n) and standard error
2. For SD1 and SD2 (pooled SD): Derive pooled SD from group 1 and 2 standard deviations (SD1/SD2) and sample sizes (n1/n2)
3. From Range: Derive SD from range (using maximum and minimum optionally to derive range) and sample size, n
4. From Percentile: Derive the SD from the percentile and the difference in upper and lower percentile values (using upper and lower values optionally).
5. From Coefficient of Variation: Derive the log-scale SD and mean from the coefficient of variation and observed mean respectively

6. From Upper Confidence Limit: Derive the upper limit of the SD from the sample size, observed standard deviation and confidence level.
7. From SD1, SD2, Correlation: Derive the SD for the differences from the first condition SD, the second condition SD and the correlation
8. For Cluster Sampling: Derive the "cluster" SD from the between-cluster and within-cluster variances, the intra-cluster correlation and subjects per cluster.
9. For specified x values: Derive the SD from a set of x data values. See section 3.3
10. Of residuals: Derive the SD of the residuals and the dependent (Y) variable from the regression coefficient, correlation coefficient and independent (X) variable SD.

All of these assistants work in the same or similar way as nQuery design tables except for the "For specified x values" option. Use the guide given in chapter 1 and the individual Home and help cards provided for guidance on how to use each table.

Options 1 - 4 provide alternative methods to derive a standard deviation for one or two sample independent means test and Option 6 provides the alternative estimate for the standard deviation by using the upper confidence limit . Option 5, 7, 8 and 10 provide help to derive the adjusted standard deviation for specific designs or testing methods. Option 5 would be used if the analysis will be of the log-transformed data, Option 7 would be used for paired means testing or analysis, Option 8 would be used if the study will use cluster randomisation and Option 10 is used if the analysis is using linear regression.

## 3.3 Data Entry Assistant

nQuery provides a data entry assistant to allow users to derive commonly used design parameters based on data which may be available to the user.

The Data Entry Assistant has three main components: the Data Entry table, the Transformation menu and the Insert Values menu.

### 3.3.1 Data Entry Table

The data entry table is the main table on the left-hand side of the Data Entry Assistant. It consists of 20 columns and 500 rows. Each column corresponds to a data set and each row corresponds to an individual subject's data value in that column's data set.

As a user fills a column, four commonly used design parameters will be derived for the data set. These are the sample size (N), the mean, the sample standard deviation

and population standard deviation, $\sigma(x)$. The difference between the sample and populations standard deviations is the denominator in the estimate for each is N-1 and N respectively.

### 3.3.2 Transformations Menu

The transformations menu gives the option to create a new column containing the transformed values for an entered data set. Transformations are commonly used when data is non-normal but has a distribution which would be amenable to being transformed into a normally distributed data using a common transformation.

The Transformations menu is found to the right of the Data Entry table and can be opened by selecting the Transformations tab at the top of the menu window.

The three transformations are a square root transformation (Sqrt(X)), a base 10 log transformation (log10(X)) and the natural log transformation (ln(X)). In addition, you can conduct these transformations to the data with a fixed value (A) added to each data value before transformation. In this case, the user is required to enter a value of A in the menu.

To use the transformations menu, select the relevant column in the data entry table, select the transformation required, enter a value for A if a (X+A) transformation is selected and select "Add".

This will generate the transformed data (with summary statistics) in the next empty column to the right of the selected data entry column. The column title will include the name of the transformation and the column transformed and any transformed columns will be highlighted in orange.

### 3.3.3 Insert Values Menu

The insert values menu gives the option to create a number column using an evenly spaced number of data values in a specified range. This can be useful to generate a large dataset easily.

The insert values menu is found to the right of the Data Entry table and can be opened by selecting the Insert Values tab at the top of the menu window.

To generate a column using the insert values menu, specify a From value and To value in the relevant cells. These correspond to the minimum and maximum values of the range of interest respectively.

The user can then activate either the Specify Increment or Specify num. values option by selecting the radio button to the left of each option.

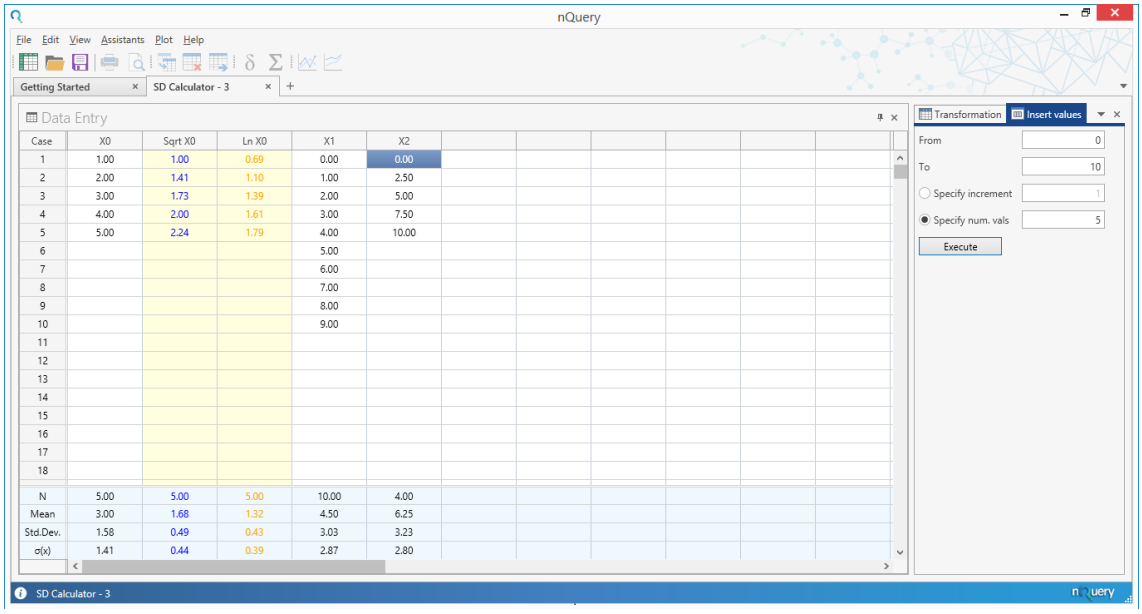
If the Specify Increment option is activated, the entered value will be increments in which the data will increase going from the "From" to the "To" value. For example,

if From = 0 and To = 10 then Specify Increment = 2 will give the five values of [0, 2, 4, 6, 8].

If the Specify num. values option is activated, the entered value will be the number of data values which will be taken between the "From" and "To" value. For example, if From = 0 and To = 10 then Specify num. values = 5 will give the five values of [0, 2.5, 5, 7.5, 10].

Once the required values are entered for "From", "To" and "Specify Increment"/"Specify num. values", select "Execute" and that data will be generated on the currently selected Data Entry Table cell.

An example of Data Entry table using the above options is shown in Figure 3.7.




| Case | X0 | Sqrt X0 | Ln X0 | X1 | X2 |
|---|---|---|---|---|---|
| 1 | 1.00 | 1.00 | 0.69 | 0.00 | 0.00 |
| 2 | 2.00 | 1.41 | 1.10 | 1.00 | 2.50 |
| 3 | 3.00 | 1.73 | 1.39 | 2.00 | 5.00 |
| 4 | 4.00 | 2.00 | 1.61 | 3.00 | 7.50 |
| 5 | 5.00 | 2.24 | 1.79 | 4.00 | 10.00 |
| 6 | | | | 5.00 | |
| 7 | | | | 6.00 | |
| 8 | | | | 7.00 | |
| 9 | | | | 8.00 | |
| 10 | | | | 9.00 | |
| 11 | | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | | | |
| 17 | | | | | |
| 18 | | | | | |
| N | 5.00 | 5.00 | 5.00 | 10.00 | 4.00 |
| Mean | 3.00 | 1.68 | 1.32 | 4.50 | 6.25 |
| Std.Dev. | 1.58 | 0.49 | 0.43 | 3.03 | 3.23 |
| σ(x) | 1.41 | 0.44 | 0.39 | 2.87 | 2.80 |

**Figure 3.7:** Data Entry Assistant Example

In column 1 is a user entered set of values (X0), in column 2 is Sqrt(X) transformed data for column 1, in column 3 is ln(X+1) (i.e. A=1 with natural log transform) transformed data for column, in column 4 is Insert Values menu data generated with From = 0, To = 10, Specify Increment = 1 and in column 5 is Insert Values menu data generated with From = 0, To = 10, Specify num. values = 5.

## 3.4 Cumulative Distribution Function Assistants

nQuery provides access to the most commonly used cumulative and inverse cumulative distribution functions used by the nQuery solver algorithms. To open a table for a specific statistical distribution, select Distribution Function from the Assistants

file menu. This will open a radio button menu of the available statistical distributions. To open a distribution table, select the relevant radio button to the left of the distribution name and select "OK".

There are distribution assistants provided for the following statistical distributions:

1. z (Gaussian/Normal)
2. t (Central)
3. t (Non-central)
4. $\chi^2$(Central) *(Chi-Square)*
5. $\chi^2$(Non-central) *(Chi-Square)*
6. F (Central)
7. F (Non-central)
8. Cumulative binomial

The distribution tables work the same as standard nQuery design tables with the solvers being for the cumulative probability of P(X<x_p) (all tables), the percentile test statistic (e.g. F_p, t_p) (all tables except cumulative binomial) and the non-centrality parameter (non-central tables only).

## 3.5 Survival Parameter Conversion Assistant

The survival parameter conversion assistant provides a tool to convert between the three most commonly used survival parameters for an exponential survival curve.

STT00-1 / Conversion to alternate rates for exponential survival curves

| | 1 | 2 | 3 |
|---|---|---|---|
| **Time t** | | | |
| **Group 1 Proportion $\pi_1$ at Time t** | | | |
| **Group 1 Median Survival, $med_1$** | | | |
| **Group 1 Exponential Parameter, $\lambda_1$** | | | |
| **Group 2 Proportion $\pi_2$ at Time t** | | | |
| **Group 2 Median Survival, $med_2$** | | | |
| **Group 2 Exponential Parameter, $\lambda_2$** | | | |
| **Hazard Ratio, $h=\ln(\pi_1)/\ln(\pi_2)=med_2/med$** | | | |

**Figure 3.8:** Survival Parameter Converter

The three parameterizations available are the proportion surviving at time t, the median survival and the exponential parameter. In the survival parameter conversion assistant, the three survival parameters can be specified for a two group design individually for each group. In addition, there are rows for the Time t and the Hazard ratio. The Time t is required to calculate the proportion surviving at time t from the median survival or exponential parameter and vice-versa. The hazard ratio is calculated after the median survival/exponential parameter are specified in each group.

Note that the exponential parameter will automatically calculate the median survival when specified and vice-versa in each group.

Many nQuery survival tables (e.g. STT2, STT6) contain an Effect Size side-table based on this assistant but where this is not present the Survival Parameter Conversion Assistant is available.

## 3.6 Posterior Error Rate Calculator Assistant

The Posterior Error Rate Calculator is an assistant which allows conversion between frequentist Type I (Significance Level) and II (1 - Power) errors and Bayesian False Positive ($\beta^*$) and False Negative ($\alpha^*$) errors given a prior probability against the null hypothesis, H0. This is based on theory from [Lee and Zelen, 2000]. The Posterior Error Rate Calculator is shown in Figure 3.9.

BOT0-1 / Posterior Error Rate Calculator

| | 1 | 2 | 3 |
|---|---|---|---|
| **Posterior False Negative Error, α*** | | | |
| **Posterior False Positive Error, β*** | | | |
| **Prior Probability against $H_0$, θ** | | | |
| **Significance Level, α** | | | |
| **Power (%), 1 - β** | | | |

**Figure 3.9:** Posterior Error Rate Calculator

The Posterior Error Rate Calculator can be used to calculate any two of the Posterior False Negative Error, Posterior False Positive Error, Prior Probability against $H_0$, Significance Level or Power given all three other inputs are given in a column.

This completes the general user guidance for using nQuery for sample size and power calculations. What follows is a series of chapters dedicated to specialized topics where additional guidance and background will assist in using the referenced features and design tables.

*Note: Posterior Error Rate Calculator Assistant is currently only available for users with an nQuery Advanced Plus license. Active packages are displayed in the Packages section in the Home tab. To purchase additional packages, see* `www.statsols.com`.

# 4 Group Sequential Designs, Interim Monitoring & Unblinded Sample Size Re-estimation

## 4.1 Introduction

nQuery is designed for the calculation of power and sample size for both fixed term and group sequential design studies. nQuery will have options available for all users for planning a group sequential design for two means, two proportions and two survival curves (assuming equal follow-up).

nQuery Adapt (nQuery Advanced Pro license required) users will have access to additional group sequential designs and an interim monitoring and unblinded sample size re-estimation tool for the most commonly used group sequential design. This section will cover the basic theory behind these concepts and will give detailed description of how to design a group sequential trial in nQuery and use the nQuery Adapt tool for interim monitoring and unblinded sample size re-estimation.

Much of theory and guidance here is also relevant for the Multi-arm Multi-Stages (MAMS) Group Sequential Design tables (MGT6, PGT4).

## 4.2 Group Sequential Test Design

### 4.2.1 Background

Group sequential designs are an extension of fixed period designs in which data from the trial is analyzed at one or more stages prior to the conclusion of the trial. The trial can then be stopped early if there is strong evidence for or against the proposed treatment being effective based on data up to that point. Group sequential designs are one of the most widely used types of adaptive trial and provide the opportunity to stop a trial early for efficacy (strong interim evidence against the null hypothesis) and/or futility (strong interim evidence for the null hypothesis) and thus significantly reduce the economic and ethical costs over the equivalent fixed period design.

*Note: Group Sequential Design is currently only available for users with an nQuery Advanced Pro license except for the Group Sequential Design tables for Two Means*

*(MTT12), Two Proportions (PTT12) and Two Survival - Equal Followup (STT12) which are in nQuery Advanced (Core). Active packages are displayed in the Packages section in the Home tab. To purchase additional packages, see* `www.statsols.com`.

### 4.2.2 Group Sequential Design Theory

In nQuery, the group sequential design tables' power and sample size calculations are performed using the Lan-DeMets alpha spending function approach [Demets and Lan, 1984, Demets and Lan, 1994]; for estimating boundary values. Building on the work of Lan and DeMets; Pampallona, Tsiatis, and Kim [Pampallona et al., 1995, Pampallona et al., 2001] later put forward the concept of using a beta spending approach to construct boundaries for futility where the evidence for the null is strong. These boundary values indicate the interim test values (test statistic, effect size, p-value) which would lead to the trial stopping early based on interim data. Boundary values can be estimated in a number of ways with nQuery providing bounds based on the O'Brien-Fleming [O'Brien and Fleming, 1979], Pocock [Pocock, 1977], Hwang-Shih-DeCani [Hwang et al., 1990] and Power Family spending functions. Each of these spending functions spend a certain proportion of the alpha error (Type I/efficacy) and/or beta error (Type II/futility) at each analysis or 'look' and then make the needed adjustments to the sample size and final errors to preserve the overall Type I and Type-2 errors. The "spent" alpha and beta values used at each look are calculated based upon the test hypothesis, the spending function chosen, the number of looks to be taken during the course of the study as well as the overall Type I and Type-2 error rates. For a full introduction to group sequential methods, we recommend *Group Sequential Methods Applications to Clinical Trials* by Jennison & Turnbull (2000) [Jennison and Turnbull, 1999].

#### 4.2.2.1 Spending Functions

There are four spending functions available to the user in nQuery for the efficacy and futility bounds. Note that nQuery also provides the option to manually input boundary values. As standard all alpha spending functions have the properties that the error spent at the start of the trial equals zero and the error spent at the final analysis equals the original desired error level i.e. the desired test significance level for alpha spending, one minus the power (as a proportion) for beta spending. Functionally the alpha and beta spending functions are the same.

The spending functions for alpha spending are summarised in Table 4.1 where $\alpha(\tau)$ is the cumulative alpha spent at the specified look, $\alpha$ is overall alpha error, $\tau$ is the information time (usually the sample size up to that look as a proportion of the total sample size ), $z_{1\text{-}\alpha/2}$ is the inverse standard normal cumulative distribution assessed at "1-$\alpha$/2", $\Phi$ is the standard normal cumulative distribution function, $\rho$ is the Power Family parameter and $\gamma$ is the Hwang-Shih-DeCani gamma parameter.

**Table 4.1:** Spending Functions

| Spending Function | Form |
|---|---|
| O'Brien-Fleming | $\alpha(\tau) = 2\left(1 - \Phi\left(\frac{z_{1-\alpha/2}}{\sqrt{\tau}}\right)\right)$ |
| Pocock | $\alpha(\tau) = \alpha ln(1 + (e-1)\tau)$ |
| Power Family | $\alpha(\tau) = \alpha\tau^{\rho}$ |
| Hwang-Shih-DeCani | $\alpha(\tau) = \alpha\left[\frac{\left(1-e^{-\gamma\tau}\right)}{(1-e^{-\gamma})}\right]$ |

Most spending functions spend less error at the earlier looks, with the O'Brien-Fleming spending function being more conservative than the Pocock spending function (Power Family/Hwang-Shih-DeCani will depend on their free parameter). This is usually a desired characteristic as it means that the results of any interim analysis will only be considered significant (and thus ending the trial) at an early stage with an extreme result. It also means that the final analysis will be more comparable in terms of sample size and significance boundaries to an equivalent fixed term design.

#### 4.2.2.2 Boundaries

The boundaries in nQuery represent the critical values at each look above or below which the trial would end early. These boundaries are usually constructed using the alpha and beta spending functions, though users are given the option of entering these manually. For the spending function approach, nQuery will automatically generate boundaries for the early rejection of the null hypothesis (if an efficacy alpha spending function is active), early finding for the null hypothesis (if a futility beta spending function is active) or both (if futility and efficacy are both active using a combination of both the alpha and beta spending functions).

Once these critical boundary statistics are generated, they can be compared during interim monitoring to the interim test statistics to decide whether to end the trial early. Essentially, if a test statistic crosses an efficacy boundary then it can be concluded that the experimental treatment shows a statistically significant effect and the trial can be stopped with rejection of the null hypothesis. If the test statistic crosses a futility boundary then this indicates with high probability that an effect will not be found, that the trial can be terminated by rejecting the alternative hypothesis. For futility bounds there are two options; either to have the boundaries binding, or non-binding. With binding boundaries, if the test statistic crosses the futility boundary, the test must be stopped, otherwise the type-1 error may become inflated. The reason for this is that there is an interaction between the efficacy and futility boundaries in their calculation that could cause the efficacy boundary to shift. In the case of non-binding boundaries; the efficacy boundaries are calculated as normal, that is, as if the futility boundaries did not exist. This eliminates the danger of inflating the Type I error if the futility boundary is overruled. The downside of

the non-binding case is that it may increase the required sample size relative to the binding case.

In nQuery, boundary values are given on the standardized scale (i.e. Z-statistic scale) and will usually equal the treatment difference divided by its standard error. Examples for two independent sample design case are given in Table 4.2.2.2.

**Table 4.2:** Two Sample Group Sequential Standardized Statistics

| Means | Proportions (Pooled) | Survival (Approx.) |
|---|---|---|
| $Z = \frac{\mu_2 - \mu_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ | $Z = \frac{\pi_2 - \pi_1}{\sqrt{\frac{\bar{\pi}(1-\bar{\pi})}{n_1} + \frac{\bar{\pi}(1-\bar{\pi})}{n_2}}}$ | $Z = \frac{ln(HR)}{\sqrt{\frac{1}{E_1} + \frac{1}{E_2}}}$ |

For early stopping, we simply calculate the Z-statistic based on the current interim data and stop early if it is above the "upper" bound or below the "lower" (2-sided) or "futility" (1-sided) bound. Note that in nQuery; "upper", "lower" and "futility" are defined relative to the direction of the pre-specified effect size. For example, if the initial design assumes a negative difference ($\mu_1 > \mu_2$) then positive "upper" interim tests statistics would be those in which the interim difference is also negative and vice-versa. Similarly, "lower" results would be those in the opposite direction. Futility bounds can be above or below zero but the same guidance applies for the Z-statistic calculation.

#### 4.2.2.3 Calculating Power or Sample Size

To calculate the power or sample size, we can use the drift parameter. This is the standardized Z-statistic defined above for the relevant test for the final look in the group sequential design. It also equals $\triangle\sqrt{I_x}$ where $\triangle$is the treatment effect (e.g. mean difference) and $I_x$ is the maximum total information of the group sequential design.

To calculate the sample size, nQuery calculates the value for the drift parameter based on the values for the significance level, whether a one-sided or two-sided analysis is being used, the power, the number of looks and the spending function. The drift parameter is then set to equal to the standardized Z-statistic for the final look and via re-arrangement or iterative search, the sample size can be found.

To calculate the power, nQuery calculates the drift parameter by calculating the standardized Z-statistic based on the relevant information including the sample size. This drift is used to reverse-calculate the power via algorithm [Jennison and Turnbull, 1999] using the significance level, whether a one-sided or two-sided analysis is being used, the number of looks and the spending function.

### 4.2.3 Group Sequential Design in nQuery

nQuery provides an intuitive interface to make planning a group sequential design and finding the appropriate sample size or power easy. This section will outline the steps to plan a group sequential trial in nQuery.

#### 4.2.3.1 Background

In nQuery, the vast majority of user actions and interface options are the same as for fixed term designs. This section will focus on the major additional issues associated with group sequential design in nQuery. See previous chapters for information on the shared table elements.

In nQuery, the inputs required for group sequential design can be split into two categories: fixed term parameters, group sequential trial (GST) parameters. The fixed term parameters are those which would be required to calculate the sample size or power for the equivalent fixed term design and thus will depend on the statistical test and data type. The GST parameters are those required to define the group sequential design such as the spending functions, number of interim looks etc. and will be effectively the same across different designs and data types. We will cover how both of these are entered and how to use nQuery for group sequential design in the following example.

#### 4.2.3.2 Group Sequential Design Example

**Main Table** The main table is used to enter the fixed term parameters including the significance level, effect size and power among others. These will be very similar or the same as those for an equivalent fixed term design. In this example, the group sequential design is for a two sample t-test and thus the effect size is characterised by the mean difference and per-group standard deviations.

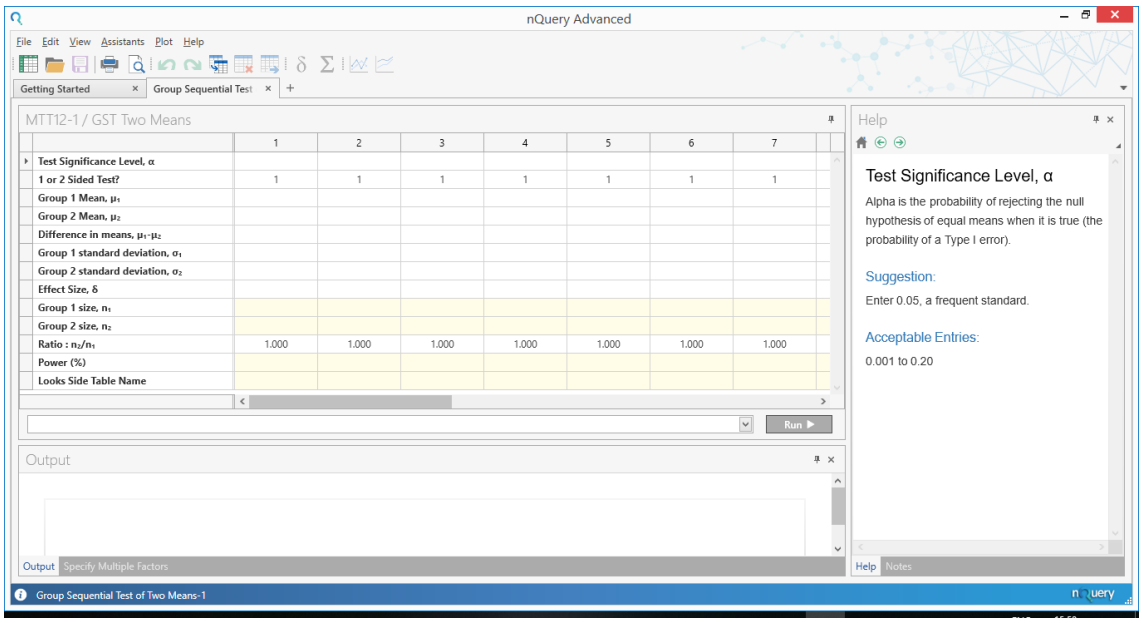



**Figure 4.1:** Two Means Group Sequential Design

The first thing we need to do is enter these fixed term design parameters. As these are effectively the same as for a fixed term design, please refer to previous chapters for guidance on filling these values. Here we will assume a treatment difference of -1 and per-group standard deviations of 2.5 for a 1-sided 5% significance test.

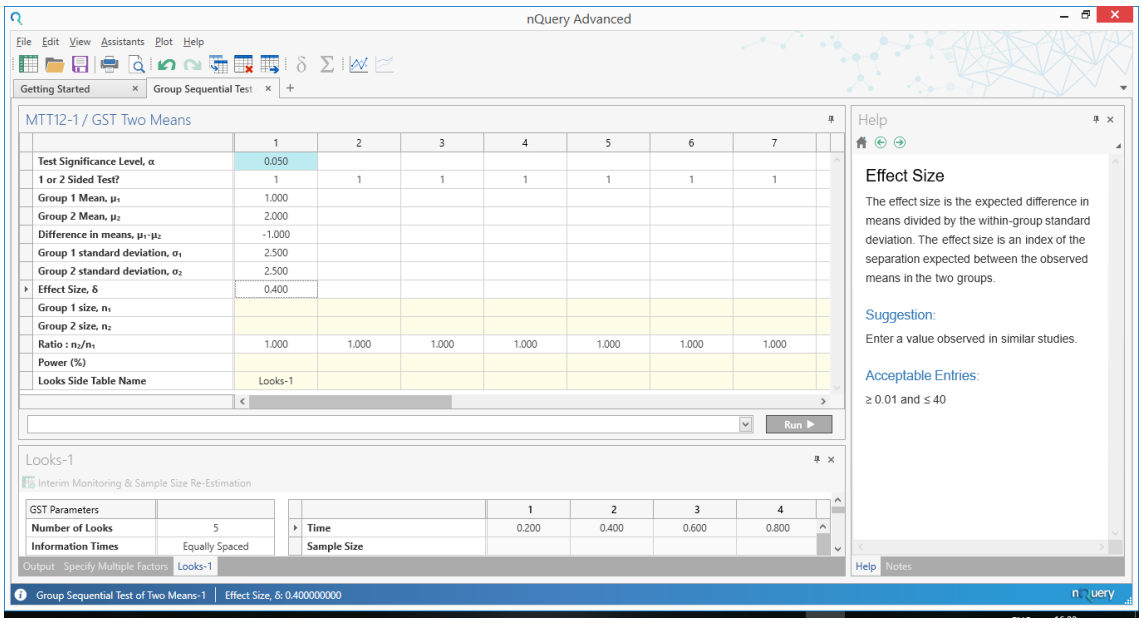



**Figure 4.2:** Example Fixed Term Parameters

We will leave the power and sample size empty for now as we want to define our

group sequential design before calculating the sample size. To do this we will need to use the Looks side-table.

**Looks Side-table** The Looks side-table will automatically open below the main table for a specific column when any cell in the column is selected. It can also be manually opened using the "Compute Effect Size" options from the toolbar or Assistants file menu. For more details on what side-tables are and how they work in general, refer to subsection 1.7.4.

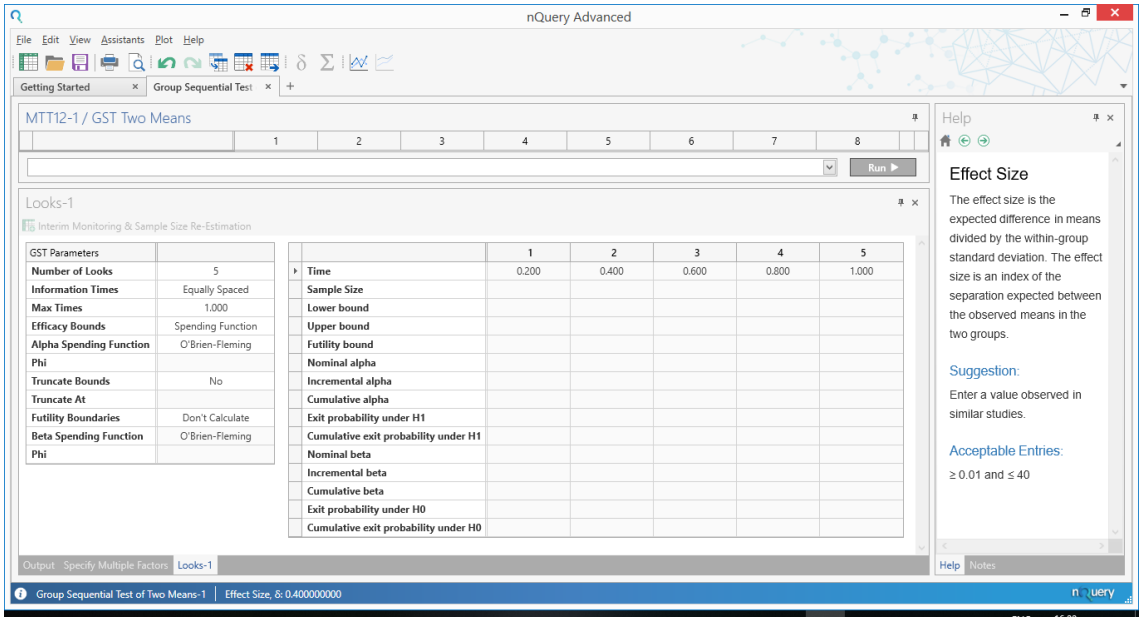



**Figure 4.3:** Looks Side-table

The Looks side-table can be split into two parts: the GST Parameters column (on the left) and the Looks table (on the right). The GST Parameters column defines the important features of the group sequential design. The Looks table will contain the important characteristics of the complete group sequential design as well as allowing some more complex study designs.

**GST Parameters** In the GST Parameters, there are 11 inputs settable to define the group sequential design operating characteristics. These can be summarised as follows:

1. Number of Looks ($\geq 2$): Set the total number of looks in group sequential design. As this is the total number of looks, the interim looks will equal this value minus one.

2. Information Times (Equally Spaced, User Input): Set whether the interim looks should be equally spaced or should be based on the user in-putted times (section 4.2.3.2)
3. Max Times (>0): Set the total maximum time. This will automatically re-scale the Time row in the Looks table when set. Note this is disabled when Information Times is set to User Input.
4. Efficacy Bounds (Spending Function, User Input, Don't Calculate): Set how the efficacy bounds will be calculated. Spending function will activate the Alpha Spending Function option, User Input will require editing the Upper Bound and Lower Bound rows of the Looks table (section 4.2.3.2), Don't Calculate will have no efficacy bounds and will only be usable if Futility Bounds is not set to Don't Calculate.
5. Alpha Spending Function (O'Brien-Fleming, Pocock, Power Family, Hwang-Shih-DeCani): Set the spending function used to calculate the efficacy bounds. For the Power Family and Hwang-Shih-DeCani options, the user will need to enter a value for the parameter row below this.
6. Power/HSD Parameter (>0 if Power Family, <3 if Hwang-Shih-DeCani): The free parameter for the Power Family or Hwang-Shih-DeCani efficacy spending function. Sometimes given the generic term of "Phi" in nQuery. See Table 4.1 for details.
7. Truncate Bounds (No, Yes): Set whether you want the efficacy bounds to be truncated at a specific user-defined value. If Yes is selected the Truncate at row will be active below.
8. Truncate at (>1): If Truncate Bounds is set to Yes, this defines the value at which the efficacy bounds are truncated. If an upper bound value in the Looks table is higher than this value then it will be set to this value. For lower bounds, it will be set to the minus version of this value.
9. Futility Bounds (Non-binding, Binding, Don't Calculate): Set how the futility bounds will be calculated. Non-binding and binding will activate the Alpha Spending Function option, Don't Calculate will have no futility bounds and will only be usable if Efficacy Bounds is not set to Don't Calculate. Non-binding allows futility bound contravention without error inflation, while binding does not allow this.
10. Beta Spending Function (O'Brien-Fleming, Pocock, Power Family, Hwang-Shih-DeCani): Set the spending function used to calculate the futility bounds. For the Power Family and Hwang-Shih-DeCani options, the user will need to enter a value for the parameter row below this.
11. Power/HSD Parameter (>0 if Power Family, <3 if Hwang-Shih-DeCani): The free parameter for the Power Family or Hwang-Shih-DeCani futility spending function. Sometimes given the generic term of "Phi" in nQuery. See Table 4.1 for details.

More detail on these are provided in the help cards and in subsection 4.2.2. In nQuery, the default values are for 5 look design with an efficacy bound using an O'Brien Fleming spending function. In this example, we will set these to have a two look equally spaced design (i.e. one interim analysis at 50% of the total sample size), an O'Brien-Fleming spending function for the efficacy bound and a Power Family spending function with a parameter equal to 1 (which is similar to Pocock spending function) for the non-binding futility bound.

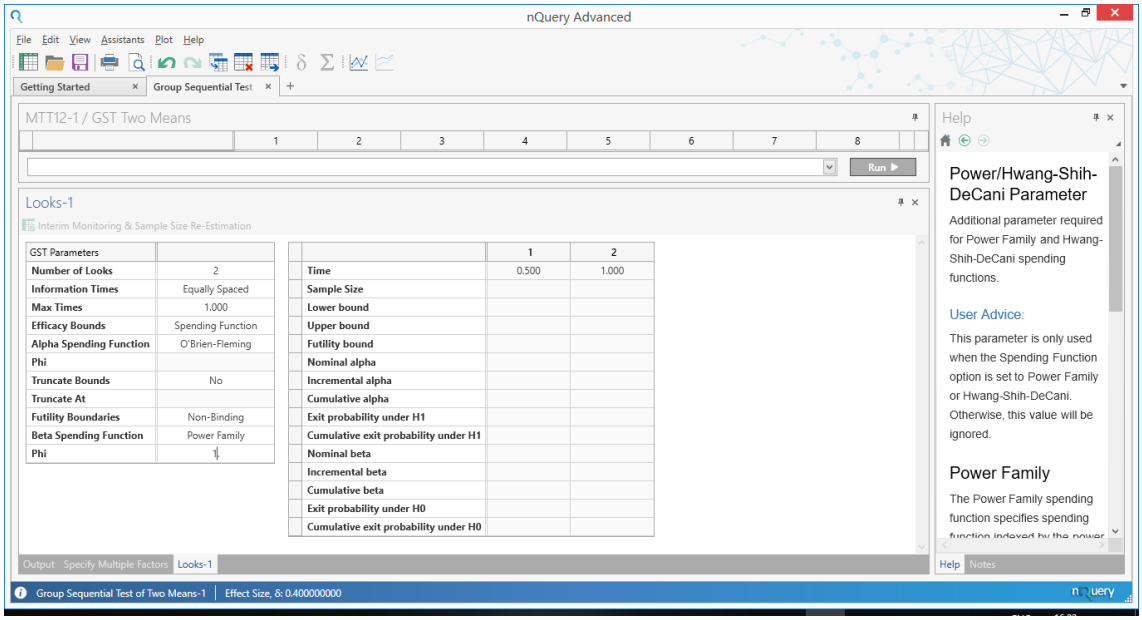



**Figure 4.4:** Example GST Parameters

**Looks Table** The Looks Table provides the information on the group sequential bounds and number of other useful parameters and additional information that characterises the complete group sequential design. In this table, each column corresponds to that Look. In this example, column 1 is the single interim look and column 2 is the final analysis. The Looks table contains the following rows:

1. Time: The Time at which each interim analysis will occur. By default, this will be the proportion of the total information at that look. The values in Time can be changed using the Max Times option (where each of these will be multiplied by the value in Max Times) or can be entered manually if Information Times is set to User Input
2. Sample Size: The cumulative sample size required at each look. The final entry will equal the calculated total sample size.
3. Lower Bound: The Lower Bound for efficacy below which we would end the trial early. The bounds are on the standardized test statistic scale. This will

only be used for 2-sided designs with an active efficacy bound. These will be entered manually if Efficacy Bounds is set to User Input.

4. Upper Bound: The Upper Bound for efficacy above which we would end the trial early. The bounds are on the standardized test statistic scale. This will be only used with an active efficacy bound. These will be entered manually if Efficacy Bounds is set to User Input.
5. Futility Bound: The Futility Bound below which we would end the trial early for futility (can ignore if non-binding). The bounds are on the standardized test statistic scale. This will only be used for 1-sided designs with an active futility bound.
6. Nominal Alpha: This is the value of alpha (i.e. p-value) for these boundaries if they were used in a single stand-alone test.
7. Incremental Alpha: This is the amount of alpha (type I error) that is spent at this interim test since the last look. It is close to nominal alpha but differs slightly due to being adjusted for multiple testing.
8. Cumulative Alpha: This is total amount of alpha (type I error) that has been spent up to and including this look.
9. Exit Probability under H1: This is the percentage chance that the trial will stop, given the specified alternative hypothesis in the main table is true. This is chance for both the efficacy and futility bounds.
10. Cumulative Exit Probability under H1: This is the cumulative percentage chance that the trial will stopped at this or any previous look, given the specified alternative hypothesis in the main table is true. This is chance for both the efficacy and futility bounds.
11. Nominal Beta: This is the value of beta for these boundaries if they were used in a single stand-alone test.
12. Incremental Beta: This is the amount of beta error (type II error) that is spent at this interim test since the last look. It is close to nominal beta but differs slightly due to being adjusted for multiple testing.
13. Cumulative Beta: This is total amount of beta error (type II error) that has been spent up to and including this look.
14. Exit Probability under H0: This is the percentage chance that the trial will stop, given the specified null hypothesis (usually difference of zero) in the main table is true. This is chance for both the efficacy and futility bounds.
15. Cumulative Exit Probability under H0: This is the cumulative percentage chance that the trial will stopped at this or any previous look, given the specified null hypothesis (usually difference of zero) in the main table is true. This is chance for both the efficacy and futility bounds.

More information on these can be found in the table help cards when these rows are selected or in subsection 4.2.2. Unless a User Input option is set, this table will be read-only and will fill automatically once a power or sample size calculation occurs in the main table. In this example, set power to 80% in the main table. This will give a sample size of 87 in each group and following Looks table (Figure 4.5)

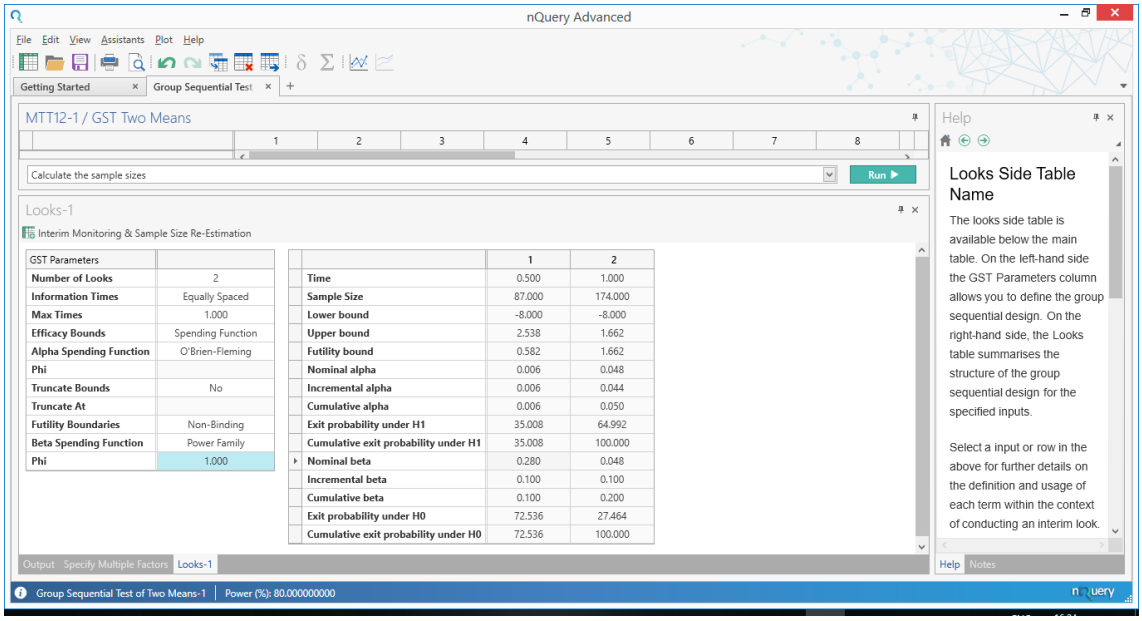



**Figure 4.5:** Example Looks Table

For practical purposes, the most important values are the Upper Bound, Futility Bound and Nominal Alpha. These are the values we will compare the interim test statistics (or p-values for nominal alpha) to decide whether to end the trial early or at the final analysis whether to reject the null hypothesis. The additional information provides useful context on the amount of error spent at each look and how likely the trial will end under the specified alternative hypothesis (i.e. the effect size in main table) or the null hypothesis.

As mentioned previously, certain elements of the Looks table will have to be set manually if Information Times or Efficacy Bounds are set to User Input. An example of these being set manually is given in Figure 4.6.

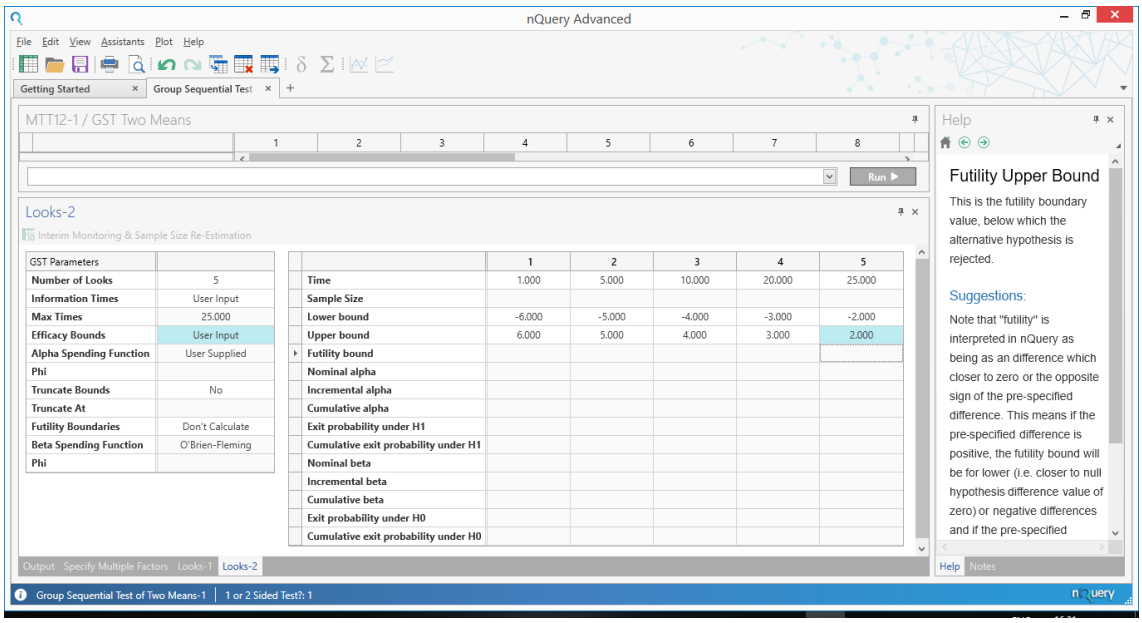



**Figure 4.6:** Custom Looks Table

## 4.3 Interim Monitoring and Unblinded Sample Size Re-estimation

### 4.3.1 Background

With nQuery Adapt, users will have access to the interim monitoring and sample size re-estimation tool. This tool will provide the opportunity to increase the sample size at an interim analysis if the interim information suggests a "promising" result but which is under-powered to find the current interim effect size if it were the true effect size. In group sequential designs and other similar designs, access to the interim data provides the opportunity to improve a study to better reflect the updated understanding of the study. One way to improve a group sequential design would be to use the interim effect size estimate not only to decide to whether to stop a trial early but to increase the sample size if the interim effect size is promising. This optionality gives the trialist the chance to power for a more optimistic effect size, thus reducing up-front costs, while still being confident of being able to find for a smaller but clinically relevant effect size by increasing sample size if needed.

*Note: Interim Monitoring and Unblinded Sample Size Re-estimation is currently only available for users with an nQuery Advanced Pro license. Active packages are displayed in the Packages section in the Home tab. To purchase additional packages, see* `www.statsols.com`.

## 4.3.2 Unblinded Sample Size Re-estimation Theory

nQuery provides unblinded sample size re-estimation methods based on two main approaches: Chen-DeMets-Lan method [Chen et al., 2004] and Cui-Hung-Wang method [Cui et al., 1999]. The differences between this will be explored below but both of these methods are based on using the conditional power to define whether a result is "promising" and whether to increase and how much to increase the sample size.

### 4.3.2.1 Conditional Power

Conditional power is the probability that the trial will reject the null hypothesis at a subsequent look given the current test statistic and the assumed parameter values, which are usually assumed to equal their interim estimates. For "promising" trials where the conditional power falls between a lower bound, a typical value would be 50%, and the initial target power the sample size can be increased to make the conditional power equal the target study power.

The conditional power is calculated by assuming a set of "true" values for the fixed term parameters, the interim standardized test statistic and the proportion of information (i.e. sample size) used at the interim analysis. The generalized formula for the conditional power at look "k" for ending at look "k+1" is as follows [Jennison and Turnbull, 1999]:

$$CP\left(k\right) = \Phi\left(\frac{Z_k\sqrt{I_k} - z_{k+1}\sqrt{I_{k+1}} + \theta(I_{k+1} - I_k)}{\sqrt{I_{k+1} - I_k}}\right)$$

where $Z_k$ is the interim standardized test statistic at time k, $I_k$ is the information at time k, $z_{k+1}$is the target test statistic at look k+1, $I_{k+1}$ is the information at time k+1 and $\theta$ is the "true" parameter value of interest

For unblinded sample size re-estimation, we will assume k is set to specified sample size re-estimation look and k+1 is the final look. This means the conditional power in nQuery is an approximation for the chance of rejecting the null at any subsequent look if k is not the penultimate look but practically there will be a negligible difference between this value and the "true" conditional power.

For unblinded sample size re-estimation, the information at a given look will equal the reciprocal of the squared standard error based on the "true" parameter values (e.g. for two means, the within-group standard deviations and sample sizes: $\left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right]$) and $\theta$ is the "true" parameter of interest (e.g. for two means, the mean difference). These "true" parameters are assumptions made by the user. Two common suggestions would be set these to either the interim estimates of these parameters or to set these to the assumed "true" values from the original design. It will be discussed later but it should be noted that the value of the interim test statistic, $Z_k$, will be treated the same for the unweighted (Wald) statistic and weighted (Cui-Hung-Wang) statistic.

#### 4.3.2.2 Unblinded Sample Size Re-estimation Methods

Unblinded sample re-estimation increases the sample size in response to a "promising" result (i.e. effect size) given the interim data. Note that sample size decreases are also allowable but practically are greatly disfavoured by most sponsors and regulators. A "promising" result is one which is clinically relevant but which the current sample size is unlikely to be found significant given the current study design and trajectory. The definition of "promising" is tied typically to conditional power, though alternative metrics such as Bayesian Predictive Power have been put forward. The sample size re-estimation is tied to the conditional power (and corresponding effect sizes) falling within a range defined as "promising". This range is defined by the study design but a typical range would be between 50% and the target power (e.g. 80% for our group sequential example). If the conditional power is based on the interim analysis, we can increase the sample size until the conditional power reaches the target level (typically the original target power level).

Thus by increasing the sample size, this "promising" result can be adequately powered for given the current available information to give the appropriate chance for finding significance at the final analysis (or any subsequent). However by allowing the study sample size to be changed based on the unblinded interim data there is the chance of increasing the overall error rate or introducing bias into a study. Two strategies for ensuring that the error rate is maintained will be discussed in this section: Chen-DeMets-Lan and Cui-Hung-Wang. Note that these methods change the method for calculating the standardized test statistics but that the group sequential bounds are unaffected and can be used as for a classic group sequential design.

**Chen-DeMets-Lan** The Chen-DeMets-Lan method [Chen et al., 2004] is a method of unblinded sample size re-estimation that requires minimal change from a group sequential trial. This method works by setting out a number of criteria related to the allowable conditions for increasing the sample size that ensures that the normal group sequential methods (bounds, test statistics etc.) can be used while ensuring the overall error rate is not inflated versus the original group sequential design. The two primary conditions which are required to ensure no error inflation are:

1. Sample Size Re-estimation is at the penultimate look
2. The conditional power at the penultimate look must lie between 50% and the original target power.

Chen-DeMets-Lan show that if these two conditions are met then the Type I error rate after increasing the sample size until the conditional power equals the original target power will not exceed the original target Type I error. For some conditions, the Type I error rate is lower than the original target Type I error but the conservativeness is very minor over this range. In addition, Gao, Ware and Mehta [Gao et al., 2008] and Mehta and Pocock [Mehta and Pocock, 2011] extended the work of Chen-DeMets-Lan to derive the exact lower bound for the conditional power which ensures

Type I error control for any given group sequential or equivalent fixed term analysis. In basic terms, this is done by iteratively searching over the full range of conditional powers (and corresponding effect sizes), calculating the required sample size increase for each conditional power and then calculating the appropriate weighted test statistic (Cui-Hung-Wang) that ensures error control. For the range where this weighted statistic is lower than the original critical unweighted test statistic, the sample size can be increased while maintaining the Type I error.

After the sample size increase occurs, the final analysis is conducted as if it were a standard group sequential design and uses the pre-existing group sequential bound values and Wald test statistic calculation (subsection 4.2.2).

**Cui-Hung-Wang** The Cui-Hung-Wang method [Cui et al., 1999] is an extension of the generalized adaptive design approach of Muller and Schaefer for the case of sample size re-estimation. Under the Cui-Hung-Wang method, the user has full flexibility over which look sample size increases can occur, how many sample size increases occur, how large the sample size increase can be and the range of "promising" conditional powers allowable for sample size increases. This robustness to in-study changes gives the greatest operational flexibility possible regarding sample size. However, this method requires the usage of a weighted test statistic which differs from the classic standardized test statistic used in group sequential designs.

The Cui-Hung-Wang method uses a weighted test statistic based on the weighted combination of the incremental test statistics. Previously when discussing test statistics, we have been referring to the cumulative test statistics which are based on all the interim data up to that point. Incremental test statistics are those calculated based only on the data between the last look and the current look. However, the cumulative and incremental test statistics can be easily related for the traditional group sequential trial as follows:

$$Z_{Cumulative} = \frac{\sqrt{w_1}Z_1+\sqrt{w_2}Z_2+...+\sqrt{w_{k-1}}Z_{k-1}+\sqrt{w_k}Z_k}{\sqrt{w_1+w_2+...+w_{k-1}+w_k}}$$

where $Z_{1,2...,k-1,k}$ are the respective incremental test statistics at each look and $w_{1,2...,k-1,k}$ are the weights for each test statistics. The weights must sum to one.

For the classic group sequential test statistic, the weights will equal the proportion of the total sample size used for each incremental test statistic (e.g. for equally spaced 4 total look design they would equal 0.25, 0.25, 0.25, 0.25). The Cui-Hung-Wang and Chen-DeMets-Lan method differ in how these weights are affected by an increase in sample size during the study. Note that for both statistics, the bounds for the classic group sequential trial are used as before for accepting and rejecting the null hypothesis or for ending the trial early for efficacy or futility.

**For the Chen-DeMets-Lan method, the weights used in this calculation are based on the actual (i.e. updated/increased) sample size for each**

**incremental test statistic.** Practically, this means that if a sample size occurs at the penultimate look, the weight placed on the final incremental test statistic will be increased so that the weight for the final cohort of data is proportional to the increased number of subjects between the penultimate and final look. Note that since the cumulative test statistic is equivalent to this version of the weighted sum of incremental test statistics, we can typically ignore the incremental tests statistics for practical purposes when using the Chen-DeMets-Lan or classic group sequential methods.

**For the Cui-Hung-Wang method, the weights used in this calculation are based on the original proposed sample size for each incremental test statistic.** This means that if a sample size increase occurs, subjects which are recruited after the sample size increase will be weighted less than subjects before the sample size re-estimation. However, this formulation ensures that the Type I error is retained regardless of the size of the sample size increase, when the sample size increase occurs and what range of conditional powers are considered "promising".

This robustness means that the Cui-Hung-Wang statistic provides a far greater level of operational control over the sample size re-estimation procedure. However, this comes at the cost of having to use a different test statistic than for the classic group sequential design and the down-weighting of subjects after the sample size increase. Note the weighting issue is a source of debate and we refer to relevant papers for further details.

### 4.3.3 Unblinded Sample Size Re-estimation and Interim Monitoring in nQuery

This section will show how to conduct an unblinded sample size re-estimation and interim monitoring for a group sequential design in nQuery. We will cover two examples: one for the Chen-DeMets-Lan method and one for the Cui-Hung-Wang method.

#### 4.3.3.1 Opening the Interim Monitoring and Sample Size Re-estimation Table

To open the Interim Monitoring and Sample Size Re-estimation table, we first must complete a group sequential design sample size or power calculation in a group sequential table with this feature available (see `www.statsols.com` for details on which tables have this feature in nQuery Adapt). In this section, we will take the completed example from subsubsection 4.2.3.2.

In short, this was a group sequential design with fixed term parameters of a mean difference of -1, common within-group standard deviations of 2.5, a one-sided 5% significance level and 80% power. The group sequential design was for a 2 look

design (1 interim analysis), the interim look at 50% of subjects analysed, O'Brien-Fleming efficacy bounds and Power Family futility bounds with the Power Family parameter set to 1. Given this the overall sample size required was 174 (87 per group).

To open the Interim Monitoring & Sample Size Re-estimation tool, we select the "Interim Monitoring & Sample Size Re-estimation" button at the top of the Looks window. This option will be greyed out if a sample size or power calculation has not been completed in that column or if you do not have an nQuery Adapt license (see `www.statsols.com` and section 8.2 for details on purchasing and activating nQuery Adapt). This button is highlighted in Figure 4.7.

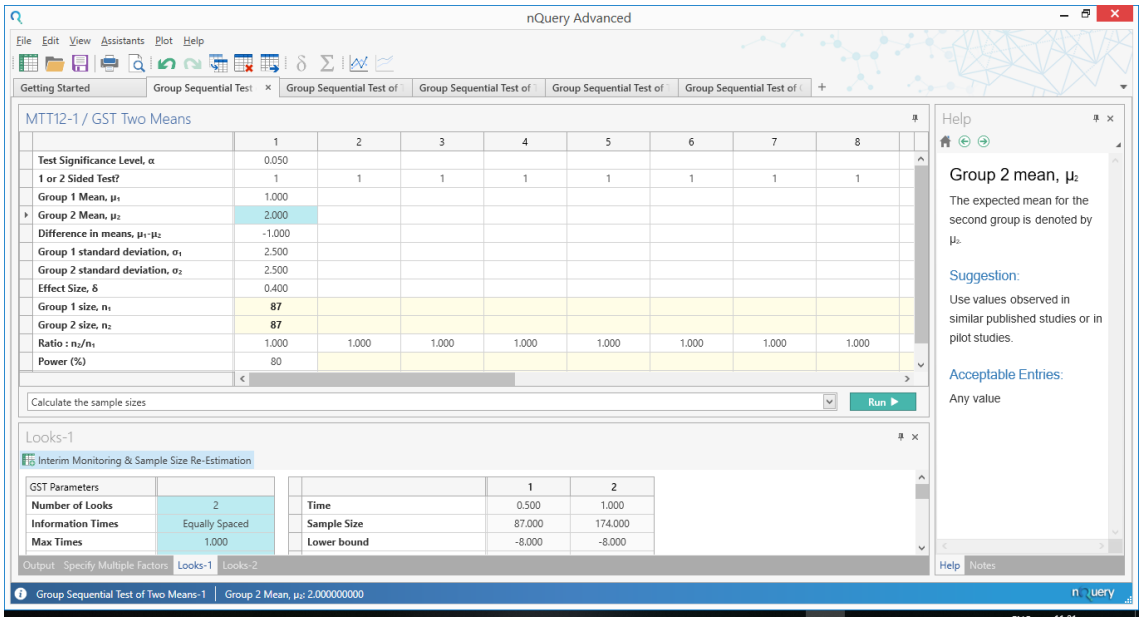



**Figure 4.7:** Opening Interim Monitoring and Sample Size Re-estimation Table

#### 4.3.3.2 Interim Monitoring and Sample Size Re-estimation Table Introduction

The interim monitoring and sample size re-estimation table provides an intuitive approach for setting up a proposed sample size re-estimation and then allowing interim monitoring with that sample size re-estimation in mind. The basic structure of the interim monitoring and sample size re-estimation tables can be split into two main parts: the sample size re-estimation rules column and the interim monitoring tool. We will refer to the former as the "SSR Rules" and the latter as the "Monitoring Table" for short. We will summarise each of these in the following section.

When first opened the table will look as per Figure 4.8 for a two means group sequential design.

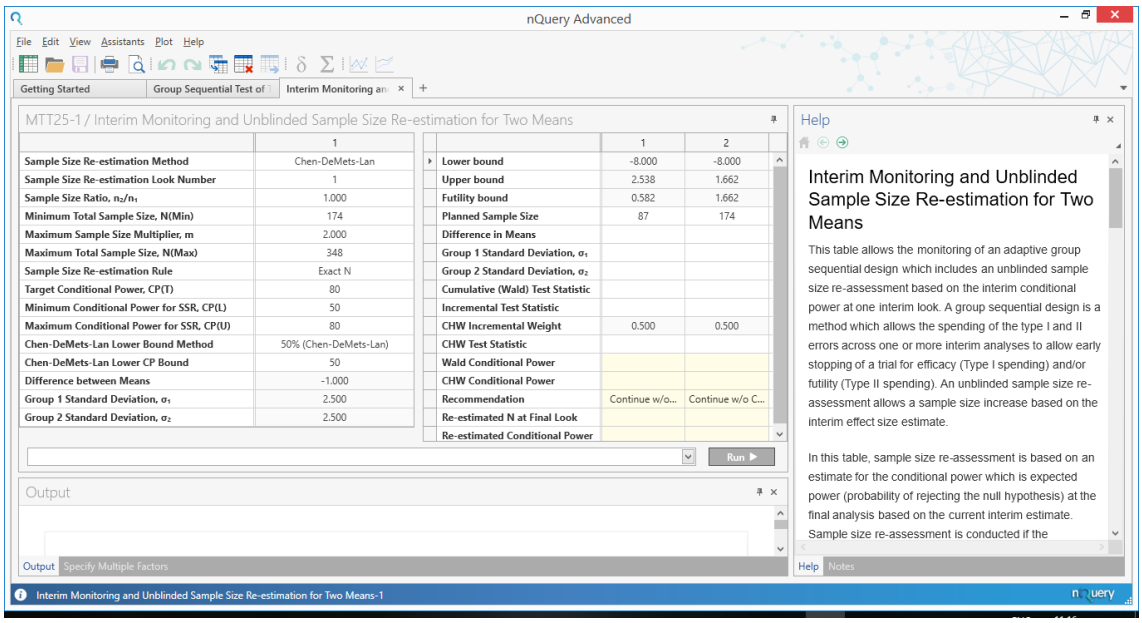



**Figure 4.8:** Interim Monitoring and Sample Size Re-estimation Table for Two Means GST

**Setting up the Sample Size Re-estimation Rules** To set up the rules that will define when and where a sample size re-estimation will occur, nQuery provides a large number of options to tailor the sample size re-estimation to the requirements of the user. These are set in the SSR Rules column (on the left in the main window). There are 12 inputs that can be edited by the user to define the sample size re-estimation rules. These are summarised below:

1. Sample Size Re-estimation Method (Chen-DeMets-Lan, Cui-Hung-Wang, None): Define the sample size re-estimation method desired for this study. The Chen-DeMets-Lan and Cui-Hung-Wang methods are summarised in subsubsection 4.3.2.2. In short, the Chen-DeMets-Lan method uses the standard group sequential test statistic but restricts sample size re-estimation to the penultimate look and for a certain range of conditional powers. The Cui-Hung-Wang method uses a different weighted test statistic from the group sequential design, based on the incremental test statistics weighted by the original planned sample size, and can be conducted for any look or conditional power. None will disable any sample size re-estimation (all subsequent options in this column can be ignored) and the tool will operate as a standard group sequential design monitoring table. This will default to the Chen-DeMets-Lan method.
2. Sample Size Re-estimation Look Number (If Chen-DeMets-Lan: Looks-1, if Cui-Hung-Wang: < Looks): The look at which a sample size re-estimation will occur. Note that nQuery Adapt allows only a single sample size re-estimation. For the Chen-DeMets-Lan method, this must equal the total number of looks (including the final analysis) minus one. For the Cui-Hung-Wang method, this

must be less than the total number of looks. Defaults to the number of looks minus one.

3. Sample Size Ratio (>0): The ratio between the group 1 and group 2 sample sizes. This will be assumed to be constant across all looks. Defaults to the sample size ratio from the original group sequential design.
4. Minimum Total Sample Size (>2): The minimum total sample size allowable in this study. Changing this will automatically update the Planned Sample Size row in the Monitoring Table to reflect this value. Defaults to the sample size from the original group sequential design.
5. Maximum Sample Size Multiplier (>1): A multiplier for the minimum total sample size versus the maximum total size. Changing this will automatically update the Maximum Total Sample Size to reflect this value. Defaults to 2 (maximum sample size double minimum sample size).
6. Maximum Total Sample Size (>Minimum Total Sample Size): The maximum total sample size allowable in this study. If a sample size re-estimation occurs, this will be maximum allowable total sample size that will occur even if the target conditional power is not reached with this value. Defaults to twice the minimum total sample size.
7. Sample Size Re-estimation Rule (Exact N, Maximum N): The rule for how much the sample size should be increased by if the interim statistic gives a conditional power within the Minimum and Maximum Conditional Power for SSR value for the Sample Size Re-estimation Look Number column. Exact N will increase the total sample size until either the conditional power equals the Target Conditional Power or the Maximum Total Sample Size is reached. Max N will always increase the sample size to the Maximum Total Sample Size. The Max N rule may over-power the study but reduces the operational risk of study participants reverse-calculating the effect size using the "Exact N" calculated to restore the conditional power to the original target power.
8. Target Conditional Power (>Minimum Conditional Power for SSR, <100): This is the conditional power targetted by the sample size re-estimation increase if the Exact N rule is active in the Sample Size Re-estimation Rule row. The sample size is increased until this value is reached or the maximum sample size is reached. Defaults to the target power from the original group sequential design.
9. Minimum Conditional Power for SSR (<Target Conditional Power & Maximum Conditional Power for SSR): This is the minimum conditional power that would be considered "promising" and lead to a sample size re-estimation at the specified Look Number. For the Chen-DeMets-Lan method, this needs to be greater than the Chen-DeMets-Lan Lower CP Bound. Defaults to 50%.
10. Maximum Conditional power for SSR (>Minimum Conditional Power for SSR, <100): This is the maximum conditional power that would be considered

"promising" and lead to a sample size re-estimation at the specified Look Number. Conditional powers above this value would have sufficient power and then a sample size increase would not be necessary. For the Chen-DeMets-Lan method, this should not be set higher than the power from the original group sequential design. Defaults to the target power from the original group sequential design.

11. Chen-DeMets-Lan Lower Bound Method (50% (Chen-DeMets-Lan), Derived (Mehta-Pocock)): Select the method used to the derive the allowable Minimum Conditional Power for SSR which can be used without leading to error inflation. Chen-DeMet-Lan sets this to 50% which holds for all group sequential designs. Derived (Mehta-Pocock) derives the lower bound which ensures the Type I error never exceeds the level for the equivalent Cui-Hung-Wang test statistic and this value will always fall between 0% and 50%. Defaults to 50% (Chen-DeMets-Lan).

In addition to these parameters, the SSR Rules column will also include the fixed term parameters used to define the study effect size. These values can be useful for reference and if the user wishes to base conditional power calculations on these original design parameters rather than the interim estimates for these parameters. For the two means case, the relevant parameters are the difference in means and group 1 and group 2 standard deviations. Once we have set our SSR Rules, we can use the Monitoring Table to start entering our interim statistics and stop our trial early, implement a sample size increase if needed or find for or against the null hypothesis at our final look. More detail on all these rows is given in the Home and Help cards of the table.

#### 4.3.3.3 Interim Monitoring Table

After the SSR Rules are set we need to enter our interim results to decide what action should be made at each look based on those results. To do this nQuery provides the Monitoring Table (on the right in the main window) which contains 16 rows and a number of columns equal to the total number of looks in the original group sequential design.

In this table each column corresponds to the respective interim look (e.g. column 1 is look 1, column 2 is look 2 etc.). To use this table we enter the required inputs in each column sequentially (i.e. complete column 1 before moving onto column 2 and so on) until the conditional power is calculated and Recommendation is made based on those inputs. We fill in each column in turn until either a recommendation is made to stop the trial early (ignorable if below non-binding futility bound) or the final column is filled and a recommendation is made to find for efficacy or futility.

There are 16 rows per column in the Monitoring Table for the Two Means case. These are summarised below:

1. Lower Bound (Read-only): The Lower Bound for efficacy below which we would end the trial early. The bounds are on the test statistic scale. These are inherited from the original group sequential design.
2. Upper Bound (Read-only): The Upper Bound for efficacy above which we would end the trial early. The bounds are on the standardized test statistic scale. These are inherited from the original group sequential design.
3. Futility Bound (Read-only): The Futility Bound below which we would end the trial early for futility (can ignore if non-binding). The bounds are on the test statistic scale. These are inherited from the original group sequential design.
4. Planned Sample Size (>2, >Planned Sample Size in Current Column - 1): The cumulative sample size analyzed at the current look. By default, this will equal the sample size at each look from the original group sequential design. This value can be changed in each column manually using the Minimum Total Sample option in the SSR rules or can be edited manually at any point by the user. This value will also be over-written by the cumulative test statistic side-table if the default value for the Cumulative Interim Total Sample Size (which defaults to this value) is changed and the Transfer Estimates option is set to Yes. If a sample size re-estimation is recommended these values will automatically update such that the sample size increase will happen between the Sample Size Re-estimation Look and subsequent look and all other Planned Sample Sizes will increase accordingly.
5. Difference in Means ($\neq$0, Unique to Two Means): The "true" assumed value for the difference in means used for the conditional power calculation. Usually set to either the interim estimate for this parameter or the assumed parameter value from the original design. The Cumulative Test Statistic side-table will automatically set this to the Interim Difference Between Means if the Transfer Estimates option is set to Yes. The original design value for this parameter is available at the bottom of the SSR Rules column.
6. Group 1 Standard Deviation (>0, Unique to Two Means): The "true" assumed value for the group 1 standard deviation used for the conditional power calculation. Usually set to either the interim estimate for this parameter or the assumed parameter value from the original design. The Cumulative Test Statistic side-table will automatically set this to the Interim Group 1 Standard Deviation if the Transfer Estimates option is set to Yes. The original design value for this parameter is available at the bottom of the SSR Rules column.
7. Group 2 Standard Deviation (>0, Unique to Two Means): The "true" assumed value for the group 2 standard deviation used for the conditional power calculation. Usually set to either the interim estimate for this parameter or the assumed parameter value from the original design. The Cumulative Test Statistic side-table will automatically set this to the Interim Group 2 Standard

Deviation if the Transfer Estimates option is set to Yes. The original design value for this parameter is available at the bottom of the SSR Rules column.

8. Cumulative (Wald) Test Statistic (Any Value): The standardized test statistic based on the interim data (see subsection 4.2.2) This is equivalent to the standard Z statistic for this data and is alternatively called the unweighted statistic (although its weighting is simply based on the empirical sample size rather than the initial sample size). This is the test statistic used for conditional power when the Chen-DeMets-Lan method or None options are selected from the Sample Size Re-estimation Method row in the SSR Rules column. This can be calculated using the Cumulative Test Statistic side-table based on the interim estimates for relevant effect size parameters (for two means, the mean difference and group 1 and 2 standard deviations). If this value falls above the Upper Bound or below the Lower Bound (2-sided), the trial should be ended early for efficacy. If this value falls below the Futility Bound, the trial can be ended early for futility.

9. Incremental Test Statistic (Any Value): The standardized test statistic (see subsection 4.2.2) based on the incremental data i.e. for the data between the last look and the current look. This is equivalent to the standard Z statistic for this incremental data. This incremental test statistics for all columns up to and including the current column are used to derive the Cui-Hung-Wang (CHW) Test statistic in combination with CHW Incremental Weights. The CHW test statistic is then used for conditional power when the Cui-Hung-Wang method is selected from the Sample Size Re-estimation Method row in the SSR Rules column. This can be calculated using the Incremental Test Statistic side-table based on the incremental estimates for relevant effect size parameters (for two means, the mean difference and group 1 and 2 standard deviations).

10. CHW Incremental Weight (Read-only): The incremental weights used in combination with the incremental test statistics to derive the Cui-Hung-Wang (CHW) Test Statistic. These will equal the original proposed incremental sample size used between each look and its prior look.

11. CHW Test Statistic (Any Value): The Cui-Hung-Wang Test weighted test statistic used for conditional power calculations when the Cui-Hung-Wang method selected from the Sample Size Re-estimation row of the SSR Rules. This will typically be auto-calculated using the incremental test statistic and CHW Incremental Weight column values (see section 4.3.2.2). However, these values can be entered manually if desired. If this value falls above the Upper Bound or below the Lower Bound (2-sided), the trial should be ended early for efficacy. If this value falls below the Futility Bound, the trial can be ended early for futility.

12. Wald Conditional Power (Read-only): The conditional power based on the "true" parameter estimates and the Cumulative (Wald) Test Statistic value

based on the interim data. This will only be calculated if the Chen-DeMets-Lan and None options are selected from Sample Size Re-estimation Method row in SSR Rules. For the sample size re-estimation look column, if the conditional power falls between the Minimum Conditional Power for SSR and Maximum Conditional Power for SSR then a sample size re-estimation will automatically activate.

13. CHW Conditional Power (Read-only): The conditional power based on the "true" parameter estimates and the CHW Test Statistic value based on the interim data. This will only be calculated if the Cui-Hung-Wang option is selected from Sample Size Re-estimation Method row in SSR Rules. For the sample size re-estimation look column, if the conditional power falls between the Minimum Conditional Power for SSR and Maximum Conditional Power for SSR then a sample size re-estimation will automatically activate.

14. Recommendation (Read-only): This gives the recommendation for this look based on the relevant test statistic and conditional power. There are four recommendations that can be made at a given interim look: a) Stop for Efficacy - Stop trial early due to strong evidence for efficacy due to test statistic being beyond Upper (1-sided or 2-sided) or Lower Efficacy Bounds (2-sided only) b) Stop for Futility - Stop trial early due to strong evidence for futility due to test statistic being below Futility bound (1-sided only) c) Continue without (w/o) change - Continue until next look with no changes due to test statistic not being in the early stopping region or the conditional power falling within the lower and upper SSR conditional power bounds d) Add N & Continue - Add required additional sample size for "promising" design and continue trial until next look as conditional power fell between lower and upper SSR conditional power bound. At the final look, two recommendations can be made: "Find for Efficacy" (for 1-sided, test statistic above upper bound; for 2-sided, test statistic outside lower to upper bound range) or "Find for Futility" (for 1-sided, test statistic below upper bound; for 2-sided, test statistic inside lower to upper bound range) .

15. Re-estimate N at Final Look (Read-only): The recommended total sample size after a sample size re-estimation occurs. This will be empty unless a sample size re-estimation has been recommended in this column (Add N & Continue in Recommendation row). The increase in sample size will automatically place the additional sample size between the current look and the next look, with the Planned Sample Size row automatically updated to reflect this. This pattern can be edited manually if desired in the Planned Sample Size row in the subsequent columns.

16. Re-estimated Conditional Power (Read-only): The conditional power based on the interim test statistic and "true" parameter values based on the sample size increase calculated in the Re-estimate N at Final Look row. This will be empty unless a sample size re-estimation has been recommended in this

column (Add N & Continue in Recommendation row). For Exact N, this will equal the target conditional power approximately unless the maximum sample size is insufficient to restore the conditional power to the target conditional power.

More detail on these is given in the Home and Help cards of the table. Note that the difference in means and standard deviation rows are unique to the two mean design. Other designs will have the appropriate parameters instead but the basic workflow and assumptions are the same for other designs.

Note that in nQuery; "upper", "lower" and "futility" are defined relative to the direction of the pre-specified effect size. For example, if the initial design assumes a negative difference ($\mu_1 > \mu_2$) then positive "upper" interim tests statistics would be those in which the interim difference is also negative and vice-versa. Similarly, "lower" results would be those in the opposite direction. Futility bounds can be above or below zero but the same guidance applies for the Z-statistic calculation.

#### 4.3.3.4 Monitoring Side-tables

In addition to these tables, nQuery provides two side-tables which make calculating and entering the required inputs easier for the "true" parameter values and the test statistics required in the Monitoring Table. These two side-tables are the Cumulative Test Statistic side-table and the Incremental Test Statistic Side-Table.

These side-tables allow the user to enter relevant interim/incremental effect size estimates and calculate the relevant test statistic and transfer this into the Monitoring Table. The Cumulative Test Statistic side-table will automatically open in the window if the "true" parameter values or Cumulative (Wald) Test Statistic rows are selected in a column and the Incremental Test Statistic side-table will automatically open in the window below the main table if the Incremental Test Statistic row is selected. Functionally, these work the same as side-tables seen in other tables (see section 3.1).

The calculations in both side-tables are identical and based on the relevant standardized test statistic calculations from subsection 4.2.2. They differ in that the Cumulative Test Statistic side-table requires the user input the cumulative interim estimates (i.e. based on all the data up to and including the current look) and the Incremental Test Statistic side-table requires the user input the incremental estimates (i.e. based only on the data from the last look to the current look). In addition, the Cumulative Test Statistic side-table gives the option to transfer the relevant interim effect size parameters and sample size into the relevant rows of the Monitoring Table by setting the Transfer Estimates option to Yes.

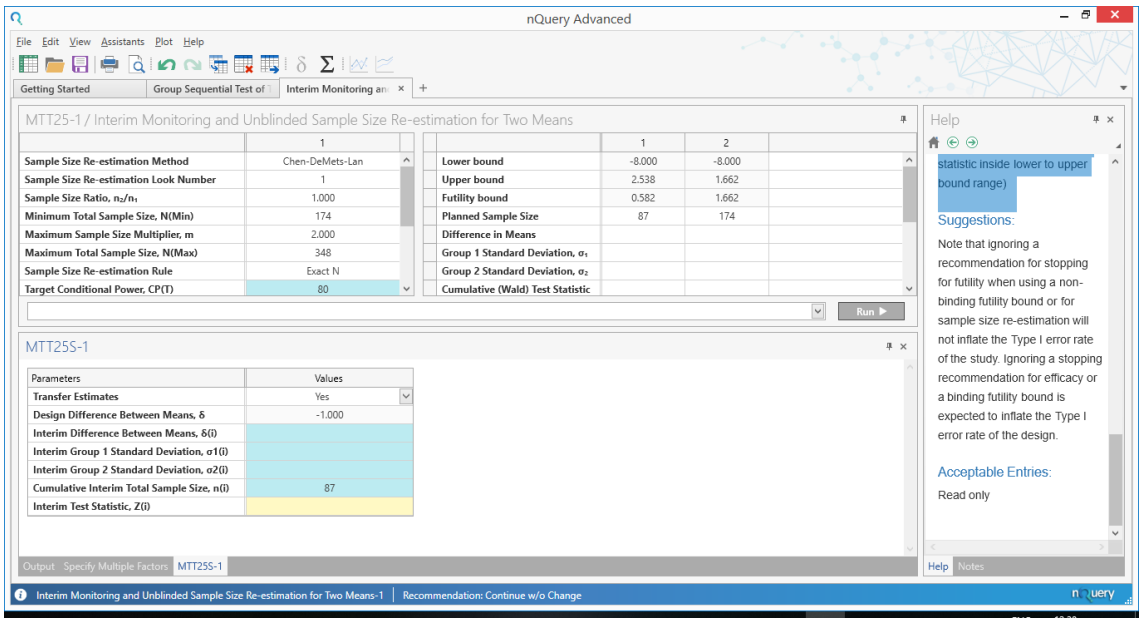



**Figure 4.9:** Cumulative Test Statistic Side-Table for Two Means

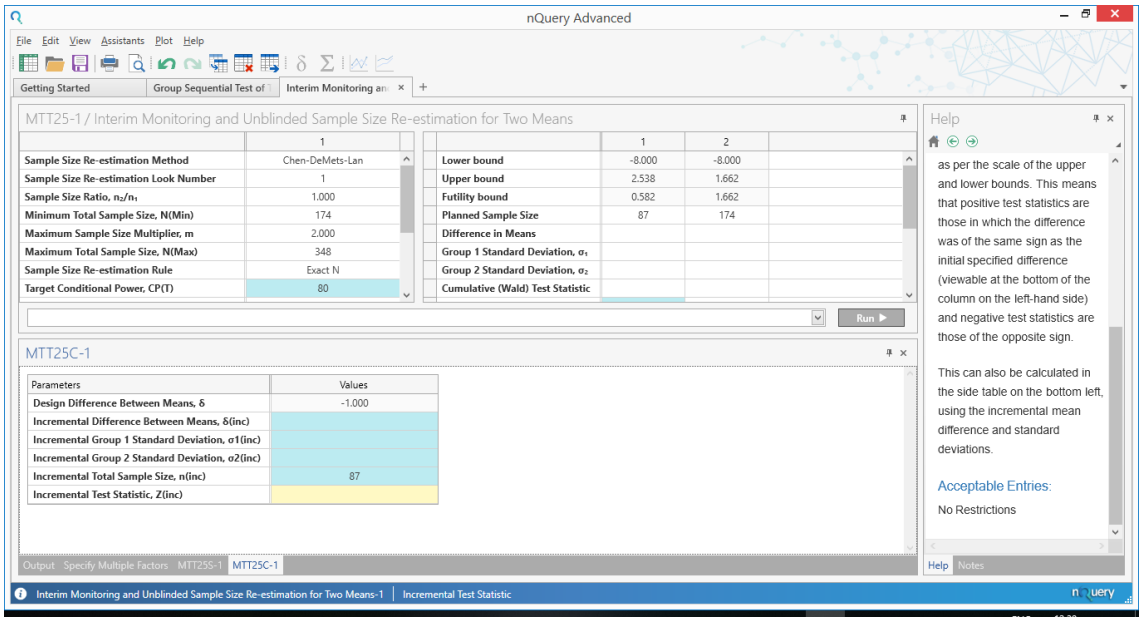



**Figure 4.10:** Incremental Test Statistic Side-Table for Two Means

The Cumulative Test Statistic side-table can help calculate the cumulative (Wald) test statistic for the Chen-DeMets-Lan method while also automatically assigning the "true" parameter values for conditional power as being equal to their interim estimates. The Incremental Test Statistic side-table can help calculate the incremental test statistics needed to calculate the Cui-Hung-Wang (CHW) test statistic

in combination with the pre-set CHW Incremental Weights. The functionality of both will be explored further in the worked example that follows.

Note that in nQuery; "upper", "lower" and "futility" are defined relative to the direction of the pre-specified effect size. For example, if the initial design assumes a negative difference ($\mu_1 > \mu_2$) then positive "upper" interim tests statistics would be those in which the interim difference is also negative and vice-versa. Similarly, "lower" results would be those in the opposite direction. Futility bounds can be above or below zero but the same guidance applies for the Z-statistic calculation. Thus these side-tables will provide the Design effect (e.g. difference in means). Cumulative/Incremental effects (e.g. differences) which have the same sign as the design effect will give a positive test statistic and those of the opposite sign will give a negative test statistic.

### 4.3.4 Interim Monitoring and Sample Size Re-estimation Worked Examples

#### 4.3.4.1 Chen-DeMets-Lan Worked Example

For this example, we will use the group sequential design from subsubsection 4.2.3.2 and have a Chen-DeMets-Lan sample size re-estimation based on the default rules used by the Interim Monitoring and Sample Size Re-estimation table. For reference, this was a group sequential design with fixed term parameters of a mean difference of -1, common within-group standard deviations of 2.5, a one-sided 5% significance level and 80% power. The group sequential design was for a 2 look design (1 interim analysis), the interim look at 50% of subjects analysed, O'Brien-Fleming efficacy bounds and Power Family futility bounds with the Power Family parameter set to 1. This gave an overall sample size of 174 (87 per group).

These default rules will give a sample size re-estimation at the penultimate look (i.e. Look 1 for two look design), the minimum sample size will equal our original sample size (174), the maximum sample size will be twice our original sample size (348), we will only increase sample size if needed until the target conditional power (equalling our initial target power of 80%) is reached and will only increase the sample size if the conditional power is between 50% (Chen-DeMets-Lan lower bound) and 80% (target power) at Look 1. Before we enter our interim data, our table will look as per Figure 4.8.

We will now enter the relevant inputs in column 1 of the monitoring table. Let us assume that the interim data suggests a 25% reduction from our initial estimate of the effect size (i.e. -0.75 mean difference, assuming standard deviations are the same). We could enter the relevant mean difference and standard deviations manually and also calculate the cumulative test statistic manually and enter it here. However, we will use the Cumulative Test Statistic Side-Table to calculate these automatically instead. When we select the "Difference in Means" row in column 1, this side-table

will open automatically in the window below the main table. In this side-table, we leave the Transfer Estimates option as Yes so that our entered interim estimates will be set to be the true values for the conditional power calculations. We will also leave the Cumulative Sample Size as its default value of the Planned Sample Size for this column. We then enter our interim estimates for the difference in means and the group 1 and 2 standard deviations. In this case, these are -0.75, 2.5 and 2.5 respectively. The side-table will automatically calculate the Cumulative Test Statistic and transfer it and the interim parameter estimates into the main table.

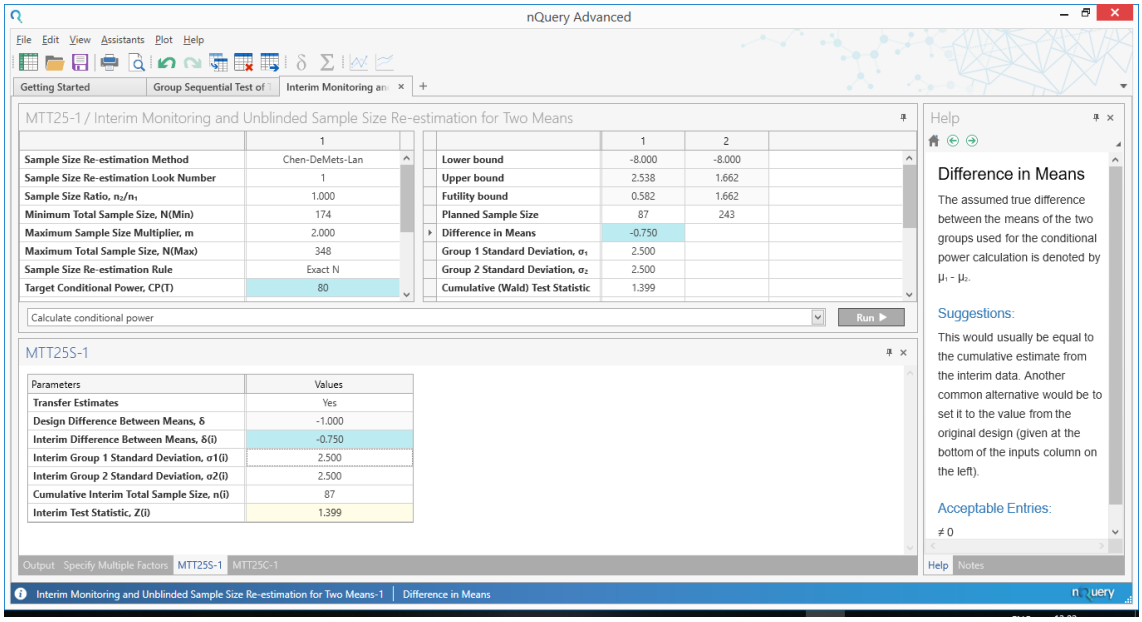



**Figure 4.11:** Chen-DeMets-Lan Side-Table Example

Note that in nQuery; "upper", "lower" and "futility" are defined relative to the direction of the pre-specified effect size. For example, if the initial design assumes a negative difference ($\mu_1 > \mu_2$) then positive "upper" interim tests statistics would be those in which the interim difference is also negative and vice-versa. Similarly, "lower" results would be those in the opposite direction. Futility bounds can be above or below zero but the same guidance applies for the Z-statistic calculation. Thus these side-tables will provide the Design effect (e.g. difference in means). Cumulative/Incremental effects (e.g. differences) which have the same sign as the design effect will give a positive test statistic and those of the opposite sign will give a negative test statistic. Thus, in this case entering -0.75 returns an interim test statistic of 1.399 but a value of 0.75 would have returned a test statistic of -1.399.

Once these values have been transferred, the conditional power is automatically calculated in this column in the Monitoring Table. In this case, these interim results give a conditional power of 67.28% which falls between our Conditional Power Sample Size Re-estimation range of 50% to 80%. Thus nQuery gives a recommend-

ation of “Add N & Continue” in the Recommendation row in column 1 and nQuery automatically calculates the required total sample size to increase the conditional power to 80%. In this case, this corresponds to an increase in the total sample size from 176 to 243, giving an updated conditional power of 80.06%. The Planned Sample Size in column 2 is updated automatically to reflect this.

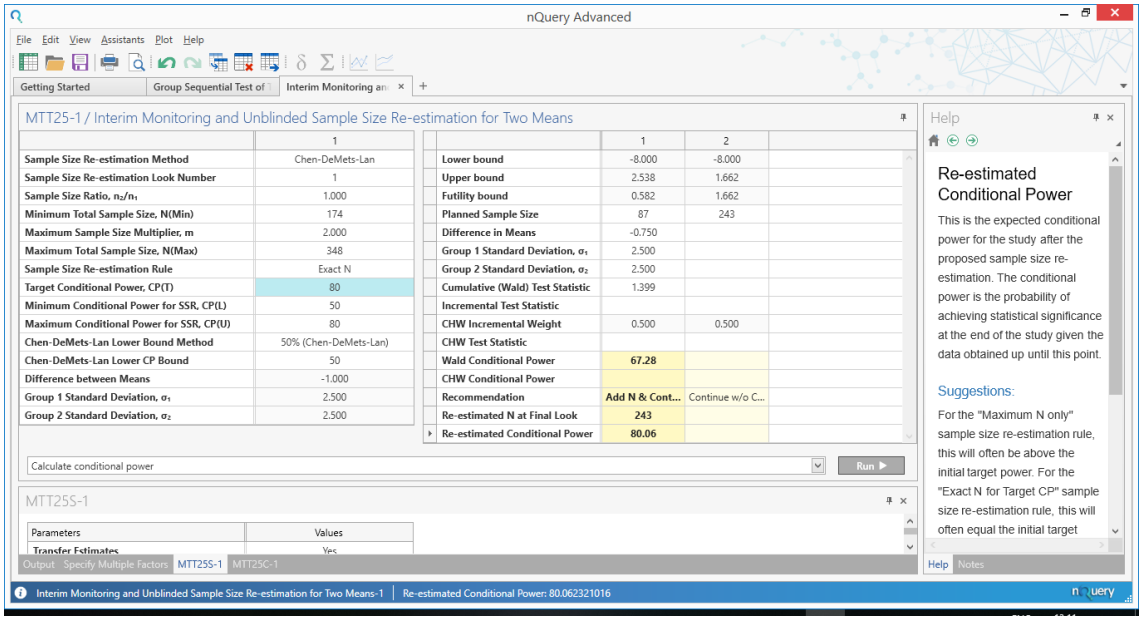



**Figure 4.12:** Chen-DeMets-Lan Sample Size Re-estimation Example

We can now complete our interim monitoring by entering our estimates at the end of our (now larger) study. Assume that the final analysis has the same difference in means (-0.75) and per-group standard deviations (2.5). We can use the Cumulative Test Statistic Side-table as before and calculate a test statistic of 2.338. As this test statistic value is greater than the Upper Bound (1.662) at the final analysis, we find for efficacy for this study at the final look.

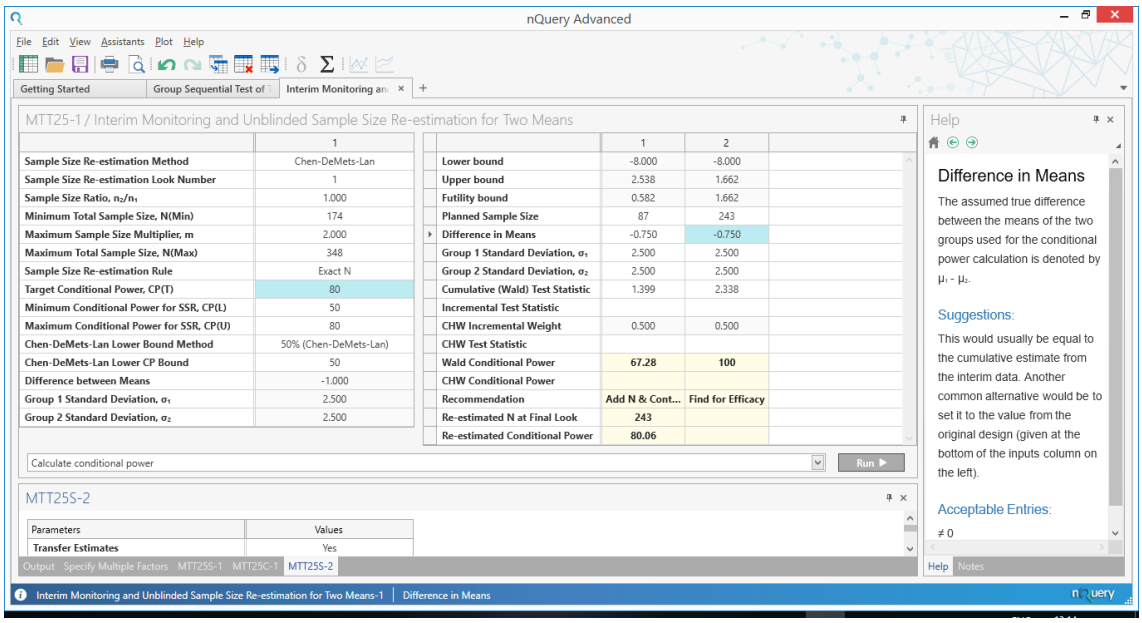



**Figure 4.13:** Chen-DeMets-Lan Completed Design Example

#### 4.3.4.2 Cui-Hung-Wang Worked Example

For this example, we will assume the same group sequential design.

For reference, this was a group sequential design with fixed term parameters of a mean difference of -1, common within-group standard deviations of 2.5, a one-sided 5% significance level and 80% power. The group sequential design was for a 2 look design (1 interim analysis), the interim look at 50% of subjects analysed, O'Brien-Fleming efficacy bounds and Power Family futility bounds with the Power Family parameter set to 1. This gave an overall sample size of 174 (87 per group).

In this case we assumed a 50% reduction in our difference in means versus the original study design (-0.5 difference in means).

For this sample size re-estimation we will also make some additional changes to the previous example. In this example we will increase the sample size multiplier to 3 (maximum N equal to 522), use a Max N rule for sample size increases (i.e. N always increased to the maximum N) and decrease the lower bound for the conditional power to 30%. This will give SSR Rules inputs as in Figure 4.14.

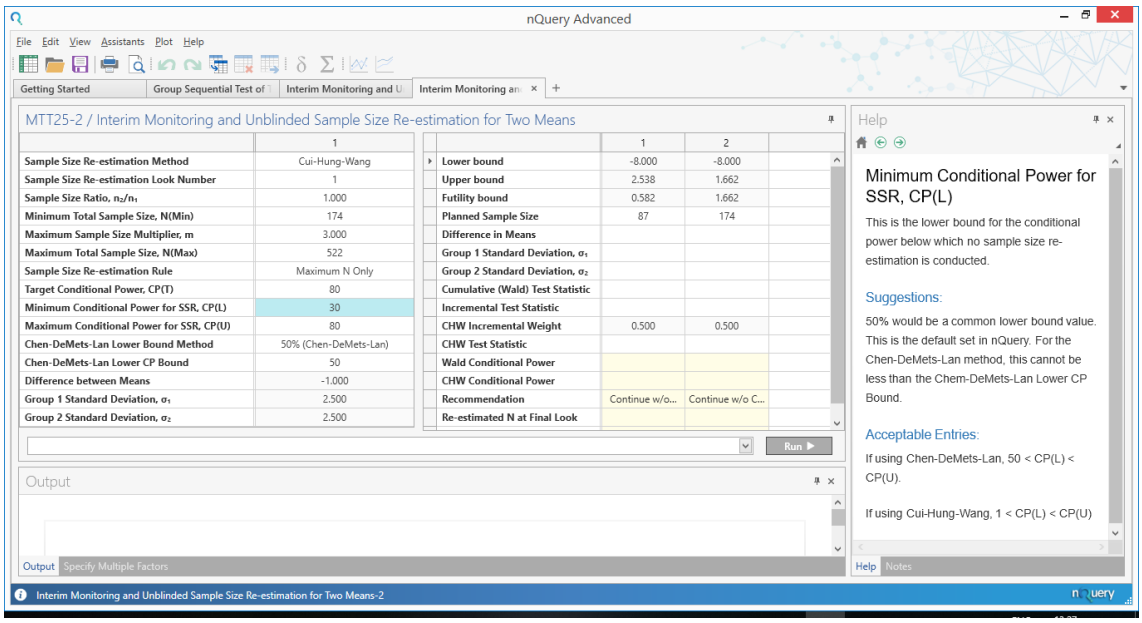



**Figure 4.14:** Cui-Hung-Wang Example Rules Setup

We will now enter the relevant inputs in column 1 of the monitoring table. In this case, we will enter our interim estimates for the design parameters in column 1 (difference in means of -0.5, per-group standard deviations equal to 2.5). We will then use the Incremental Test Statistic Side-Table to calculate the Incremental Test Statistic. When we select the "Incremental Test Statistic" row in column 1, this side-table will open automatically in the window below the main table. In this side-table, we will leave the Incremental Sample Size as its default value of the Planned Sample Size for this column. For designs with a greater number of columns this would equal the Planned Sample Size in the current column minus the Planned Sample Size in the prior column. We then enter our incremental estimates for the difference in means and the group 1 and 2 standard deviations. In this case, these are -0.5, 2.5 and 2.5 respectively. The side-table will automatically calculate the Incremental Test Statistic and transfer it into the main table.

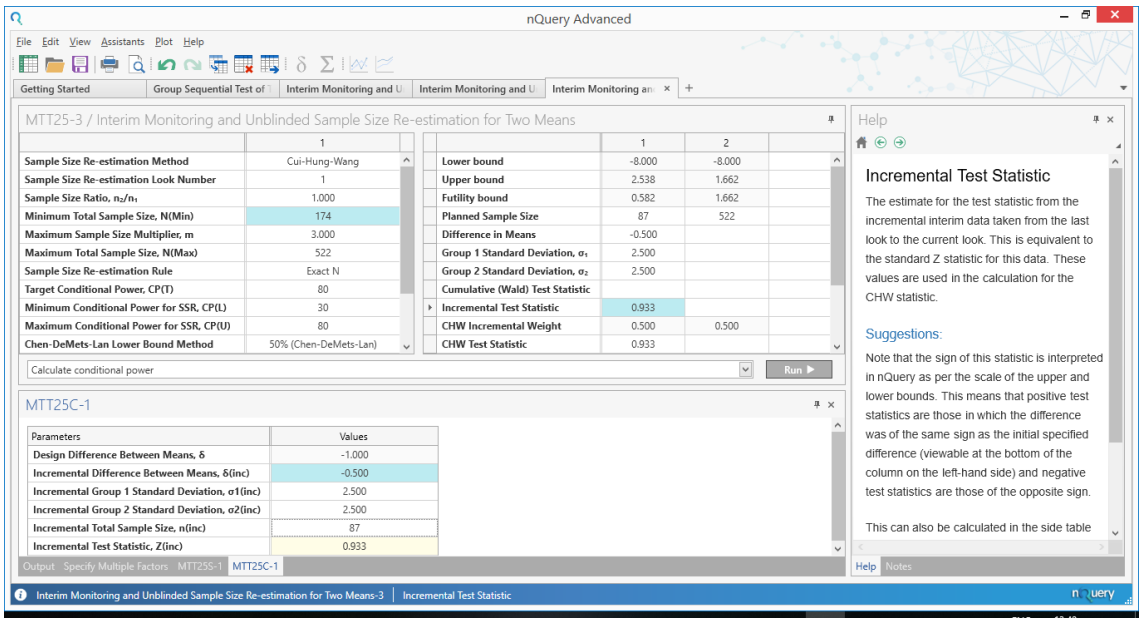



**Figure 4.15:** Cui-Hung-Wang Side-Table Example

Note that in nQuery; "upper", "lower" and "futility" are defined relative to the direction of the pre-specified effect size. For example, if the initial design assumes a negative difference ($\mu_1 > \mu_2$) then positive "upper" interim tests statistics would be those in which the interim difference is also negative and vice-versa. Similarly, "lower" results would be those in the opposite direction. Futility bounds can be above or below zero but the same guidance applies for the Z-statistic calculation. Thus these side-tables will provide the Design effect (e.g. difference in means). Cumulative/Incremental effects (e.g. differences) which have the same sign as the design effect will give a positive test statistic and those of the opposite sign will give a negative test statistic. Thus, in this case entering -0.5 returns an interim test statistic of 0.933 but a value of 0.5 would have returned a test statistic of -0.933. It is important to note that for the looks before the sample size re-estimation look that the Cumulative and CHW tests statistics should be identical. You may want to calculate both at each look to confirm this.

Once these values have been transferred, the conditional power is automatically calculated in this column in the Monitoring Table. In this case, these interim results give a conditional power of 31.38% which falls between our Conditional Power Sample Size Re-estimation range of 30% to 80%. Note that no sample size re-estimation would have occurred with the default range. Thus nQuery gives a recommendation of "Add N & Continue" in the Recommendation row in column 1 and nQuery automatically sets this sample size to the maximum sample size of 522. In this case, this gives an updated conditional power of 75.24%. Note that this conditional power is lower than the target conditional power. With the Exact N rule, the result would have been the same in this case as the target conditional power was

not reached with this maximum sample size. The Planned Sample Size in column 2 is updated automatically to reflect this.

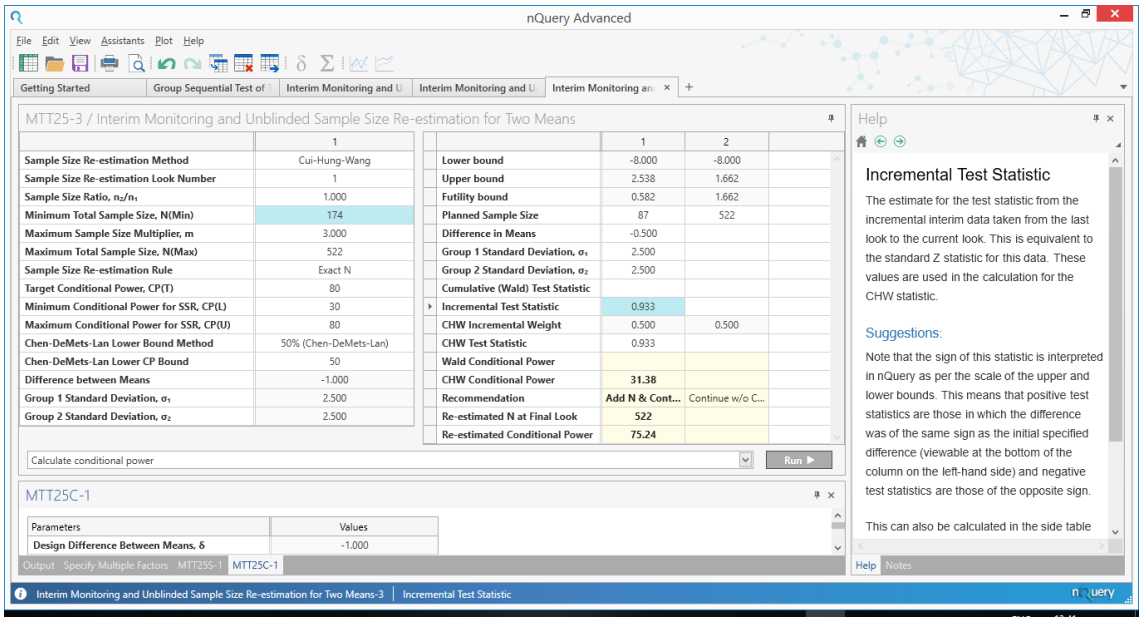



**Figure 4.16:** Cui-Hung-Wang Sample Size Re-estimation Example

We can now complete our interim monitoring by entering our estimates at the end of our (now larger) study. Assume that the final analysis has the same difference in means (-0.5) and per-group standard deviations (2.5). We can enter these parameter values in column 2 and use the Incremental Test Statistic Side-table as before and calculate a CHW test statistic of 2.134 based on an incremental test statistic of 2.086. As this test statistic value is greater than the Upper Bound (1.662) at the final analysis, we find for efficacy for this study at the final look.

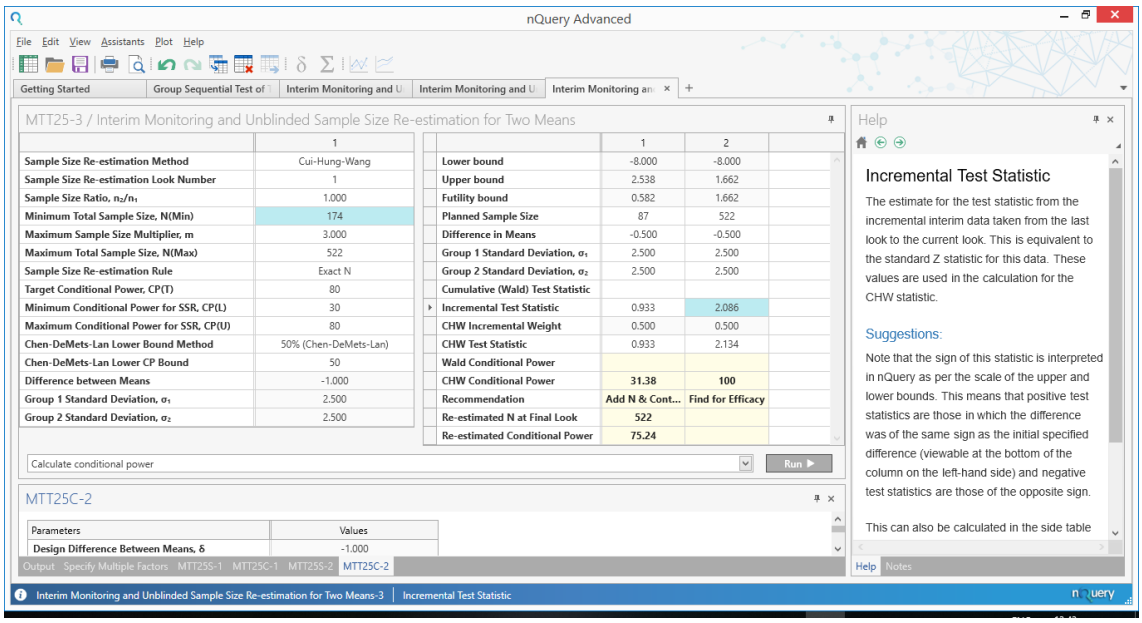



**Figure 4.17:** Cui-Hung-Wang Completed Design Example

# 5 Multiple Comparisons Procedure - Modelling (MCP-Mod)

nQuery has the facility to calculate the sample size or power needed for the "Proof-of-Concept" (PoC) stage of an MCP-Mod trial. As the MCP-Mod methodology has two main purposes (PoC, dose-finding (DF)) different criteria can be used to calculate the required sample size or power.

One approach is to concentrate on the PoC step of the procedure, calculating the sample size needed to ensure a minimum power level for detecting PoC under an assumed placebo and maximum dose-response and an assumed set of dose-response models. The focus here and in nQuery is on this approach.

Another strategy is to focus on the dose-estimating part of the method. This would find the sample size needed for a prespecified minimum precision (i.e. confidence interval width) for the dose-response estimates. This approach is not explored here.

This section will give a background to MCP-Mod, the underlying theory and derivations and the methods employed in nQuery. A demonstration of how to design an MCP-Mod trial in nQuery is also included.

*Note: MCP-Mod is currently only available for users with an nQuery Advanced Pro license. Active packages are displayed in the Packages section in the Home tab. To purchase additional packages, see* `www.statsols.com`.

## 5.1 Multiple Comparison Procedures and Modelling (MCP-Mod) Design

### 5.1.1 Background

MCP-Mod is a type of Phase II trial design methodology used in dose-response studies. Phase II studies have the following two main goals:

1. Proof-of-Concept (PoC): The goal here is to establish that changes in dose lead to significance changes in the efficacy (and/or safety) endpoint and that at least one dose is of clinical interest.
2. Dose-finding (DF): The second goal is to select one or more effective (and safe) doses for evaluation in confirmatory (Phase III) clinical trials or equivalent studies.

Historically, the proof-of-concept and dose finding objectives have often been evaluated in separate trials with proof-of-concept established first and then a dose-finding study used to select doses of interest for further evaluation. In this framework, the proof-of-concept trial was referred to as the Phase IIa trial and the dose-finding trial was referred to as the Phase IIb trial, with the proof-of-concept and dose-response studies using different statistical methods. For proof-of-concept Phase IIa trials a common approach was the multiple comparison procedure (MCP) and for dose-finding Phase IIb trials a common approach was parametric dose-response modelling (Mod).

Under the Multiple Comparison Procedure (MCP) strategy, one evaluates the statistical significance of a contrast test between doses while preserving the family wise error rate (FWER). PoC is established when at least one contrast is statistically significant. If PoC is established, the minimum effective dose (MED) is found. The MED is the lowest dose that is considered statistically (significant) and clinically superior to placebo. The MCP approach regards the dose as a qualitative factor and generally makes few, if any, assumptions about the underlying dose response relationship. However, inferences about the target dose are restricted to the discrete, possibly small, set of doses used in the trial.

The modeling approach is primarily used to estimate the true dose-response curve. Doses that achieve desired clinical effects are estimated via inverse regression methods, which can also be used to evaluate the precision (i.e. confidence intervals) of the estimated doses. The dose is taken to be a quantitative factor, allowing greater flexibility for target dose estimation including potentially doses not evaluated directly in the study. However, the validity of the modelling approach strongly depends on choosing the appropriate parametric dose-response model (e.g. linear, Emax, Beta). Note that modelling can also be used for PoC by testing the significance of the parametric fitted dose-response model versus a flat dose-response null model. However, this approach is less robust than MCP due to the sensitivity to model choice as mentioned previously.

The MCP-Mod methodology was designed to provide a unified strategy to the analysis of data from dose-response studies which combines strengths of the multiple comparison procedure and modelling approaches by having the flexibility of modeling for dose estimation, while preserving the robustness to model misspecification associated with MCP.

The basic idea behind MCP-Mod was first proposed in [Tukey, 1985]. They proposed the use of several trend tests simultaneously and to adjust the resulting p-values for multiplicity. [Bretz et al., 2005] formalized the MCP-Mod procedure and so users are recommended to see this paper for more details on the basic methodology. A more detailed description of the method and it's practical implementation, including power analysis, can be found in [Pinheiro et al., 2006].

## 5.1.2 MCP-Mod Theory

### 5.1.2.1 MCP-Mod Model

The notation used to represent the dose-response model is presented here. Let Y denote the observed response for a given set of patients assigned to one of a set of $k$ doses which we will denote as $d$ where $d = d_1, \ldots, d_k$. We will consider the following dose-response model:

$$Y_{ij} = \mu_{d_i} + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma^2), i = 1, \ldots, k.j = 1, \ldots, n_i$$

Where $\mu_{d_i} = f(d_i, \theta)$ denotes the mean response at dose $d_i$, $\theta$ represents the vector of model parameters, $i$ represents the dose group and $j$ represents the patient within dose group $i$. Let $\mu = \mu_{d_1}, \ldots, \mu_{d_k}$ denote the mean dose response vector.

### 5.1.2.2 MCP-Mod Test (MCP)

Instead of prespecifying a single dose-response model, MCP-Mod uses a set of candidate models covering a suitable range of dose-response shapes. Each of the models in the candidate set is tested using appropriate contrasts and employing an MCP that preserves the FWER. PoC is established if at least one of the model tests is significant. Once PoC is verified, the "best" model(s) among the candidate set is chosen to fit the data and to produce estimated doses using modeling techniques, while still ensuring the appropriate FWER. The "best" model can be evaluated using a model fit statistic such as the lowest MCP p-value or the AIC.

The procedure requires that a set of M candidate models be chosen. Let $f_m(d, \theta_m)$ and $f_m^0(d, \theta_m^0), m = 1, \ldots, M$ denote the model and standardized model functions respectively, where $\theta_m$ and $\theta_m^0$ are the parameters of model $m$ under the unstandardized and standardized model functions respectively.

The hypothesis of interest for each of the dose-response models is $H_0^m : c_m^{'}\mu = 0$, where $c_m = (c_m 1, \ldots, c_m k)$ is the optimal contrast vector for model $m$, and $\sum_{i=1}^k c_m i = 0$. Therefore each candidate models is tested using the following contrast test,

$$T_m = \frac{\sum_{i=1}^k c_{mi} \bar{Y}_i}{S\sqrt{\sum_{i=1}^k c_{mi}^2 / n_i}}, m = 1, \ldots, M.$$

where $\bar{Y}_i = \sum_{i=1}^k Y_{ij}/n_i$ is the arithmetic mean response in dose group $i$, $S^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / v$ is the pooled variance estimator, with $v = \sum_i n_i - k$ degrees of freedom.

Each contrast test can thus be considered as a decision procedure determining whether the given dose-response shape is statistically significant, based on the observed data. The contrast coefficients, $c_m$ are optimal in the sense that they maximize the power to detect the underlying model.

The final detection of a significant dose-response signal (i.e demonstrating PoC), is based on the maximum contrast test statistic, $T_{max} = maxT_1, \ldots, T_M$. Under the null hypothesis of no dose effect the mean response at each dose should be equal. Under this assumption and the dose-response model formulation, it follows that the contrast test statistics follow a multivariate t-distribution with $v$ degrees of freedom. PoC is hence established if

$$T_{max} \geq q_{1-\alpha}$$

where $q_{1-\alpha}$ is the α significance level multiplicity adjusted critical value (i.e., the relevant quantile of the central multivariate t distribution).

#### 5.1.2.3 MCP-Mod Power and Sample Size

Sample size is based on MCP contrast test for the models in the candidate set outlined in subsubsection 5.1.2.2. This method is outlined in Pinheiro et al (2006). Under this sample size procedure, the power of the MCP procedure is determined by the distribution of the maximum contrast test statistic under the alternative hypothesis that the $m^{th}$ dose-response model is the true response.

Under the alternative hypothesis, these contrast test statistics have a non-central multivariate t-distribution. Given the contrast test outlined previously the power for each candidate model $m$ in MCP procedure is defined as

$$(1-\beta)_m = P\left(max_l T_l \geq q_{1-\alpha} | \mu = \mu_m\right) = 1 - P\left(T_1 < q_{1-\alpha} \ldots T_m < q_{1-\alpha} | \mu = \mu_m\right)$$

For sample size determination, it is useful to consider the vector of power values per candidate model. It follows from the model specification and the properties of the multivariate t distribution that the contrast test statistics $T_l$ are jointly distributed as non-central multivariate t distribution with $k(n-1)$ degrees of freedom, the correlation matrix derived from the contrast coefficients (see Pinehiero et al 2006) and per model non-centrality parameters of $\delta_{ml} = \sqrt{n} c_l' \mu_m / \sigma$.

The calculation of quantiles of the non-central multivariate t-distribution is required and in nQuery the calculation of these quantiles is done using the Separation-of-Variables Method as discussed in Genz and Bretz (2002).

For practical purposes, a single summary measure of power is needed for sample size determination. Multiple options are available but the chosen function should be monotonically increasing. In nQuery, three choices of summary function are given. These are the Mean (unweighted), Minimum, and Median. nQuery provides the per-model power so the manual calculation of other summary functions is possible.

#### 5.1.2.4 Dose-Response Models

The list of available dose-response models is given below. Note that most of these models require the pre-specification by the user of 1-3 additional parameters to fully define the model. The definition of each model including these additional parameters is given below with additional parameters given in brackets with model name e.g. **Emax*(ED50)*** indicates ED50 is required unique input for Emax model. These additional parameters alongside the placebo effect, maximum effect, standard deviation and specified dose-levels given by the user elsewhere (see section 5.2) define the model effects.

Note some common terminology including $d$ (dose response at specific dose), $E_0$ (placebo dose response for most models, basal effect for logistic model), $E_{max}$ (maximum dose response, for models other than the Beta and Quadratic models this will equal response at last dose). $f(d, \theta)$ is the unstandardized model and $f^0(d, \theta^0)$ is standardized model.

- **Linear**: This model assumes a positive linear relationship between $E_0$ and $E_{max}$ on the original dose scale (i.e. between response at placebo and largest dose).$\delta$ is the linear slope parameter and this is derived automatically from the model definitions via simultaneous equations.

$$f(d, \theta) = E_0 + \delta d$$

$$f^0\left(d, \theta^0\right) = d$$

- **Linear Log-Dose (*Off*):** This model assumes a positive linear relationship between $E_0$ and $E_{max}$ on the log-dose scale (i.e. between response at placebo and largest dose). The $Off$ parameter is provided to prevent issues with doses equal to zero. This is typically set to a small value such as 0.01 times the maximum dose. $\delta$ is the linear slope parameter and this is derived automatically from the model definitions via simultaneous equations.

$$f(d, \theta) = E_0 + \delta \log(d + Off)$$

$$f^0\left(d, \theta^0\right) = \log(d + Off)$$

- **Emax (*ED50*)**: A positive monotonic concave dose-response curve model. Also known as the hyperbolic Emax model, in contrast to the more flexible sigmoidal Emax model (see below). $ED_{50}$ is the expected dose at which we would expect 50% the maximum dose response to occur.

$$f(d, \theta) = E_0 + E_{max}\frac{d}{ED_{50} + d}$$

$$f^0\left(d, \theta^0\right) = \frac{d}{ED_{50} + d}$$

- **Sigmoidal Emax (*ED50, h*)**: A positive monotonic concave dose-response curve model. It is similar to the Emax model (see above) but has an additional parameter, $h$, (often called the Hill parameter/coefficient) which characterizes the slope of the dose-response curve at the $ED_{50}$ dose. A Hill parameter value of one is equivalent to the (hyperbolic) Emax model with values lower than one implying a shallower slope than the Emax model and values higher than one implying a steeper slope than the Emax model.

$$f(d,\theta) = E_0 + E_{max}\frac{d^h}{ED_{50} + d^h}$$

$$f^0\left(d,\theta^0\right) = \frac{d^h}{ED_{50} + d^h}$$

- **Logistic (*ED50, δ*)**: A positive monotonic concave dose-response curve model. It is closely related to the sigmoidal Emax model which is equivalent to a logistic model on the log(dose) scale, with this Logistic model being on the original dose scale. It requires the specification of ED50 (the dose giving half the maximum dose response effect) and $\delta$ which controls for the slope of the curve. Higher $\delta$ imply a steeper dose-response curve.

$$f(d,\theta) = E_0 + E_{max}/\left\{1 + \exp\left[\left(ED_{50} - d\right)/\delta\right]\right\}$$

$$f^0\left(d,\theta^0\right) = 1/\left\{1 + \exp\left[\left(ED_{50} - d\right)/\delta\right]\right\}$$

- **Exponential (*δ*)**: A positive monotonic convex dose-response curve model. It requires the specification of the parameter $\delta$ which controls the convexity of the model with higher values implying a faster increase in response with increase in dose. $E_1$ is a slope parameter and this is derived automatically from the model definitions via simultaneous equations.

$$f(d,\theta) = E_0 + E_1\left(\exp\left(\frac{d}{\delta}\right) - 1\right)$$

$$f^0\left(d,\theta^0\right) = \exp\left(\frac{d}{\delta}\right) - 1$$

- **Beta (*a, b, Scale*)**: A flexible model that can be used to model non-monotonic dose-response curves. It requires the specification of three parameters. "a" and "b" are the first and second shape parameters for the Beta model and the distributional assumptions of the Beta distribution in terms of shape can be applied here. The scale parameter is used to characterise the range of the Beta function (e.g. scale of 100 would re-scale Beta Distribution to [0,100] rather than the standard [0,1]). The scale parameter should be greater than the largest dose with a value of 1.2 times the maximum dose

being a useful default.

$$f\left(d,\theta\right)=E_0+E_{max}B\left(a,b\right)\left(\frac{d}{Scale}\right)^a\left(1-\frac{d}{Scale}\right)^b,B\left(a,b\right)=\frac{\left(a+b\right)^{a+b}}{a^ab^b}$$

$$f^0\left(d,\theta^0\right)=B\left(a,b\right)\left(\frac{d}{Scale}\right)^a\left(1-\frac{d}{Scale}\right)^b$$

- **Quadratic (δ)**: A flexible model which assumes a quadratic dose-response relationship and can be used to model non-monotonic dose-response curves. It requires only the specification of the parameter δ. δ represents a ratio of the quadratic model parameters, $\beta_1$and $\beta_2$,and is defined below. Note that the concave version of the quadratic model has been implemented in nQuery which imposes the restriction that $\beta_2$is negative. Note that, through the formulation below, specifying a negative value for δ will result in a positive value being found for $\beta_1$and vice versa.

$$f\left(d,\theta\right)=E_0+\beta_1d+\beta_2d^2$$

$$f^0\left(d,\theta^0\right)=d+\delta d^2,\delta=\frac{\beta_2}{|\beta_1|}\&\beta_2<0$$

$\beta_2$ here represents the curvature of the dose-response relationship and $\beta_1$represents the slope of the curve. The values for the$\beta_1$and $\beta_2$parameters of the quadratic model are found through the following equations:

$$\beta_2=-|4\delta^2(ME-PE)|$$

$$\beta_1=\frac{\beta_2}{\delta}$$

where $ME$ represents the maximum response effect expected to be observed in the dose-range and $PE$ represents the effect at placebo.

## 5.2 MCP-Mod Demonstration

### 5.2.1 Background

In nQuery the layout for MCP-Mod is similar to that of most other tables in nQuery. See chapter 1 for an introduction to using nQuery. This section will focus on the additional issues associated with the MCP-Mod table by demonstrating a detailed example of using the table in nQuery.

In nQuery, the inputs required for the MCP-Mod design can be split into three separate categories.

The first of these categories consists of the general test design parameters such as power and the placebo effect. These are equivalent to the main table inputs seen in standard nQuery tables.

The other two categories are the dose levels being assessed in the study and the candidate models selected and their per-model inputs (see subsubsection 5.1.2.4) and outputs e.g. power. These are both contained in the mandatory side-table. See section 3.1 for general guidance on side-tables.

In this section we will consider an example which relates to a real Phase II design in the development program of a drug for the indication of generalized anxiety disorder (GAD). Five active doses are to be used in the study: 10, 25, 50, 100, and 150mg, with an additional placebo arm (corresponding to a 0mg dose). This is the same example considered in section 2 of Pinheiro et al. (2006). Further details will be provided as we complete the example.

### 5.2.2 Main Table (General Design Parameters)

The main table (as discussed in previous chapters of this manual) is used to enter the common parameters which apply to each model. These are the test significance level, whether to use a one or two sided test, the placebo effect, the maximum treatment effect and common standard deviation (equivalent to the residual standard deviation in our model and practically the standard deviation of responses within a dose-level).

Alongside these common parameters, the number of doses and number of models need to be specified. This will set the number of columns in the doses input table and candidates models specification table contained in the side table. See subsection 5.2.3 for details.

Figure 5.1 illustrates how the main table will appear on the initial opening of the table in nQuery.

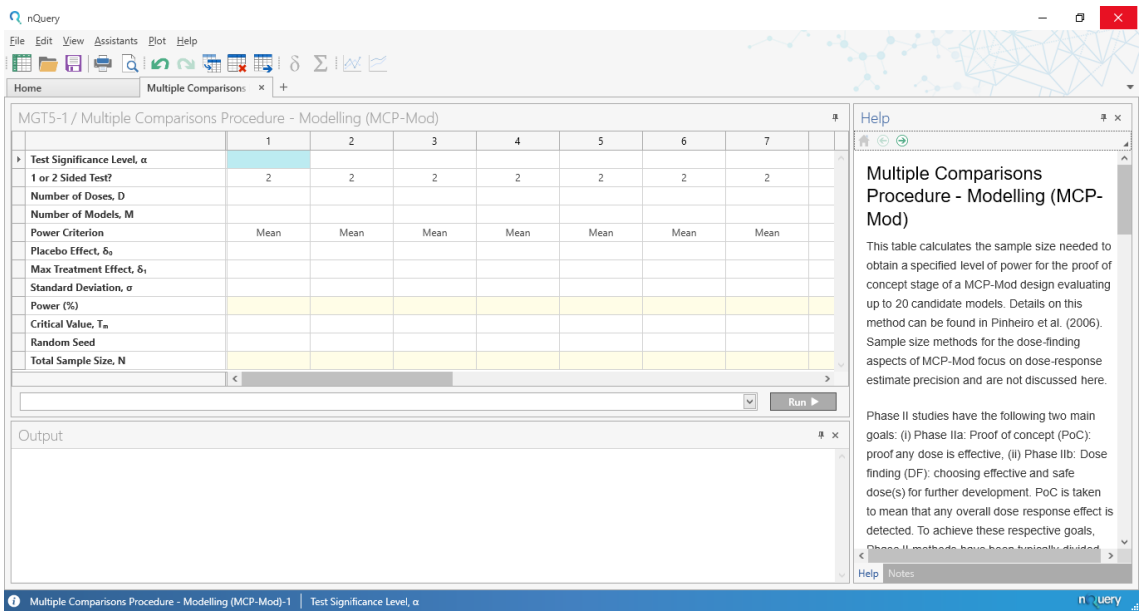



**Figure 5.1:** MCP-Mod Design Main Table

The first step required is to specify the desired significance level (α), whether a one or two sided test is required and the number of doses and candidate models of interest. Select the power criteria (mean, minimum or median power) and then specify the expected placebo effect, the maximum effect, and the common standard deviation. If one wishes to solve for the total sample size then specify the target power. If one wishes to solve for power, then specify the total sample size here or specify the individual sample sizes per dose group in the Dose Level Side table, as discussed in section 5.2.3.1 below.

In this example, considered here we will assume a significance level of 0.05, 80% power, a residual standard deviation of 1, a one sided test, a placebo Effect of 0, a maximum treatment effect of 0.4, 6 doses (5 doses and the placebo dose), a candidate set of 6 models and use the "Mean" power criteria. The power value will be left empty for now and entered after the side-table section is complete.

Optionally, the user can also enter a random seed and critical value. The random seed is used for the simulations used for the multivariate t-distribution method implemented in nQuery. If left blank this will default to a value based on the system time. Setting this option to a specific value allows the user to generate the exact same results for a given set of inputs. In this example, we will set the seed to 1234 to allow exact replication of this example.

The critical value is the critical multivariate t-statistic for the multivariate t-distribution CDF ($q_{1-\alpha}$). To use the optimal t-statistic set the test significance level to desired level and this will make the Critical Value read-only. To use a custom test statistic do not enter the significance level and enter the critical value. As we have entered the significance level, the optimal contrasts will be used here.

For more information on these parameters, see the help cards in the software. The Critical Value parameter will be calculated automatically for us. The completed main table is shown in Figure 5.2.

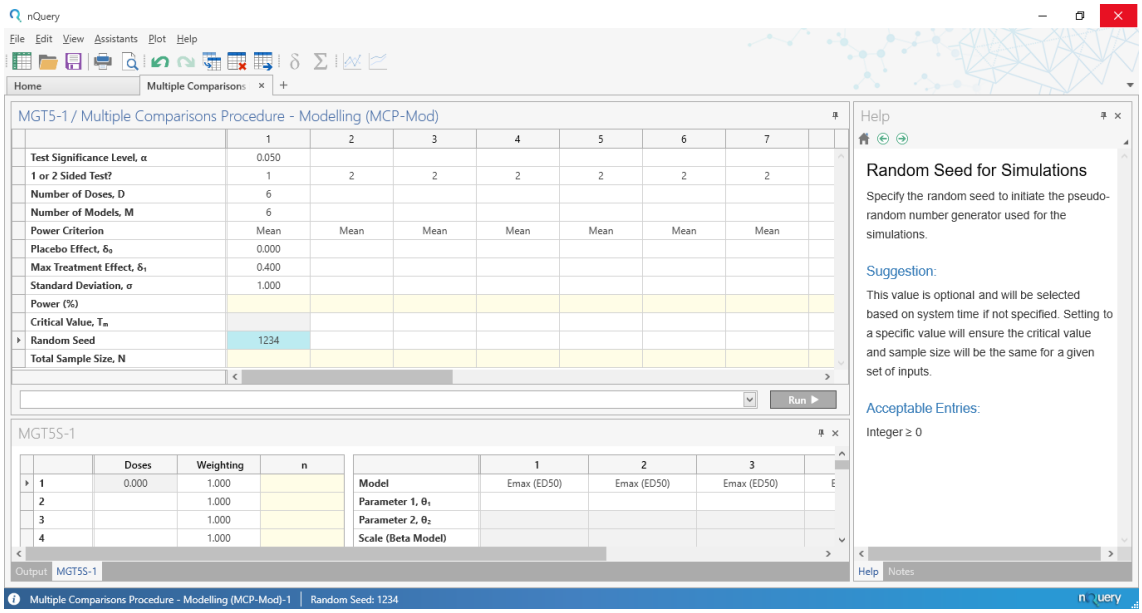



**Figure 5.2:** Completed MCP-Mod Main Table

### 5.2.3 Side-Table (Dose Levels and Candidate Models)

After the number of doses and number of models is specified in a column, the MCP-Mod side-table will automatically appear below the main table window. See subsection 5.2.3 for general guidance on side-tables. In this example, upon opening the side-table will be as per Figure 5.3.

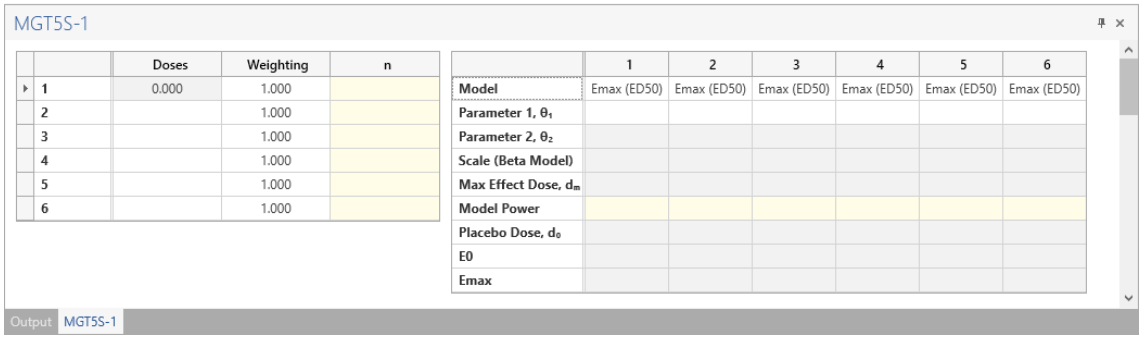



**Figure 5.3:** MCP-Mod Side-Table

The table on the left-hand side is the Doses Table where the dose levels of interest

will be entered. On the right-hand side is the Model Specification Table where candidate models are selected and model specific inputs and outputs are provided.

#### 5.2.3.1 Doses Table

On the left-hand side of the side-table, the planned dose levels table is provided as per Figure 5.4. This table specifies which doses will be considered in this trial and sample size weightings applied to each dose level.

| | Doses | Weighting | n |
|---|---|---|---|
| 1 | 0.000 | 1.000 | |
| 2 | | 1.000 | |
| 3 | | 1.000 | |
| 4 | | 1.000 | |
| 5 | | 1.000 | |
| 6 | | 1.000 | |

**Figure 5.4:** Dose Levels Table

In this table, enter the proposed study dose levels (how much/concentration of the proposed treatment given to each dose level) in the first "Doses" column. Note that the first row entry in the Doses column is automatically set to 0 as this represents the placebo dose (i.e. no treatment given) which is required for this method. It is recommended that doses are entered in increasing order down the column but nQuery will automatically adjust if a non-increasing set of dose levels is given. In this example we will assume the following set of doses: 0, 10, 25, 50, 100, 150 (mg).

Next, enter the sample size weightings per dose level in the second "Weighting" column. By default all entries in this column will equal one, corresponding to equal sample size per dose level. The proportion of the total sample size in a given dose level is given by $W_i / \sum W_i$ where $W_i$ is the weighting per dose level.

If solving for power, one may enter the individual dose group sample sizes in the third column. Otherwise, once a sample size determination is complete the third column, "n", will automatically be filled with the per-dose sample sizes. For this example, assume equal sample size per column and leave the default "1's" inputs. The final dose levels tables will be filled as per Figure 5.5.

| | Doses | Weighting | n |
|---|---|---|---|
| 1 | 0.000 | 1.000 | |
| 2 | 10.000 | 1.000 | |
| 3 | 25.000 | 1.000 | |
| 4 | 50.000 | 1.000 | |
| 5 | 100.000 | 1.000 | |
| 6 | 150. | 1.000 | |

**Figure 5.5:** Completed MCP-Mod Dose Levels Table

#### 5.2.3.2 Model Specification Table

On the right-hand side of the side-table, the model specification table is provided as per Figure 5.6. This table specifies which candidate set of models will be considered and where any additional parameters required per-model are inputted. After a sample size determination, it contains the per-model outputs including model power, minimum effect dose level ($d_0$), maximum effect dose level ($d_m$), the minimum dose effect ($E_0$) and maximum dose effect ($E_{max}$). Note that the maximum and minimum dose levels will automatically be calculated for most models after the Dose Levels table is filled. The major exceptions are the Beta model and the Quadratic model where the maximum effect dose level is not known pre-calculation.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Model** | Emax (ED50) | Emax (ED50) | Emax (ED50) | Emax (ED50) | Emax (ED50) | Emax (ED50) |
| **Parameter 1, $\theta_1$** | | | | | | |
| **Parameter 2, $\theta_2$** | | | | | | |
| **Scale (Beta Model)** | | | | | | |
| **Max Effect Dose, $d_m$** | 150.000 | 150.000 | 150.000 | 150.000 | 150.000 | 150.000 |
| **Model Power** | | | | | | |
| **Placebo Dose, $d_0$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **E0** | | | | | | |
| **Emax** | | | | | | |

**Figure 5.6:** Model Specification Table

In this table, the desired candidate models are selected in each column using the drop down menu in the Model row of the model specification table. After a model is selected, 0 to 3 (equal to the number of parameters given in brackets beside model name) of the rows will become available to edit depending on the model selected. Column cells that require inputs for calculations will be white and those not required

for the given model will be grey. Note that the same model can be selected multiple times, characterized with different model specific parameters, except for the Linear model which has no additional inputs. Details on the models available and the model specific parameters required are given in subsubsection 5.1.2.4.

In this example we will consider the following set of candidate models, with the corresponding parameter values assumed for each model given in brackets after each model:

- Linear
- Emax (ED50 = 25)
- Logistic (ED50 = 50, δ = 10.88111)
- Exponential (δ = 85)
- Beta (a = 0.33, b = 2.31, Scale = 200)
- Beta (a = 1.39, b = 1.39, Scale = 200)

The completed model specification table is given in Figure 5.7.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Model** | Linear | Emax (ED50) | Logistic (ED50, δ) | Exponential (δ) | Beta (a, b, Scale) | Beta (a, b, Scale) |
| **Parameter 1, $\theta_1$** | | 25.000 | 50.000 | 85.000 | 0.330 | 1.390 |
| **Parameter 2, $\theta_2$** | | | 10.881 | | 2.310 | 1.390 |
| **Scale (Beta Model)** | | | | | 200.000 | 200.000 |
| **Max Effect Dose, $d_m$** | 150.000 | 150.000 | 150.000 | 150.000 | 150.000 | 150.000 |
| **Model Power** | | | | | | |
| **Placebo Dose, $d_0$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **E0** | | | | | | |
| **Emax** | | | | | | |

**Figure 5.7:** Completed Model Specification Table Example

In this example, after the final parameter which is required in the model specification table is entered (e.g. "Scale" in column 6), the calculation will not run as the Power row has not been filled in the main table. However, if the main table had been fully filled the calculation would occur automatically. Note that if you do not want every edit to this side-table to automatically start the calculation, leaving one or more of the mandatory inputs in the main table empty is recommended. Once a calculation is complete the per-model power, placebo dose, maximum effect dose, E0 and Emax will be calculated automatically in this table. Note that some models will also calculate additional parameters such as the slope parameter for the linear and exponential models.

### 5.2.4 Results

In this example, the last input is 80% in the Power row in the main table After this is entered, the calculation will start. After the calculation is complete the entire table in nQuery will be as per Figure 5.8. Note that this method (due to the computation of multivariate t-distribution) may take some time to run.

MGT5-1 / Multiple Comparisons Procedure - Modelling (MCP-Mod)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Test Significance Level, α | 0.050 | | | | | | | |
| 1 or 2 Sided Test? | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Number of Doses, D | 6 | | | | | | | |
| Number of Models, M | 6 | | | | | | | |
| Power Criterion | Mean | Mean | Mean | Mean | Mean | Mean | Mean | Mean |
| Placebo Effect, $\delta_0$ | 0.000 | | | | | | | |
| Max Treatment Effect, $\delta_1$ | 0.400 | | | | | | | |
| Standard Deviation, σ | 1.000 | | | | | | | |
| Power (%) | 80.45 | | | | | | | |
| Critical Value, $T_m$ | 2.151 | | | | | | | |
| Random Seed | 1234 | | | | | | | |
| Total Sample Size, N | 372 | | | | | | | |

Calculate the required sample size  Run ▶

MGT5S-1

| Doses | Weighting | n |
|---|---|---|
| 0.000 | 1.000 | 62.000 |
| 10.000 | 1.000 | 62.000 |
| 25.000 | 1.000 | 62.000 |
| 50.000 | 1.000 | 62.000 |
| 100.000 | 1.000 | 62.000 |
| 150.000 | 1.000 | 62.000 |

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Model | Linear | Emax (ED50) | Logistic (ED50, δ) | Exponential (δ) | Beta (a, b, Scale) | Beta (a, b, Scale) |
| Parameter 1, $\theta_1$ | 0.003 | 25.000 | 50.000 | 85.000 | 0.330 | 1.390 |
| Parameter 2, $\theta_2$ | | | 10.881 | 0.083 | 2.310 | 1.390 |
| Scale (Beta Model) | | | | | 200.000 | 200.000 |
| Max Effect Dose, $d_m$ | 150.000 | 150.000 | 150.000 | 150.000 | 25.000 | 100.000 |
| Model Power | 78.820 | 77.094 | 91.535 | 76.182 | 82.445 | 76.624 |
| Placebo Dose, $d_0$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| E0 | 0.000 | 0.000 | -0.004 | 0.000 | 0.000 | 0.000 |
| Emax | | 0.467 | 0.404 | | 0.400 | 0.400 |

Output  MGT5S-1

**Figure 5.8:** Completed MCP-Mod Example

This example results in a sample size of 372 and thus a per dose group sample size of 62 as per [Pinheiro et al., 2006].The exact mean power is automatically updated to 80.44% for this sample size. The per-model powers and other parameters are given in the model specification table in the side-table. The per-model powers ranged from 76.139% for the exponential model to 91.522% for the logistic model. Note the maximum effect dose is the final dose for all models except the Beta models, with the first Beta model having its maximum effect (i.e. the maximum dose effect specified in the main table) at the 25mg dose and the second Beta model having its maximum effect at the 100mg dose. These are found automatically by evaluating the Beta model at each specified dose level and finding the maximum effect dose from the results.

Using a custom critical value will have a shorter calculation time so the optimal

critical value can be used in other examples directly if the significance level and model specification were unchanged.

# 6 nQuery Predict

nQuery Predict is a tool to predict the timing of key clinical trial enrollment and event milestones both before or during a trial. Predictions are generated via simulation using a variety of statistical models based on interim trial data or pre-trial estimates of accrual, event and dropout rates.

This chapter will consist of two sections. In section 6.1, enrollment and event milestone prediction is described at a high level with a technical background provided on how simulations are generated for all the available methods in nQuery Predict. In section 6.2, fully worked demonstrations are provided for each of the main enrollment and event milestone prediction types with guidance on setting up a prediction, the user interface and the main results and outputs for a milestone prediction.

*Note: nQuery Predict is currently only available for users with an nQuery Advanced Expert license. Active packages are displayed in the Packages section in the Home tab. To purchase additional packages, go to* `www.statsols.com`.

## 6.1 Trial Milestone Prediction

nQuery Predict focusses on prediction for two primary milestone types: enrollment and events. Enrollment prediction focusses on when the required number of subjects will be enrolled into the study. Event prediction focusses on when the required number of events in a survival (time-to-event) trial will occur in a study. Terms which are used directly in the nQuery Predict user interface are **highlighted in bold.**

### 6.1.1 Enrollment Prediction

#### 6.1.1.1 Background

The majority of clinical trials have a pre-specified target for enrollment, often derived using a sample size determination calculation. After an enrollment target is set there is a practical requirement for trialists and sponsors to estimate how long it will take to enroll the target sample size. This requires knowledge of the expected accrual rate and accrual pattern using data from prior trials or expert elicitation. Statistical methods can be applied using this knowledge to make informed inferences about the

likely length of the accrual period and resulting study length. These inferences can then be used in the logistics of planning a trial.

However, it is very common for pre-trial projections of accrual rates to be inaccurate. To combat this, interim data can be used to generate more accurate estimates of the accrual rate and pattern and this information can be used to update the expected accrual period and study length. This provides trialists with the knowledge they may need to adjust stakeholder expectations or to get a trial back on track by making on-going changes to the trial e.g. by opening new sites (e.g. hospitals).

nQuery Predict uses simulation from common parametric statistical distributions to model the recruitment process and generate estimates of how long recruitment will likely need to continue to achieve a **Target Sample Size**. Enrollment prediction is available for the following scenarios:

- Enrollment Prediction using Subject-level Data Only
- Enrollment Prediction using Subject and Site-level Data
- Enrollment Prediction using Summary Data

Data-level predictions (**Subject-level Data Only, Subject and Site-level Data**) assume the predictions are being made while the trial is on-going based on interim trial data. Data-level prediction assume the user is providing a subject-level dataset where each column corresponds to a required subject-level user input and each row corresponds to an individual subject. If site-level data is being used, each column corresponds to required site-level user input and each row corresponds to an individual site.

**Summary Data** (also known as **Fixed Parameters**) prediction can be made pre-trial or while the trial is on-going. **Summary Data** assumes the user inputs fixed single parameter estimates to set up the prediction.

For enrollment prediction using **Subject-level Data Only**, the data is required to include the column **Arrival Time** for each subject up until the current time (i.e. time of last accrual).

For enrollment predicting using **Subject and Site-level Data**, the subject data is required to include the **Arrival Time** and **Site ID** for each subject. The site data is required to include the **Site ID** for each site, the expected future enrollment rate per site (**Accrual Rate/Site**), the **Enrollment Cap** per site (i.e. maximum number of subjects that can be recruited from that site) and **Site Initiation Time** (i.e. opening time) for currently **Opened Sites**. Optionally, the **Start Time** and **End Time** for the **Site Initiation Period** for currently **Unopened Sites** can be provided. Note that all **Site ID** values contained in the subject-level data must be contained in the **Site ID** values in the site-level data but not vice-versa (as site-level data can contain unopened sites or open sites which have not enrolled a subject at the **Current Time**).

For Summary Data, the **Current Sample Size** and **Current Time** are required. For pre-trial prediction, these are both set to zero.

The following statistical models are available for generating enrollment times:

- Poisson

The **Poisson** model assumes a constant event rate where the timing of each event is independent from each other. The **Poisson** model is assumed to be a global enrollment process for the **Subject-level Data Only** and **Summary Data** cases. For **Subject and Site-level Data**, an individual Poisson process is specified for each individual site.

#### 6.1.1.2 Technical Details

**Poisson Enrollment Generation** Poisson enrollment generation is created by sampling the expected inter-arrival time between each accrual. These can be shown to follow the exponential distribution for a Poisson process [Ahrens and Dieter, 1972]. The probability distribution function (pdf) and generating model are as follows :

$$f(x) = \lambda \exp(-\lambda x)$$

$$x = -ln(u)/\lambda$$

$$\lambda = \frac{n}{t}$$

where $f(x)$ is the pdf, $\lambda$ is the Poisson rate, $x$ is the random inter-arrival time generated, $u$ is a value generated from a standard uniform distribution ($U(0,1)$) and $n$ is the sample size accrued up to current time, $t$. The definition of $\lambda$ given in the third equation above is the default value given in nQuery Predict for the Poisson accrual rate.

For subject-level data only or Summary Data, the number of inter-arrival times generated equals the **Target Sample Size** minus the **Current Sample Size**. The total length of the remaining accrual period after the **Current Time** is given by the sum of all the inter-arrival times generated to simulate the necessary number of subjects. Each subject's arrival time after the **Current Time** equals the cumulative sum of inter-arrival times until and including that subject's inter-arrival time. These times relative to the study start are calculated by adding the **Current Time** to them e.g. **Total Accural Period Length = Cumulative Sum(x) + Current Time.**

For subject and site-level data, inter-arrival times are generated for each site up to the smaller of the **Enrollment Cap** minus the **Current Sample Size** in a site or the total number of required simulated subjects (**Target Sample Size - Current Sample Size**). The effective arrival time for each subject is the cumulative sum

within each site of the inter-arrival times up until and including that subject's inter-arrival time. The effective arrival times across all sites are then compared with only the lowest required number of arrival times (i.e. target sample size minus current sample size) used.

**Site Opening Time Generation** For unopened sites, the site opening time is generated using a uniform distribution between the specified site initiation start time and end time. This gives a site initiation times as follows:

$$OT = ST + (ET - ST)u$$

where $OT$ is the simulated opening time for a site, $ST$ is the specified site initiation window start time, $ET$ is the specified site initiation window end time and $u$ is a value generated from a standard uniform distribution ($U(0, 1)$).

## 6.1.2 Events Prediction

### 6.1.2.1 Background

In survival analysis (also referred to as time-to-event analysis), the primary endpoint of interest is the length of time until a subject experiences an event of interest. Common events of interest are death (overall survival - OS) or disease progression (progression free survival - PFS).

Unlike other common endpoints, the follow up per subject is unknown a priori since a subject's follow up time will depend on when the endpoint of interest occurs for a subject i.e. their survival time. Importantly, the inference at the end of the study will primarily be based on the subjects who have had the event of interest since subjects who dropout or are censored will not have an explicit time to event. Therefore the study length will depend on when the required number of events occurs. This milestone **Target Number of Event**s is often derived from a power analysis for example. In survival analysis the power is related to the number of events with the sample size an estimate of the number of subjects needed to achieve that number of events given study constraints such as accrual and event rates.

From a milestone event prediction perspective, this requires the modelling of both the enrollment process (unless enrollment is complete at the time of an interim analysis) and the survival process. For the survival process, there are number of potential models available. In addition, competing processes which prevent a subject from having the event of interest in the future, such as dropout or competing events, may also be modelled.

The choice of administrative right-censoring for subject who do not have the event of interest or any competing event during the study is also an important consideration.

Two common choices are an event-driven (also known as variable follow up) design and a fixed follow up design. In the event-driven design, the study is continued until the required number of events has occurred at which point the remaining subjects in the study are right-censored. In a fixed follow up design, each subject has a fixed maximum follow up period and if they do not have the event during this period then they are administratively right-censored.

Given the additional complexity of generating simulations for the enrollment, event and dropout/competing event processes, a wider number of scenarios and models are available for events prediction. Enrollment simulation, where needed, is conducted as per subsection 6.1.1. This gives the following scenarios for events prediction:

- Blinded Events Prediction with Enrollment Ongoing using Subject-level Data Only
- Blinded Events Prediction with Enrollment Ongoing using Subject and Site-level Data
- Blinded Events Prediction with Enrollment Ongoing using Summary Data
- Unblinded Events Prediction with Enrollment Ongoing using Subject-level Data Only
- Unblinded Events Prediction with Enrollment Ongoing using Subject and Site-level Data
- Unblinded Events Prediction with Enrollment Ongoing using Summary Data
- Blinded Events Prediction with Enrollment Complete using Subject-level Data
- Blinded Events Prediction with Enrollment Complete using Summary Data
- Unblinded Events Prediction with Enrollment Complete using Subject-level Data
- Unblinded Events Prediction with Enrollment Complete using Summary Data

Data-level predictions (**Subject-level Data Only, Subject and Site-level Data**) assume the predictions are being made while the trial is on-going based on interim trial data. Data-level prediction assume the user is providing a subject-level dataset where each column corresponds to a required subject-level user input and each row corresponds to an individual subject. If site-level data is being used, each column corresponds to required site-level user input and each row corresponds to an individual site.

**Summary Data** (also known as **Fixed Parameters**) prediction can be made pre-trial or while the trial is on-going. **Summary Data** assumes the user inputs fixed single parameter estimates to set up the prediction.

**Blinded Events** and **Unblinded Events** refers to whether the treatment assignment for each subject is assumed to be known (and will be used for event milestone

prediction) by the trialist. **Unblinded**, also known as a comparative analysis, assumes treatment assignment is known and **Blinded Events**, also known as non-comparative, assumes treatment assignment is unknown. If **Blinded Events** is selected, the event (and dropout) model will assume a "global" event process for all subjects. If **Unblinded Events** is selected, separate event process models are used to individually model the two treatment groups.

**Enrollment Ongoing** indicates that enrollment is still happening at the point that the current prediction is being made. **Enrollment Complete** indicates that the enrollment process is complete at the time the current prediction is being made. **Enrollment Complete** means that **Subject and Site-level Data** cannot be used as site-level data is only used for modelling the enrollment process and has no input on the event model.

If using **Subject-level Data** (with or without site-level data), columns for the **Arrival Time** per subject, the **Status Indicator** and **Time on Study** are required for each subject.

The **Arrival Time** is time from the start of the study until when a subject entered the study and is the same as for the Enrollment Prediction case.

The **Status Indicator** column indicates whether a subject has the **Event Status, Dropout Status** or **Censored Status** at the current time. The value in the **Status Indicator** column which corresponds to each status is set by the user and can be manually inputted for any status that has not occurred at the **Current Time** (see subsection 6.2.7 for details).

**Event Status** indicates a subject has had the event of interest (e.g. death) at the current time.

**Dropout Status** indicates a subject had dropped out at the current time. **Dropout Status** can also be used to indicate subjects who have had a "competing" event which prevents the subject from having the event of interest in the future.

The **Censored Status** (also referred to as **Available Status**) indicates a subject who has neither had the event of interest or a dropout event and therefore would be censored if the study was to end at the **Current Time**. These subjects are still available to have the event of interest (or dropout) and therefore will have a predicted event (and dropout) time simulated for them.

The **Time on Study** column gives the value for how long a subject was in the study. The definition will depend on whether they have a **Event/Dropout Status** or **Censored Status**. For subjects who have the **Event Status** or **Dropout Status** for their **Status Indicator**, the **Time on Study** will be the length of time from their **Arrival Time** until they have had an event or dropped out. For subjects who have the **Censored Status** for their **Status Indicator**, the **Time on Study** is the length of time from their **Arrival Time** until the **Current Time**.

If using **Subject-level Data** (with or without **Site-level Data**) alongside **Unblinded Events,** a **Treatment ID** column is required which indicates which treat-

ment group each subject is in. The two groups are named the **Control Group** and **Treatment Group** and the value in the **Treatment ID** column which corresponds to each of these groups is set by the user (see subsection 6.2.7 for details).

As per enrollment prediction, if using **Subject and Site-level Data** when **Enrollment Ongoing** then the subject data is required to include the **Site ID** column. The site data is required to include the **Site ID** for each site, the expected future enrollment rate per site (**Accrual Rate/Site**), the **Enrollment Cap** per site (i.e. maximum number of subjects that can be recruited from that site) and **Site Initiation Time** (i.e. opening time) for currently **Opened Sites**. Optionally, the **Start Time** and **End Time** for the **Site Initiation Period** for currently **Unopened Sites** can be provided. Note that all **Site ID** values contained in the subject-level data must be contained in the **Site ID** values in the site-level data but not vice-versa (as site-level data can contain unopened sites or open sites which have not enrolled a subject at the **Current Time**).

For **Summary Data** with **Blinded Events**, the **Current Sample Size, Current Number of Events, Current Number of Dropouts** and **Current Time** are required. The **Current Number Censored** (i.e. available) is automatically calculated for the user by subtracting the **Current Events** and **Current Number of Dropouts** from the **Current Sample Size.**

For **Summary Data** with **Unblinded Events**, the **Current Sample Size (Control), Current Sample Size (Treatment), Current Number of Events (Control), Current Number of Events (Treatment), Current Number of Dropouts (Control), Current Number of Dropouts (Treatment)** and **Current Time** are required. **(Current/Treatment)** in brackets indicates the treatment group. The **Current Number Censored (Control)** (i.e. available) is automatically calculated for the user by subtracting the **Current Number of Events (Control)** and **Current Number of Dropouts (Control)** from the **Current Sample Size (Control).** The **Current Number Censored (Treatment)** (i.e. available) is automatically calculated for the user by subtracting the **Current Number of Events (Treatment)** and **Current Number of Dropouts (Treatment)** from the **Current Sample Size (Treatment).**

The following models are available for generating survival times:

- Exponential
- Piecewise Exponential (available only for Subject-level Data Only, Subject and Site-level Data)
- Weibull (available only for Subject-level Data Only, Subject and Site-level Data)

The **Exponential** model assumes a constant hazard (event) rate over time where events are independent from time on study. The **Piecewise Exponential** model assumes a piecewise survival curve where there are pre-specified time periods which have individual differing exponential hazard rates. The **Weibull** model is a more

flexible model than the **Exponential** which allows the hazard rate to change over time and depend on how long a subject has been in the study (see subsubsection 6.1.2.2 for details).

The following models are available for generating dropout times:

- Exponential
- Piecewise Exponential (available only for Subject-level Data Only, Subject and Site-level Data)

The dropout models work the same as the event models.

For the **Enrollment Ongoing** case, the same models are used as for **Enrollment Prediction** above. See subsection 6.1.1 for details.

#### 6.1.2.2 Technical Details

**Constant Exponential Event Prediction** Exponential event times are generated directly from the exponential distribution [Ahrens and Dieter, 1972]. Note that in nQuery Predict this refers to the model where **Response Distribution = Exponential** and **Number of Hazard Pieces = 1** at the **Event & Dropout Information** step of **Setup** (see subsection 6.2.7)

The probability distribution function (pdf) and generating model are as follows:

$$f(x) = \lambda \exp(-\lambda x)$$

$$x = -ln(u)/\lambda$$

$$\lambda_i = \frac{E_i}{F_i}$$

where $f(x)$ is the pdf, $\lambda$ is the exponential hazard rate, $x$ is the random survival time generated, $u$ is a value generated from a standard uniform distribution ($U(0,1)$), $E_i$ is the number of events that have occurred in the group of interest and $F_i$ is the total follow up time (i.e. sum of follow up) for all subjects in the group of interest up until the current time.

The definition of $\lambda$ given in the third equation above is the default value given in nQuery Predict for the event **Hazard Rate**. For **Blinded Events**, the total events and total follow up for all subjects is used to calculate a global exponential hazard rate applied to all subjects. For **Unblinded Events**, the events and total follow up for each group are used to calculate an exponential hazard rate for each group separately.

The survival time generated here, $x$, is survival time from the **Current Time** when the prediction is generated. If interim data was used, for subjects who were still available to have the event in that data (**Censored Status**) their time-to-event will equal the survival time generated using this model plus their **Time on Study** up until the current time. For the time of event relative to the study start (rather than time-to-event after **Arrival Time**), add the time-to-event (i.e. **Time on Study** + $x$) and **Arrival Time.**

**Piecewise Exponential Event Prediction** Piecewise exponential event times are generated using the memoryless independent increments method. Note that in nQuery Predict this refers to the model where **Response Distribution = Exponential** and **Number of Hazard Pieces > 1** at the **Event & Dropout Information** step of **Setup** (see subsection 6.2.7)

In the piecewise exponential model, there are $K$ sequential independent time periods (the first of which starts at the current study time (defined as ($\tau_1 = 0$)) with an independent exponential hazard rate which applies over each pre-specified time period. Each time period has a length equal to $\tau_{i+1} - \tau_i$.

$$P = 1, 2, ..., K$$

$$\mathrm{T} = \tau_1, \tau_2, ...\tau_K, \tau_{K+1}; \tau_1 = 0, \tau_{K+1} = \infty$$

$$\Lambda = \lambda_1, \lambda_2, ..., \lambda_K$$

$$L = l_1 = \tau_2 - \tau_1, l_2 = \tau_2 - \tau_1, ...l_K = \tau_K - \tau_{K+1} = \infty$$

$$X = x_1, x_2, ..., x_K$$

$$x_i = -ln(u_i)/\lambda_i$$

$$\begin{cases} x = \tau_i + x_i & x_i \leq L_i \\ Repeat_{i+1} & x_i > L_i \end{cases}$$

where $P$ is sequential periods from 1 to $K$, $T$ is the start time ($\tau_i$) of each time period (first period starts at 0, $K+1$ "starts" at infinity), $\Lambda$ is the exponential hazard rates ($\lambda_i$) assigned to each time period starting at time $\tau_i$, $L$ is the length of the each time period (this equals the start time of the next period minus the start time of

the current period $(\tau_{i+1} - \tau_i)$), $X$ is the random survival time generated $(x_i)$ for each period, $u_i$ is a value generated from a standard uniform distribution $(U(0,1))$ for each period and $Repeat_{i+1}$ indicates that the step above should be repeated for the $(i+1)^{th}$ case.

In the memoryless method, an exponential survival time $(x_1)$ is first randomly generated for the first time period $(P = 1)$ and its corresponding exponential hazard rate $(\lambda_1)$. If the generated survival time $(x_1)$ for the first period is less than or equal to the length of the first time period $(l_1 = \tau_2 - \tau_1 = \tau_2)$ i.e. $(x_1 \leq l_1)$ then we stop the algorithm and set the overall survival time equal to the start time of the first period plus the generated survival time for the first period i.e. $(x = \tau_1 + x_1 = x_1)$.

However, if the first survival time exceeds the length of the first time period i.e. $(x_1 > l_1)$, it is assumed the subject survived at least up until the start of the next (second) period $(\tau_2)$ and the algorithm continues. Next for the second time period $(P = 2)$, an exponential survival time is generated for the second time period $(x_2)$ with its corresponding exponential hazard rate $(\lambda_2)$. If the generated survival time $(x_2)$ for the second period is less than or equal to the length of the first time period $(l_2 = \tau_3 - \tau_2)$ i.e. $(x_2 \leq l_2)$ then the algorithm stops and the overall survival time is set equal to the start time of the second period plus the generated survival time for the second period i.e. $(x = \tau_2 + x_2)$.

However, if the second survival time exceeds the length of the second time period i.e. $(x_2 > l_2)$, it is assumed the subject survived at least up until the next $(P = 3)$ period's start time $(\tau_3)$and the algorithm continues. These steps are continued until either the algorithm stops due to the generated survival time in a given period being less than or equal to the length of that time period i.e. $(x_i \leq l_i)$ or until the $K^{th}$ period is reached at which point the overall survival equals the final period start time plus the final generated survival time i.e. $(x = \tau_K + x_K)$.

In nQuery Predict the piecewise exponential default is the same as the **Constant Exponential**, with time period starts $(\tau_i)$ that increase from zero in increments of one for each additional number of hazard pieces (time periods) specified with the hazard rate equal to the default from the constant exponential hazard rate case (see section 6.1.2.2) in each period.

The survival time generated here $(x)$ is the survival time from the **Current Time** i.e. the time when the prediction is generated. If interim data was used, for subjects who were still available to have the event in that data (**Censored Status**) their time-to-event will equal the survival time generated using this model plus their **Time on Study** up until the current time. For the time of event relative to the study start (rather than time-to-event after **Arrival Time**), add the time-to-event (i.e. **Time on Study** + $x$) and **Arrival Time.**

Note that there are alternative methods proposed for the generation of piecewise exponential survival times such as the CDF method ([Walke, 2010]).

**Weibull Event Prediction** Weibull survival times are generated from Weibull distribution. As the hazard rate changes over time for a subject, the calculation of survival time will differ for subjects who were already in the study and available to have the event (**Status Censored**) at the time of the current prediction (conditional) and subjects who will enter at or after the current prediction time (unconditional).

For subjects who are in the study at the time of the prediction, the conditional survival time is calculated as follows:

$$x = \frac{1}{\mu}\left[(\mu T)^{\nu} - \ln(1-u)\right]^{\frac{1}{\nu}} - T$$

where $x$ is the generated survival time from the current time, $\mu$ is Weibull scale parameter, $\nu$ is the Weibull shape parameter, $T$ is the time spent on the trial (**Time on Study**) up until the current time and $u$ is a value generated from a standard uniform distribution ($U(0,1)$).

For subjects who are enrolled at or after the current prediction time, the unconditional survival time is calculated as follows:

$$x = \frac{1}{\mu}\left(-\ln(1-u)\right)^{\frac{1}{\nu}}$$

where $x$ is the generated survival time from the current time, $\mu$ is Weibull scale parameter, $\nu$ is the Weibull shape parameter and $u$ is a value generated from a standard uniform distribution ($U(0,1)$).

Note that if the Weibull shape parameter ($\nu$) is equal to one then the Weibull distribution reduces to an exponential distribution with rate equal to the Weibull scale parameter ($\mu$). When the shape parameter is less than one, this indicates that the event rate reduces over time and when the shape parameter is greater than one, this indicates that the event rate will increase over time.

nQuery Predict will provide the best fit Weibull model by default based on the interim subject-level data. The two parameter estimates can be derived via maximum likelihood estimation using numerical optimization over the Weibull likelihood function below:

$$L(\mu,\nu) = \prod_{i=1}^{n}\left[(\mu\nu)(\mu t_j)^{\nu-1}\right]^{\delta_j} e^{-(\mu t_j)^{\nu}}$$

where $L(\mu,\nu)$ is the likelihood function, $\mu$ is Weibull scale parameter, $\nu$ is the Weibull shape parameter, $t_j$ is the length of time a subject has been in the study (**Time on Study**) and $\delta_j$ is an indicator function which equals 1 for subjects who had the event of interest (**Status Event**) and zero otherwise. Note that nQuery applies optimization over the log-likelihood, an example of which can be found in [Nielsen, 2011].

**Dropout Generation** The exponential and piecewise exponential models are available for the dropout process. These work the same as for the event generation methods outlined above.

## 6.2 nQuery Predict Demonstration

This section will demonstrate how to create milestone predictions using nQuery Predict. The demonstration is split into nine subsections as follows:

1. Creating an nQuery Predict Workspace (subsection 6.2.1)
2. nQuery Predict Workspace Layout (subsection 6.2.2)
3. Dataset Management (subsection 6.2.3)
4. nQuery Predict Example Datasets (subsection 6.2.4)
5. Select the Type of Prediction (see subsection 6.2.5)
6. Enrollment Prediction Demonstration (subsection 6.2.6)
7. Event Prediction Demonstration (subsection 6.2.7)
8. Results Output Summary (Reports/Plots/Tables) (subsection 6.2.8)
9. Other Features (subsection 6.2.9)

**Creating an nQuery Predict Workspace** (subsection 6.2.1) gives an overview of how to create a new nQuery Predict workspace both with and without interim data.

**nQuery Predict Workspace Layout** (subsection 6.2.2) gives an overview for the layout of an nQuery Predict workspace and the primary purpose of each of the main user interface elements.

**Dataset Management** (subsection 6.2.3) gives an overview of how to open, edit, upload and remove datasets within a nQuery Predict workspace.

**nQuery Predict Example Datasets** (subsection 6.2.4) gives an overview of the example subject (SubjectData.csv) and site-level (SiteData.csv) datasets provided with nQuery and a basic introduction to the required fields needed in nQuery Predict, alongside other important information such as range restrictions and optional values.

**Select the Type of Prediction/Setup Introduction** (see subsection 6.2.5) gives an overview of Step 1 of the Setup stage of a prediction where the **Target, Input** and **Enrollment Status** are specified. A brief introduction to the **Setup** stage of the nQuery Predict workspace is also provided

**Enrollment Prediction Demonstration** (subsection 6.2.6) gives an overview of enrollment milestone prediction using Subject Data Only, using Subject + Site Data or using Summary Data inputs.

**Event Prediction Demonstration** (subsection 6.2.7) gives an overview of event milestone prediction for survival (time-to-event) analysis using Subject Data Only, using Subject + Site Data or using Summary Data inputs and with or without enrollment being complete.

**Results Output Summary** (subsection 6.2.8) gives a high-level overview of the outputs from prediction including the **Reports, Plots** and **Tables** available. Note that prediction-specific details for different prediction types are provided in the **Prediction Demonstration** sections above.

**Other Features** (subsection 6.2.9) gives an overview of other features available in an nQuery Predict workspace such as exporting results and error checking.

The first five subsections are common across all predictions. Enrollment Prediction Demonstration and Event Prediction Demonstration will have different **Setup** steps and results (**Reports/Plots/Tables**) depending on choices at the **Select Type of Prediction** step. These are broken up further into predictions using an **Input** of **Subject Data Only, Subject + Site Data** or **Summary Data** (Fixed Parameters), and **Enrollment Status** (Event Prediction Only). **Results Summary** and **Other Features** have both common and prediction-specific elements which are both highlighted.

### 6.2.1 Creating an nQuery Predict Workspace

The nQuery Predict Workspace is a tab in which a user will setup, run and view the results of their milestone prediction. To create a new nQuery Prediction Workspace the user can select the "Create Prediction" button in the Home Screen tab in top-left Home quadrant (1), select the Predict menu in the file menu at the top of the software window (2) or select the Predict toolbar icon in the toolbar (3). These three options are highlighted in Figure 6.1.

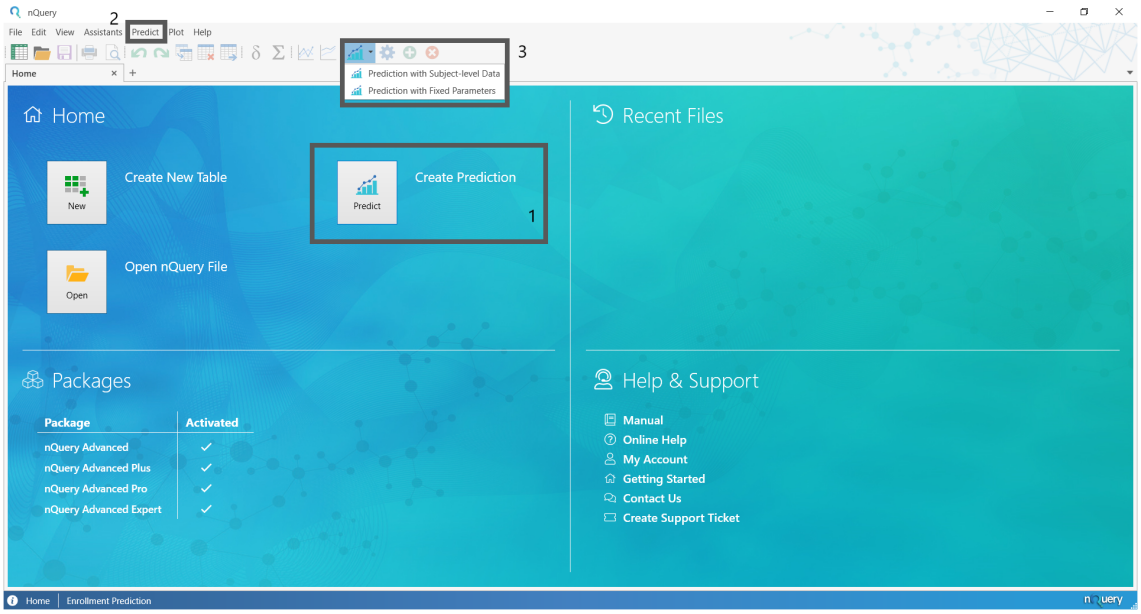



**Figure 6.1:** nQuery Predict Workspace Creation Options

If the Predict > New Prediction file menu or New Prediction toolbar option are used the user is presented with the option to either create a "Prediction with Subject-level Data" or "Prediction with Fixed Parameters". These two methods are described below.

#### 6.2.1.1 Prediction with Subject-level Data

If "Prediction with Subject-level Data" is selected, the "New Prediction - Import Datasets" window is opened. This window is also shown if the "Create Prediction" button is selected from the Home tab. This is shown in Figure 6.2.

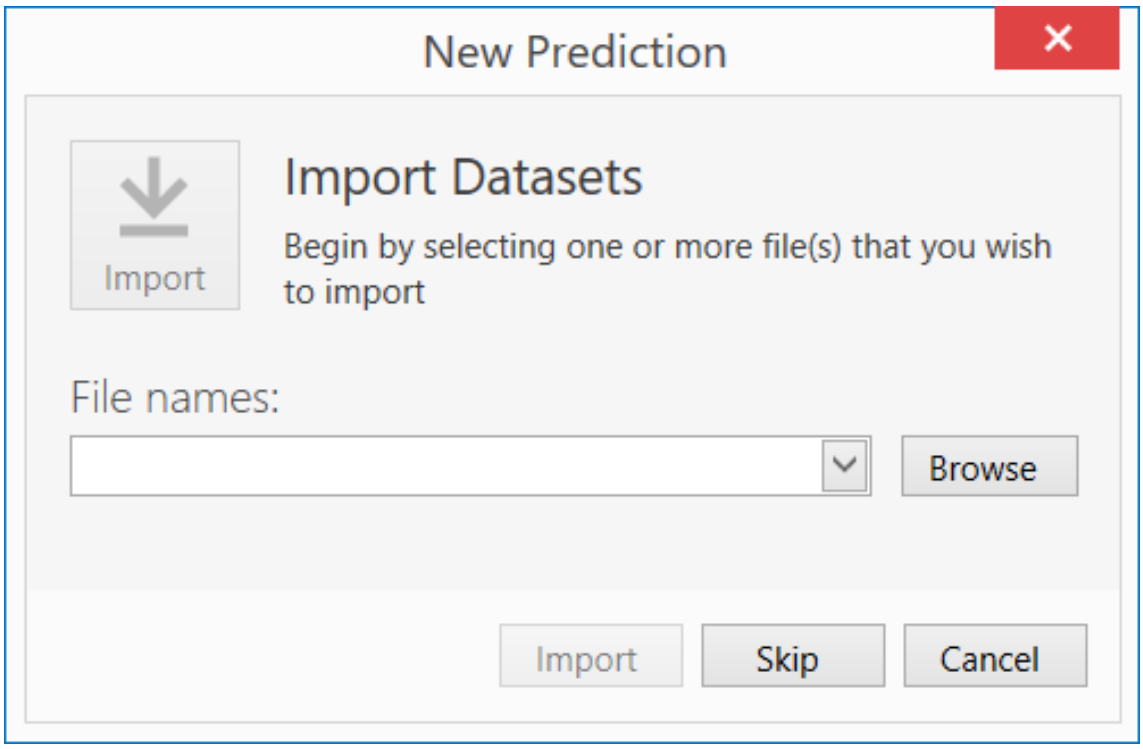



**Figure 6.2:** New Prediction - Import Datasets Window

In the "New Prediction - Import Datasets" window, the user can import csv files containing the required interim data for milestone prediction. To import datasets, select the Browse button to open the Open window, shown in Figure 6.3.

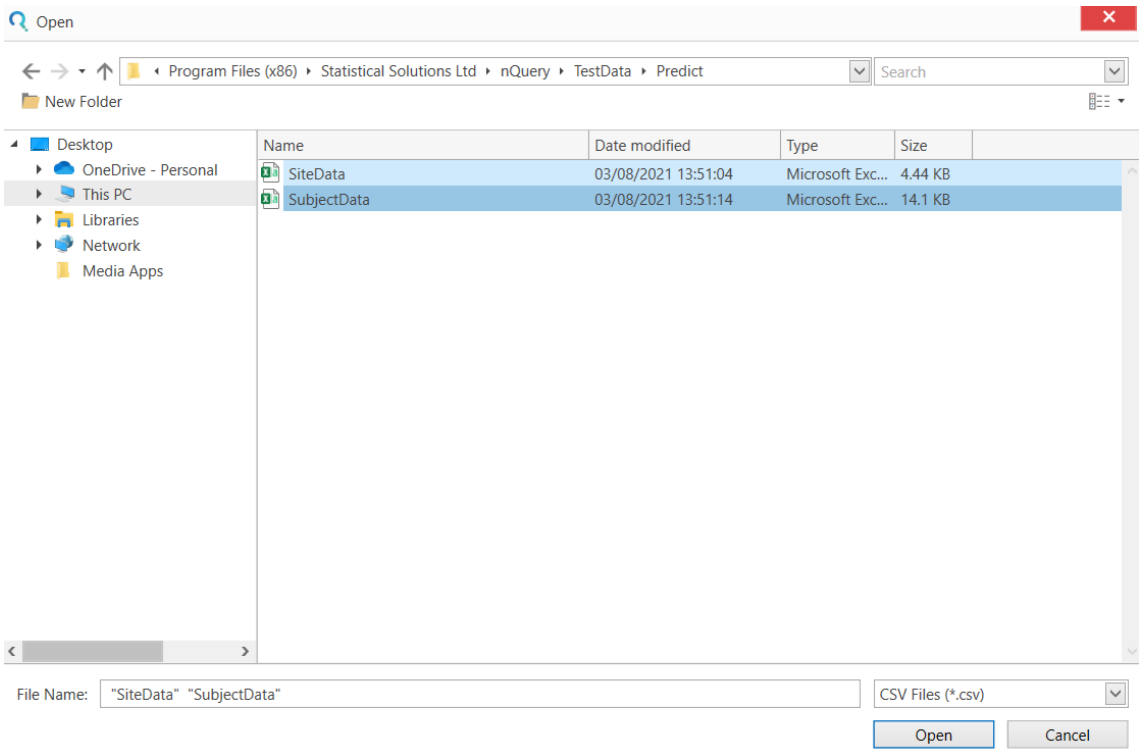



**Figure 6.3:** Open Window for nQuery Predict Data

nQuery Predict comes with an example subject-level dataset (named SubjectData.csv) and site-level dataset (named SiteData.csv). These simulated datasets were created based on available summary data from a real clinical trial. These datasets can be found in the nQuery installation folder which by default is C:\Program Files (x86)\Statistical Solutions Ltd\nQuery\TestData\Predict. All demonstration examples in the remainder of this chapter will use these example files.

To find a file the user can use the Navigation Pane on the left of the Open window, enter the file location directly in the Address Bar at the top or use the Search field in the top-right to find a file within the current folder and subfolders.

To open a file, the user first selects the required file(s) in the main window. To select multiple files, select a first file then select another file while holding down SHIFT (to select all files between the first and second file) or CTRL (to select multiple individual files). All files currently selected will be displayed in the File Name field at the bottom of the Open window. When the required files are selected, select Open in the bottom-right to return to the "New Prediction - Import Datasets" window.

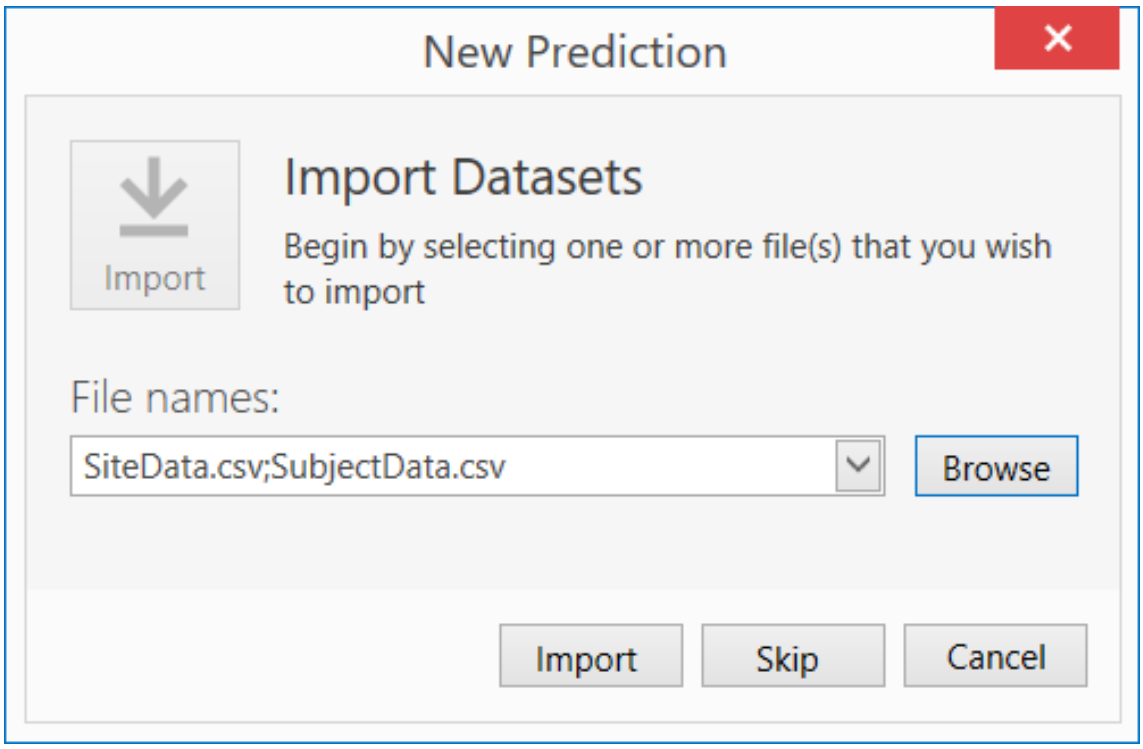



**Figure 6.4:** New Prediction - Import Datasets Complete Example

To create the nQuery Predict workspace, select the Import button. The new nQuery Predict workspace will be as per Figure 6.5.

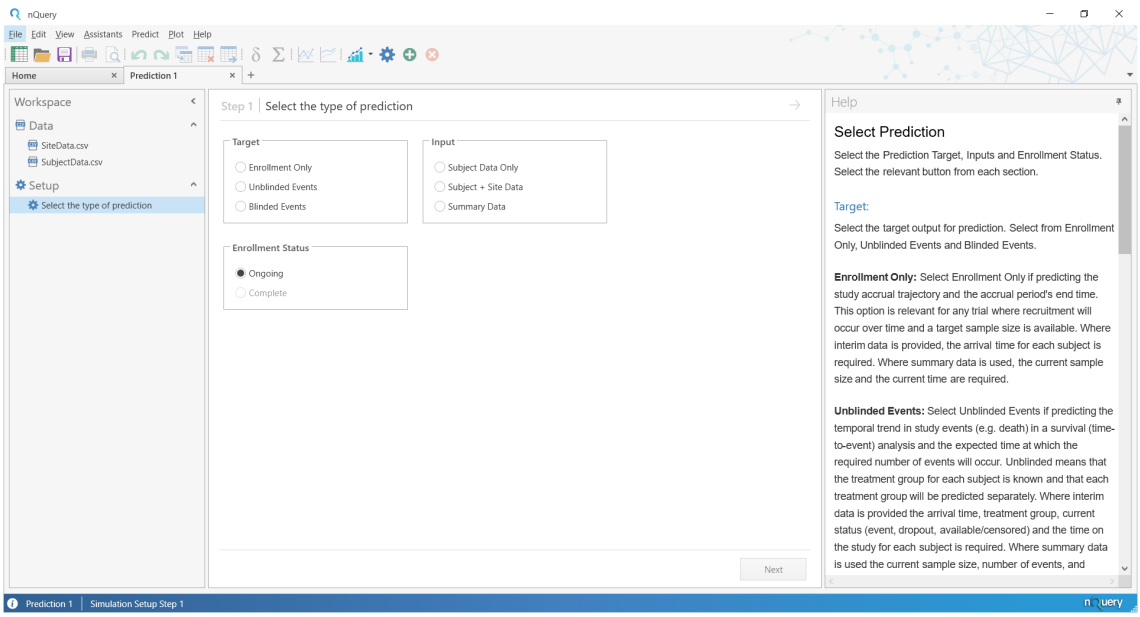



**Figure 6.5:** New nQuery Predict Workspace with Data

#### 6.2.1.2 Prediction with Fixed Parameters

If “Prediction with Fixed Parameters” is selected an nQuery Predict workspace will immediately open. Selecting the Skip button on the “New Prediction - Import Datasets” window (see Figure 6.2) will also open an empty workspace.

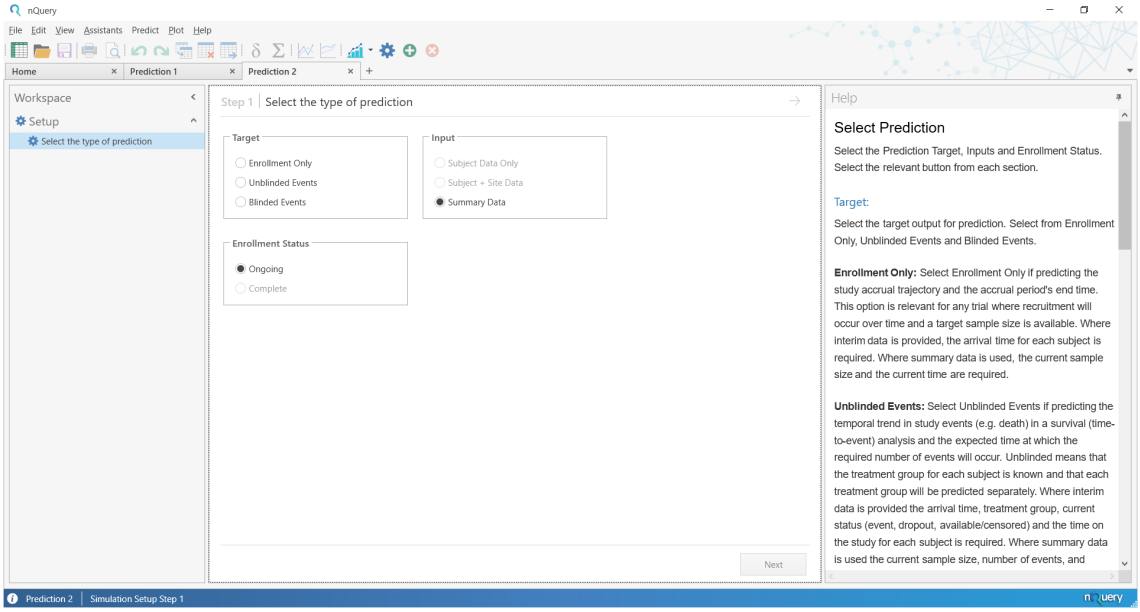



**Figure 6.6:** New nQuery Predict Workspace without Data

If this method is used the workspace will not contain any data and therefore the Subject Data Only and Subject + Site Data Input Options will not be available (see subsection 6.2.5). To add datasets after a workspace is open, the Add Dataset option from the Predict file menu or toolbar can be used (see subsection 6.2.3).

### 6.2.2 nQuery Predict Workspace Layout

#### 6.2.2.1 nQuery Predict Workspace Layout Overview

An nQuery Predict workspace will consist of three primary user interface elements: the Workspace Navigation Bar (1), the Main Display Window (2) and the Help Window (3). These are numbered in Figure 6.7.

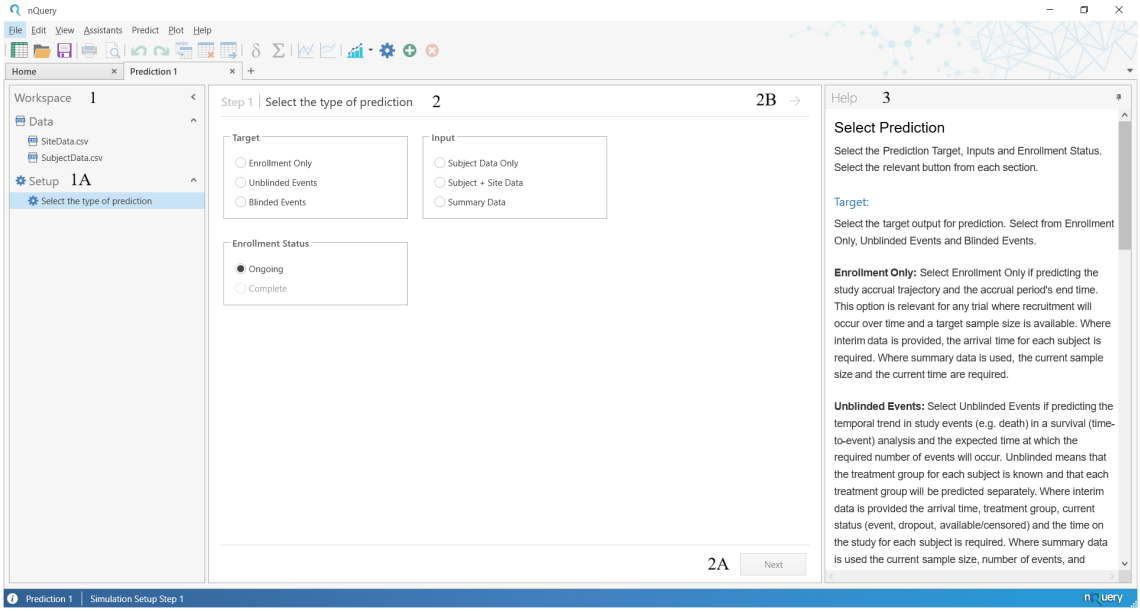



**Figure 6.7:** nQuery Predict Workspace Layout

The Workspace Navigation Bar (1) is used to navigate to a specific element (**Data, Setup** step**, Report, Plot, Table**) of the current nQuery Predict Workspace at any time. Elements are organized under headers with an example header (**Setup**) highlighted as per 1A in Figure 6.7. A specific element can be selected from under a given header. Header elements can opened and collapsed by selecting the header name. In Figure 6.7, the **Select the Type of Prediction** step can be selected from the **Setup** header or the **SubjectData.csv** or **SiteData.csv** datasets can be selected from the **Data** header.

The following elements are potentially available, depending on the current prediction stage, in the Workspace Navigation Bar: **Data, Setup, Reports, Plots, Tables.** Details for **Data** are given in subsection 6.2.3. Details for **Setup** are given in subsection 6.2.5, subsection 6.2.6 and subsection 6.2.7. Details for **Reports, Plots** and **Tables** are given in subsection 6.2.8, subsection 6.2.6 and subsection 6.2.7.

The Main Display Window (2) displays the current element selected and is where the user can input values and view results. During the **Setup** stage, the user can select the Next button (2A) or Right-Arrow buttons (2B) to move to the next step (note subsequent **Setup** steps will have a corresponding adjacent Back and Left-Arrow buttons to go back to prior step) and users can select and enter values in the main fields (e.g. for the current **Select the Type of Prediction** step, radio buttons can be select from the **Target, Input** and **Enrollment Status** fields). Details on what is displayed in the Main Display Window for a given element will be shown in the sections below.

The Help Window (3) provides context and guidance on all of the elements currently

displayed in Main Display Window. The Help will include definitions, context, suggested values and tips.

#### 6.2.2.2 Editing the nQuery Predict Workspace Layout

There are three primary ways to change the default layout of an nQuery Predict Workspace: collapsing the Workspace Navigation Bar, unpinning the Help Window and changing the width of the Main Display Window.

To collapse the Workspace Navigation Bar, select the button to the right of the main "Workspace" header at the top-right of the Workspace Navigation Bar. When the Workspace Navigation Bar is collapsed, elements can be shown and selected by clicking the named bar (name will be the same as the current element shown in the Main Display Window) on the left-hand side of the nQuery Predict Workspace. To uncollapse the Workspace Navigation Bar, select the button at the top of the named bar on the left-hand side.

To unpin the Help window, select the button on the right of Help Window header in the top-right. When the Help Window is unpinned, the Help Window will not be displayed unless the cursor is placed over the Help section in the top-right of the nQuery Predict Workspace after which the Help Window will be displayed until a different element is selected in the nQuery Predict Workspace. To repin the Help Window, select the button when the Help Window is open.

To change the width of the Main Display Window, move the cursor over the left-hand edge (if the Workspace Navigation Bar is not collapsed) or right-hand edge (if the Help Window is displayed) of the Main Display Window until the cursor changes to a two-way arrow (↔) and click and drag the cursor to change the Main Display Window width.

An example workspace where the Workspace Navigation Bar is collapsed and Help Window is unpinned and unselected is shown in Figure 6.8, with the user interface items referenced above highlighted in yellow.

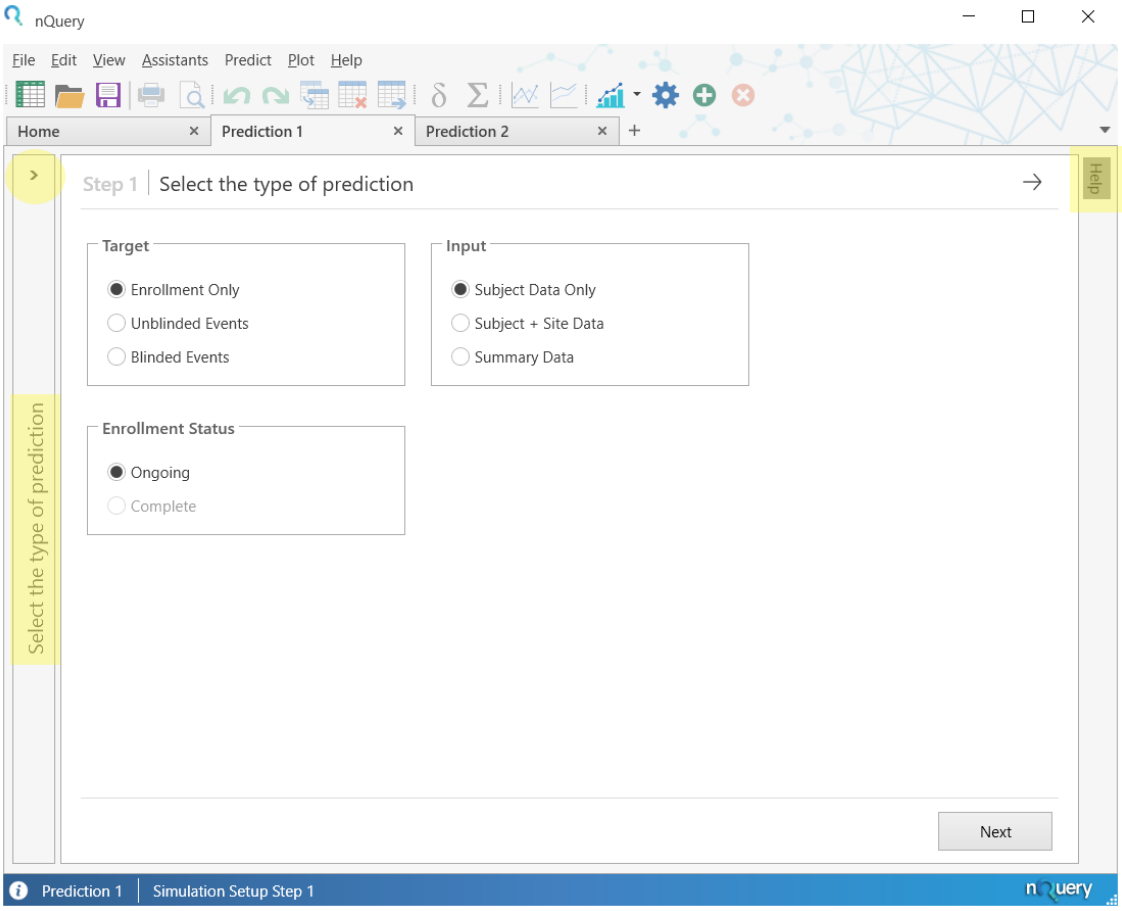



**Figure 6.8:** Collapsed nQuery Predict Workspace

## 6.2.3 Dataset Management

### 6.2.3.1 Viewing and Editing Datasets

To view any dataset in the current nQuery Predict workspace, select that dataset from the **Data** header in the Workspace Navigation Bar (subsection 6.2.2) on the left-hand side of the workspace which has been highlighted in Figure 6.9.

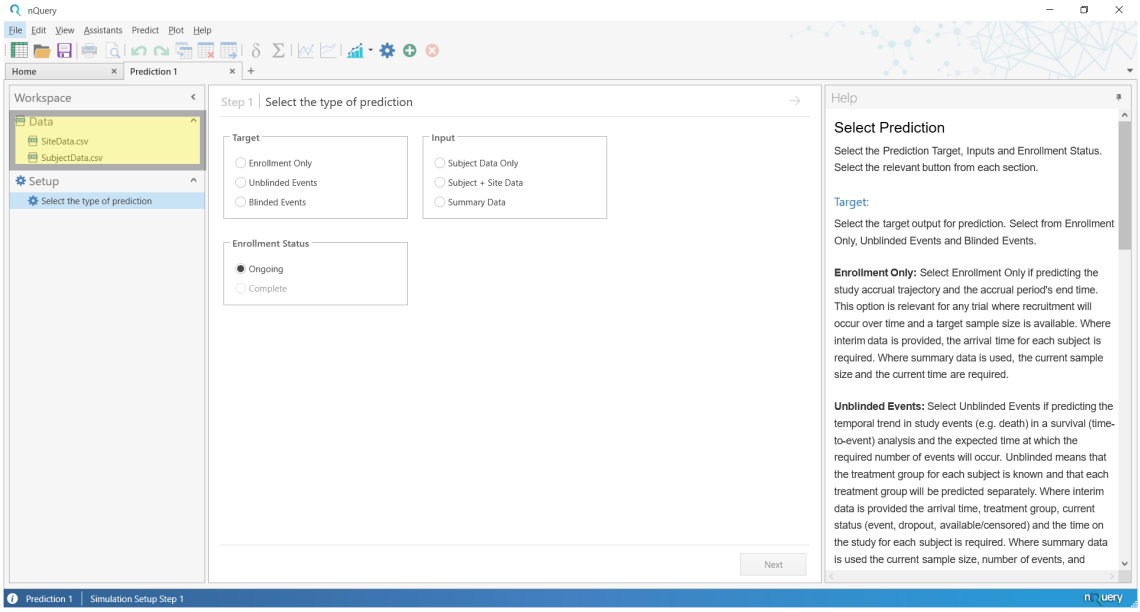
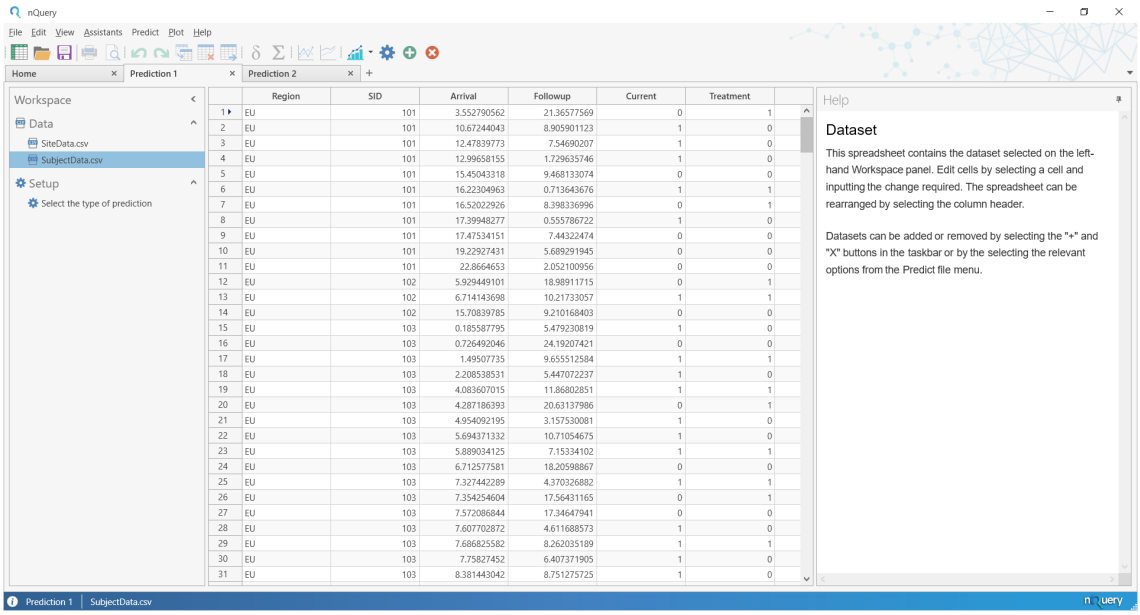



**Figure 6.9:** nQuery Predict Workspace with Highlighted Data field

After selecting a dataset from the Data field, the selected dataset will be displayed in the main nQuery Predict workspace window. An example with the SubjectData.csv file is shown in Figure 6.10.




| | Region | SID | Arrival | Followup | Current | Treatment |
|---|---|---|---|---|---|---|
| 1 | EU | 101 | 3.552790562 | 21.36577569 | 0 | 1 |
| 2 | EU | 101 | 10.67244043 | 8.905901123 | 1 | 0 |
| 3 | EU | 101 | 12.47839773 | 7.54690207 | 1 | 0 |
| 4 | EU | 101 | 12.99658155 | 1.729635746 | 1 | 0 |
| 5 | EU | 101 | 15.45043318 | 9.468133074 | 0 | 0 |
| 6 | EU | 101 | 16.22304963 | 0.713643676 | 1 | 1 |
| 7 | EU | 101 | 16.52022926 | 8.398336996 | 0 | 1 |
| 8 | EU | 101 | 17.39948277 | 0.555786722 | 1 | 0 |
| 9 | EU | 101 | 17.47534151 | 7.44322474 | 0 | 0 |
| 10 | EU | 101 | 19.22927431 | 5.689291945 | 0 | 0 |
| 11 | EU | 101 | 22.8664653 | 2.052100956 | 0 | 0 |
| 12 | EU | 102 | 5.929449101 | 18.98911715 | 0 | 1 |
| 13 | EU | 102 | 6.714143698 | 10.21733057 | 1 | 1 |
| 14 | EU | 102 | 15.70839785 | 9.210168403 | 0 | 0 |
| 15 | EU | 103 | 0.185587795 | 5.479230819 | 1 | 0 |
| 16 | EU | 103 | 0.726492046 | 24.19207421 | 0 | 0 |
| 17 | EU | 103 | 1.49507735 | 9.655512584 | 1 | 1 |
| 18 | EU | 103 | 2.208538531 | 5.447072237 | 1 | 0 |
| 19 | EU | 103 | 4.083607015 | 11.86802851 | 1 | 1 |
| 20 | EU | 103 | 4.287186393 | 20.63137986 | 0 | 1 |
| 21 | EU | 103 | 4.954092195 | 3.157530081 | 1 | 0 |
| 22 | EU | 103 | 5.694371332 | 10.71054675 | 1 | 0 |
| 23 | EU | 103 | 5.889034125 | 7.15334102 | 1 | 1 |
| 24 | EU | 103 | 6.712577581 | 18.20598867 | 0 | 0 |
| 25 | EU | 103 | 7.327442289 | 4.370326882 | 1 | 1 |
| 26 | EU | 103 | 7.354254604 | 17.56431165 | 0 | 1 |
| 27 | EU | 103 | 7.572086844 | 17.34647941 | 0 | 0 |
| 28 | EU | 103 | 7.607702872 | 4.611688573 | 1 | 0 |
| 29 | EU | 103 | 7.686825582 | 8.262035189 | 1 | 1 |
| 30 | EU | 103 | 7.75827452 | 6.407371905 | 1 | 0 |
| 31 | EU | 103 | 8.381443042 | 8.751275725 | 1 | 0 |

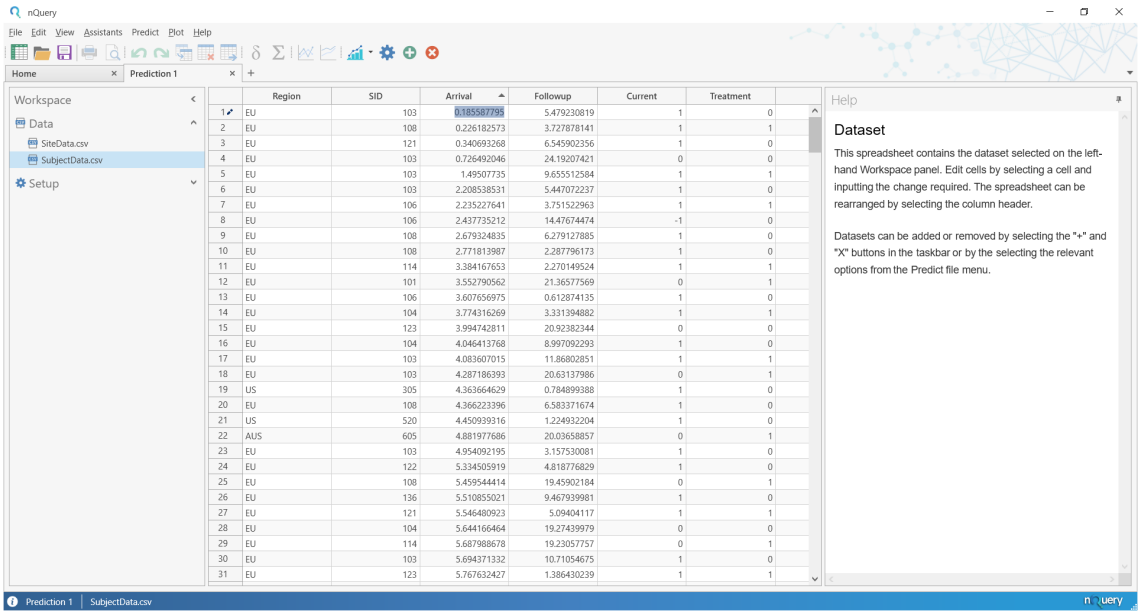
**Figure 6.10:** nQuery Predict Workspace Displaying SubjectData Example Dataset

When a dataset is displayed, the user can edit cells and sort columns. To edit a cell, double-click a cell and enter the new required input. Note that nQuery automatically

detects whether a column is a non-numeric category, integer or real numeric value based on the values in that column. This is to reduce the number of options shown during the Subject-level Data (Step 2 in **Setup** when **Input = Subject Data Only** or **Subject + Site Level Data**) and Site-level Data (Step 3 in **Setup** if **Subject + Site Data** selected from **Input**) drop-down menus. If the user enters a real value into a category or integer column, an out of range error will be displayed in that cell (see subsection 2.2.1).

To sort a column from low to high for a numeric column or A to Z for a alphanumeric column, select the column header containing the column name at the top of the required column. To sort the column from high to low (or Z to A) select the column header a second time. When the dataset has been sorted by a column, an up (if low to high/A to Z) or down arrow (if high to low/Z to A) will be shown to the right of the column header.

An example where the SubjectData has been sorted by the **Arrival** column from low to high and an edit is underway in the first row cell is shown in Figure 6.11.




| | Region | SID | Arrival | Followup | Current | Treatment |
|---|---|---|---|---|---|---|
| 1 | EU | 103 | 0.185587795 | 5.479230819 | 1 | 0 |
| 2 | EU | 108 | 0.226182573 | 3.727878141 | 1 | 1 |
| 3 | EU | 121 | 0.340693268 | 6.545902356 | 1 | 0 |
| 4 | EU | 103 | 0.726492046 | 24.19207421 | 0 | 0 |
| 5 | EU | 103 | 1.49507735 | 9.655512584 | 1 | 1 |
| 6 | EU | 103 | 2.208538531 | 5.447072237 | 1 | 0 |
| 7 | EU | 106 | 2.235227641 | 3.751522963 | 1 | 1 |
| 8 | EU | 106 | 2.437735212 | 14.47674474 | -1 | 0 |
| 9 | EU | 108 | 2.679324835 | 6.279127885 | 1 | 0 |
| 10 | EU | 108 | 2.771813987 | 2.287796173 | 1 | 0 |
| 11 | EU | 114 | 3.384167653 | 2.270149524 | 1 | 1 |
| 12 | EU | 101 | 3.552790562 | 21.36577569 | 0 | 1 |
| 13 | EU | 106 | 3.607656975 | 0.612874135 | 1 | 0 |
| 14 | EU | 104 | 3.774316269 | 3.331394882 | 1 | 1 |
| 15 | EU | 123 | 3.994742811 | 20.92382344 | 0 | 0 |
| 16 | EU | 104 | 4.046413768 | 8.997092293 | 1 | 0 |
| 17 | EU | 103 | 4.083607015 | 11.86802851 | 1 | 1 |
| 18 | EU | 103 | 4.287186393 | 20.63137986 | 0 | 1 |
| 19 | US | 305 | 4.363664629 | 0.784899388 | 1 | 0 |
| 20 | EU | 108 | 4.366223396 | 6.583371674 | 1 | 0 |
| 21 | US | 520 | 4.450939316 | 1.224932204 | 1 | 0 |
| 22 | AUS | 605 | 4.881977686 | 20.03658857 | 0 | 1 |
| 23 | EU | 103 | 4.954092195 | 3.157530081 | 1 | 0 |
| 24 | EU | 122 | 5.334505919 | 4.818776829 | 1 | 0 |
| 25 | EU | 108 | 5.459544414 | 19.45902184 | 0 | 1 |
| 26 | EU | 136 | 5.510855021 | 9.467939981 | 1 | 0 |
| 27 | EU | 121 | 5.546480923 | 5.09404117 | 1 | 1 |
| 28 | EU | 104 | 5.644166464 | 19.27439979 | 0 | 0 |
| 29 | EU | 114 | 5.687988678 | 19.23057757 | 0 | 1 |
| 30 | EU | 103 | 5.694371132 | 10.71054675 | 1 | 0 |
| 31 | EU | 123 | 5.767632427 | 1.386430239 | 1 | 1 |

**Figure 6.11:** nQuery Predict Workspace with SubjectData Column Sort and Cell Edit

#### 6.2.3.2 Adding and Removing Datasets

After an nQuery Predict workspace is open, datasets can be added or removed from the workspace by using the **Add Dataset** and **Remove Dataset** options. These two options are available from the Predict file menu or by selecting the **Add Dataset** ⊕ or **Remove Dataset** ⊗ buttons from the Toolbar menu.

Selecting **Add Dataset** will open the "New Prediction - Import Datasets" window. See subsubsection 6.2.1.1 for details on how the Import Datasets window works. The only difference for the **Add Dataset** option is the lack of the **Skip** button as this is not required for an existing workspace.

Selecting **Remove Dataset** while you have a dataset selected in the current workspace will remove that dataset from the workspace's Data field. Note that there is no additional prompt before the dataset is removed.

### 6.2.4 nQuery Predict Example Datasets

Before providing a demonstration of using nQuery Predict to create milestone predictions, a brief overview of the example datasets is provided here.

These datasets can be found in the nQuery installation folder which was specified during installation. By default, the datasets would be located in the following folder: C:\Program Files (x86)\Statistical Solutions Ltd\nQuery\TestData\Predict

These datasets are artificial but have been created to have summary characteristics similar to data from a real clinical trial.

The simulated data was based on the data available at the interim analysis of a group sequential survival analysis which had a **Target Number of Events** of 374 events, with subjects followed until this target events was reached at which point any remaining subjects were administratively right-censored. The one interim analysis was planned when 50% of the **Target Number of Events** was reached i.e. 187 events.

To achieve this **Target Number of Events**, it was assumed a **Target Sample Size** of 460 would be needed. The study plan was to accrue subjects over the first 30 months of the trial (giving an expected **Accrual Rate** of 460/30 = 15.3333). There were 127 sites (hospitals) available for subject recruitment in the study plan.

It was expected that an overall **Study Duration** of 40 months would be needed to reach the **Target Number of Events** given that accrual process and expected **Event Hazard Rates**.

The interim analysis data simulated here occurs 24.9 (83% of planned accrual period length) months (i.e. **Current Time = 24.9**) into the study when 187 (50%) of the target events have occurred. At this point, 402 (87.4%) of the subjects have been accrued, with 118 out of the total possible 127 sites open.

#### 6.2.4.1 SubjectData.csv

The SubjectData.csv is a comma-separated values (.csv) file which has six columns and 402 rows. Each row corresponds to an individual subject and each column is a characteristic for each subject. The columns in order from left to right, alongside a description and context, are as follows:

- **Region** (Region - Any Input Type): The **Region** the subject/site is from. 3 distinct values: EU (European Union), US (United States), AUS (Australasia)
- **SID** (**Site ID** - Any Input Type, all values must be in **Site-Level Data Site ID** column): The **Site ID** of the site from which a subject was recruited. 93 distinct integer values ranging from 101 to 623. See subsubsection 6.2.4.2 for details on the example **Site-Level Dataset: SiteData.csv**
- **Arrival** (**Arrival Time** - Numeric Values $\geq 0$): The **Arrival Time** (start of the study is time = 0) in months at which a subject was recruited. 402 distinct real numeric values ranging from 0.185587795 (first arrival) to 24.91856625 (last arrival, also is **Current Time**)
- **Followup** (**Time on Study** - Numeric Values $\geq$ **Arrival Time** in row): The length of time in months a subject has been followed for since arrival (i.e. **Time on Study**). 402 distinct real numeric values ranging from 0 (just arrived) to 24.19207421 (months on study). Note that **Followup** definition depends on **Current** column, where **Followup** equals the **Time on Study** before subject had an event (**Event Status: Current = 1**) or dropped out (**Dropout Status: Current = -1**) but **Followup** equals time from **Arrival** until the **Current Time** for subjects who are still available to have the event in the future (**Censored Status : Current = 0**)
- **Current** (**Status Indicator** - Any Input Type with minimum one, maximum three distinct values: The current status of a subject (i.e. **Status Indicator**). 3 distinct integer values: -1 = Subject has dropped out at **Current Time**, 0 = Subject is available (i.e. has neither had event or dropped out) at **Current Time**, 1 = Subject has have event of interest (death) at **Current Time**
- **Treatment** (**Treatment ID** - Any Input Type with two distinct values): The treatment group (i.e. **Treatment ID)** a subject was assigned to. 2 distinct integer values: 0 = Subject is in **Control Group**, 1 = Subject is in **Treatment Group**

#### 6.2.4.2 SiteData.csv

The SiteData.csv is comma-separated values file which has seven columns and 127 rows. Each row corresponds to an individual site and each column is a characteristic for each site. The columns in order from left to right, alongside a description and context, are as follows:

- **Region** (Region - Any Input Type): The **Region** the subject/site is from. 3 distinct values: EU (European Union), US (United States), AUS (Australasia)
- **SID** (**Site ID** - Any Input Type, must include all **Site ID** values from **Subject-Level Dataset Site ID** column): The site ID for a site. 127 distinct integer values ranging from 101 to 623. Note that if **Input = Subject + Site Data**, the Site ID column for the Site Data must contain all Site IDs given in

the corresponding subject-level Site ID column. However, the Site Data Site ID can contain values not included in the Subject Data Site ID for sites which are unopened or have not recruited a subject at the **Current Time**.

- **Cap** (**Enrollment Cap** - Integer > 0): The **Enrollment Cap** in a site i.e. the maximum number of subjects that can recruited from a site. Integer values ranging from 5 to 72.
- **Open Time** (**Site Initiation Time: Opened Sites** - Numeric Value $\geq 0$) : The time at which a site opened (study start is time = 0) i.e. **Site Initiation Time** for **Open Sites**. Real numeric values ranging from 0 to 19.375. Note that 9 rows do not have a value for **Open Time**. These are the 9 unopened sites at the **Current Time**. If a given site has a **Site Initiation Time for Open Sites** and **Start and End Time for Site Initiation window for Unopened Sites**, it is assumed the site is **Open** and only this value is used.
- **Rate** (**Accrual Rate/Site** - Numeric Value $\geq 0$): The planned accrual rate (i.e. **Accrual Rate/Site**) in a site. The accrual rate is the number of subjects recruited per unit time (months in this case). Real numeric values ranging from 0.045541405 to 2.224390244. Note that for open sites which have recruited a minimum of one subject, the default accrual rate for a site given in **Accrual Options Setup** step will be based on the actual rate rather than the planned rate given here
- **Start** (**Site Initiation Time: Unopened Sites (Optional): Start Time** - **Current Time**$\leq$ Numeric Value $\leq$ **End Time**): The **Start Time** of the **Site Initiation** window for **Unopened Sites** to open. Nine real numeric values between 25.1 and 25.51. Note that 118 rows do not have a value for **Start** as these sites are already open at the **Current Time**. If a given site has a **Site Initiation Time for Open Sites** and **Start and End Time for Site Initiation window for Unopened Sites**, it is assumed the site is **Open** and this value is ignored.
- **End** (**Site Initiation Time: Unopened Sites (Optional): EndTime** - Numeric Value $\geq$ **Start Time**): The **End Time** of the **Site Initiation** window for **Unopened Sites** to open. Nine real numeric values between 25.3 and 26.11. Note that 118 rows do not have a value for **End** as these sites are already open at the **Current Time**. If a given site has a **Site Initiation Time for Open Sites** and **Start and End Time for Site Initiation window for Unopened Sites**, it is assumed the site is **Open** and this value is ignored.

If there is an input from the Site Data which is incompatible with the Subject Data provided (for example Open Time > Current Time, Enrollment Cap > subjects recruited in Site at Current Time), these will be highlighted in the **Accrual Infos/Sites** table at the **Accrual Options** step of **Setup** (see subsection 6.2.6 and subsection 6.2.7).

### 6.2.5 Select the Type of Prediction/Setup Introduction

The **Select the Type of Prediction** is the first step of the **Setup** stage of the creating a milestone prediction in nQuery Predict. The **Select the Type of Prediction** step selects the specific characteristics needed to setup the appropriate prediction for a study. There are three sections at this step: **Target, Input and Enrollment Status**. The **Select the Type of Prediction** step in the nQuery Predict workspace is shown in Figure 6.12.

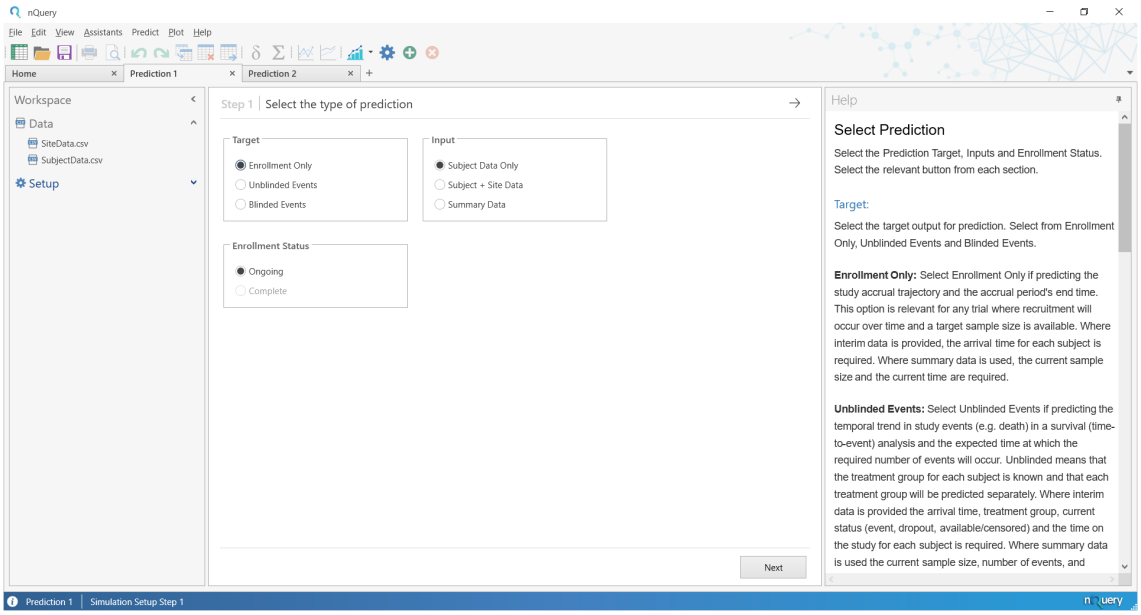



**Figure 6.12:** Select a Type of Prediction Step

The appropriate **Setup** is chosen by selecting the required radio button to the left of the appropriate option in each section. A brief introduction on using the **Setup** field and three sections of the **Setup the Type of Prediction** step are provided next.

#### 6.2.5.1 Target

The **Target** section specifies what will be the target of the milestone prediction. The three options for **Target** are as follows:

- Enrollment Only
- Unblinded Events
- Blinded Events

**Enrollment Only** will select a milestone prediction where the objective is to project the length of time (**Accrual Duration**) required to enroll a **Target Sample Size**.

When **Subject-level Data** is used (**Input = Subject Data Only** or **Subject + Site Data**), this will require an **Arrival Time** per subject. For **Summary Data Input**, this will require the **Current Sample Size** and **Current Time**

**Unblinded Events** will select a milestone prediction where the objective is to project the length of time (**Study Duration**) required to achieve a **Target Number of Events** when the treatment group assignment for each subject is known. When a **Subject-level Data** is used (**Input = Subject Data Only** or **Subject + Site Data**), this will require the **Arrival Time, Time on Study, Status Indicator (Event Status, Dropout Status, Censored Status)** and **Treatment ID (Control Group, Treatment Group)**. For **Summary Data Input,** the **Current Sample Size (Control), Current Sample Size (Treatment), Current Number of Events (Control), Current Number of Events (Treatment), Current Number of Dropouts (Control), Current Number of Dropouts (Treatment)** and **Current Time** are required. **(Current/Treatment)** in brackets indicates the treatment group.

**Blinded Events** will select a milestone prediction where the objective is to project the length of time (**Study Duration**) required to achieve a **Target Number of Events** when the treatment group assignment for each subject is unknown. When a **Subject-level Data** is used (**Input = Subject Data Only** or **Subject + Site Data**), this will require the **Arrival Time, Time on Study** and **Status Indicator (Event Status, Dropout Status, Censored Status)**. For **Summary Data Input,** the **Current Sample Size, Current Number of Events, Current Number of Dropouts** and **Current Time** are required.

#### 6.2.5.2 Input

The **Input** section specifies what will be the input type used to generate the milestone prediction. The three options for **Input** are as follows:

- Subject Data Only
- Subject + Site Data
- Summary Data

**Subject Data Only** means the milestone prediction will be based on subject-level interim data only for the enrollment and/or event process. The specific fields required in the Subject Data will depend on the choice of **Target** (see subsubsection 6.2.5.1).

**Subject + Site Data Only** means the milestone prediction will be based on subject-level and site-level interim data. The specific fields required in the Subject Data will depend on the choice of **Target** (see subsubsection 6.2.5.1) and will be the same as for **Subject Data Only** with the addition of a **Site ID** field to link the **Subject-Level Data** and **Site-Level Data**. The Site Data fields will be same for all predictions as the Site Data is only used for enrollment modelling. **Subject + Site Data** is not available if **Enrollment Status** is set to **Complete**.

**Summary Data** means the milestone prediction will be based on fixed parameter estimates made either pre-trial or based on summary statistics from an ongoing trial. The specific fields required will depend on the choice of **Target** (see subsubsection 6.2.5.1). For a pre-trial milestone prediction, enter zero for all allowable fields at the second **Fixed Parameters** step.

#### 6.2.5.3 Enrollment Status

The **Enrollment Status** section specifies if enrollment is complete in a survival analysis trial at the time that the milestone prediction will be generated. The two options for **Enrollment Status** are as follows:

- Ongoing
- Complete

**Ongoing** means that the enrollment process is still ongoing at the time at which the milestone prediction is being generated. In this case, the **Accrual Options** step of **Setup** will be included to set the **Target Number of Events** and define the **Accrual Model** used to generate future simulated subjects that will be enrolled into the study.

**Complete** means that the enrollment process is complete at the time at which the milestone prediction is being generated. In this case, the **Accrual Options** step summarizes the accrual process up to the **Current Time** but since enrollment is complete, the **Target Number of Events** and **Accrual Model** inputs are not required at the **Accrual Options** step. **Complete** is only available if **Target** equals **Unblinded Events** or **Blinded Events.**

#### 6.2.5.4 Setup Introduction

During the **Setup** stage the user will be required to specify the desired inputs to generate simulations for the milestone prediction. The specific subsequent steps will depend on the selections made at the **Select the type of Prediction** step (see subsection 6.2.6 and subsection 6.2.7 for details) but there are five broad categories of step. These are, in order inputted by the user, known as:

- Data Field Selection: Select fields from Dataset which correspond the required inputs for milestone prediction if **Input = Subject Data Only** or **Subject + Site Data**
- Fixed Parameters: Input fixed parameter estimates if **Input = Summary Data**
- Accrual Options: Select a model and edit options to specify how accrual simulations will be generated. This step will differ depending if **Input = Subject Data Only** or **Subject + Site Data**. If **Enrollment Status = Complete**, this step will provide a summary of accrual up to the **Current Time** only

- Event and Dropout Information: Select a model and edit options to specify how event and dropout simulations will be generated if **Target = Unblinded Events** or **Blinded Events**
- Simulation Controls: Edit the options for the overall simulation such as the number of simulations, random seed and additional tables desired. See subsubsection 6.2.9.4 for details

The **Setup** stage involves completing the required steps in order. To move to the next step, the user must fill in all required information and then select the Next button in the bottom left of the Main Window or the → button in the top-right of the Main Window. These options will be greyed out until all required information at a given step has been entered, though most stages after Data Field Selection/Fixed Parameters will have defaults based on fitting the models to the provided information (information on the models available and their nQuery Predict defaults is given in section 6.1). At the final **Simulation Controls** step, the Next button will be replaced with a Run button which when selected will start the milestone prediction simulation .

Once a step is complete, the user can also go forward and backwards within the **Setup** stage by using the Next/Back buttons in the bottom-right of the Main Window or the →/← buttons in the top-right of the Main Window.

The user can navigate to a specific step by selecting the desired step from under the **Setup** heading in the Workspace Navigation Bar on the left-hand side (see subsection 6.2.2 for details).

When a simulation has been completed, the Workspace Navigation Bar's **Setup** heading can be used to change any of the **Setup** inputs. To make changes return to any step of the **Setup** stage, edit the desired simulation inputs in that step, select the Next or → button at that step and then re-run the simulation after selecting Run at the final **Simulation Controls** step.

Note that changing any option at the **Select the type of prediction** step of **Setup** will reset the prediction and delete all subsequent steps and any user inputs made at those steps.

### 6.2.6 Enrollment Prediction Demonstration

Enrollment milestone prediction is the projection of how long a study will take to recruit a **Target Sample Size**. In nQuery Predict, there are three main approaches to enrollment milestone prediction:

- Enrollment Prediction using Subject-level Data only (subsubsection 6.2.6.1)
- Enrollment Prediction using Subject and Site-level Data (subsubsection 6.2.6.2)
- Enrollment Prediction using Summary Data (subsubsection 6.2.6.3)

Each of these scenarios will be demonstrated using the SubjectLevel.csv and SiteLevel.csv datasets provided with nQuery Predict. See subsection 6.2.4 for details on these datasets.

Each demonstration will start with demonstration using the default values given by nQuery Predict, followed by additional scenarios taken from the context these datasets were taken from and which use additional options available in nQuery Predict.

The additional options covered in each demo are as follows:

- Subject-level Data only: Editing **Target Sample Size** and future **Accrual Rate**
- Subject and Site-level Data: Editing **Target Sample Size**, removing and editing individual **Sites**
- Summary Data: Editing **Target Sample Size**, **Pre-Trial** enrollment prediction

For these demonstrations, it is assumed the appropriate nQuery Workspace has been created as per the **Creating a nQuery Predict Workspace** section above (subsection 6.2.1).

#### 6.2.6.1 Enrollment Prediction using Subject-Level Interim Data Only

**Setup** **Enrollment Only** milestone prediction using **Subject-level Data Only** means that the prediction will project the expected length of the accrual period needed to recruit a specified **Target Sample Size** using subject-level interim data alone.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step select **Enrollment Only** from the **Target** field and **Subject Data Only** from the **Input** field. **Enrollment Status** must be **Ongoing** for enrollment milestone prediction. These selections are shown in Figure 6.13. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

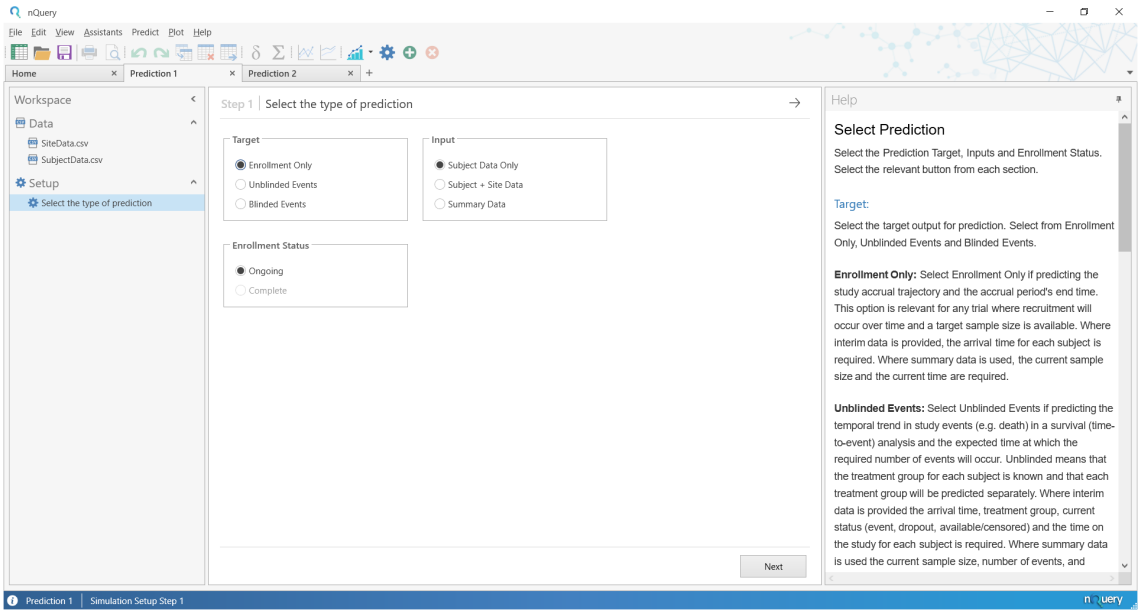



**Figure 6.13:** Select the Type of Prediction Setup Step for Enrollment Prediction using Subject-level Data

The next step is the **Subject-level Data** step. The **Subject-level Data** step is a Data Field Selection step. The **Subject-level Data** is the dataset where each row corresponds to a specific subject enrolled in the study.

Data Field Selection steps are steps in which the required dataset in the current Workspace is selected and columns in the selected dataset are assigned to the required fields for that prediction problem. The options in a given Data Field Selection field are opened by selecting the button on the right of each field which is found to the right of the field name e.g. **Select Data**. Each field will have specific requirements (e.g. numeric only, two distinct values only) and nQuery Predict only displays options which are consistent with those fields.

In this prediction there are two fields which are as follows:

- Select Dataset: Select the dataset that is the Subject-level Data from the datasets in the current workspace
- Arrival Time: The arrival time column contains the length of time since the study start (Time = 0) until a subject entered the study. Each subject's Arrival Time should be greater than or equal to zero

For this prediction, in the **Select Dataset** field select **SubjectData.csv** and in the **Arrival Time** field select **Arrival**. These selections are shown in Figure 6.14. Select the Next or → button to move to the next **Setup** step.

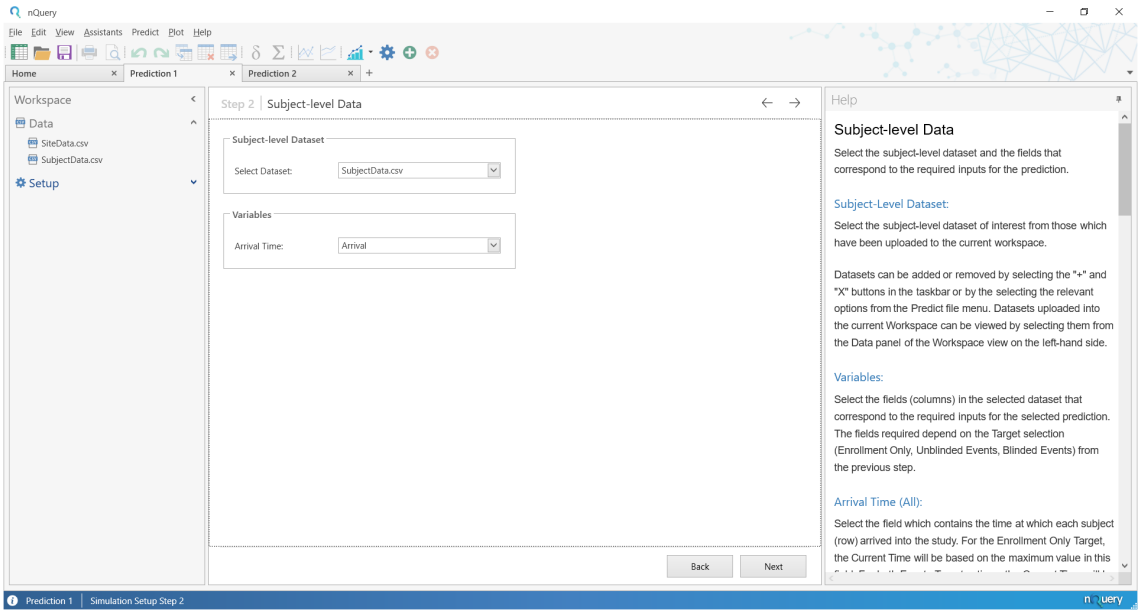



**Figure 6.14:** Subject-level Data Setup Step for Enrollment Prediction using Subject-level Data

The next step is the **Accrual Options** step which specifies the **Target Sample Size, Accrual Model** and future **Accrual Rate**. The **Accrual Options** step, and its defaults, for this demonstration are shown in Figure 6.15.

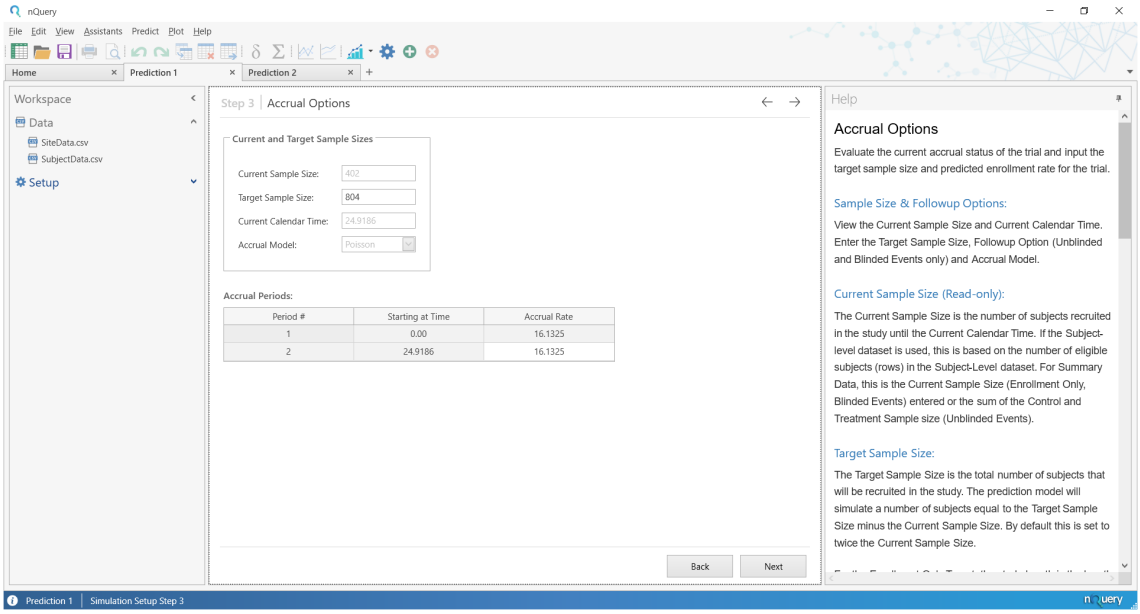



**Figure 6.15:** Accrual Options Setup Step for Enrollment Prediction using Subject-level Data

The **Accrual Options** step consists of two main elements: the **Current and Target Sample Sizes** input fields and the **Accrual Periods** table.

The **Current and Target Sample Sizes** input fields provides information on study enrollment based on the subject-level data and allows the user to edit the **Target Sample Size** and select the **Accrual Model**. The following fields are provided in this prediction:

- Current Sample Size (Read-only): Sample Size recruited into a study so far. Equals the number of rows in the Subject-level Data
- Target Sample Size (Editable, Integer > Current Sample Size): Total number of subjects that will be recruited in this study. The number of simulated subjects equals the Total Sample Size minus the Current Sample Size
- Current Time (Read-only): The length of time that has passed since the study started. Equals the maximum Arrival Time value
- Accrual Model (Editable): The statistical model that will be used to generate simulated enrollments

In nQuery Predict, the **Target Sample Size** defaults to twice the **Current Sample Size** and the **Accrual Model** defaults to **Poisson.** In this demonstration, the **Current Sample Size** equals 402 and therefore the default **Target Sample Size** is $402\,(2) = 804$.

The **Accrual Periods** table contains information on the **Accrual Rate** up the **Current Time** and allows the user to edit the future **Accrual Rate** used to generate enrollment simulations. The **Accrual Periods** has three columns ,with each row corresponding to a time period, with the columns defined as follows:

- Period #: A numeric ID for the current time period. Increases in increments of one for each subsequent row
- Starting at Time: The starting time for the current time period row. The first row will equal 0 (rate from study start to Current Time) and the second row will equal the Current Time
- Accrual Rate: The accrual rate is the average number of subjects recruited per unit time (months). The accrual rate in the first row will equal the accrual rate up to the Current Time

For the **Poisson Accrual Model**, the **Accrual Periods** table will consist of two time period rows. The first row will correspond to the information provided by the user regarding the study up until the **Current Time.** The second row will correspond to the inputs used to generate future enrollments. By default the **Accrual Rate** in the future second row will equal the **Accrual Rate** from the first row i.e. the accrual rate up until the **Current Time.** For the inputs used in this demonstration, this **Accrual Rate** column values in each row equal 16.13254936. The user can edit the **Accrual Rate** for generating future enrollments by editing the second cell of the **Accrual Rate** column.

For now, the default values will be used for **Target Sample Size** (804), **Accrual Model** (Poisson) and **Accrual Rate** (16.13254936) as per Figure 6.15. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.16. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

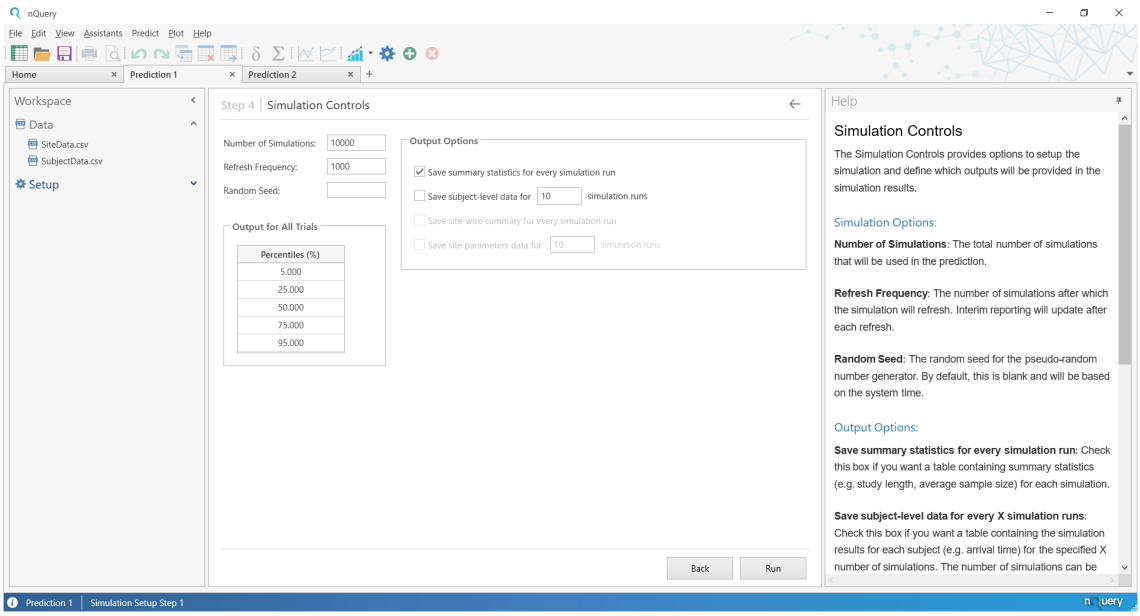



**Figure 6.16:** Simulation Controls Setup Step for Enrollment Prediction using Subject-level Data

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.17.

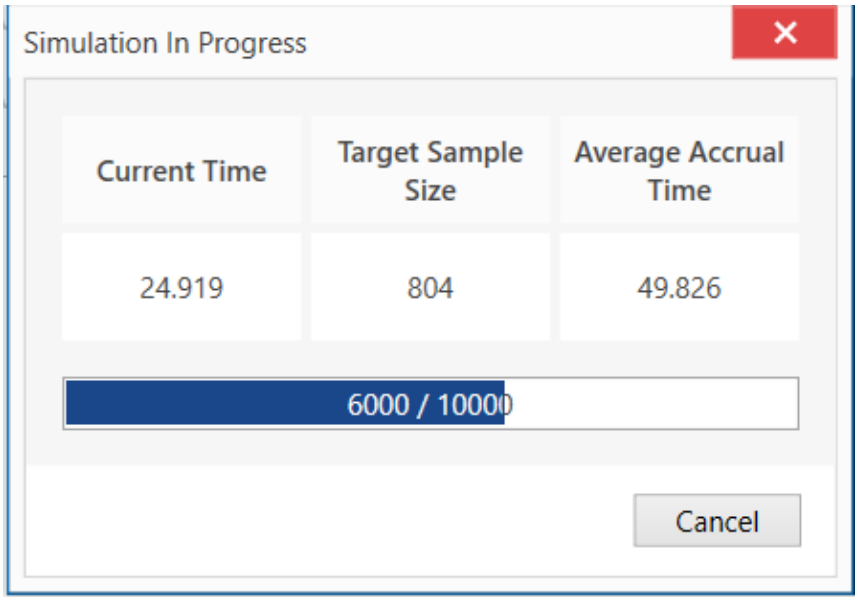



**Figure 6.17:** Simulation in Progress Window for Enrollment Prediction using Subject-level Data

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.18.

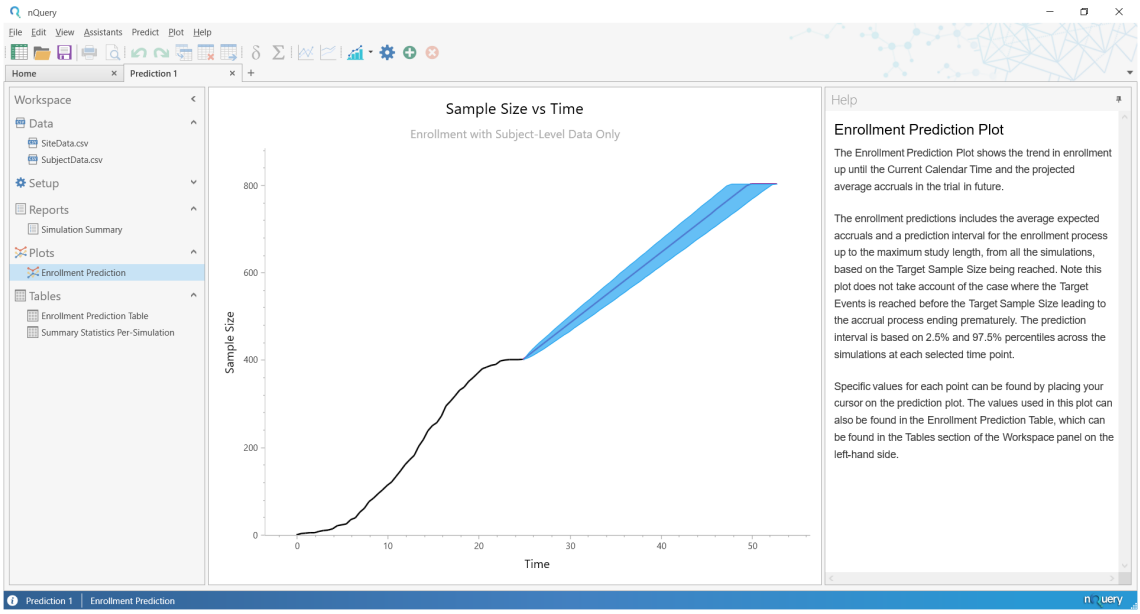



**Figure 6.18:** Enrollment Prediction Plot for Enrollment Prediction using Subject-level Data

The results of a nQuery Predict milestone prediction can be divided into three categories: **Reports, Plots** and **Tables**. A high-level summary of the results in nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.19.

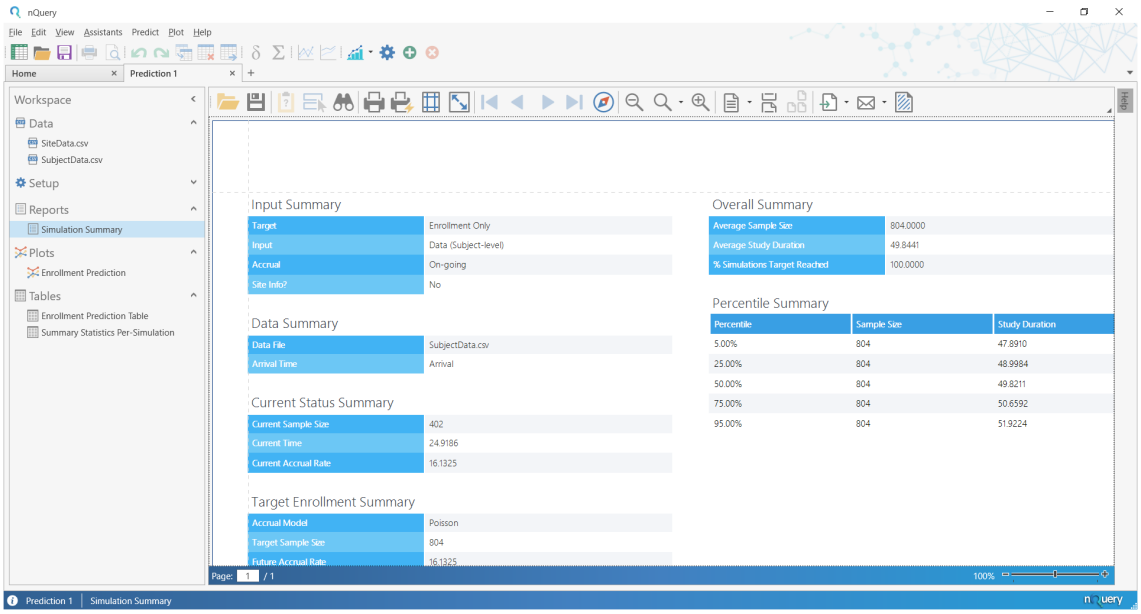



**Figure 6.19:** Simulation Summary Report for Enrollment Prediction using Subject-level Data

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.20. For this demonstration, the **Overall Summary** shows that the **Target Sample Size** of 804 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 49.8441. The **Percentile Summary** shows that 90% of predictions had an **Accrual Duration** between 47.8910 (5% Percentile) and 51.9224 (95% Percentile), 50% a duration between 48.9984 (25% Percentile) and 50.6592 (75% Percentile) and a median (50% Percentile) duration of 49.8211.

Overall Summary

| | |
|---|---|
| Average Sample Size | 804.0000 |
| Average Study Duration | 49.8441 |
| % Simulations Target Reached | 100.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration |
|---|---|---|
| 5.00% | 804 | 47.8910 |
| 25.00% | 804 | 48.9984 |
| 50.00% | 804 | 49.8211 |
| 75.00% | 804 | 50.6592 |
| 95.00% | 804 | 51.9224 |

**Figure 6.20:** Simulation Summary Report Results for Enrollment Prediction

To edit the **Simulation Summary report there a** number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment simulation from the study start to when the **Target Sample Size** was reached in all simulations. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.21.

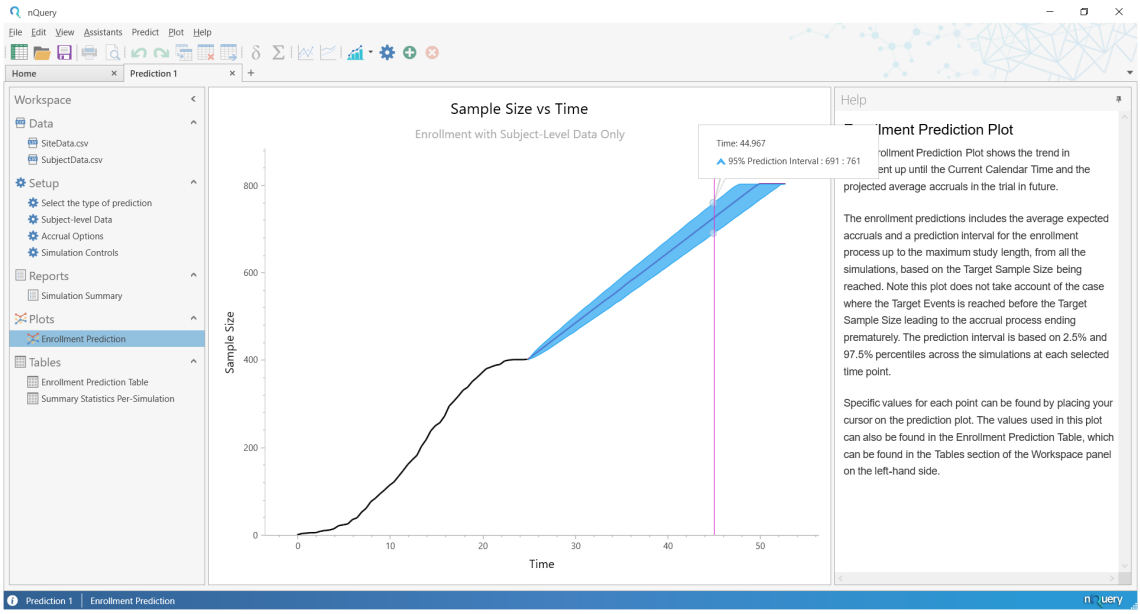
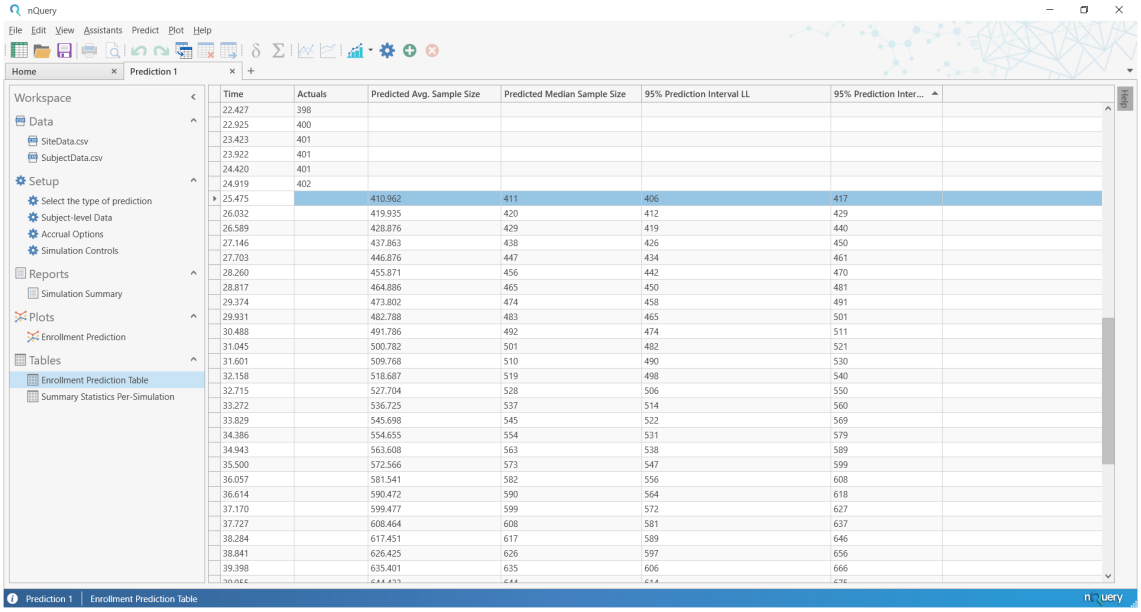



**Figure 6.21:** Enrollment Prediction Plot for Enrollment Prediction using Subject-level Data

The **Enrollment Prediction** plot consists of two parts: from Time = 0 to **Current Time** and from **Current Time** to the **Maximum Time**. The first part is a black line which plots the accrual pattern from the user inputs provided. The second part is a blue conic section with a grey line where the blue conic section plots the 95% prediction interval for the **Sample Size** enrolled at a given **Time** across all simulations and the grey line is the average **Sample Size** enrolled at a given time. The 95% prediction interval indicates the 95% of predictions had a sample size that fell between the lower and upper limit at a given time.

The **Enrollment Prediction Table** is the table used to create the **Enrollment Prediction** plot, which by default consists of ~100 rows (50 pre-simulation and 50 post-simulation). It consists of five columns: **Time, Actual, Predicted Avg. Sample Size, Predicted Median Sample Size, 95% Prediction Interval LL, 95% Prediction Interval UL**. The **Enrollment Prediction Table** for this demonstration at the start of the simulated period is given in Figure 6.22.

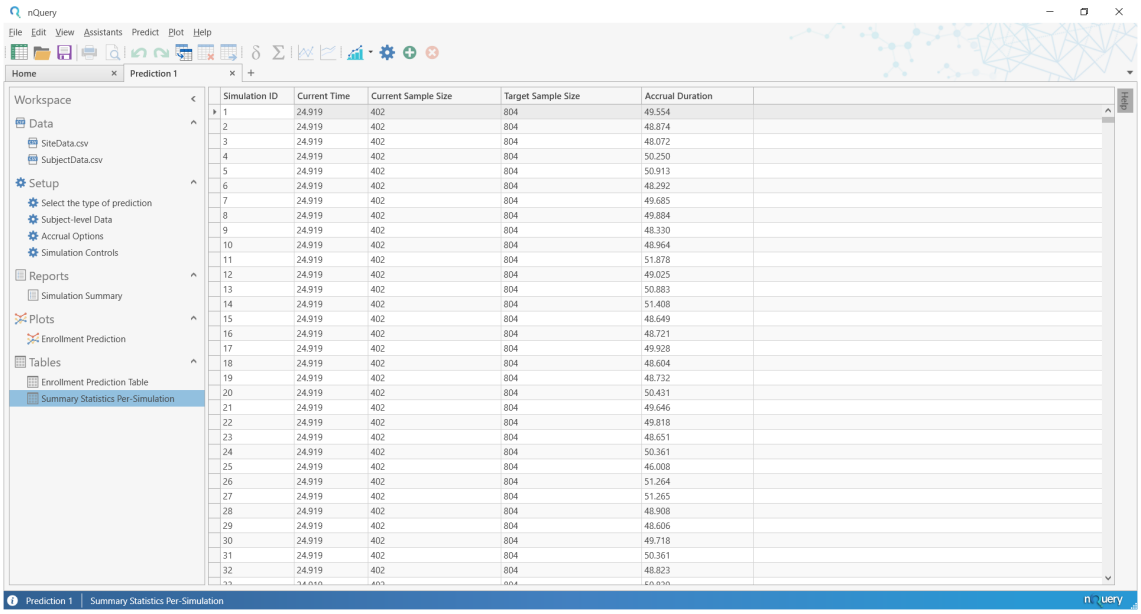



| Time | Actuals | Predicted Avg. Sample Size | Predicted Median Sample Size | 95% Prediction Interval LL | 95% Prediction Inter... |
|---|---|---|---|---|---|
| 22.427 | 398 | | | | |
| 22.925 | 400 | | | | |
| 23.423 | 401 | | | | |
| 23.922 | 401 | | | | |
| 24.420 | 401 | | | | |
| 24.919 | 402 | | | | |
| 25.475 | | 410.962 | 411 | 406 | 417 |
| 26.032 | | 419.935 | 420 | 412 | 429 |
| 26.589 | | 428.876 | 429 | 419 | 440 |
| 27.146 | | 437.863 | 438 | 426 | 450 |
| 27.703 | | 446.876 | 447 | 434 | 461 |
| 28.260 | | 455.871 | 456 | 442 | 470 |
| 28.817 | | 464.886 | 465 | 450 | 481 |
| 29.374 | | 473.802 | 474 | 458 | 491 |
| 29.931 | | 482.788 | 483 | 465 | 501 |
| 30.488 | | 491.786 | 492 | 474 | 511 |
| 31.045 | | 500.782 | 501 | 482 | 521 |
| 31.601 | | 509.768 | 510 | 490 | 530 |
| 32.158 | | 518.687 | 519 | 498 | 540 |
| 32.715 | | 527.704 | 528 | 506 | 550 |
| 33.272 | | 536.725 | 537 | 514 | 560 |
| 33.829 | | 545.698 | 545 | 522 | 569 |
| 34.386 | | 554.655 | 554 | 531 | 579 |
| 34.943 | | 563.608 | 563 | 538 | 589 |
| 35.500 | | 572.566 | 573 | 547 | 599 |
| 36.057 | | 581.541 | 582 | 556 | 608 |
| 36.614 | | 590.472 | 590 | 564 | 618 |
| 37.170 | | 599.477 | 599 | 572 | 627 |
| 37.727 | | 608.464 | 608 | 581 | 637 |
| 38.284 | | 617.451 | 617 | 589 | 646 |
| 38.841 | | 626.425 | 626 | 597 | 656 |
| 39.398 | | 635.401 | 635 | 606 | 666 |

**Figure 6.22:** Enrollment Prediction Table for Enrollment Prediction using Subject-level Data

**Time** is the time value relative to the study start time of zero. **Actual** is the actual sample size achieved at a pre-simulation **Time** row. **Predicted Avg. Sample Size** is the predicted average (mean) sample size enrolled at a post-simulation **Time** row. **Predicted Median Sample Size** is the predicted median sample size enrolled at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the sample size enrolled at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the sample size enrolled at a post-simulation **Time** row.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Target Sample Size, Accrual Duration.** The **Summary Statistics Per-Simulation Table** for this demonstration is shown in Figure 6.23.




| Simulation ID | Current Time | Current Sample Size | Target Sample Size | Accrual Duration |
|---|---|---|---|---|
| 1 | 24.919 | 402 | 804 | 49.554 |
| 2 | 24.919 | 402 | 804 | 48.874 |
| 3 | 24.919 | 402 | 804 | 48.072 |
| 4 | 24.919 | 402 | 804 | 50.250 |
| 5 | 24.919 | 402 | 804 | 50.913 |
| 6 | 24.919 | 402 | 804 | 48.292 |
| 7 | 24.919 | 402 | 804 | 49.685 |
| 8 | 24.919 | 402 | 804 | 49.884 |
| 9 | 24.919 | 402 | 804 | 48.330 |
| 10 | 24.919 | 402 | 804 | 48.964 |
| 11 | 24.919 | 402 | 804 | 51.878 |
| 12 | 24.919 | 402 | 804 | 49.025 |
| 13 | 24.919 | 402 | 804 | 50.883 |
| 14 | 24.919 | 402 | 804 | 51.408 |
| 15 | 24.919 | 402 | 804 | 48.649 |
| 16 | 24.919 | 402 | 804 | 48.721 |
| 17 | 24.919 | 402 | 804 | 49.928 |
| 18 | 24.919 | 402 | 804 | 48.604 |
| 19 | 24.919 | 402 | 804 | 48.732 |
| 20 | 24.919 | 402 | 804 | 50.431 |
| 21 | 24.919 | 402 | 804 | 49.646 |
| 22 | 24.919 | 402 | 804 | 49.818 |
| 23 | 24.919 | 402 | 804 | 48.651 |
| 24 | 24.919 | 402 | 804 | 50.361 |
| 25 | 24.919 | 402 | 804 | 46.008 |
| 26 | 24.919 | 402 | 804 | 51.264 |
| 27 | 24.919 | 402 | 804 | 51.265 |
| 28 | 24.919 | 402 | 804 | 48.908 |
| 29 | 24.919 | 402 | 804 | 48.606 |
| 30 | 24.919 | 402 | 804 | 49.718 |
| 31 | 24.919 | 402 | 804 | 50.361 |
| 32 | 24.919 | 402 | 804 | 48.823 |

**Figure 6.23:** Summary Statistics Per-Simulation Table for Enrollment Prediction using Subject-level Data

**Simulation ID** is in order of simulation from 1 up to **Number of Simulations**, **Current Time, Current Sample Size** and **Target Sample Size** summarize the user inputs and are constant, and the **Accrual Duration** is the length of time needed to achieve the **Target Sample Size** in that simulation.

**Alternative Scenarios** The demonstration has focused on the default inputs using **SubjectData.csv** up to this point. However, as noted in the data summary given in subsection 6.2.4 the study from which the example data was simulated had an original sample size target of 460 (402 enrolled in **SubjectData.csv**). Additionally, in the **Enrollment Prediction Plot** there is noticeable "kink" where the accrual rate was slowing down for the final 4-5 months before the current time.

To explore this additional information, we will investigate two additional scenarios

- Scenario A: **Target Sample Size = 460** using default **Accrual Rate**
- Scenario B: **Target Sample Size = 460** using slower "updated" **Accrual Rate**

**Scenario A** The only changes required compared to the first demonstration is to change the **Target Sample Size** to 460 at the **Accrual Options** step. The simplest way to do this is to select the **Setup** header in the Workspace Navigation Bar on the left and then select the **Accrual Options** drop down option. Then edit the **Target Sample Size** 460 in the **Current and Target Sample Sizes** section

in the top-left and select the Next or → button. At the next **Simulation Controls** step, we will add a single simulation run by selecting the checkbox beside the **Save subject-level data for (x) simulation runs** and replace the default of 10 runs with 1 runs at the "(x)" spot. An edited down summary of the **Setup** changes is provided in Figure 6.24.

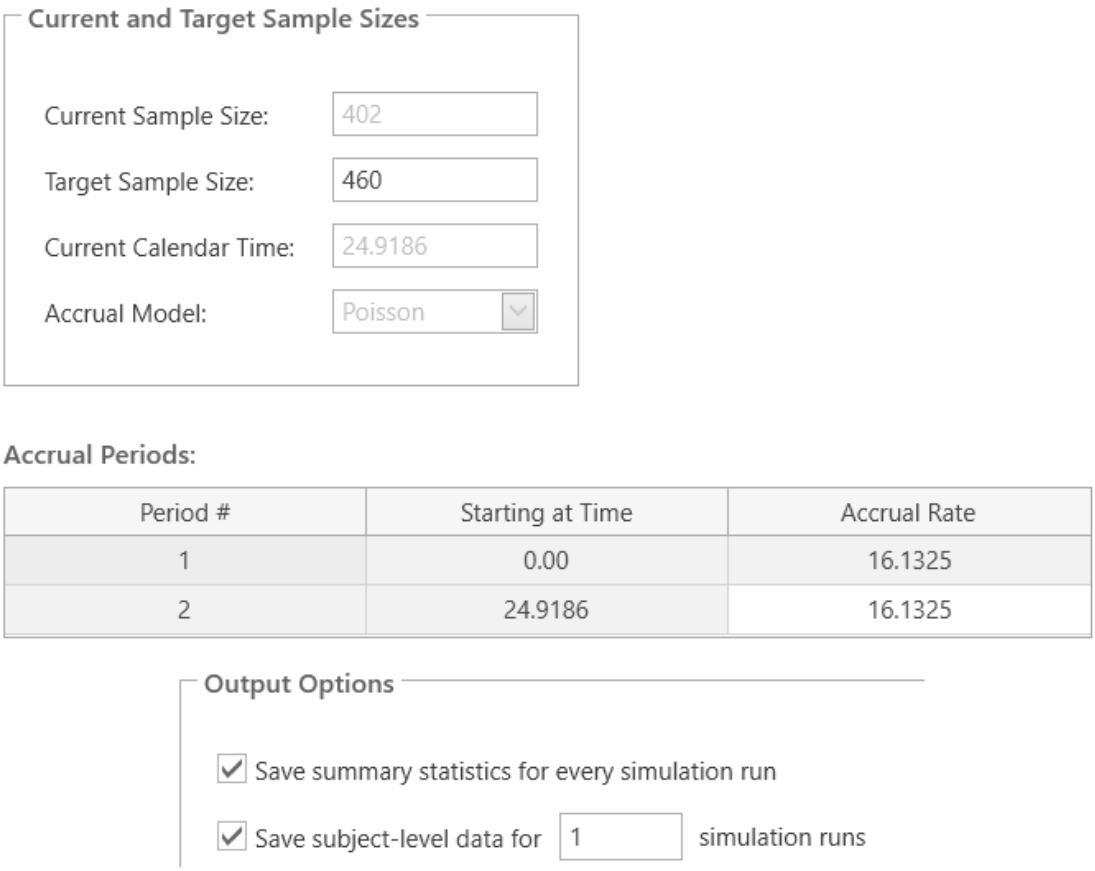



**Figure 6.24:** Scenario A Setup Changes for Enrollment Prediction using Subject-level Data

To run the simulation again with the new Scenario A, select the Run/→ button. The main results of this Scenario A simulation are shown in Figure 6.25.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Study Duration | 28.5155 |
| % Simulations Target Reached | 100.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration |
|---|---|---|
| 5.00% | 460 | 27.7693 |
| 25.00% | 460 | 28.1827 |
| 50.00% | 460 | 28.4985 |
| 75.00% | 460 | 28.8238 |
| 95.00% | 460 | 29.3238 |

**Figure 6.25:** Scenario A Results for Enrollment Prediction using Subject-level Data

For Scenario A, the **Overall Summary** shows that the **Target Sample Size** of 460 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 28.5155. The **Percentile Summary** shows that 90% of predictions had an **Accrual Duration** between 27.7693 (5% Percentile) and 29.3238 (95% Percentile), 50% a duration between 28.1827 (25% Percentile) and 28.8238 (75% Percentile) and a median (50% Percentile) duration of 28.4985.

Given that the original target accrual period length was 30 months, this indicates that if the average accrual rate over the entire period up to the current time was being maintained that the accrual period would around 1.5 months less than expected.

The **Per-Simulation Subject-level Data Table** was also generated here and is shown for Scenario A in Figure 6.26 with the first simulated subject (and their **Arrival Time**) in the first simulation (**Simulation ID = 1**) highlighted. Note that the order of the Subject ID is based on **Arrival Time** order.

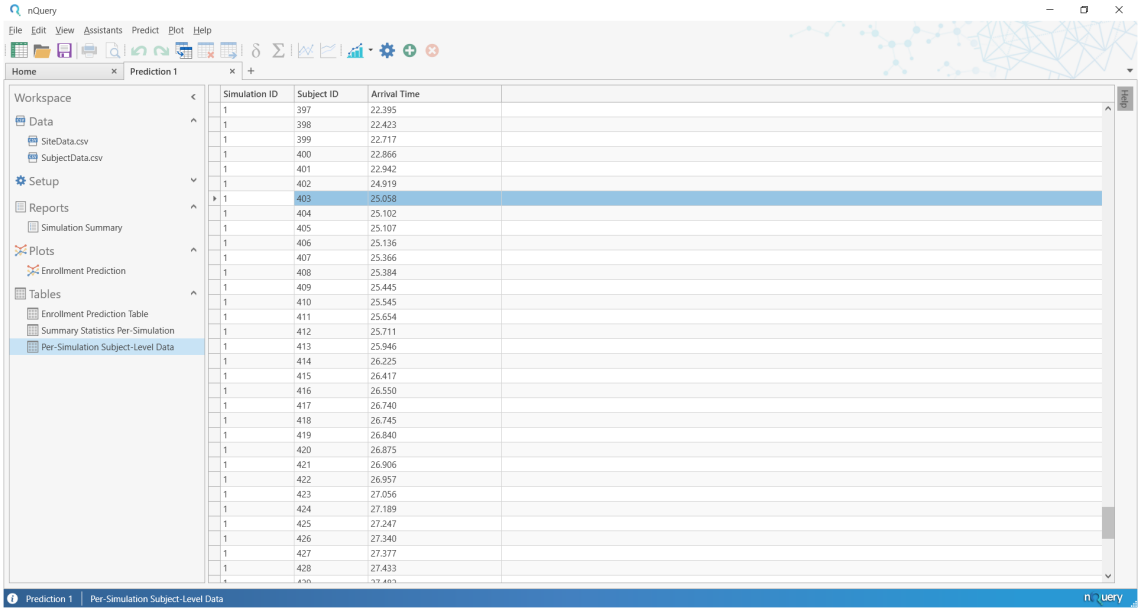



**Figure 6.26:** Per-Simulation Subject-level Data Table for Subject-level Data

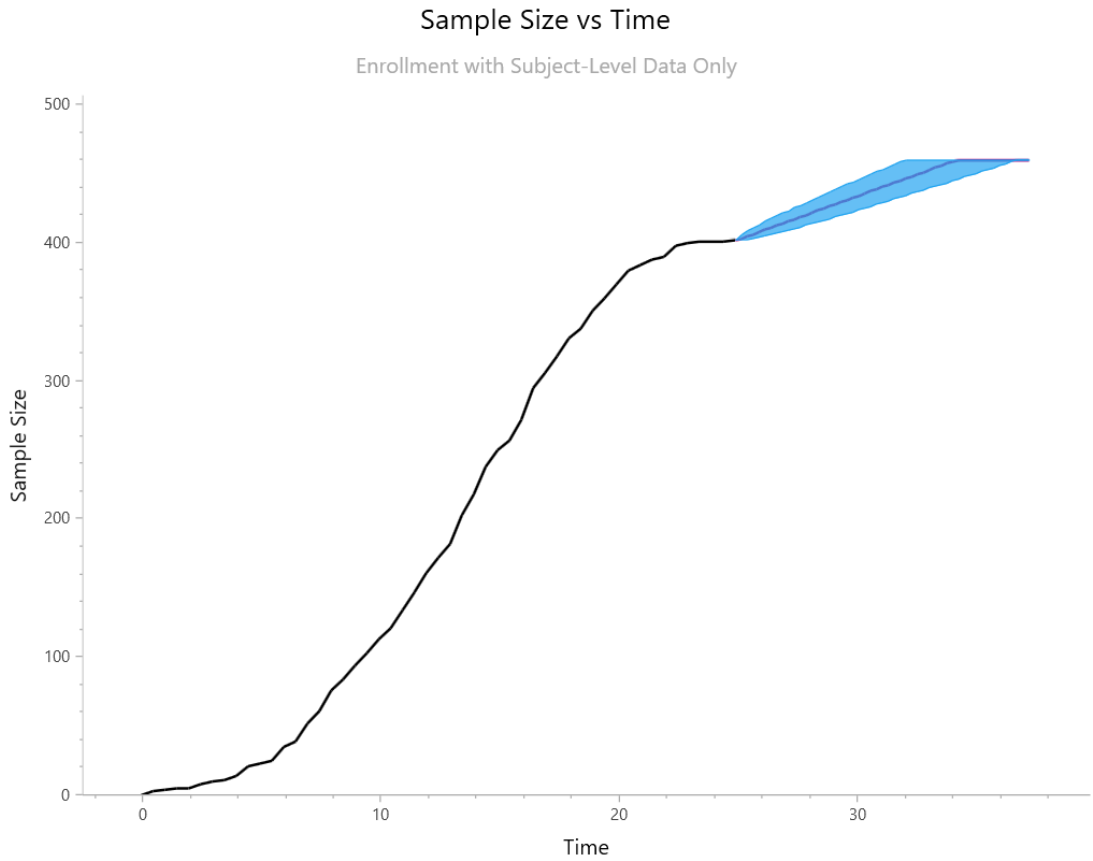
**Scenario B** Scenario A assumed the **Accrual Rate** would continue at average rate over the entirety of the period up until this simulation was made. However, there is evidence that in the last five months there was a significant reduction in the **Accrual Rate**, perhaps reflecting a wind down of the enrollment process.

To evaluate this scenario, we note that 31 subjects where recruited after 20 months of the accrual period is passed. The remaining accrual period after 20 months to the current time equals 4.9186 (24.9816 (Current Time) - 20). This gives an average accrual rate over this period of ($31/4.9186 \approx 6.3$). This is significantly lower than the average 16.13 **Accrual Rate** over the entire period up to now.

To evaluate this scenario, select **Accrual Options** from under the **Setup** header in the Workspace Navigation Bar on the left then change the **Accrual Rate** in the **Period # = 2** row to 6.3 at the **Accrual Options** step by and then select the Next or $\rightarrow$ button. At the next **Simulation Controls** step, select the Run or $\rightarrow$ button to simulate Scenario B. The changes and results for Scenario B are summarized in Figure 6.27.

Accrual Periods:

| Period # | Starting at Time | Accrual Rate |
|---|---|---|
| 1 | 0.00 | 16.1325 |
| 2 | 24.9186 | 6.30 |




Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Study Duration | 34.1373 |
| % Simulations Target Reached | 100.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration |
|---|---|---|
| 5.00% | 460 | 32.2229 |
| 25.00% | 460 | 33.3058 |
| 50.00% | 460 | 34.0726 |
| 75.00% | 460 | 34.9239 |
| 95.00% | 460 | 36.2008 |

**Figure 6.27:** Scenario B for Enrollment Prediction with Subject-level Data

For Scenario B, the **Overall Summary** shows that the **Target Sample Size** of

460 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 34.1373 with the **Percentile Summary** showing that 90% of predictions had a **Accrual Duration** between 32.2229 and 36.2008, 50% a duration between 33.3058 and 34.9239 with a median duration of 34.0726.

Given that the original target accrual period length was 30 months, this indicates that if the reduced accrual rate over approximately last five months were to continue then the accrual period may be approximately 4 months longer than expected. This may indicate that action is needed to get the trial back on track compared to Scenario A where the trial was assumed to be ahead of schedule.

#### 6.2.6.2 Enrollment Prediction using Subject and Site-level Interim Data

**Setup** **Enrollment Only** milestone prediction using **Subject + Site-level Data** means that the prediction will project the expected length of the accrual period needed to recruit a specified **Target Sample Size** using subject-level and site-level interim data.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step select **Enrollment Only** from the **Target** field and **Subject + Site Data** from the **Input** field. **Enrollment Status** must be **Ongoing** for enrollment milestone prediction. These selections are shown in Figure 6.28. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

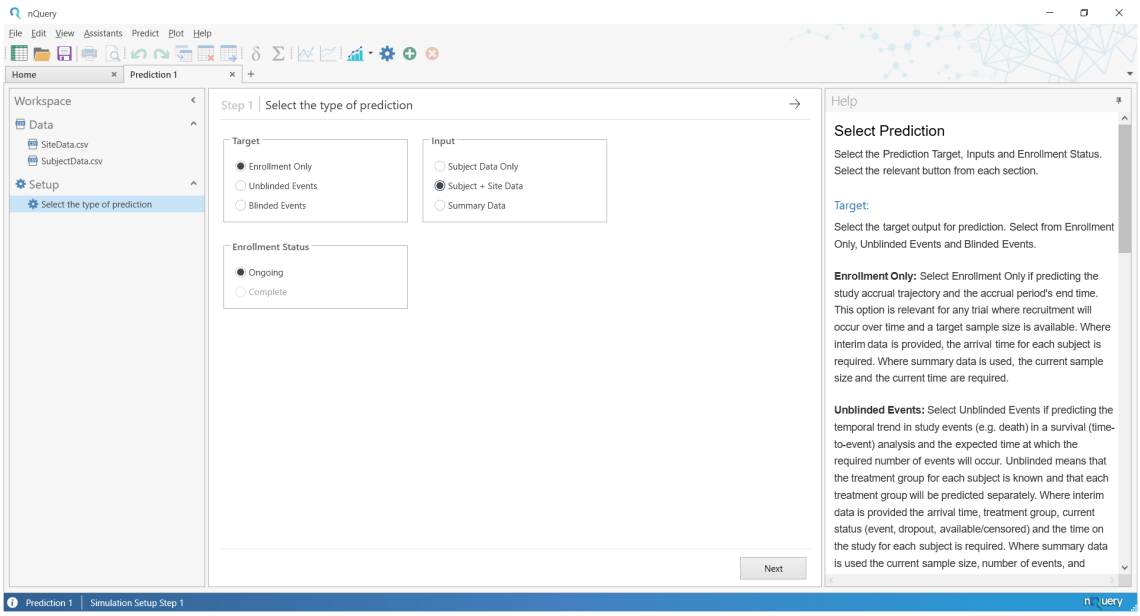



**Figure 6.28:** Select the Type of Prediction Setup Step for Enrollment Prediction using Subject + Site-level Data

The next step is the **Subject-level Data** step. The **Subject-level Data** step is a Data Field Selection step. The **Subject-level Data** is the dataset where each row corresponds to a specific subject enrolled in the study.

Data Field Selection steps are steps in which the required dataset in the current Workspace is selected and columns in the selected dataset are assigned to the required fields for that prediction problem. The options in a given Data Field Selection field are opened by selecting the button on the right of each field which is found to the right of the field name e.g. **Select Data**. Each field will have specific requirements (e.g. numeric only, two distinct values only) and nQuery Predict only displays options which are consistent with those fields.

In this prediction there are three fields which are as follows:

- Select Dataset: Select the dataset that is the Subject-level Data from the datasets in the current workspace
- Arrival Time: The arrival time column contains the length of time since the study start (Time = 0) until a subject entered the study. Each subject's Arrival Time should be greater than or equal to zero
- Site ID: Site ID of site from which a subject was recruited. Site IDs here must correspond to Site ID values contained in the Site-level Data Site ID column

For this prediction, in the **Select Dataset** field select **SubjectData.csv,** in the **Arrival Time** field select **Arrival** and in the **Site ID** field select **SID.** These selections are shown in Figure 6.29. Select the Next or → button to move to the next **Setup** step.

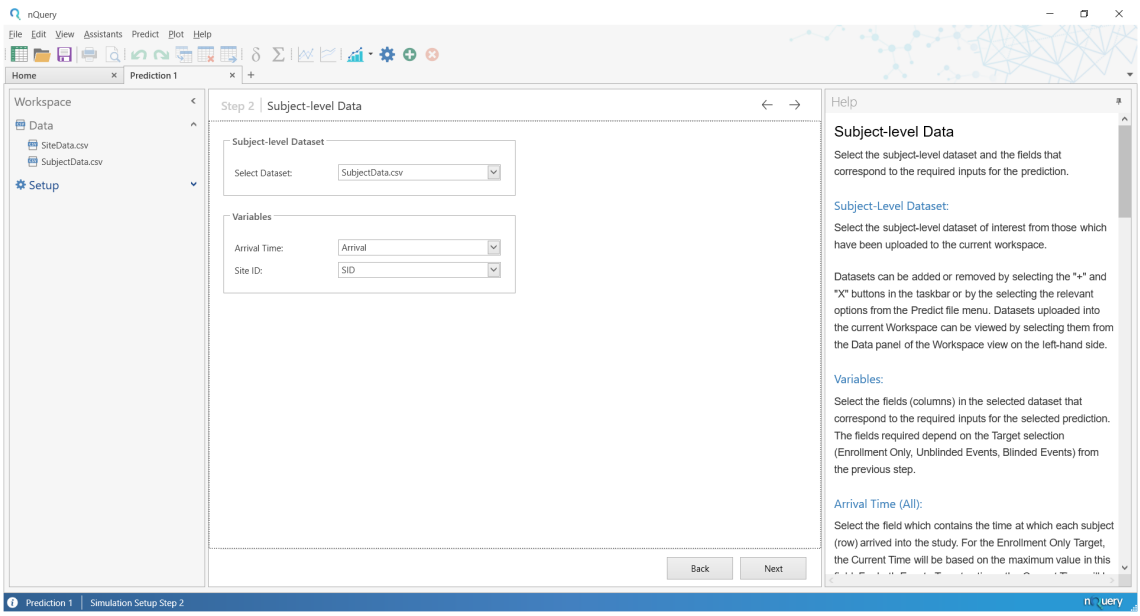



**Figure 6.29:** Subject-level Data Setup Step for Enrollment Prediction using Subject + Site-level Data

The next step is the **Site-level Data** step. The **Site-level Data** step is a Data Field Selection step as per the **Subject-level Data** step. The **Site-level Data** is the dataset where each row corresponds to a specific site.

The **Site-level Data** step fields are the same for all predictions. These fields are:

- Select Dataset: Select the dataset that is the Site-level Data from the datasets in the current workspace
- Site ID: Site ID for each site. Each row should have a distinct Site ID and Site IDs from Subject-level Data must be included in the list of Site IDs selected here
- Accrual Rate/Site: The planned accrual rate (subjects recruited per unit time) in a site. Each site's Accrual Rate/Site should be a numeric value greater than zero
- Enrollment Cap: The maximum number of subjects that can be recruited from that site. Each site's Enrollment Cap should be an integer greater than the higher of zero and the No. of Accruals (see below)
- Site Initiation Time (Opened Sites): The time at which a site was open and able to start enrolling subjects. Each site's Site Initiation Time should be a numeric value less than the Current Time
- Start Time (Unopened Sites): The start time of the window during which an unopened site could open. Optional field for unopened sites, should be numeric value greater than or equal to the Current Time
- End Time (Unopened Sites): The start time of the window during which an unopened site could open. Optional field for unopened sites, should be numeric value greater than or equal to the Start Time (Unopened Sites) for that site

For this prediction, in the **Select Dataset** field select **SiteData.csv,** in the **Site ID** field select **SID**, in the **Accrual Rate/Site** field select **Rate**, in the **Enrollment Cap** field select **Cap**, in the **Site Initiation Time** field select **OpenTime**, in the **Start Time** field select **Start** and in the **End Time** field select **End.** These selections are shown in Figure 6.30. Select the Next or → button to move to the next **Setup** step.

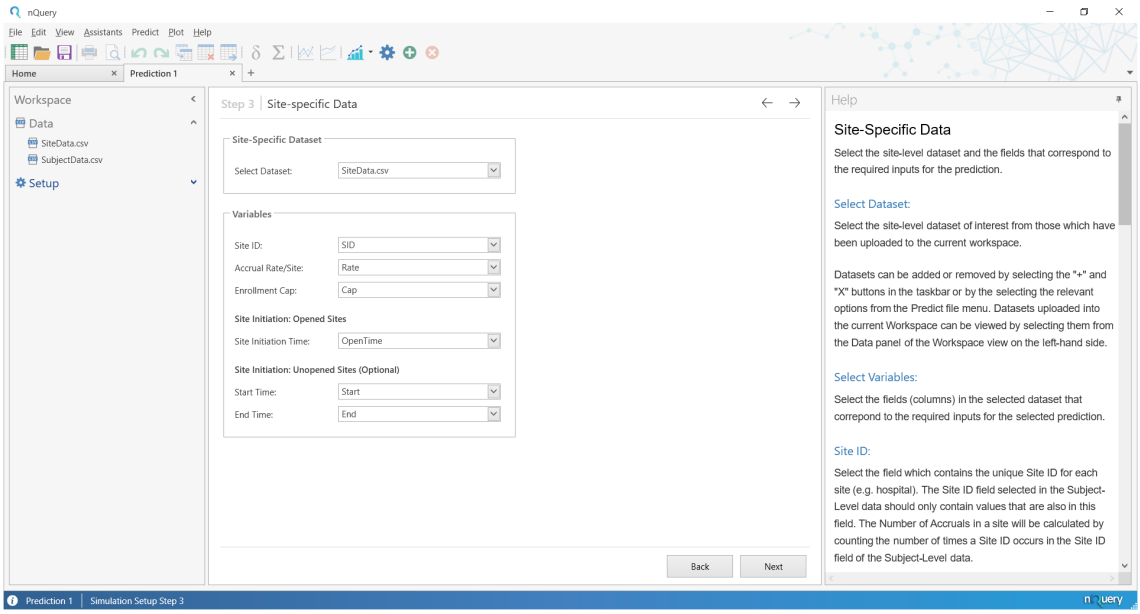



**Figure 6.30:** Site-level Data Setup Step

The next step is the **Accrual Options** step. When **Site-level Data** is used, the **Accrual Options** consists of two tabs: **Sample Size & Followup Options** and **Accrual Options/Site**. A given tab can be opened by selecting the grey tab name at the top of the main window, just below the **Step 4 | Accrual Options** name. These tabs are highlighted in Figure 6.31.

The **Sample Size & Followup Options** tab provides overall information about the trial up until the current time alongside the **Target Sample Size** and **Accrual Model.** The **Sample Size & Followup Options** tab for this demonstration is shown in Figure 6.31.

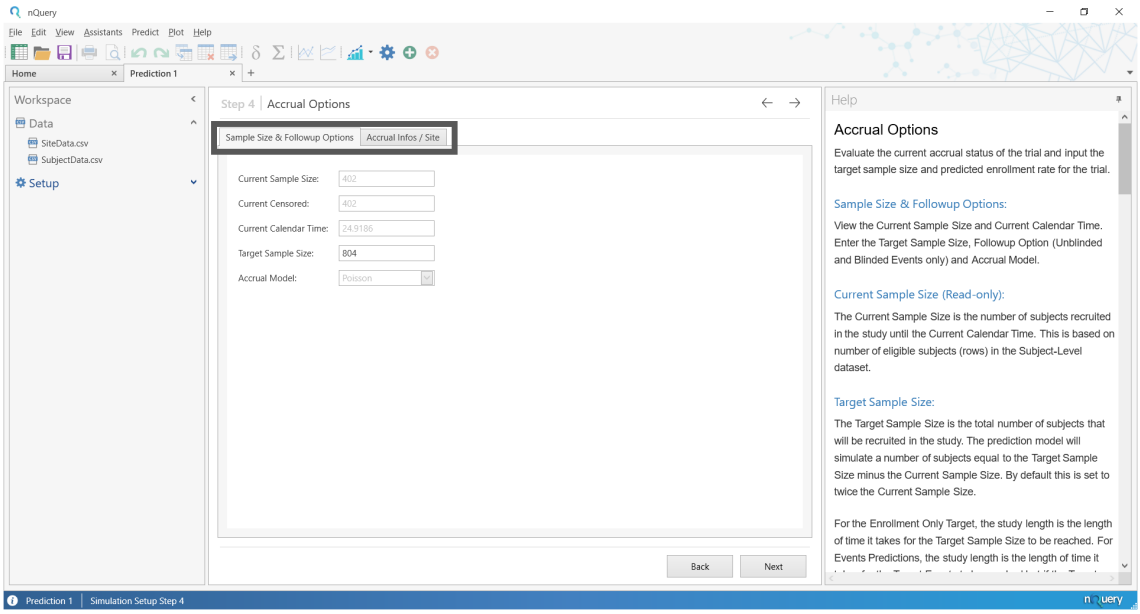



**Figure 6.31:** Accrual Options Step Sample Size & Followup Options tab for Site-level Data

The **Sample Size & Followup Options** fields are:

- Current Sample Size (Read-only): Sample Size recruited into a study so far. Equals the number of rows in the Subject-level Data
- Current Time (Read-only): The length of time that has passed since the study started. Equals the maximum Arrival Time value
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study. The number of simulated subjects equals the Total Sample Size minus the Current Sample Size
- Accrual Model (Editable): The statistical model that will be used to generate simulated enrollments

In nQuery Predict, the **Target Sample Size** defaults to twice the **Current Sample Size** and the **Accrual Model** defaults to **Poisson.** In this demonstration, the **Current Sample Size** equals 402 and therefore the default **Target Sample Size** is $402\,(2) = 804$.

To open the **Accrual Options/Site** tab select the tab at the top of the main window with that name. This is shown in Figure 6.32

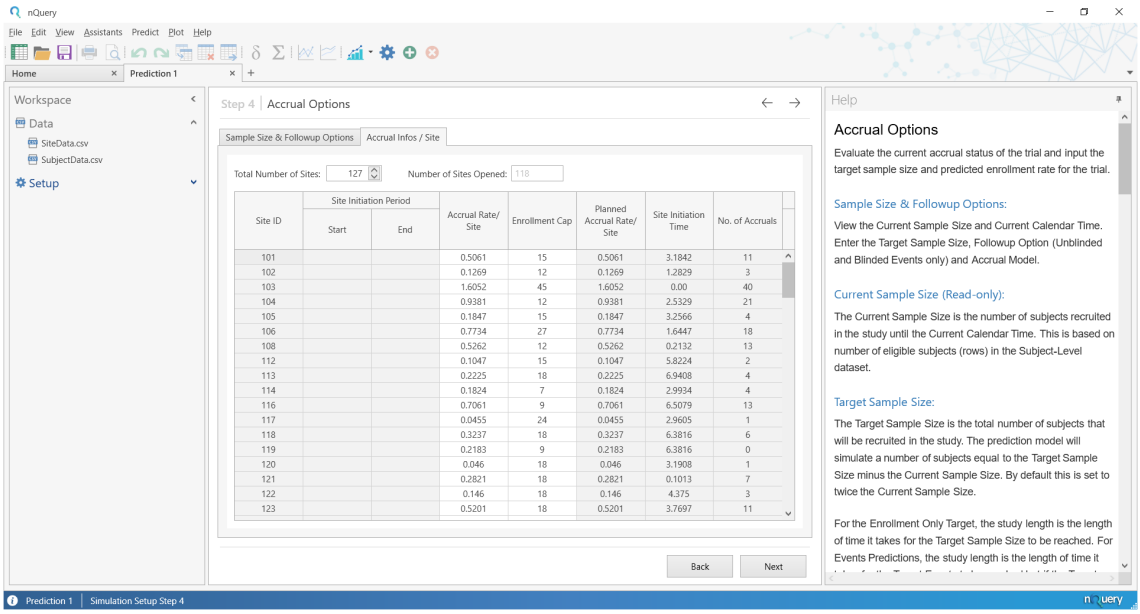



**Figure 6.32:** Accrual Options Step Accrual Options/Site tab for Site-level Data

The **Accrual Options/Site** tab has three main elements: the **Total Number of Sites** field, the **Number of Sites Open** field and the **Site Accrual Table**.

The **Total Number of Sites** and **Number of Sites Open** field are found at the top of tab window. The **Total Number of Sites** field sets the total number of sites that will be used to simulate subjects in this milestone prediction and by default is equal to the number of rows/sites in the **Site-level Data**.

The **Number of Sites Open** field is the number sites that are already open at the time the simulation is being performed and is the number of sites which had a **Site Initiation Time** in the site-level data. The **Total Number of Sites** must always be greater than or equal to the **Number of Sites Open**.

The **Site Accrual Table** consists of eight columns with each row corresponding to required information for each site. The columns are defined as follows:

- Site ID (Read-only for Open Sites, Editable for Unopened Sites): Site ID for a site. Taken directly from Site-level Data. All Site ID must be unique and include all Site ID in the subject-level Site ID field
- Site Initiation Period - Start Time (Editable for Unopened Sites, Numeric $\geq$ Current Time): Start time for window during which unopened site can open. Not used for open sites
- Site Initiation Period - End Time (Editable for Unopened Sites, Numeric $\geq$ Start Time): End time for window during which unopened site can open. Not used for open sites
- Accrual Rate/Site (Editable, Numeric $> 0$): The accrual rate (subjects recruited per unit time) that will be used to simulate subjects for that site

- Enrollment Cap (Editable, Numeric $\geq$ No. of Accruals) The maximum number of subjects that can be recruited in a site. This must be greater than the No. of Accruals in a site
- Planned Accrual Rate/Site (Read-only): The original planned accrual rate in a site. Taken from Site-level Data
- Site Initiation Time (Read-only): The site initiation time for opened sites. Taken from Site-level Data. Not used for unopened sites
- No. of Accruals (Read-only): The number of subjects that have been enrolled in an open site at the current time. Not used for unopened sites

**Site ID, Site Initiation Period (Start Time, End Time), Enrollment Cap, Planned Accrual Rate/Site** and **Site Initiation Time** are taken directly from the **Site-level Data**.

For sites which have already recruited subjects (i.e. **No. of Accruals** $> 0$), the default **Accrual Rate/Site** is calculated using the **No. of Accruals** and **Site Initiation Time** using the following relationship:

$$R = \frac{\#A}{(CT - SIT)}$$

where $R$ is the **Accrual Rate/Site**, $\#A$ is the **No. of Accruals** in a site, $CT$ is the **Current Time** and $SIT$ is the **Site Initiation Time** for that site.

For sites which are unopened sites or which have not recruited a subject (i.e. **No. of Accruals** $= 0$), the **Planned Accrual Rate/Site** from the **Site-level Data's Accrual Rate** field is used.

The **No. of Accruals** is calculated by counting the number of times each **Site ID** value occurred in the **Subject-level Data's Site ID** column.

As referenced above, the inputs for the **Site Accrual Table** for each site depend on whether a site is open or unopened. Unopened sites are automatically placed at the bottom of the **Site Accrual Table**.

The user can also add additional sites (rows) by increasing the **Total Number of Sites** field above the original value derived from the number of rows in the **Site-level Data**. This will add a blank row at the bottom of the **Site Accrual Table**. Note **Setup** cannot continue until all new rows are filled appropriately. An example of **Site Accrual Table** with the unopened sites from the **SiteLevel.csv** data alongside one blank new site by increasing **Total Number of Sites** to 128 is provided in Figure 6.33.

Sample Size & Followup Options | Accrual Infos / Site

Total Number of Sites: 128 Number of Sites Opened: 118

| Site ID | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap | Planned Accrual Rate/ Site | Site Initiation Time | No. of Accruals |
|---|---|---|---|---|---|---|---|
| | Start | End | | | | | |
| 607 | | | 0.3517 | 10 | 0.3517 | 7.8592 | 6 |
| 608 | | | 0.2815 | 9 | 0.2815 | 10.0658 | 0 |
| 609 | | | 2.2244 | 12 | 2.2244 | 19.375 | 0 |
| 610 | | | 0.1848 | 16 | 0.1848 | 8.6842 | 3 |
| 611 | | | 0.1739 | 9 | 0.1739 | 7.6645 | 3 |
| 620 | | | 0.1307 | 9 | 0.1307 | 9.2434 | 0 |
| 621 | | | 0.0567 | 18 | 0.0567 | 7.2697 | 1 |
| 622 | | | 0.052 | 18 | 0.052 | 5.6908 | 1 |
| 623 | | | 0.0517 | 10 | 0.0517 | 5.5592 | 1 |
| 144 | 25.51 | 26.11 | 0.50 | 14 | 0.50 | | |
| 145 | 25.51 | 26.11 | 0.50 | 14 | 0.50 | | |
| 347 | 25.00 | 25.30 | 0.10 | 6 | 0.10 | | |
| 368 | 25.00 | 25.75 | 0.20 | 12 | 0.20 | | |
| 369 | 25.00 | 25.75 | 0.20 | 12 | 0.20 | | |
| 420 | 25.10 | 25.20 | 0.30 | 16 | 0.30 | | |
| 423 | 25.10 | 25.20 | 0.15 | 8 | 0.15 | | |
| 443 | 25.00 | 25.50 | 0.12 | 8 | 0.12 | | |
| 543 | 25.00 | 26.00 | 0.50 | 24 | 0.50 | | |

**Figure 6.33:** Unopened and New Sites in Site Accrual Table

For this demonstration, assume the **Total Number of Sites** is reset to its default of 127. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.34. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

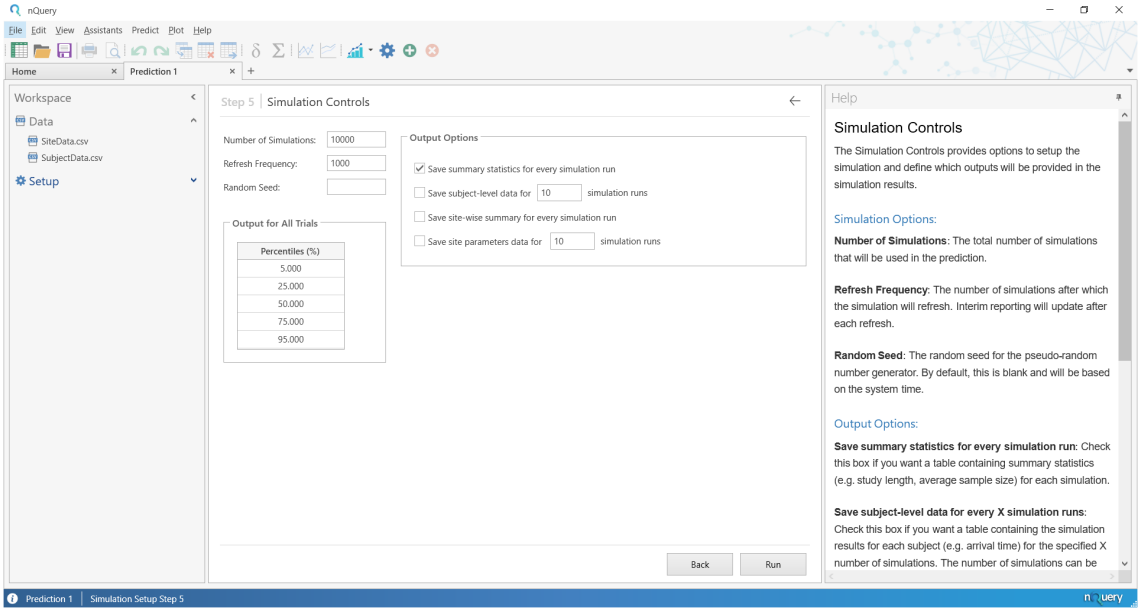



**Figure 6.34:** Simulation Controls Setup Step for Enrollment Prediction using Subject + Site-level Data

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.35.

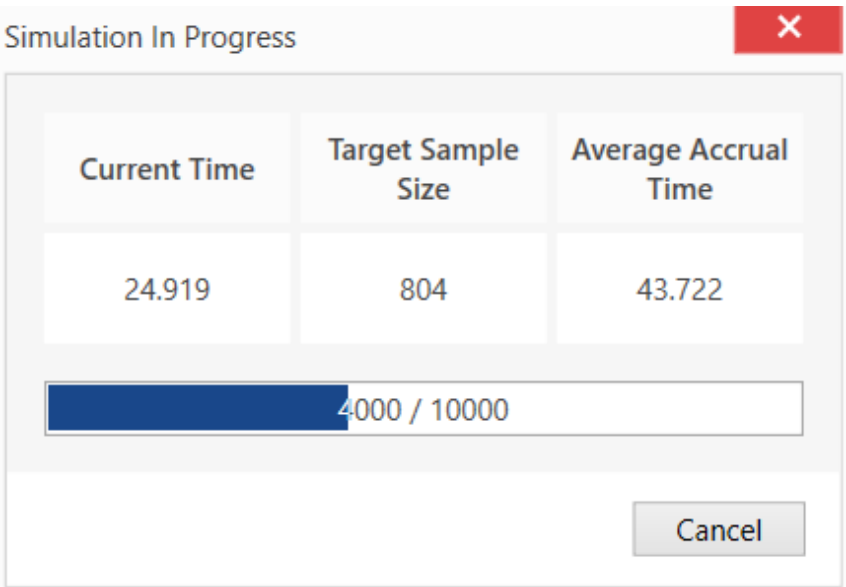



**Figure 6.35:** Simulation in Progress Window for Enrollment Prediction using Subject + Site-level Data

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.36.

**Figure 6.36:** Enrollment Prediction Plot for Enrollment Prediction using Subject + Site-level Data

The results of a nQuery Predict milestone prediction can be divided into three categories: **Reports, Plots** and **Tables**. A high-level summary of the results in nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.37.

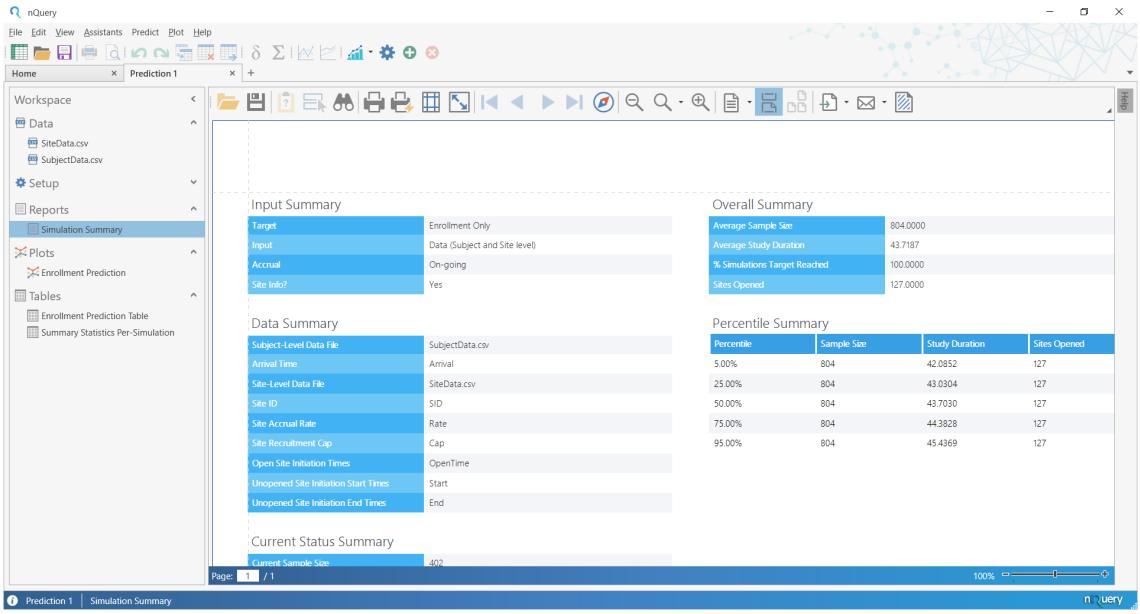



**Figure 6.37:** Simulation Summary Report for Enrollment Prediction using Subject + Site-level Data

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.38. For this demonstration, the **Overall Summary** shows that the **Target Sample Size** of 804 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 43.7187. All 127 **Sites Opened** in all simulations. The **Percentile Summary** shows that 90% of predictions had an **Accrual Duration** between 42.0852 (5% Percentile) and 45.4369 (95% Percentile), 50% a duration between 43.0304 (25% Percentile) and 44.3828 (75% Percentile) and a median (50% Percentile) duration of 43.7030.

Overall Summary

| | |
|---|---|
| Average Sample Size | 804.0000 |
| Average Study Duration | 43.7187 |
| % Simulations Target Reached | 100.0000 |
| Sites Opened | 127.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration | Sites Opened |
|---|---|---|---|
| 5.00% | 804 | 42.0852 | 127 |
| 25.00% | 804 | 43.0304 | 127 |
| 50.00% | 804 | 43.7030 | 127 |
| 75.00% | 804 | 44.3828 | 127 |
| 95.00% | 804 | 45.4369 | 127 |

**Figure 6.38:** Simulation Summary Report Results for Enrollment Prediction with Subject + Site-level Data

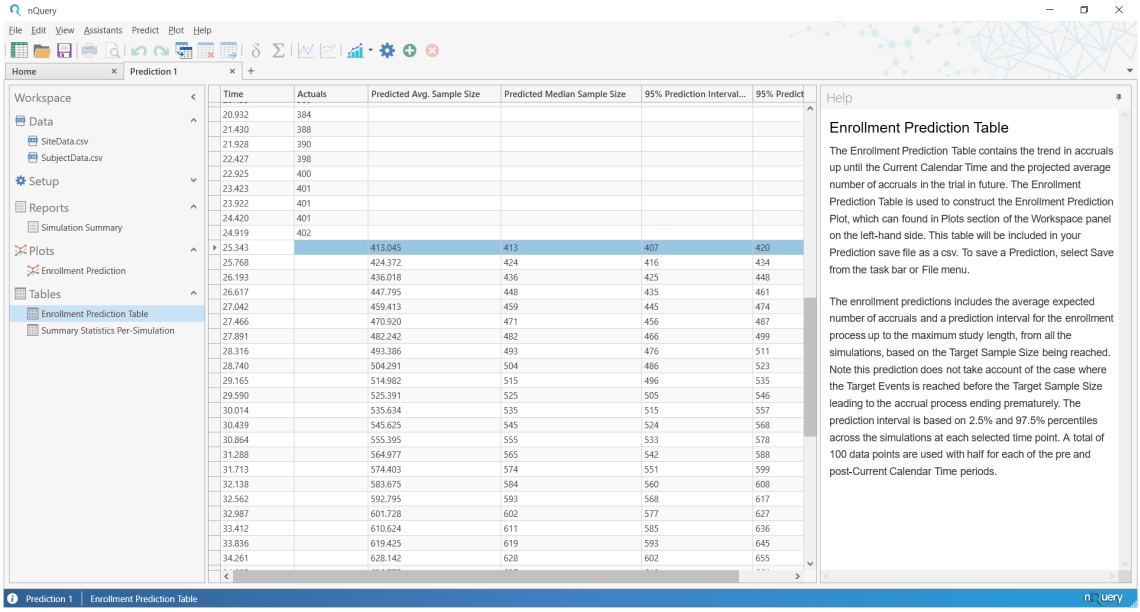
To edit the **Simulation Summary report there a** number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment simulation from the study start to when the **Target Sample Size** was reached in all simulations. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.39.

**Figure 6.39:** Enrollment Prediction Plot for Enrollment Prediction using Subject + Site-level Data

The **Enrollment Prediction** plot consists of two parts: from Time = 0 to **Current Time** and from **Current Time** to the **Maximum Time**. The first part is a black line which plots the accrual pattern from the user inputs provided. The second part is a blue conic section with a grey line where the blue conic section plots the 95% prediction interval for the **Sample Size** enrolled at a given **Time** across all simulations and the grey line is the average **Sample Size** enrolled at a given time. The 95% prediction interval indicates the 95% of predictions had a sample size that fell between the lower and upper limit at a given time.

The **Enrollment Prediction Table** is the table used to create the **Enrollment Prediction** plot, which by default consists of ~100 rows (50 pre-simulation and 50 post-simulation). It consists of five columns: **Time, Actual, Predicted Avg. Sample Size, Predicted Median Sample Size, 95% Prediction Interval LL, 95% Prediction Interval UL**. The **Enrollment Prediction Table** for this demonstration at the start of the simulated period is given in Figure 6.40.




| Time | Actuals | Predicted Avg. Sample Size | Predicted Median Sample Size | 95% Prediction Interval... | 95% Predict |
|---|---|---|---|---|---|
| 20.932 | 384 | | | | |
| 21.430 | 388 | | | | |
| 21.928 | 390 | | | | |
| 22.427 | 398 | | | | |
| 22.925 | 400 | | | | |
| 23.423 | 401 | | | | |
| 23.922 | 401 | | | | |
| 24.420 | 401 | | | | |
| 24.919 | 402 | | | | |
| 25.343 | | 413.045 | 413 | 407 | 420 |
| 25.768 | | 424.372 | 424 | 416 | 434 |
| 26.193 | | 436.018 | 436 | 425 | 448 |
| 26.617 | | 447.795 | 448 | 435 | 461 |
| 27.042 | | 459.413 | 459 | 445 | 474 |
| 27.466 | | 470.920 | 471 | 456 | 487 |
| 27.891 | | 482.242 | 482 | 466 | 499 |
| 28.316 | | 493.386 | 493 | 476 | 511 |
| 28.740 | | 504.291 | 504 | 486 | 523 |
| 29.165 | | 514.982 | 515 | 496 | 535 |
| 29.590 | | 525.391 | 525 | 505 | 546 |
| 30.014 | | 535.634 | 535 | 515 | 557 |
| 30.439 | | 545.625 | 545 | 524 | 568 |
| 30.864 | | 555.395 | 555 | 533 | 578 |
| 31.288 | | 564.977 | 565 | 542 | 588 |
| 31.713 | | 574.403 | 574 | 551 | 599 |
| 32.138 | | 583.675 | 584 | 560 | 608 |
| 32.562 | | 592.795 | 593 | 568 | 617 |
| 32.987 | | 601.728 | 602 | 577 | 627 |
| 33.412 | | 610.624 | 611 | 585 | 636 |
| 33.836 | | 619.425 | 619 | 593 | 645 |
| 34.261 | | 628.142 | 628 | 602 | 655 |

**Figure 6.40:** Enrollment Prediction Table for Enrollment Prediction using Subject + Site-level Data

**Time** is the time value relative to the study start time of zero. **Actual** is the actual sample size achieved at a pre-simulation **Time** row. **Predicted Avg. Sample Size** is the predicted average (mean) sample size enrolled at a post-simulation **Time** row. **Predicted Median Sample Size** is the predicted median sample size enrolled at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the sample size enrolled at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the sample size enrolled at a post-simulation **Time** row.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Target Sample Size, Accrual Duration.** The **Summary Statistics Per-Simulation Table** for this demonstration is shown in Figure 6.41.

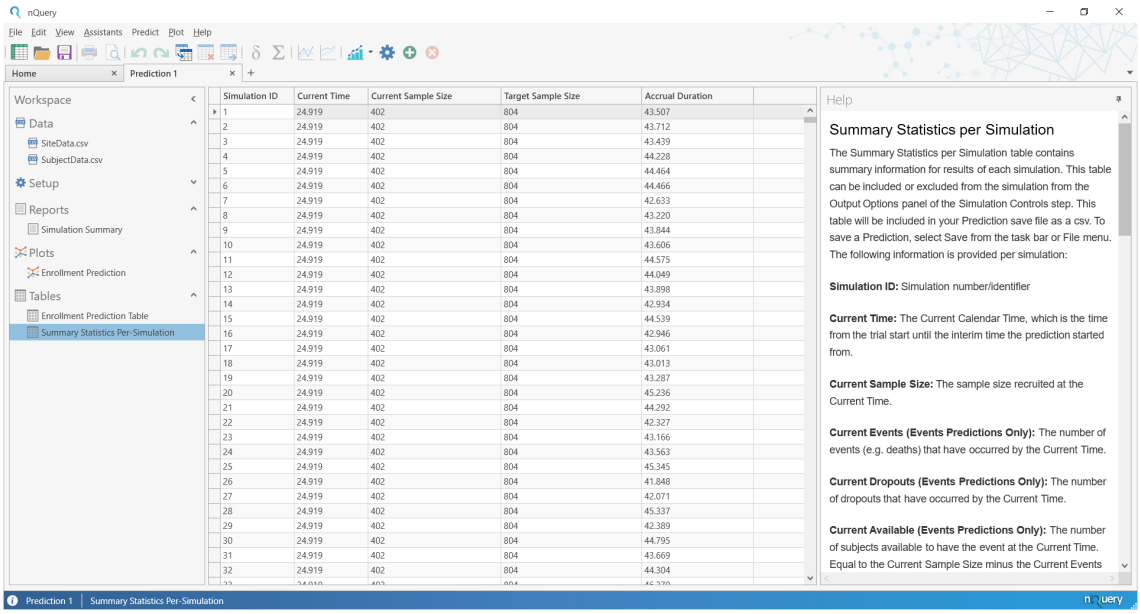



**Figure 6.41:** Summary Statistics Per-Simulation Table for Enrollment Prediction using Subject + Site-level Data

**Simulation ID** is in order of simulation from 1 up to **Number of Simulations**, **Current Time, Current Sample Size** and **Target Sample Size** summarize the user inputs and are constant, and the **Accrual Duration** is the length of time needed to achieve the **Target Sample Size** in that simulation.

**Alternative Scenarios** The demonstration has focused on the default inputs using **SubjectData.csv** up to this point. However, as noted in the data summary given in subsection 6.2.4 the study from which the example data was simulated had an original sample size target of 460 (402 enrolled in **SubjectData.csv**).

However, note that the percentage of the **Target Sample Size** enrolled $(100(402/406) = 87.5)$ at the **Current Time** of 24.9186 months is higher than the percentage of the targetted accrual period length (30 months) used so far $(100\,(24.9186/30) = 83.1\%)$. This could indicate that the planned number of new sites (9) may be higher than necessary and an adjustment may be made to reduce this number and stagger their opening times to avoid being in the process of opening sites when enrollment is complete.

To explore this additional information, we will investigate two additional scenarios

- Scenario A: **Target Sample Size = 460** using defaults for each site
- Scenario B: **Target Sample Size = 460** using less unopened sites and opening them over a longer period

**Scenario A** The change required compared to the first demonstration is to change the **Target Sample Size** to 460 in the **Sample Size & Followup Options** tab of the **Accrual Options** step. The simplest way to do this is to select the **Setup** header in the Workspace Navigation Bar on the left and then select the **Accrual Options** drop down option. Then edit the **Target Sample Size** 460 in the **Sample Size & Followup Options tab** in the top-left and select the Next or → button.

At the **Simulation Controls** step, we will add the three additional tables under the **Output Options** field by selecting the checkbox beside each. We will replace the default of 10 with 1 (at "(x)" spot) for **Save subject-level data for (x) simulation runs** and **Save site parameters data for (x) simulation runs**. An edited down summary of the **Setup** changes is provided in Figure 6.42.

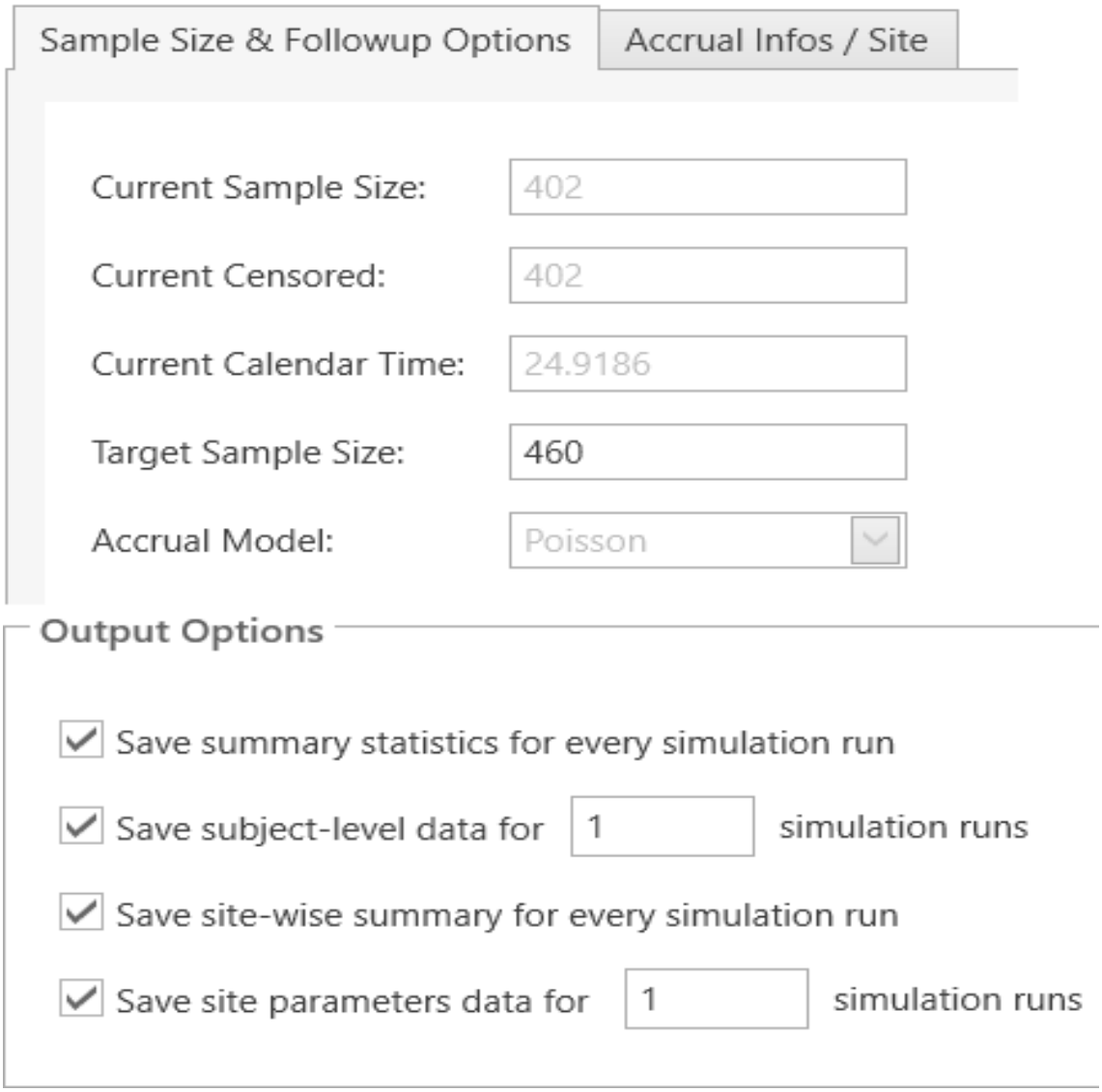
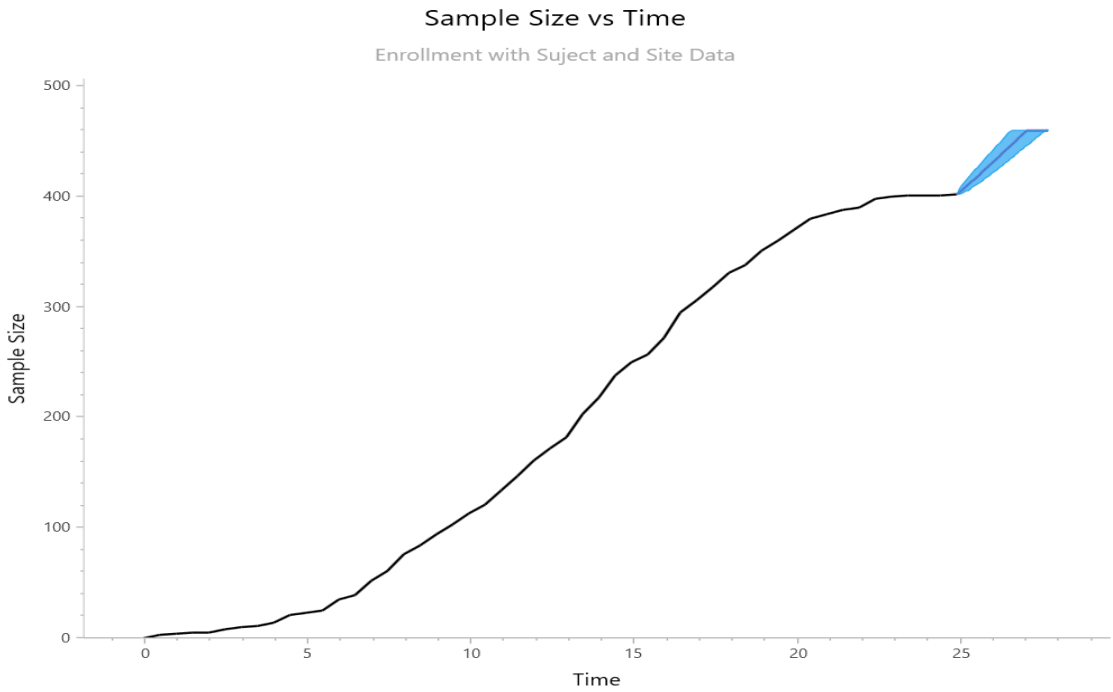



**Figure 6.42:** Scenario A Setup Changes for Enrollment Prediction using Subject + Site-level Data

To run the simulation again with the new Scenario A, select the Run/$^{\rightarrow}$ button. The main results of this Scenario A simulation are shown in Figure 6.43.




Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Study Duration | 27.0594 |
| % Simulations Target Reached | 100.0000 |
| Sites Opened | 127.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration | Sites Opened |
|---|---|---|---|
| 5.00% | 460 | 26.6307 | 127 |
| 25.00% | 460 | 26.8680 | 127 |
| 50.00% | 460 | 27.0472 | 127 |
| 75.00% | 460 | 27.2409 | 127 |
| 95.00% | 460 | 27.5243 | 127 |

**Figure 6.43:** Scenario A Results for Enrollment Prediction using Subject + Site-level Data

For Scenario A, the **Overall Summary** shows that the **Target Sample Size** of 460 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 27.0594. The **Percentile Summary** shows that 90% of predictions had an **Accrual Duration** between 26.6307 (5% Percentile) and 27.5243 (95% Percentile), 50% a duration between 26.8680 (25% Percentile) and 27.2409 (75% Percentile) and a median (50% Percentile) duration of 27.0472.

Given that the original target accrual period length was 30 months, this indicates that if the average accrual rate over the entire period up to the current time was being maintained that the accrual period would around 3 months less than expected.

The **Per-Simulation Subject-level Data Table** was also generated here and

is shown for Scenario A in Figure 6.44 with the first simulated subject (and their **Arrival Time** and **Site ID** for the site they were generated from) in the first simulation (**Simulation ID = 1**) highlighted. Note that the order of the Subject ID is based on **Arrival Time** order.

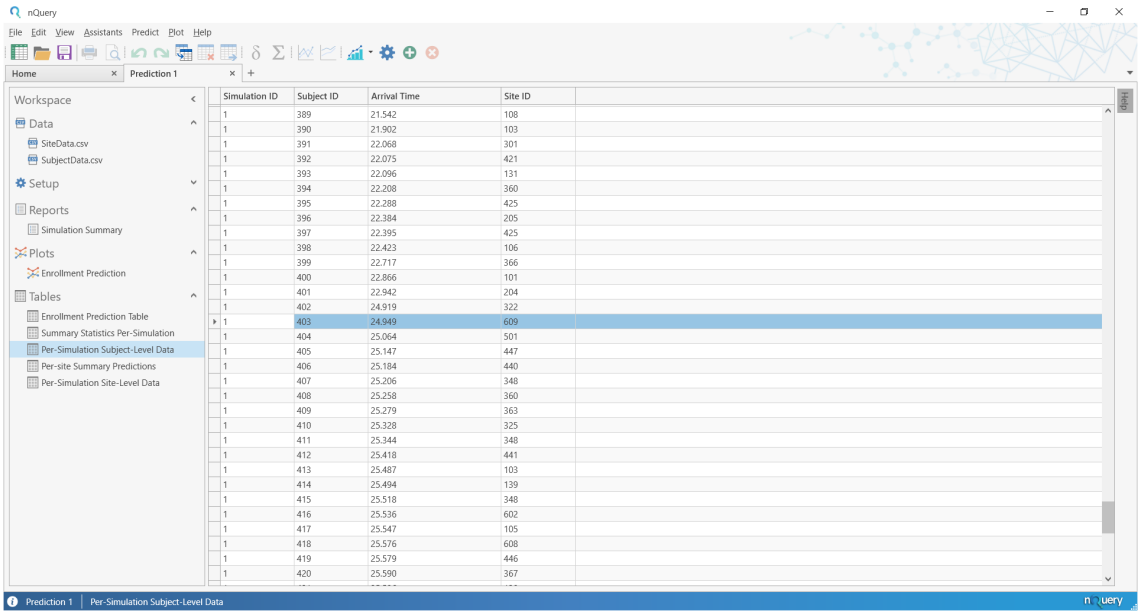
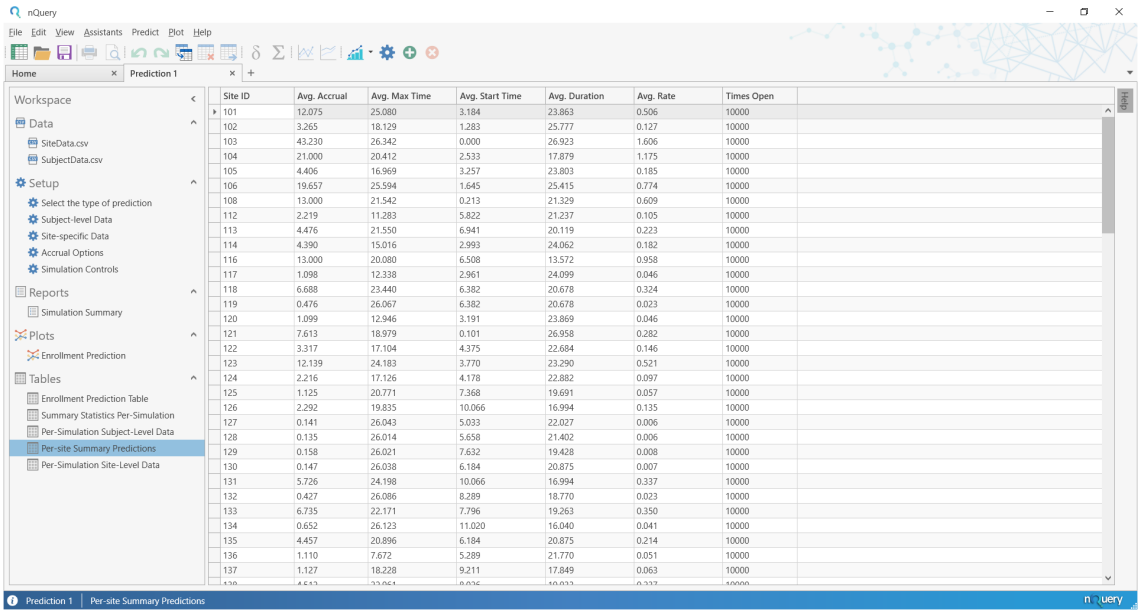



**Figure 6.44:** Per-Simulation Subject-level Data Table for Subject + Site-level Data

The **Per-Site Summary Predictions Table** was also generated here and is shown for Scenario A in Figure 6.45. The **Per-Site Summary Predictions Table** contains the following columns:

- Site ID: Site ID for a site
- Avg. Accrual: The average number of subjects enrolled from this site across all simulations
- Avg. Max Time: The average time of the final enrollment in a site
- Avg. Start Time: The average site initiation time for a site. This is from the Site-level Data for open sites and generated randomly for unopened sites
- Avg. Duration: The average length of time a site was open and enrolling subjects
- Avg. Rate: The average accrual rate for a site during the period it was open
- Times Open: The number of simulations in which the site opened before enrollment was complete

Note that **Times Opens** will equals the number of simulations except for unopened sites with any generated start times higher than a given simulation's **Accrual Duration**.

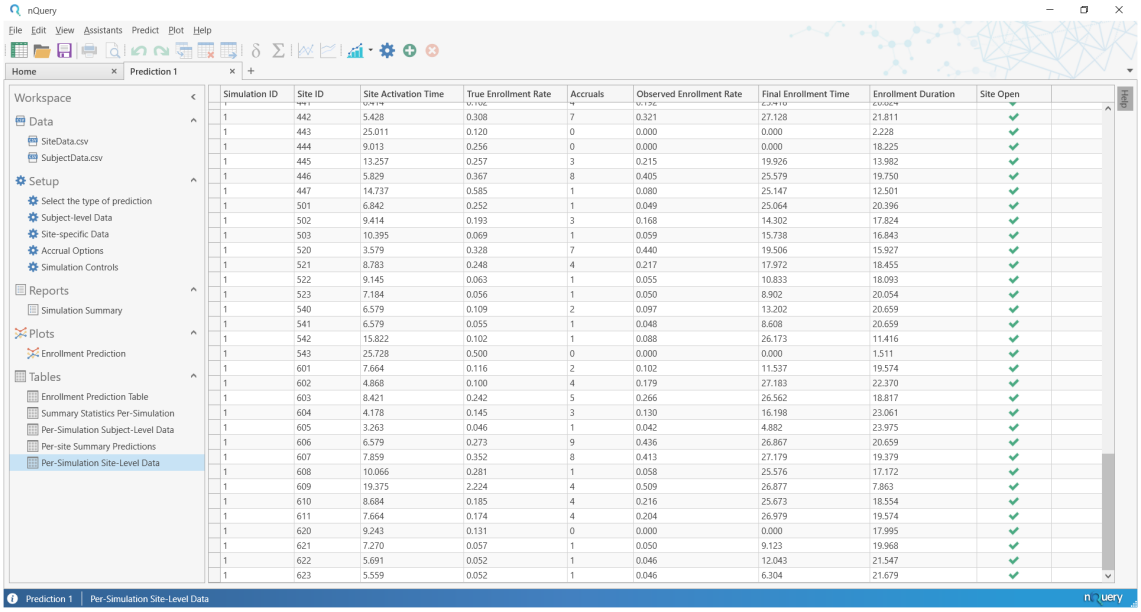



| Site ID | Avg. Accrual | Avg. Max Time | Avg. Start Time | Avg. Duration | Avg. Rate | Times Open |
|---|---|---|---|---|---|---|
| 101 | 12.075 | 25.080 | 3.184 | 23.863 | 0.506 | 10000 |
| 102 | 3.265 | 18.129 | 1.283 | 25.777 | 0.127 | 10000 |
| 103 | 43.230 | 26.342 | 0.000 | 26.923 | 1.606 | 10000 |
| 104 | 21.000 | 20.412 | 2.533 | 17.879 | 1.175 | 10000 |
| 105 | 4.406 | 16.969 | 3.257 | 23.803 | 0.185 | 10000 |
| 106 | 19.657 | 25.594 | 1.645 | 25.415 | 0.774 | 10000 |
| 108 | 13.000 | 21.542 | 0.213 | 21.329 | 0.609 | 10000 |
| 112 | 2.219 | 11.283 | 5.822 | 21.237 | 0.105 | 10000 |
| 113 | 4.476 | 21.550 | 6.941 | 20.119 | 0.223 | 10000 |
| 114 | 4.390 | 15.016 | 2.993 | 24.062 | 0.182 | 10000 |
| 116 | 13.000 | 20.080 | 6.508 | 13.572 | 0.958 | 10000 |
| 117 | 1.098 | 12.338 | 2.961 | 24.099 | 0.046 | 10000 |
| 118 | 6.688 | 23.440 | 6.382 | 20.678 | 0.324 | 10000 |
| 119 | 0.476 | 26.067 | 6.382 | 20.678 | 0.023 | 10000 |
| 120 | 1.099 | 12.946 | 3.191 | 23.869 | 0.046 | 10000 |
| 121 | 7.613 | 18.979 | 0.101 | 26.958 | 0.282 | 10000 |
| 122 | 3.317 | 17.104 | 4.375 | 22.684 | 0.146 | 10000 |
| 123 | 12.139 | 24.183 | 3.770 | 23.290 | 0.521 | 10000 |
| 124 | 2.216 | 17.126 | 4.178 | 22.882 | 0.097 | 10000 |
| 125 | 1.125 | 20.771 | 7.368 | 19.691 | 0.057 | 10000 |
| 126 | 2.292 | 19.835 | 10.066 | 16.994 | 0.135 | 10000 |
| 127 | 0.141 | 26.043 | 5.033 | 22.027 | 0.006 | 10000 |
| 128 | 0.135 | 26.014 | 5.658 | 21.402 | 0.006 | 10000 |
| 129 | 0.158 | 26.021 | 7.632 | 19.428 | 0.008 | 10000 |
| 130 | 0.147 | 26.038 | 6.184 | 20.875 | 0.007 | 10000 |
| 131 | 5.726 | 24.198 | 10.066 | 16.994 | 0.337 | 10000 |
| 132 | 0.427 | 26.086 | 8.289 | 18.770 | 0.023 | 10000 |
| 133 | 6.735 | 22.171 | 7.796 | 19.263 | 0.350 | 10000 |
| 134 | 0.652 | 26.123 | 11.020 | 16.040 | 0.041 | 10000 |
| 135 | 4.457 | 20.896 | 6.184 | 20.875 | 0.214 | 10000 |
| 136 | 1.110 | 7.672 | 5.289 | 21.770 | 0.051 | 10000 |
| 137 | 1.127 | 18.228 | 9.211 | 17.849 | 0.063 | 10000 |

**Figure 6.45:** Per-Site Summary Predictions for Scenario A

The **Per-Simulation Site-level Data Table** was also generated here and is shown for Scenario A in Figure 6.46. The **Per-Simulation Site-level Data Table** contains the following columns:

- Simulation ID: The simulation number
- Site ID: Site ID for a site
- Site Activation Time: Time after study start when site become open to enroll subjects. Taken from Site-level Data for open sites, generated for unopened sites
- True Enrollment Rate: The user specified enrollment rate for a site
- Accruals: Total number of subjects enrolled from a site in a simulation
- Observed Enrollment Rate: The actual enrollment rate (subjects enrolled per unit time) achieved in a site in a simulation
- Final Enrollment Time: Time of the final enrollment in a site in a simulation
- Enrollment Duration: The length of time a site was open in a simulation
- Site Open: Indicator of whether site opened in a simulation. Tick = Open, X = Did not open




| Simulation ID | Site ID | Site Activation Time | True Enrollment Rate | Accruals | Observed Enrollment Rate | Final Enrollment Time | Enrollment Duration | Site Open |
|---|---|---|---|---|---|---|---|---|
| 1 | 442 | 5.428 | 0.308 | 7 | 0.321 | 27.128 | 21.811 | ✓ |
| 1 | 443 | 25.011 | 0.120 | 0 | 0.000 | 0.000 | 2.228 | ✓ |
| 1 | 444 | 9.013 | 0.256 | 0 | 0.000 | 0.000 | 18.225 | ✓ |
| 1 | 445 | 13.257 | 0.257 | 3 | 0.215 | 19.926 | 13.982 | ✓ |
| 1 | 446 | 5.829 | 0.367 | 8 | 0.405 | 25.579 | 19.750 | ✓ |
| 1 | 447 | 14.737 | 0.585 | 1 | 0.080 | 25.147 | 12.501 | ✓ |
| 1 | 501 | 6.842 | 0.252 | 1 | 0.049 | 25.064 | 20.396 | ✓ |
| 1 | 502 | 9.414 | 0.193 | 3 | 0.168 | 14.302 | 17.824 | ✓ |
| 1 | 503 | 10.395 | 0.069 | 1 | 0.059 | 15.738 | 16.843 | ✓ |
| 1 | 520 | 3.579 | 0.328 | 7 | 0.440 | 19.506 | 15.927 | ✓ |
| 1 | 521 | 8.783 | 0.248 | 4 | 0.217 | 17.972 | 18.455 | ✓ |
| 1 | 522 | 9.145 | 0.063 | 1 | 0.055 | 10.833 | 18.093 | ✓ |
| 1 | 523 | 7.184 | 0.056 | 1 | 0.050 | 8.902 | 20.054 | ✓ |
| 1 | 540 | 6.579 | 0.109 | 2 | 0.097 | 13.202 | 20.659 | ✓ |
| 1 | 541 | 6.579 | 0.055 | 1 | 0.048 | 8.608 | 20.659 | ✓ |
| 1 | 542 | 15.822 | 0.102 | 1 | 0.088 | 26.173 | 11.416 | ✓ |
| 1 | 543 | 25.728 | 0.500 | 0 | 0.000 | 0.000 | 1.511 | ✓ |
| 1 | 601 | 7.664 | 0.116 | 2 | 0.102 | 11.537 | 19.574 | ✓ |
| 1 | 602 | 4.868 | 0.100 | 4 | 0.179 | 27.183 | 22.370 | ✓ |
| 1 | 603 | 8.421 | 0.242 | 5 | 0.266 | 26.562 | 18.817 | ✓ |
| 1 | 604 | 4.178 | 0.145 | 3 | 0.130 | 16.198 | 23.061 | ✓ |
| 1 | 605 | 3.263 | 0.046 | 1 | 0.042 | 4.882 | 23.975 | ✓ |
| 1 | 606 | 6.579 | 0.273 | 9 | 0.436 | 26.867 | 20.659 | ✓ |
| 1 | 607 | 7.859 | 0.352 | 8 | 0.413 | 27.179 | 19.379 | ✓ |
| 1 | 608 | 10.066 | 0.281 | 1 | 0.058 | 25.576 | 17.172 | ✓ |
| 1 | 609 | 19.375 | 2.224 | 4 | 0.509 | 26.877 | 7.863 | ✓ |
| 1 | 610 | 8.684 | 0.185 | 4 | 0.216 | 25.673 | 18.554 | ✓ |
| 1 | 611 | 7.664 | 0.174 | 4 | 0.204 | 26.979 | 19.574 | ✓ |
| 1 | 620 | 9.243 | 0.131 | 0 | 0.000 | 0.000 | 17.995 | ✓ |
| 1 | 621 | 7.270 | 0.057 | 1 | 0.050 | 9.123 | 19.968 | ✓ |
| 1 | 622 | 5.691 | 0.052 | 1 | 0.046 | 12.043 | 21.547 | ✓ |
| 1 | 623 | 5.559 | 0.052 | 1 | 0.046 | 6.304 | 21.679 | ✓ |

**Figure 6.46:** Per-Simulation Site-level Data Table for Scenario A

**Scenario B** Scenario A assumed that an additional 9 sites would open after the **Current Time**, as per the original study plan. Given that the enrollment is ahead of schedule in Scenario A, it was decided to reduce the number of planned unopened sites to 3 with Site ID 144, 145, 347 being retained.

**Site ID** 144 (expected accrual rate of 0.5 per month) would retain their current opening window of 25.51 - 26.11 months, the **Site ID 145** (expected accrual rate of 0.5 per month) window was shifted up by one month to 26.51-27.11 months and the smaller **Site ID** 347 (expected accrual rate of 0.1 per month) shifted up to a reserve position with an opening window of 28-30 months and was expected to be unneeded.

To evaluate this scenario, select **Accrual Options** from under the **Setup** header in the Workspace Navigation Bar. In **Accrual Options**, open the **Accrual Infos/Sites** tab. Change the **Total Number of Sites** to 121. In the **Site Accrual Table**, scroll down to the unopened sites.

The **Site ID** = 144 row is unchanged. In the **Site ID** = 145 row, change the **Start** column value to 26.51 and the **End** column value to 27.11. In the **Site ID** = 347 row, change the **Start** column to 28 and the **End** column to 30. The updated **Accrual Infos/Sites** tab for Scenario B is shown in Figure 6.47.

Sample Size & Followup Options | Accrual Infos / Site

Total Number of Sites: 121 Number of Sites Opened: 118

| Site ID | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap | Planned Accrual Rate/ Site | Site Initiation Time | No. of Accruals |
|---|---|---|---|---|---|---|---|
| | Start | End | | | | | |
| 542 | | | 0.102 | 24 | 0.102 | 15.8224 | 0 |
| 601 | | | 0.1159 | 7 | 0.1159 | 7.6645 | 2 |
| 602 | | | 0.0997 | 27 | 0.0997 | 4.8684 | 2 |
| 603 | | | 0.2425 | 45 | 0.2425 | 8.4211 | 4 |
| 604 | | | 0.1446 | 9 | 0.1446 | 4.1776 | 3 |
| 605 | | | 0.0462 | 7 | 0.0462 | 3.2632 | 1 |
| 606 | | | 0.2726 | 30 | 0.2726 | 6.5789 | 5 |
| 607 | | | 0.3517 | 10 | 0.3517 | 7.8592 | 6 |
| 608 | | | 0.2815 | 9 | 0.2815 | 10.0658 | 0 |
| 609 | | | 2.2244 | 12 | 2.2244 | 19.375 | 0 |
| 610 | | | 0.1848 | 16 | 0.1848 | 8.6842 | 3 |
| 611 | | | 0.1739 | 9 | 0.1739 | 7.6645 | 3 |
| 620 | | | 0.1307 | 9 | 0.1307 | 9.2434 | 0 |
| 621 | | | 0.0567 | 18 | 0.0567 | 7.2697 | 1 |
| 622 | | | 0.052 | 18 | 0.052 | 5.6908 | 1 |
| 623 | | | 0.0517 | 10 | 0.0517 | 5.5592 | 1 |
| 144 | 25.51 | 26.11 | 0.50 | 14 | 0.50 | | |
| 145 | 26.51 | 27.11 | 0.50 | 14 | 0.50 | | |
| 347 | 28.00 | 30.00 | 0.10 | 6 | 0.10 | | |

**Figure 6.47:** Scenario B Accrual Infos/Sites Tab

Next or → button. At the next **Simulation Controls** step, select the Run or → button to simulate Scenario B. The results for Scenario B are summarized in Figure 6.48.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Study Duration | 27.1852 |
| % Simulations Target Reached | 100.0000 |
| Sites Opened | 119.8660 |

Percentile Summary

| Percentile | Sample Size | Study Duration | Sites Opened |
|---|---|---|---|
| 5.00% | 460 | 26.7222 | 119 |
| 25.00% | 460 | 26.9806 | 120 |
| 50.00% | 460 | 27.1730 | 120 |
| 75.00% | 460 | 27.3722 | 120 |
| 95.00% | 460 | 27.6869 | 120 |

**Figure 6.48:** Scenario B for Enrollment Prediction with Subject + Site-level Data

For Scenario B, the **Overall Summary** shows that the **Target Sample Size** of 460 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 27.1852 with the **Percentile Summary** showing that 90% of predictions had an **Accrual Duration** between 26.7222 and 27.6869, 50% a duration between 26.9806 and 27.3722 with a median duration of 27.1730.

Note that number of sites ranged from 119 to 120 with an average of 119.8660 sites opened across all simulations. In the **Per-Site Summary Predictions Table** for this demonstration Site 144 opened in every simulation (0.681 subjects enrolled on average), Site 145 opened in 8547 simulations (0.222 subjects enrolled on average) and Site 346 opened 3 times (<0.001 subjects enrolled on average).

Given that Scenario A had an **Average Study Duration** of 27.0594 this indicates the new sites had a minimal impact on reaching the **Target Sample Size** and perhaps no further sites were needed or that sites could be closed. In nQuery Predict "Closing" opened sites at the **Current Time** could be achieved by setting the **Enrollment Cap** equal to **No. of Accruals** in the **Site Accrual Table.**

#### 6.2.6.3 Enrollment Prediction using Summary Data

**Setup** **Enrollment Only** milestone prediction using **Summary Data** means that the prediction will project the expected length of the accrual period needed to recruit a specified **Target Sample Size** using summary data.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step select **Enrollment Only** from the **Target** field and **Summary** from the **Input** field. **Enrollment Status** must be **Ongoing** for enrollment milestone prediction. These selections are shown in Figure 6.49. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

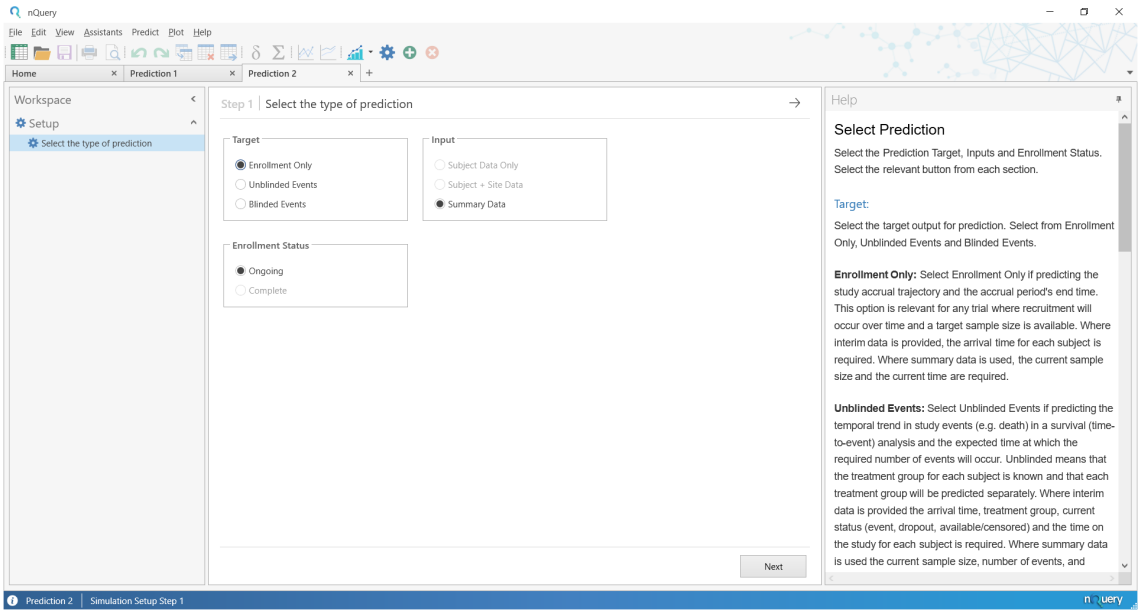



**Figure 6.49:** Select the Type of Prediction Setup Step for Enrollment Prediction using Summary Data

The next step is the **Fixed Parameters** step. For this demonstration, there is a **Current Sample Size and Calendar Time** section which contains two fields which are as follows:

- Current Sample Size: The sample size enrolled into the study at the Current Time
- Current Time: The length of time the study has been on-going since the study began

For this demonstration, inputs are selected which mirror those from **Subject-Data.csv** used in the examples above. This would give a **Current Sample Size** of 402 and a **Current Time** of 24.9186. The **Fixed Parameters** step with these values inputted is shown in Figure 6.50. After all fields all filled, select the Next or → button to move to the next **Setup** step.

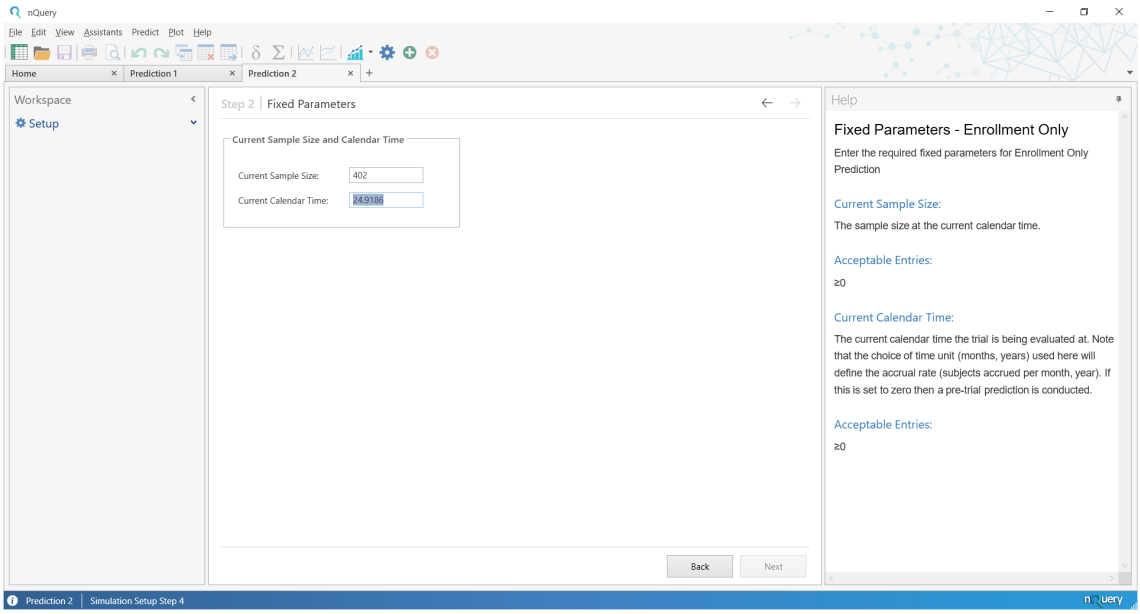



**Figure 6.50:** Fixed Parameters Setup Step for Enrollment Prediction using Summary Data

The next step is the **Accrual Options** step which specifies the **Target Sample Size, Accrual Model** and future **Accrual Rate**. The **Accrual Options** step, and its defaults, for this demonstration are shown in Figure 6.51.

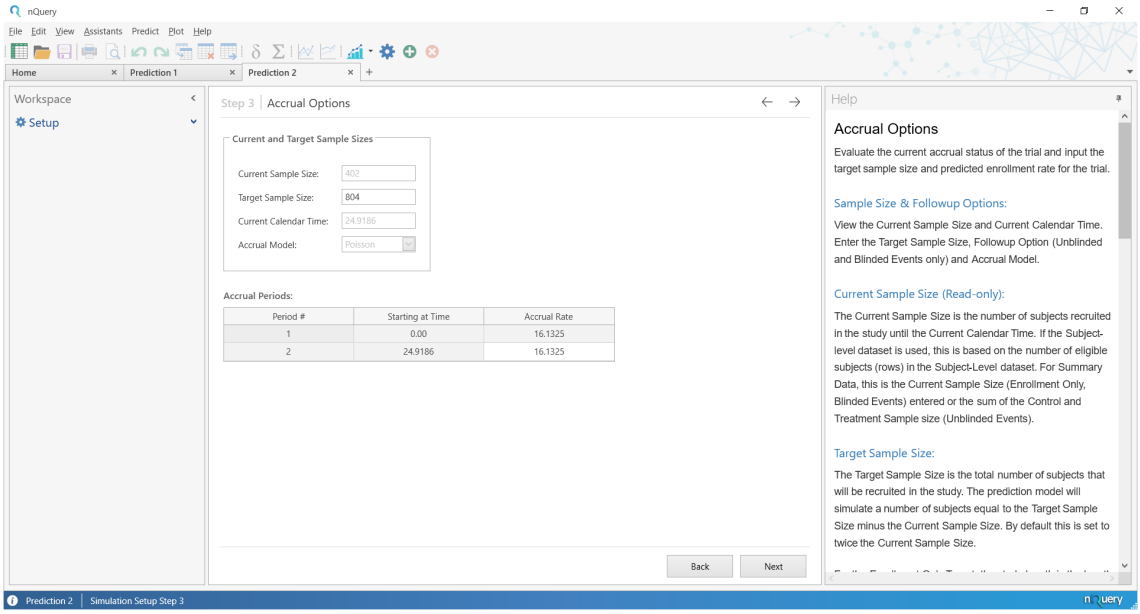



**Figure 6.51:** Accrual Options Setup Step for Enrollment Prediction using Summary Data

The **Accrual Options** step consists of two main elements: the **Current and Target Sample Sizes** input fields and the **Accrual Periods** table.

The **Current and Target Sample Sizes** input fields provides information on study enrollment based on the **Fixed Parameters** step inputs and allows the user to edit the **Target Sample Size** and select the **Accrual Model**. The following fields are provided in this prediction:

- Current Sample Size (Read-only): Sample Size recruited into a study so far. Taken from Fixed Parameters step
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study. The number of simulated subjects equals the Total Sample Size minus the Current Sample Size.
- Current Time (Read-only): The length of time that has passed since the study started. Taken from Fixed Parameters step
- Accrual Model (Editable): The statistical model that will be used to generate simulated enrollments

In nQuery Predict, the **Target Sample Size** defaults to twice the **Current Sample Size** and the **Accrual Model** defaults to **Poisson.** In this demonstration, the **Current Sample Size** equals 402 and therefore the default **Target Sample Size** is $402\,(2) = 804$.

The **Accrual Periods** table contains information on the **Accrual Rate** up the **Current Time** and allows the user to edit the future **Accrual Rate** used to generate enrollment simulations. The **Accrual Periods** has three columns ,with each row corresponding to a time period, with the columns defined as follows:

- Period #: A numeric ID for the current time period. Increases in increments of one for each subsequent row
- Starting at Time: The starting time for the current time period row. The first row will equal 0 (rate from study start to Current Time) and the second row will equal the Current Time
- Accrual Rate: The accrual rate is the average number of subjects recruited per unit time (months). The accrual rate in the first row will equal the accrual rate up to the Current Time

For the **Poisson Accrual Model**, the **Accrual Periods** table will consist of two time period rows. The first row will correspond to the information provided by the user regarding the study up until the **Current Time.** The second row will correspond to the inputs used to generate future enrollments. By default the **Accrual Rate** in the future second row will equal the **Accrual Rate** from the first row i.e. the accrual rate up until the **Current Time.** For the inputs used in this demonstration, this **Accrual Rate** column values in each row equal 16.13254936 (**Current Sample Size/Current Time**). The user can edit the **Accrual Rate**

for generating future enrollments by editing the second cell of the **Accrual Rate** column.

For now, the default values will be used for **Target Sample Size** (804), **Accrual Model** (Poisson) and **Accrual Rate** (16.13254936) as per Figure 6.51. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.52. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

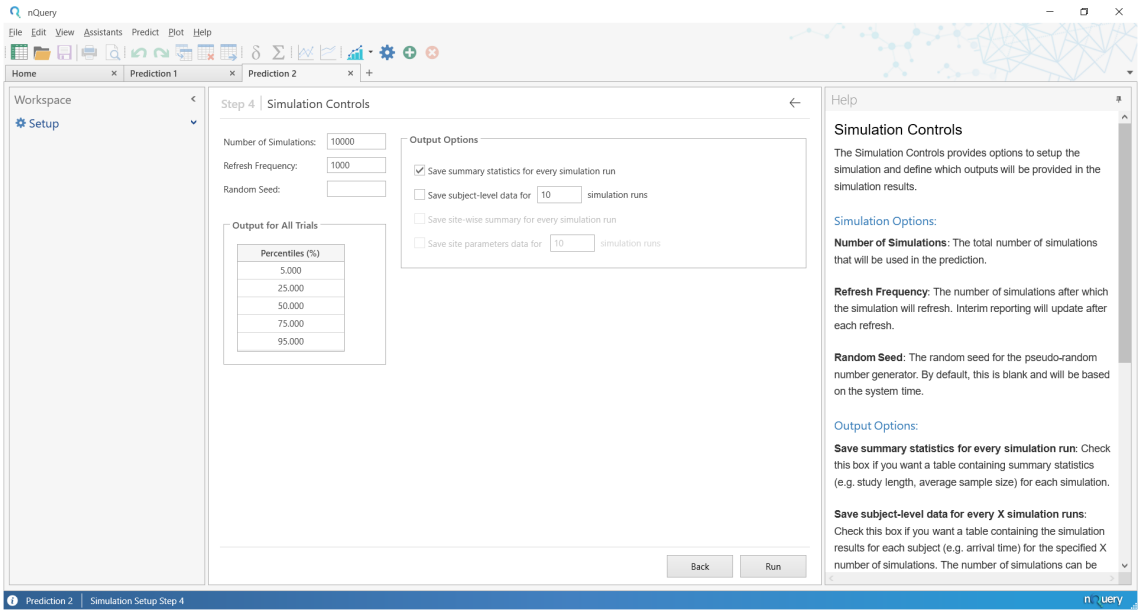



**Figure 6.52:** Simulation Controls Setup Step for Enrollment Prediction using Summary Data

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.53.

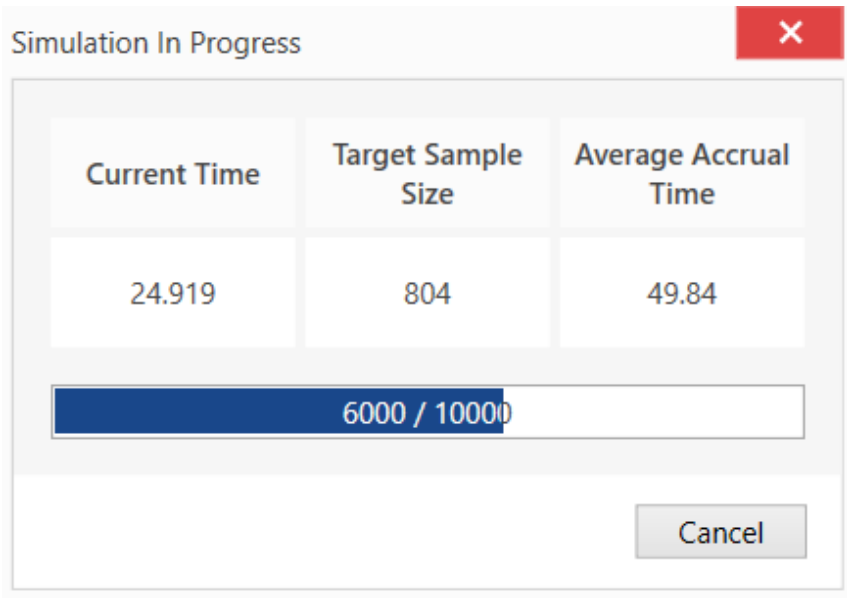



**Figure 6.53:** Simulation in Progress Window for Enrollment Prediction using Summary Data

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.54.

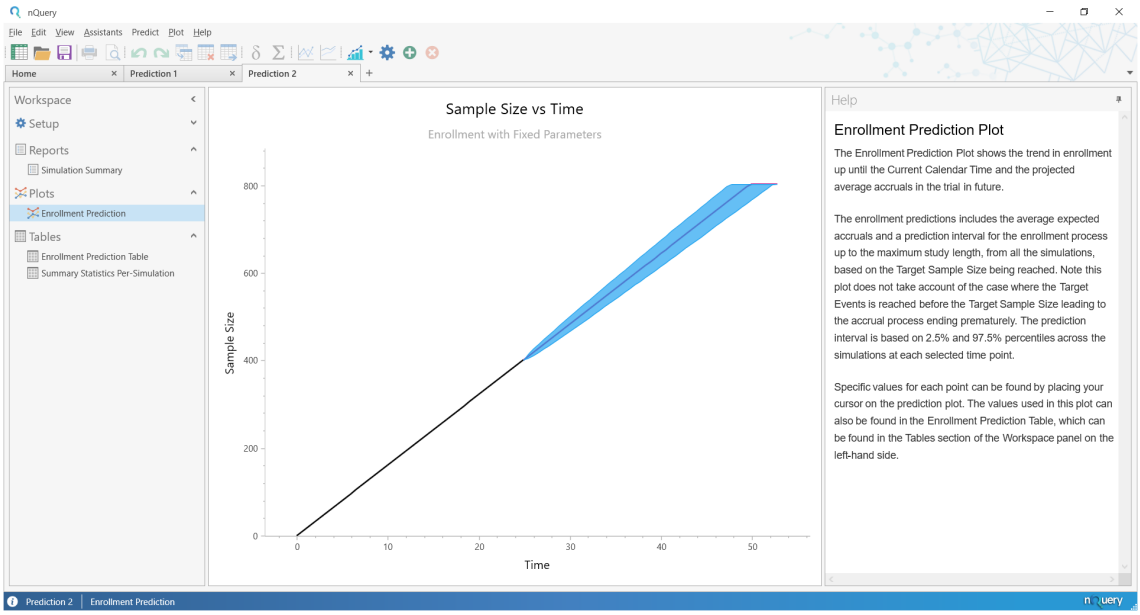



**Figure 6.54:** Enrollment Prediction Plot for Enrollment Prediction using Summary Data

The results of a nQuery Predict milestone prediction can be divided into three categories: **Reports, Plots** and **Tables**. A high-level summary of the results in nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.55.

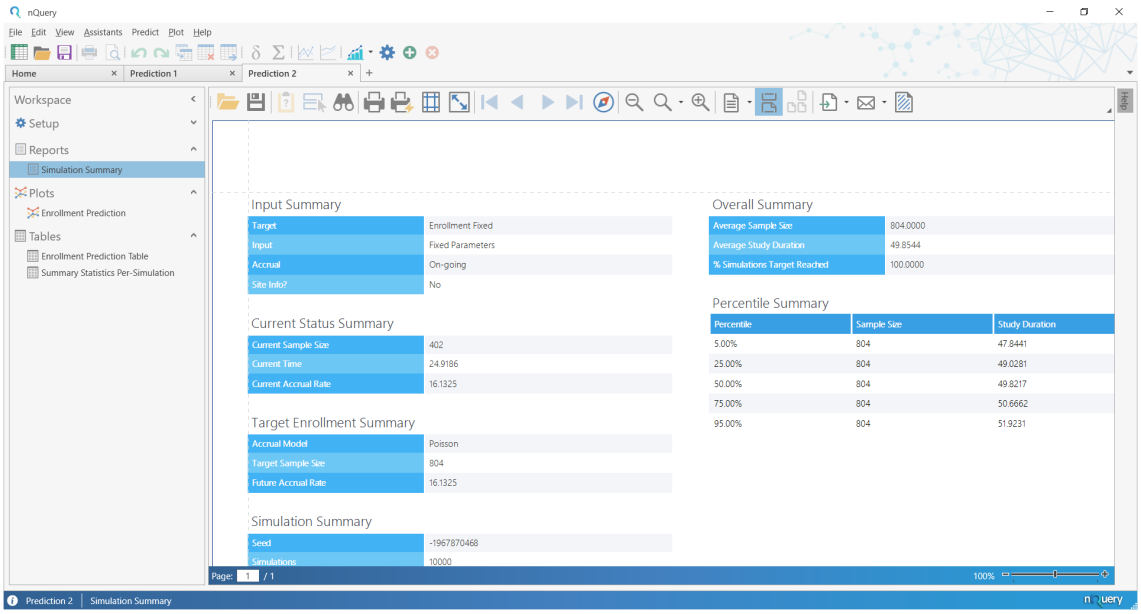



**Figure 6.55:** Simulation Summary Report for Enrollment Prediction using Summary Data

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.56. For this demonstration, the **Overall Summary** shows that the **Target Sample Size** of 804 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 49.8544. The **Percentile Summary** shows that 90% of predictions had a **Accrual Duration** between 47.8441 (5% Percentile) and 51.9231 (95% Percentile), 50% a duration between 49.0281 (25% Percentile) and 50.6662 (75% Percentile) and a median (50% Percentile) duration of 49.8217.

Overall Summary

| | |
|---|---|
| Average Sample Size | 804.0000 |
| Average Study Duration | 49.8544 |
| % Simulations Target Reached | 100.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration |
|---|---|---|
| 5.00% | 804 | 47.8441 |
| 25.00% | 804 | 49.0281 |
| 50.00% | 804 | 49.8217 |
| 75.00% | 804 | 50.6662 |
| 95.00% | 804 | 51.9231 |

**Figure 6.56:** Simulation Summary Report Results for Enrollment Prediction using Summary Data

To edit the **Simulation Summary report there a** number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment simulation from the study start to when the **Target Sample Size** was reached in all simulations. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.57.

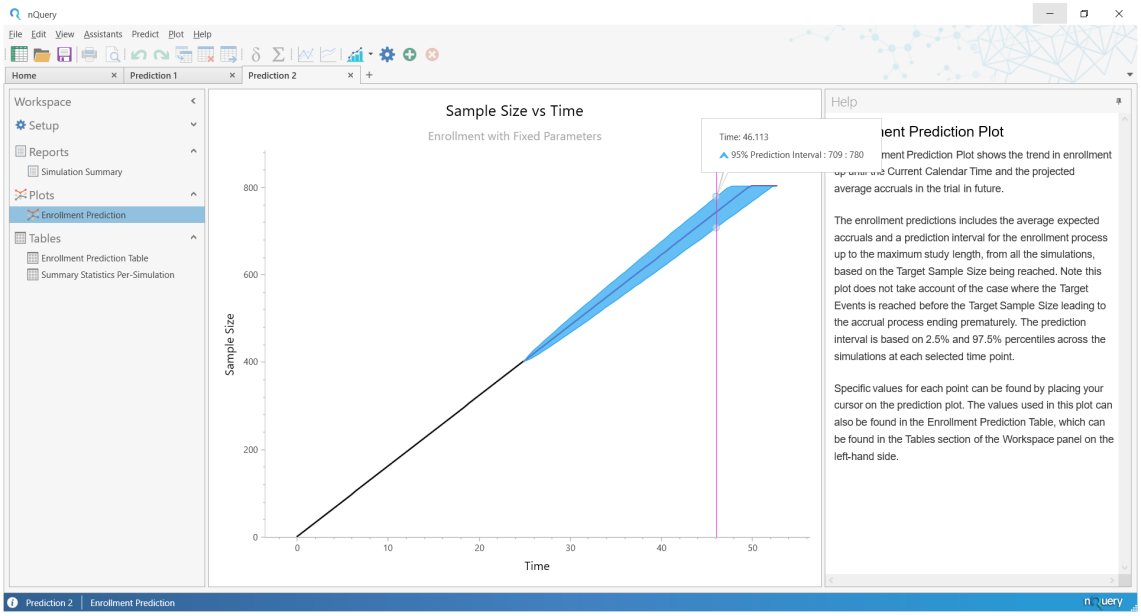
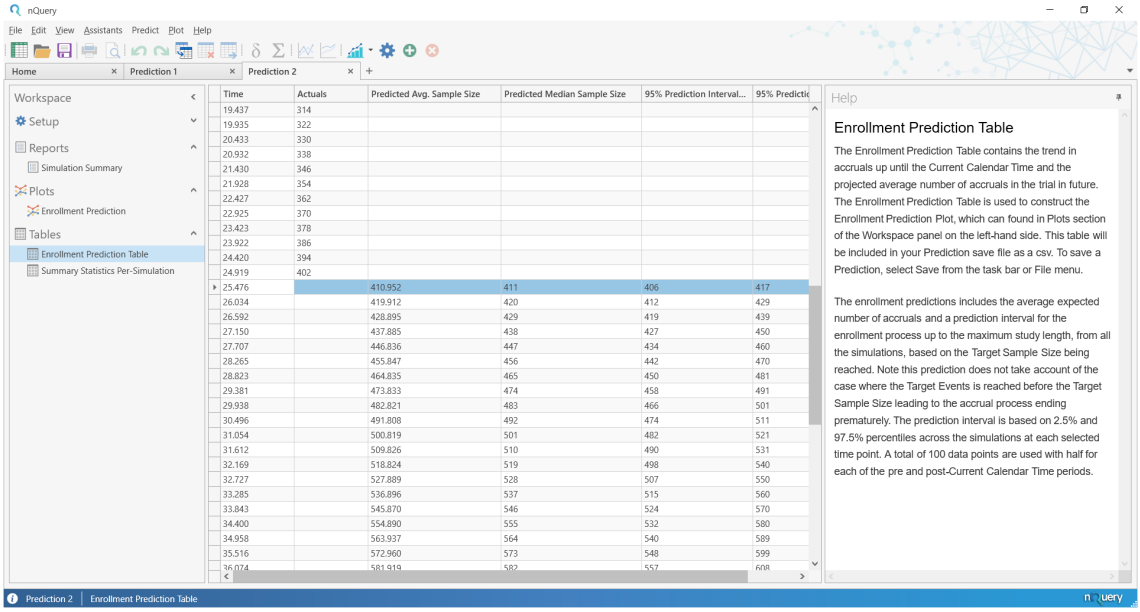



**Figure 6.57:** Enrollment Prediction Plot for Enrollment Prediction using Summary Data

The **Enrollment Prediction** plot consists of two parts: from Time = 0 to **Current Time** and from **Current Time** to the **Maximum Time**. The first part is a black line which gives the hypothetical accrual pattern based on the **Fixed Parameters** step, in this case a constant linear increase in **Sample Size.**

The second part is a blue conic section with a grey line where the blue conic section plots the 95% prediction interval for the **Sample Size** enrolled at a given **Time** across all simulations and the grey line is the average **Sample Size** enrolled at a given time. The 95% prediction interval indicates the 95% of predictions had a sample size that fell between the lower and upper limit at a given time.

The **Enrollment Prediction Table** is the table used to create the **Enrollment Prediction** plot, which by default consists of ~100 rows (50 pre-simulation and 50 post-simulation). It consists of five columns: **Time, Actual, Predicted Avg. Sample Size, Predicted Median Sample Size, 95% Prediction Interval LL, 95% Prediction Interval UL**. The **Enrollment Prediction Table** for this demonstration at the start of the simulated period is given in Figure 6.58.




| Time | Actuals | Predicted Avg. Sample Size | Predicted Median Sample Size | 95% Prediction Interval... | 95% Predicti |
|---|---|---|---|---|---|
| 19.437 | 314 | | | | |
| 19.935 | 322 | | | | |
| 20.433 | 330 | | | | |
| 20.932 | 338 | | | | |
| 21.430 | 346 | | | | |
| 21.928 | 354 | | | | |
| 22.427 | 362 | | | | |
| 22.925 | 370 | | | | |
| 23.423 | 378 | | | | |
| 23.922 | 386 | | | | |
| 24.420 | 394 | | | | |
| 24.919 | 402 | | | | |
| 25.476 | | 410.952 | 411 | 406 | 417 |
| 26.034 | | 419.912 | 420 | 412 | 429 |
| 26.592 | | 428.895 | 429 | 419 | 439 |
| 27.150 | | 437.885 | 438 | 427 | 450 |
| 27.707 | | 446.836 | 447 | 434 | 460 |
| 28.265 | | 455.847 | 456 | 442 | 470 |
| 28.823 | | 464.835 | 465 | 450 | 481 |
| 29.381 | | 473.833 | 474 | 458 | 491 |
| 29.938 | | 482.821 | 483 | 466 | 501 |
| 30.496 | | 491.808 | 492 | 474 | 511 |
| 31.054 | | 500.819 | 501 | 482 | 521 |
| 31.612 | | 509.826 | 510 | 490 | 531 |
| 32.169 | | 518.824 | 519 | 498 | 540 |
| 32.727 | | 527.889 | 528 | 507 | 550 |
| 33.285 | | 536.896 | 537 | 515 | 560 |
| 33.843 | | 545.870 | 546 | 524 | 570 |
| 34.400 | | 554.890 | 555 | 532 | 580 |
| 34.958 | | 563.937 | 564 | 540 | 589 |
| 35.516 | | 572.960 | 573 | 548 | 599 |
| 36.074 | | 581.919 | 582 | 557 | 608 |

**Figure 6.58:** Enrollment Prediction Table for Enrollment Prediction using Summary Data

**Time** is the time value relative to the study start time of zero. **Actual** is the actual sample size achieved at a pre-simulation **Time** row. **Predicted Avg. Sample Size** is the predicted average (mean) sample size enrolled at a post-simulation **Time** row. **Predicted Median Sample Size** is the predicted median sample size enrolled at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the sample size enrolled at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the sample size enrolled at a post-simulation **Time** row.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Target Sample Size, Accrual Duration.** The **Summary Statistics Per-Simulation Table** for this demonstration is shown in Figure 6.59.

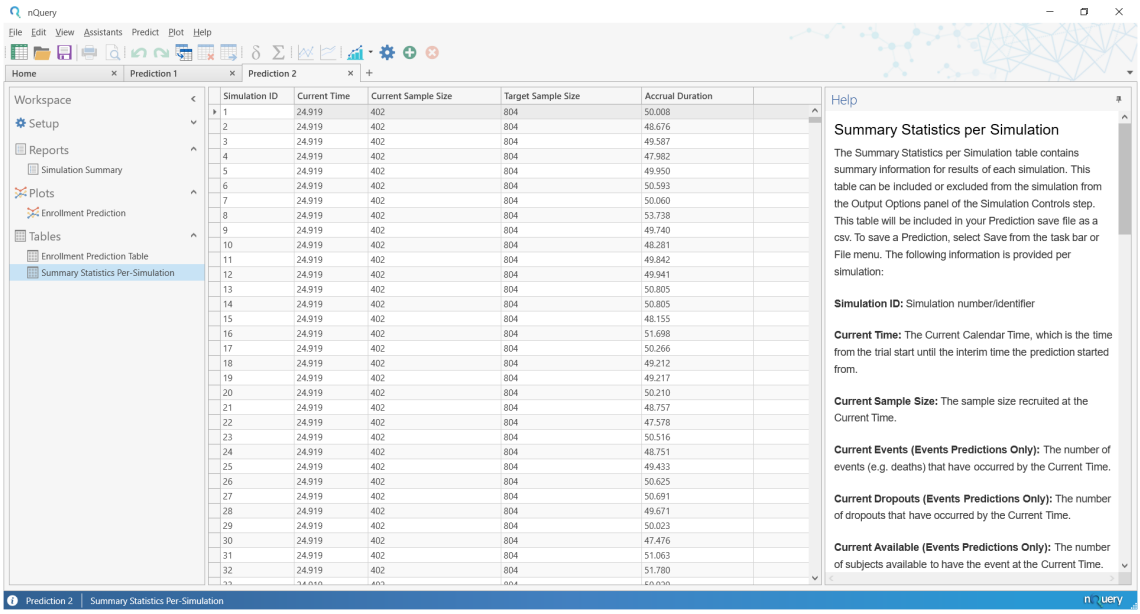



**Figure 6.59:** Summary Statistics Per-Simulation Table for Enrollment Prediction using Summary Data

**Simulation ID** is in order of simulation from 1 up to **Number of Simulations**, **Current Time, Current Sample Size** and **Target Sample Size** summarize the user inputs and are constant, and the **Accrual Duration** is the length of time needed to achieve the **Target Sample Size** in that simulation.

**Alternative Scenarios** The demonstration has focused on the default inputs using **Fixed Parameters** based on **SubjectData.csv** up to this point. However, as noted in the data summary given in subsection 6.2.4, the study from which the example data was simulated had an original sample size target of 460 (402 enrolled in **SubjectData.csv**). In addition, the pre-trial enrollment plan (460 enrolled over 30 months) can be simulated in nQuery Predict.

To explore this additional information, we will investigate two additional scenarios

- Scenario A: **Target Sample Size = 460** using default **Accrual Rate**
- Scenario B: Pre-trial plan of 460 enrolled in 30 months

**Scenario A** The only changes required compared to the first demonstration is to change the **Target Sample Size** to 460 at the **Accrual Options** step. The simplest way to do this is to select the **Setup** header in the Workspace Navigation Bar on the left and then select the **Accrual Options** drop down option. Then edit the **Target Sample Size** 460 in the **Current and Target Sample Sizes** section in the top-left and select the Next or $\rightarrow$ button. At the next **Simulation Controls**

step, we will add a single simulation run by selecting the checkbox beside the **Save subject-level data for (x) simulation runs** and replace the default of 10 runs with 1 runs at the "(x)" spot. An edited down summary of the **Setup** changes is provided in Figure 6.60.

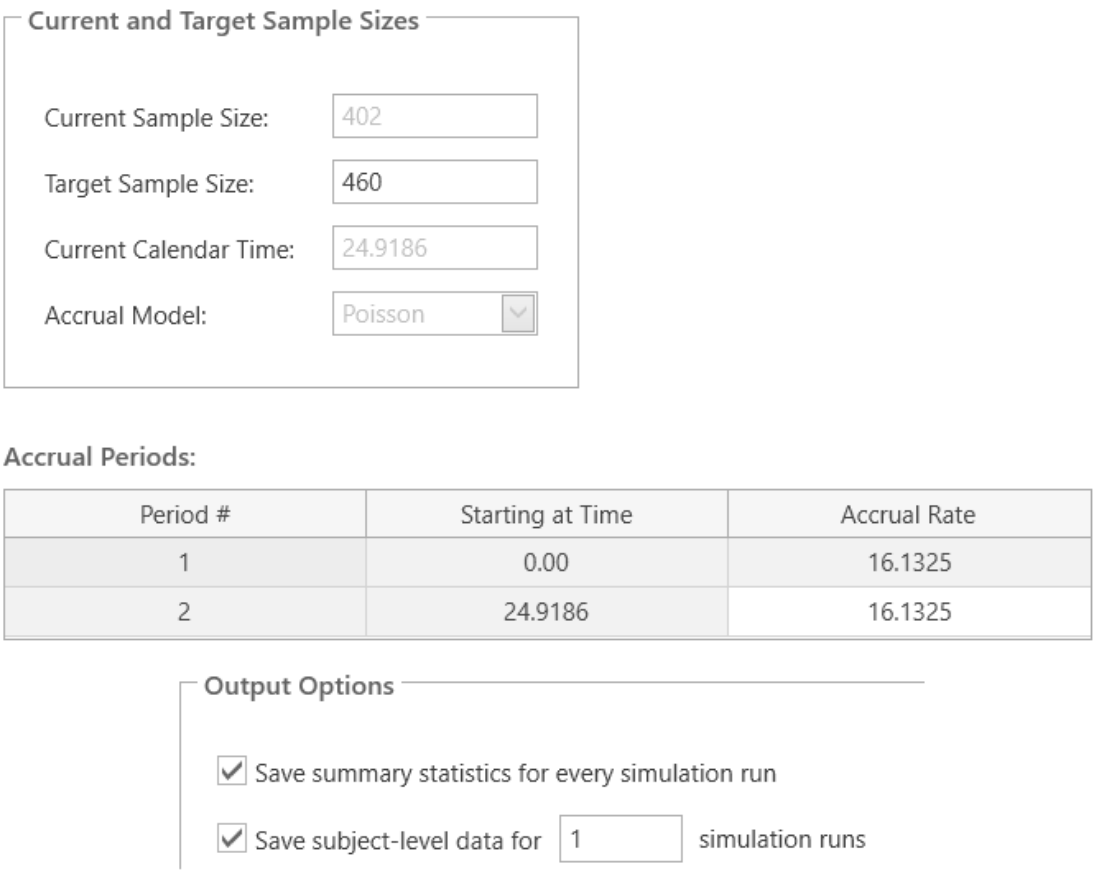



**Figure 6.60:** Scenario A Setup Changes for Enrollment Prediction using Summary Data

To run the simulation again with the new Scenario A, select the Run/$^{\rightarrow}$ button. The main results of this Scenario A simulation are shown in Figure 6.61.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Study Duration | 28.5144 |
| % Simulations Target Reached | 100.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration |
|---|---|---|
| 5.00% | 460 | 27.7499 |
| 25.00% | 460 | 28.1831 |
| 50.00% | 460 | 28.4979 |
| 75.00% | 460 | 28.8238 |
| 95.00% | 460 | 29.3381 |

**Figure 6.61:** Scenario A Results for Enrollment Prediction using Summary Data

For Scenario A, the **Overall Summary** shows that the **Target Sample Size** of 460 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 28.5144. The **Percentile Summary** shows that 90% of predictions had an **Accrual Duration** between 27.7499 and 29.3381, 50% a duration between 28.1831 and 28.8238 and a median duration of 28.4979.

Given that the original target accrual period length was 30 months, this indicates that if the average accrual rate over the entire period up to the current time was being maintained that the accrual period would around 1.5 months less than expected.

The **Per-Simulation Subject-level Data Table** was also generated here and is shown for Scenario A in Figure 6.62 with the first simulated subject (and their **Arrival Time**) in the first simulation (**Simulation ID = 1**) highlighted. Note that the order of the Subject ID is based on **Arrival Time** order. Note that the **Subject ID** field starts at 403 as the hypothetical subjects enrolled before the **Current Time** are not retrospectively simulated.

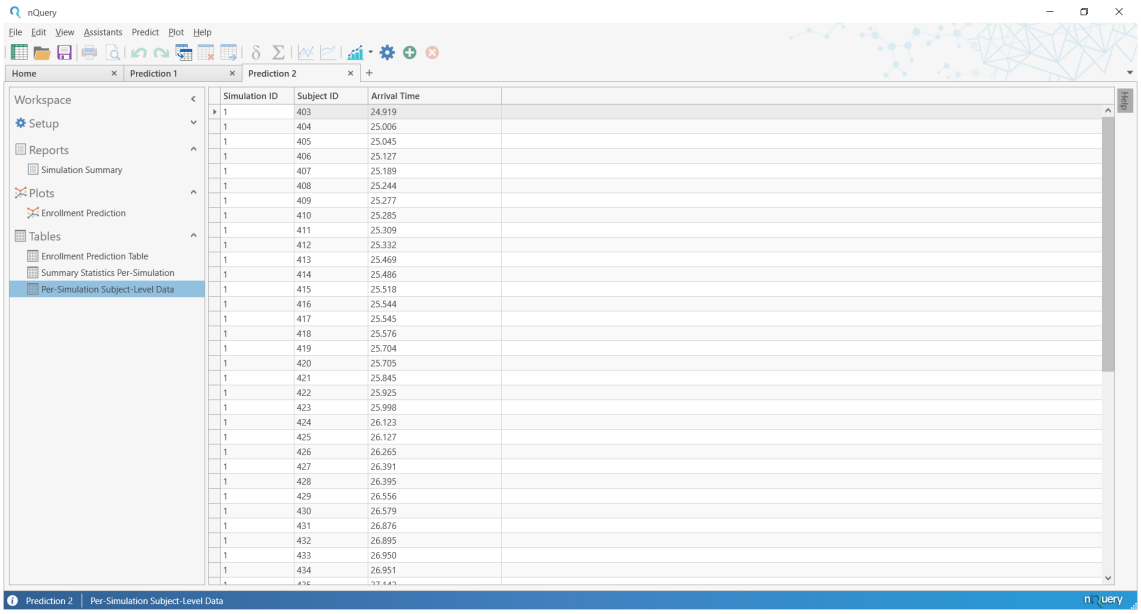



**Figure 6.62:** Per-Simulation Subject-level Data Table for Summary Data

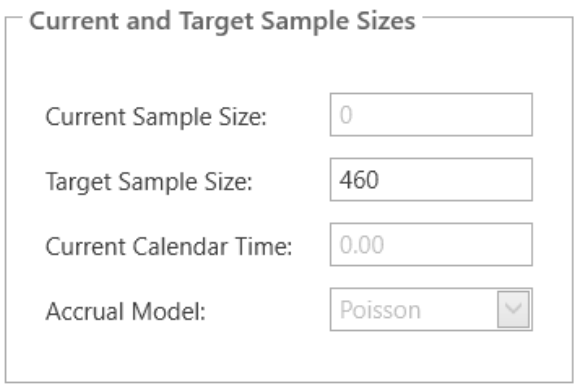
**Scenario B** Before the trial began, the planned accrual period length was 30 months with a **Target Sample Size**

To replicate this pre-study plan, create a new empty nQuery Predict Workspace by selecting **Predict > New Prediction >Prediction with Fixed Parameters** from the file menu.

In the **Select the Type of Prediction** step, select **Enrollment Only** from **Target** and **Summary Data** from **Input** as before. Select the Next or → button to continue.

In the **Fixed Parameters** step, set **Current Sample Size** and **Current Time** to zero. Select the Next or → button to continue.

In the **Accrual Options** step, set the **Target Sample Size** to 460 in the **Current and Target Sample Sizes** field. In the **Accrual Periods** table, set the **Period #2 Accrual Rate** equal to 15.3333 (Target Sample Size/Planned Accrual Period Length = 460/30). These **Accrual Options** inputs are shown in Figure 6.63. Select the Next or → button to continue.

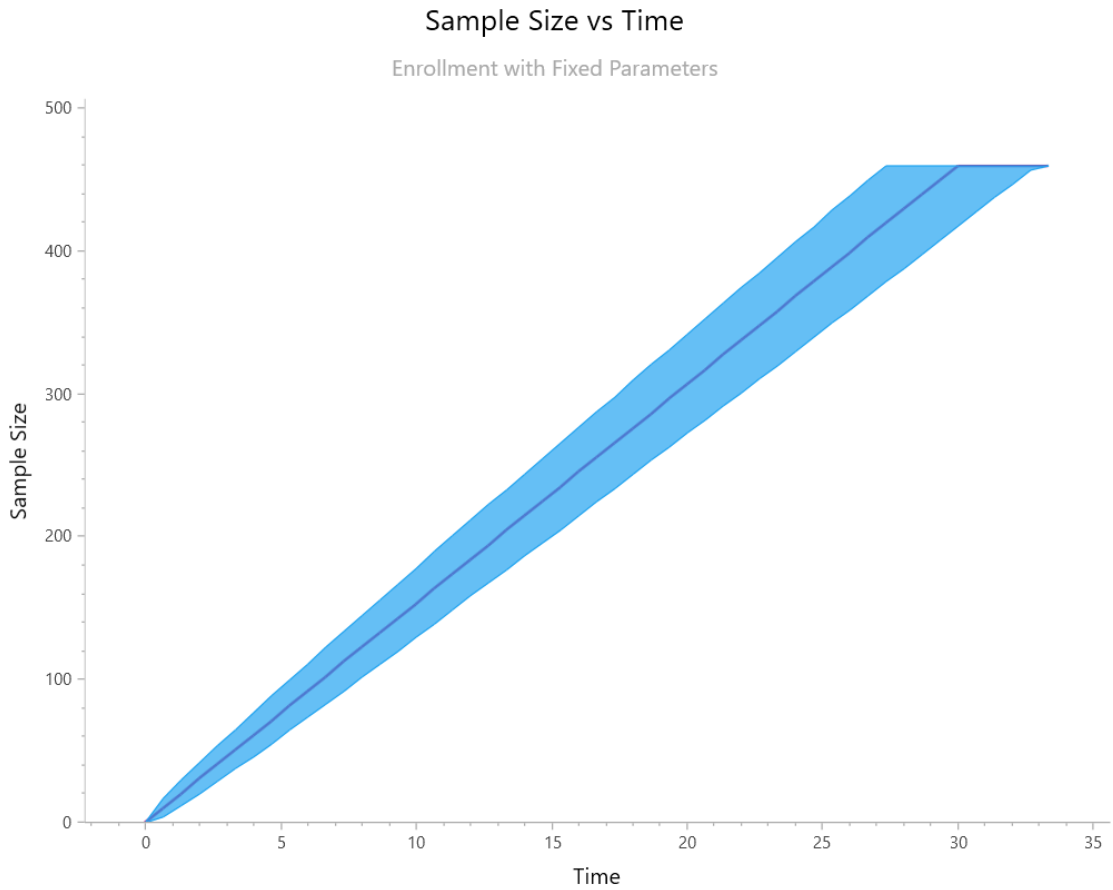



Accrual Periods:

| Period # | Starting at Time | Accrual Rate |
|---|---|---|
| 1 | 0.00 | 0.00 |
| 2 | 0.00 | 15.3333 |

**Figure 6.63:** Accrual Options Step for Pre-Trial Enrollment Only Scenario B

At the **Simulation Controls** step, select the Run or → button to simulate Scenario B. The results for Scenario B are summarized in Figure 6.64.




Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Study Duration | 30.0025 |
| % Simulations Target Reached | 100.0000 |

Percentile Summary

| Percentile | Sample Size | Study Duration |
|---|---|---|
| 5.00% | 460 | 27.7095 |
| 25.00% | 460 | 29.0589 |
| 50.00% | 460 | 29.9660 |
| 75.00% | 460 | 30.9309 |
| 95.00% | 460 | 32.4217 |

**Figure 6.64:** Pre-Trial Enrollment Only Prediction Results

For Scenario B, the **Overall Summary** shows that the **Target Sample Size** of 460 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 30.0025 with the **Percentile Summary** showing that 90%

of predictions had an **Accrual Duration** between 27.7095 and 32.4217, 50% a duration between 29.0589 and 30.9309 with a median duration of 29.9660. This gives a baseline understanding of the likely outcome for the enrollment process pre-trial under a simple **Poisson** process.

### 6.2.7 Events Prediction Demonstration

Event milestone prediction is the projection of how long a survival (time-to-event) study will take to recruit a **Target Number of Events.** In nQuery Predict, there are two main approaches to event milestone prediction:

- Blinded Events Prediction
- Unblinded Events Prediction

Blinded events prediction is where prediction is made without knowledge of treatment group assignment. Unblinded events prediction is where prediction is made with knowledge of treatment group assignment.

Within each approach, examples will be shown for when **Input** is **Subject Data Only, Subject + Site Data** or **Summary Data** and for when **Enrollment Status** is **Ongoing** or **Complete.** The combinations shown in the examples in each sections will be as follows:

- Blinded Events Prediction with Subject Data Only with Enrollment Ongoing (subsubsection 6.2.7.1)
- Unblinded Events Prediction with Subject + Site Data with Enrollment Ongoing (subsubsection 6.2.7.2)
- Blinded Events Prediction with Subject Data Only with Enrollment Complete (subsubsection 6.2.7.3)
- Unblinded Events Prediction with Summary Data with Enrollment Ongoing (subsubsection 6.2.7.4)
- Blinded Events Prediction with Summary Data Pre-Trial (subsubsection 6.2.7.5)

These five examples will cover the major changes in workflow depending on the other options selected at the **Select the Type of Prediction** step.

Each of these scenarios will be demonstrated using the SubjectLevel.csv and SiteLevel.csv datasets provided with nQuery Predict. See subsection 6.2.4 for details on these datasets.

Each demonstration will start with demonstration using the default values given by nQuery Predict, followed by additional scenarios taken from the context these datasets were taken from and which use additional options available in nQuery Predict.

The additional options covered in each demo are as follows:

- Blinded Events, Subject Data Only, Enrollment Ongoing: Editing **Target Sample Size**, **Weibull Event Model**, **Fixed Followup** design
- Unblinded Events, Subject + Site Data, Enrollment Ongoing: Editing **Target Sample Size**, removing and "closing" sites, **Weibull Event Model**
- Blinded Events, Subject Data Only, Enrollment Complete: Editing **Target Number of Events**, **Piecewise Exponential Event Model**
- Unblinded Events, Summary Data, Enrollment Ongoing: Editing **Exponential Model Hazard Rates**
- Blinded Events, Summary Data, Pre-Trial: **Unblinded Events Pre-Trial**, editing **Exponential Model Hazard Rates**

For these demonstrations, it is assumed an nQuery Workspace with the two example datasets has been created as per the **Creating a nQuery Predict Workspace** section above (subsection 6.2.1).

#### 6.2.7.1 Blinded Events Prediction with Subject Data Only and Enrollment Ongoing

**Setup** **Blinded Events** milestone prediction using **Subject-level Data Only** means that the prediction will project the expected length of the accrual period and study needed to achieve a specified **Target Number of Events** using subject-level interim data alone.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step select **Blinded Events** from the **Target** field, **Subject Data Only** from the **Input** field and **Ongoing** from the **Enrollment Status** field.

These selections are shown in Figure 6.65. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

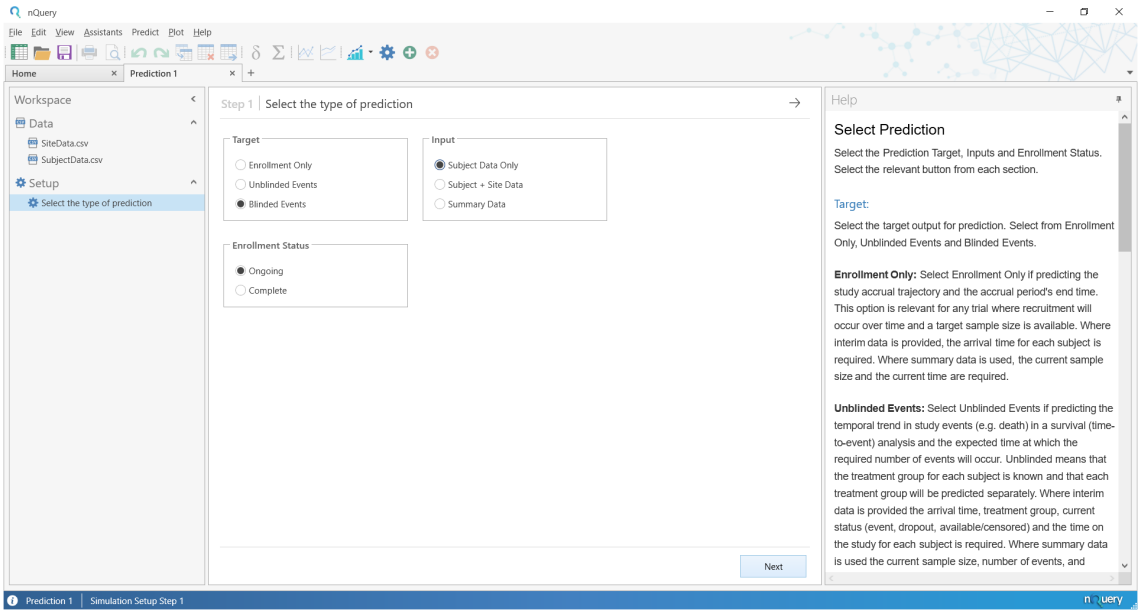



**Figure 6.65:** Select the Type of Prediction Setup Step for Blinded Events using Subject-level Data

The next step is the **Subject-level Data** step. The **Subject-level Data** step is a Data Field Selection step. The **Subject-level Data** is the dataset where each row corresponds to a specific subject enrolled in the study.

Data Field Selection steps are steps in which the required dataset in the current Workspace is selected and columns in the selected dataset are assigned to the required fields for that prediction problem. The options in a given Data Field Selection field are opened by selecting the button on the right of each field which is found to the right of the field name e.g. **Select Data**. Each field will have specific requirements (e.g. numeric only, two distinct values only) and nQuery Predict only displays options which are consistent with those fields.

In this prediction there are four fields which are as follows:

- Select Dataset: Select the dataset that is the Subject-level Data from the datasets in the current workspace
- Arrival Time: The arrival time column contains the length of time since the study start (Time = 0) until a subject entered the study. Each subject's Arrival Time should be greater than or equal to zero
- Status Indicator: The status (event, dropout, censored/active) for each subject. This Status Indicator column should have a minimum of one alphanumeric unique value and a maximum of three unique values. It has the following sub-fields
  - Event Status: Enter value in Status Indicator column that corresponds to subject having had an event

  - Dropout Status: Enter value in Status Indicator column that corresponds to a subject having dropped out
  - Censored Status: Enter value in Status Indicator column that corresponds to a subject having neither had the event or dropped out

- Time on Study: The length of time after Arrival Time that a subject was in the study. Must be numeric value greater than or equal to zero. The definition depends on Status Indicator (see below)

When the **Status Indicator** field is selected, the **Event Status**, **Dropout Status** and **Censored Status** rows become active. These sub-fields are used to select which values in the **Status Indicator** field correspond to each subject status of interest. The **Censored Status** is also referred to as the **Available Status** as the subject would have been administratively censored if the study ended at the **Current Time** but will be available to have the event of interest in the future during the event milestone prediction.

If the nQuery Predict default values of 1 = Event, -1 = Dropout, 0 = Censored is used in the **Status Indicator** field in the **Subject-level Data** then the **Status** sub-fields will be filled automatically with those assignments. Otherwise, manually enter values by either inputting a value directly into the field or selecting the button on the right of each **Status** sub-field and selecting from the list of unique values in the selected **Status Indicator** field. If there are less than three unique values in the **Status Indicator** field (for example if there have been no dropouts in the **Subject-Level Data** i.e. zero rows with **Dropout Status**) then the user must manually enter the value they would like to have assigned for that **Status** outcome. Note that appropriate defaults will be provided for the rate of previously unused **Status** category in subsequent steps e.g. Dropout Hazard Rate(s) = 0.

The definition of **Time on Study** will depend on the **Status Indicator** for that subject. If the **Status Indicator** is equal to **Event Status** or **Dropout Status**, the **Time on Study** is the length of time a subject was followed for after their **Arrival Time** before they had an event/dropped out. If the **Status Indicator** is equal to **Censored Status** , the **Time on Study** is the length of time a subject was followed for after their **Arrival Time** until the **Current Time.**

For this prediction, in the **Select Dataset** field select **SubjectData.csv**. Then in the **Arrival Time** field select **Arrival**, in the **Status Indicator** field select **Current** (assign 1 to **Event Status**, -1 to **Dropout Status**, 0 to **Censored Status**) and in the **Time on Study** field select **Followup.** These selections are shown in Figure 6.66. Select the Next or → button to move to the next **Setup** step.

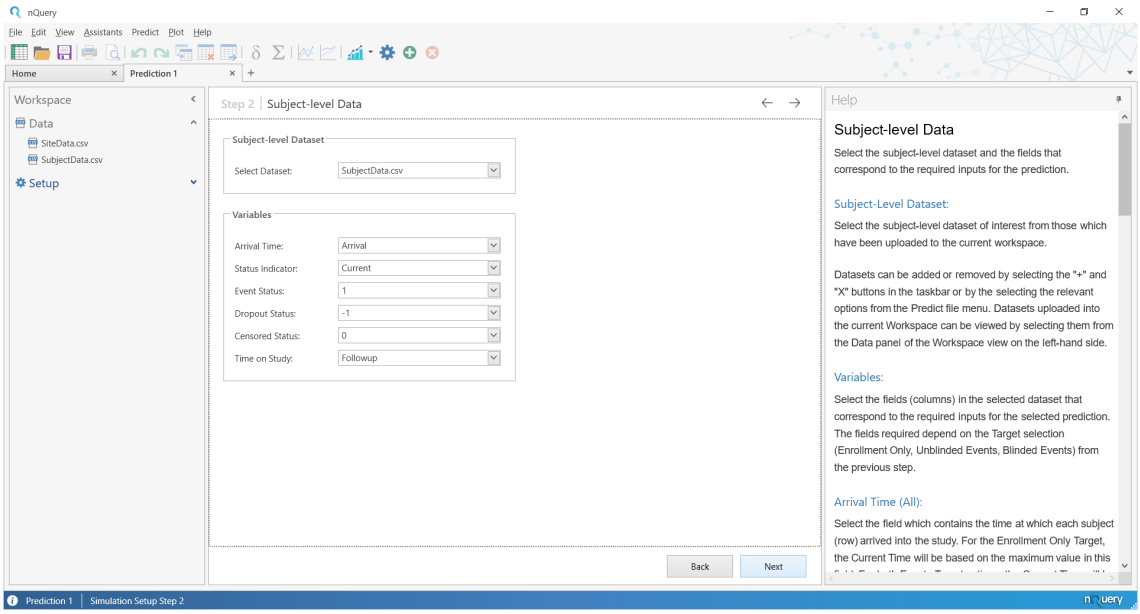



**Figure 6.66:** Subject-level Data Setup Step for Blinded Events Prediction using Subject-level Data

The next step is the **Accrual Options** step which specifies the **Target Sample Size, Accrual Model, Followup Option** and future **Accrual Rate**. The **Accrual Options** step, and its defaults, for this demonstration are shown in Figure 6.67.

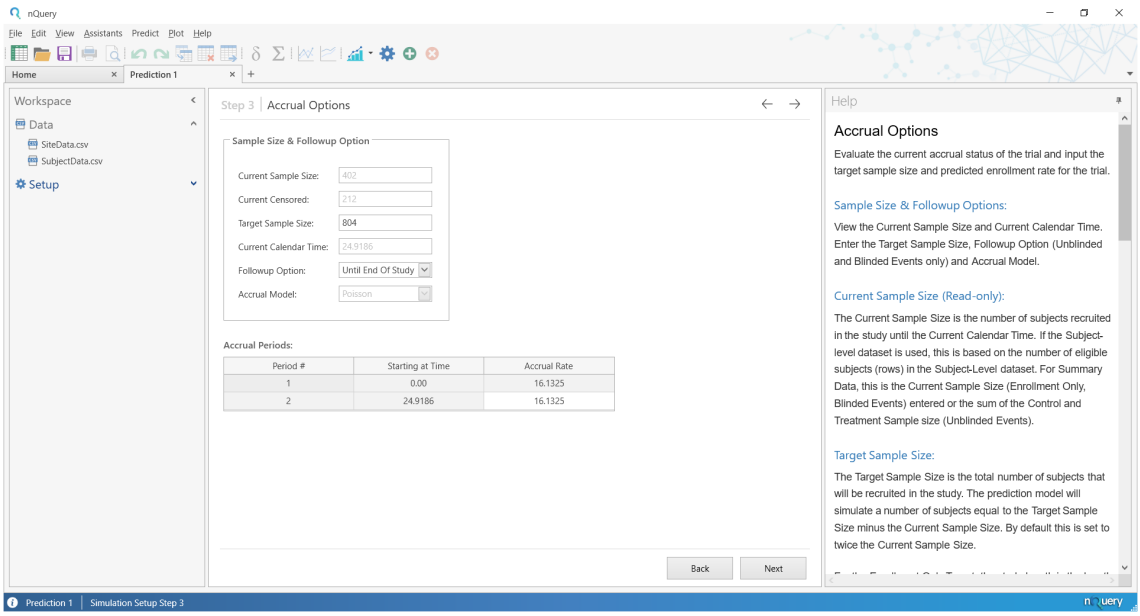



**Figure 6.67:** Accrual Options Setup Step for Blinded Events Prediction using Subject-level Data

The **Accrual Options** step consists of two main elements: the **Sample Size & Followup Option** input fields and the **Accrual Periods** table.

The **Sample Size & Followup Option** input fields provides information on enrollment process, censored status and current time based on the subject-level data and allows the user to edit the **Target Sample Size** and select the **Followup Option** and **Accrual Model**. The following fields are provided in this prediction:

- Current Sample Size (Read-only): Sample Size recruited into a study so far. Equals the number of rows in the Subject-level Data
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study. The number of simulated subjects equals the Total Sample Size minus the Current Sample Size
- Current Time (Read-only): The length of time that has passed since the study started. Equals the maximum of the sum of the Arrival Time + Time on Study values for each subject
- Followup Option: Specifies whether subjects who do not have the event/dropout will be administratively censored at the end of the study or after a fixed follow up period
  - Followup Time (Followup Option = For Fixed Period only): Length of time a subject is followed for until administratively censored if they do not have the event/dropout
- Accrual Model (Editable): The statistical model that will be used to generate simulated enrollments

In nQuery Predict, the **Target Sample Size** defaults to twice the **Current Sample Size**, **Followup Option** is set to **Until End of Study** and the **Accrual Model** defaults to **Poisson.** In this demonstration, the **Current Sample Size** equals 402 and therefore the default **Target Sample Size** is $402\,(2) = 804$.

If **Followup Option** is set to **Until End of Study** then subjects who do not have the event or dropout will be followed until the end of the study when the **Target Number of Events** is reached, at which point they will be administratively right-censored.

If **Followup Option** is set to **For Fixed Period** then subjects who do not have the event or dropout will be followed until they have a **Time on Study** equal to the value specified in the **Followup Time** field, which appears automatically below the **Followup Option** field when **For Fixed Period** is selected, at which point a subject will be administratively right-censored. Note that for subjects whose **Followup Time** is less than length of the time a subject was on study before the

study ended when the **Target Number of Events** was reached, they will be administratively censored at the study end rather than after their **Followup Time**.

The **Accrual Periods** table contains information on the **Accrual Rate** up the **Current Time** and allows the user to edit the future **Accrual Rate** used to generate enrollment simulations. The **Accrual Periods** has three columns ,with each row corresponding to a time period, with the columns defined as follows:

- Period #: A numeric ID for the current time period. Increases in increments of one for each subsequent row
- Starting at Time: The starting time for the current time period row. The first row will equal 0 (rate from study start to Current Time) and the second row will equal the Current Time
- Accrual Rate: The accrual rate is the average number of subjects recruited per unit time (months). The accrual rate in the first row will equal the accrual rate up to the Current Time

For the **Poisson Accrual Model**, the **Accrual Periods** table will consist of two time period rows. The first row will correspond to the information provided by the user regarding the study up until the **Current Time.** The second row will correspond to the inputs used to generate future enrollments. By default the **Accrual Rate** in the future second row will equal the **Accrual Rate** from the first row i.e. the accrual rate up until the **Current Time.** For the inputs used in this demonstration, this **Accrual Rate** column values in each row equal 16.13254936. The user can edit the **Accrual Rate** for generating future enrollments by editing the second cell of the **Accrual Rate** column.

For now, the default values will be used for **Target Sample Size** (804), **Followup Option** (Until End of Study) **Accrual Model** (Poisson) and **Accrual Rate** (16.13254936) as per Figure 6.67. Select the Next or → button to move to the next **Setup** step.

The next step is the **Event and Dropout Information** step which specifies the **Target Number of Events,** event model (and associated statistical parameters) and dropout model (and associated statistical parameters). The **Event and Dropout Information** step on the default **Events Model** tab is shown in Figure 6.68.

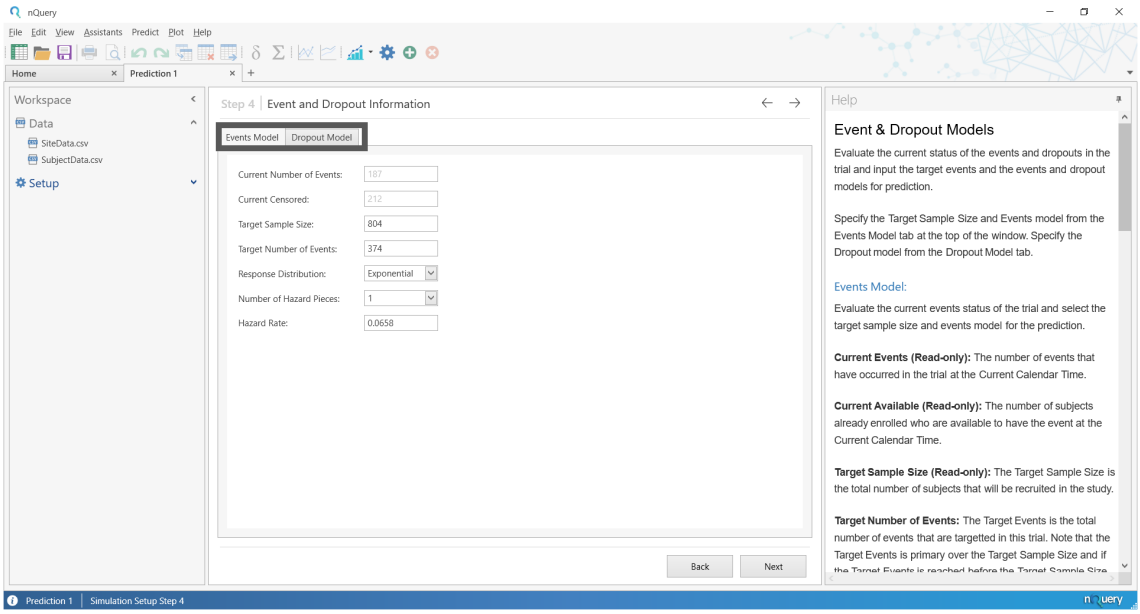



**Figure 6.68:** Event and Dropout Information Setup Step for Blinded Events Prediction using Subject-level Data

The **Event and Dropout Information** step consists of two tabs: the **Events Model** tab and the **Dropout Model** tab. These tabs have been highlighted in Figure 6.68 and are opened by selecting the required tab name at the top of the main window, below the step title. For **Blinded Events**, it is assumed that survival and dropout times will be simulated using a single "global" process which is specified by the user.

The **Events Model** tab provides fields which provide information on the current number of subject who have had an event, could have an event going forward and allows the input of the **Target Number of Events, Target Sample Size, Response Distribution** for the events model and additional parameters needed for that specific model. The following fields are provided in this prediction:

- Current Events (Read-only): The number of subjects who had the event of interest at the current time. Equals the number of subjects with the Event Status in the Status Indicator field
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study. The number of simulated subjects equals the Total Sample Size minus the Current Sample Size
- Target Number of Events (Editable): The required number of subjects needed to have the event of interest for the study to end

- Response Distribution: The statistical model that will be used to generate survival times for subjects available to have the event
- Number of Hazard Pieces (Response Distribution = Exponential only): Number of hazard pieces used to specify the piecewise exponential model. Select value between 1 and 10
- Hazard Rate (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The exponential hazard rate for the constant exponential survival model specified above

In nQuery Predict, the **Target Sample Size** at this step is inherited from the prior **Accrual Options** step, the **Target Number of Events** defaults to twice the **Current Events**, the default **response Distribution** is **Exponential** with the **Number of Hazard Pieces** set to 1 and the default **Hazard Rate** set the expected event rate from the provided information (see subsubsection 6.1.2.2 for details on the survival distributions used in nQuery Predict, see section 3.5 for details on how to calculate a exponential **Hazard Rate** using a median survival or survival probability)**.**

In this demonstration, the **Current Number of Events** is 187 and therefore the default **Target Number of Events** is $187\,(2) = 374$ giving an estimated exponential **Hazard Rate** of 0.0658. Note that if the **Target Number of Events** is reached before the **Target Sample Size** has been enrolled then the it is assumed the enrollment process will stop early and the total sample size achieved may be less than the **Target Sample Size** specified.

In nQuery, there are effectively three primary survival distributions that survival times can be generated from:

- Constant Exponential (Response Distribution = Exponential, Number of Hazard Pieces =1): Input Hazard Rate
- Piecewise Exponential (Response Distribution = Exponential, Number of Hazard Pieces > 1): Input table of Starting at Time and Hazard Rate columns
- Weibull (Response Distribution = Weibull): Input Scale Parameter and Shape Parameter

Details on how survival times are generated from each of these distributions is provided in subsubsection 6.1.2.2, alongside how the default values provided in nQuery Predict are estimated from the provided information. The change in the **Events Model** tab for each of these choices is shown in Figure 6.69.

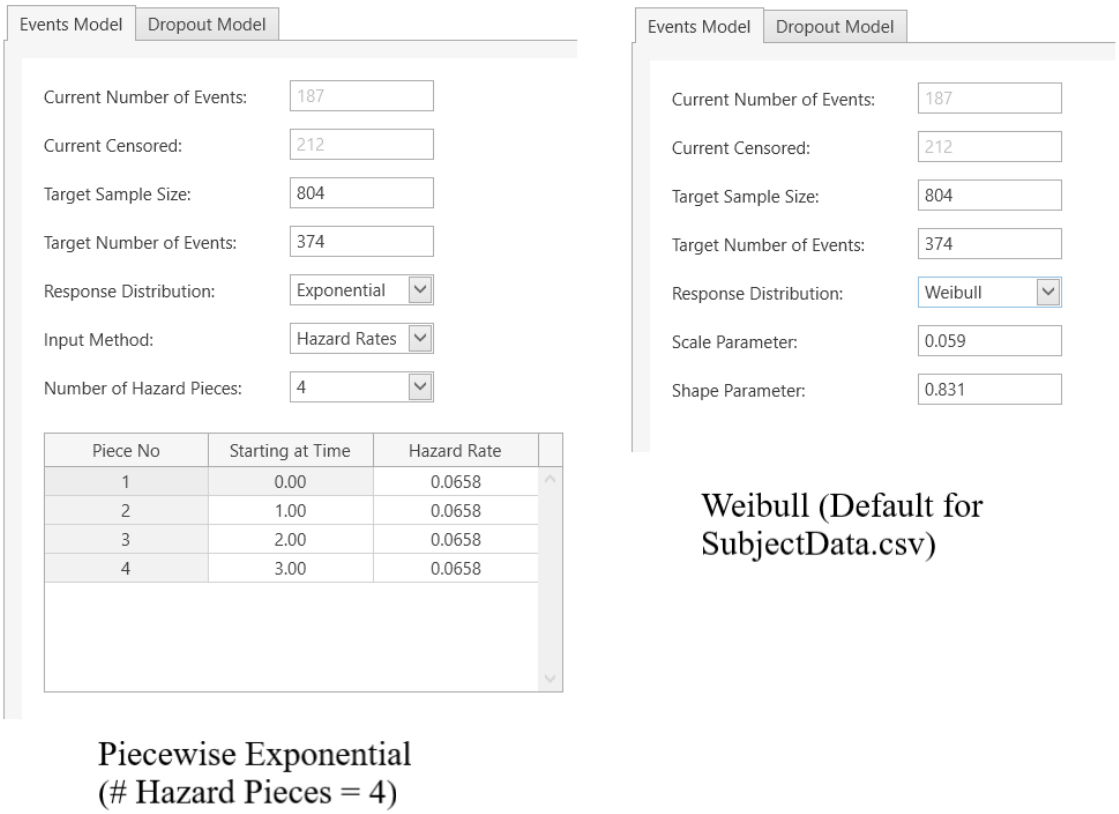



**Figure 6.69:** Piecewise Exponential (4 Pieces) and Weibull Defaults for SubjectData.csv

The piecewise exponential table will have a number of rows equal to **Number of Hazard Pieces.** The **Piece No** column indicates the hazard piece number of a row. The **Starting at Time** column sets the start time for when the **Hazard Rate** value in that row takes effect. The **Hazard Rate** column sets the exponential hazard rate in effect from the start time of the current hazard piece to the start time in the next row (see section 3.5 for details on how to calculate a exponential **Hazard Rate** using a median survival or survival probability and the Compute Effect Size side table for STT3 may also be useful). By default the piecewise exponential assumes a constant exponential model **Hazard Rate** (same **Hazard Rate** as for the **Number of Hazard Pieces** = 1 case) with **Starting at Time** being values increasing by 1 up to the final row.

The Weibull distribution has two parameters: the Weibull **Scale Parameter** and Weibull **Shape Parameter**. The default for these is found by fitting the survival data from the subject-level data (see subsubsection 6.1.2.2 for details).

The **Dropout Model** tab is opened by selecting the **Dropout Model** tab name at the top of the main window. For this demonstration, the **Dropout Model** tab is shown in Figure 6.70.

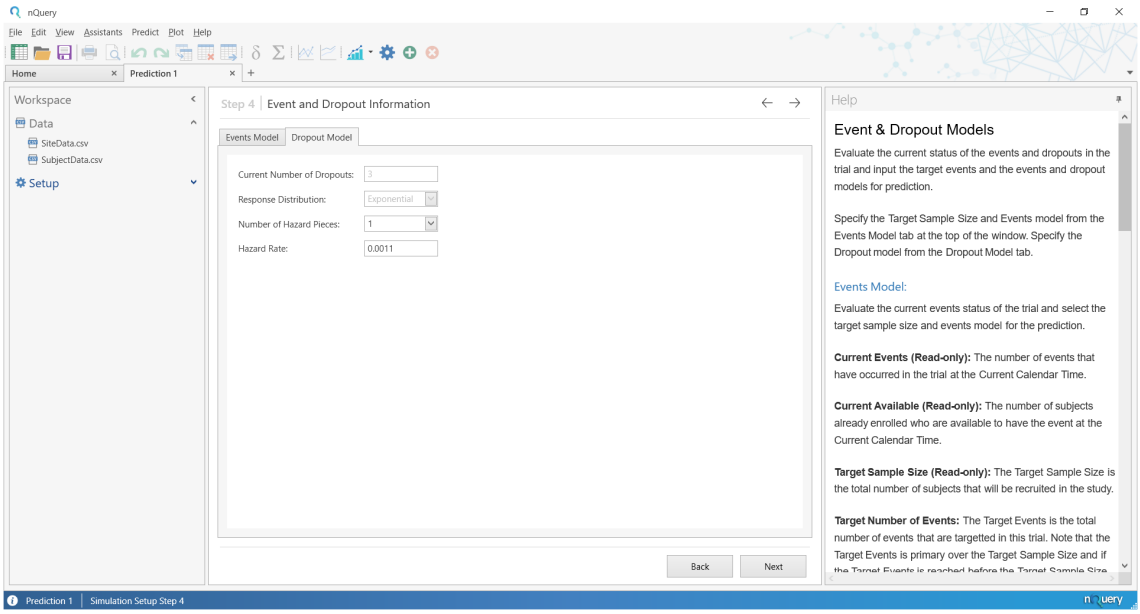



**Figure 6.70:** Dropout Model tab for Blinded Events with Subject-level Data

For **Dropout Model**, the **Current Number of Dropouts** (number of subjects who have dropped out up to the **Current Time**), **Response Distribution** and **Number of Hazard Pieces** are available. For dropout, the only **Response Distribution** option is **Exponential**. The inputs and defaults for dropout **Exponential Response Distribution**, for both the case where **Hazard Pieces = 1** (Constant Exponential) or **Hazard Pieces > 1** (Piecewise Exponential), are the same as from the **Events Model** described above and therefore will not be replicated here. In this demonstration, the constant exponential **Hazard Rate** for dropout calculated from the information provided equals 0.0011.

In this demonstration the default for all inputs (**Target Sample Size = 804, Target Number of Events = 374, Events Model Exponential Response Distribution (1 Hazard Piece) Hazard Rate = 0.0658, Dropout Model Exponential Response Distribution (1 Hazard Piece) Hazard Rate = 0.0011**) are used. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.71. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

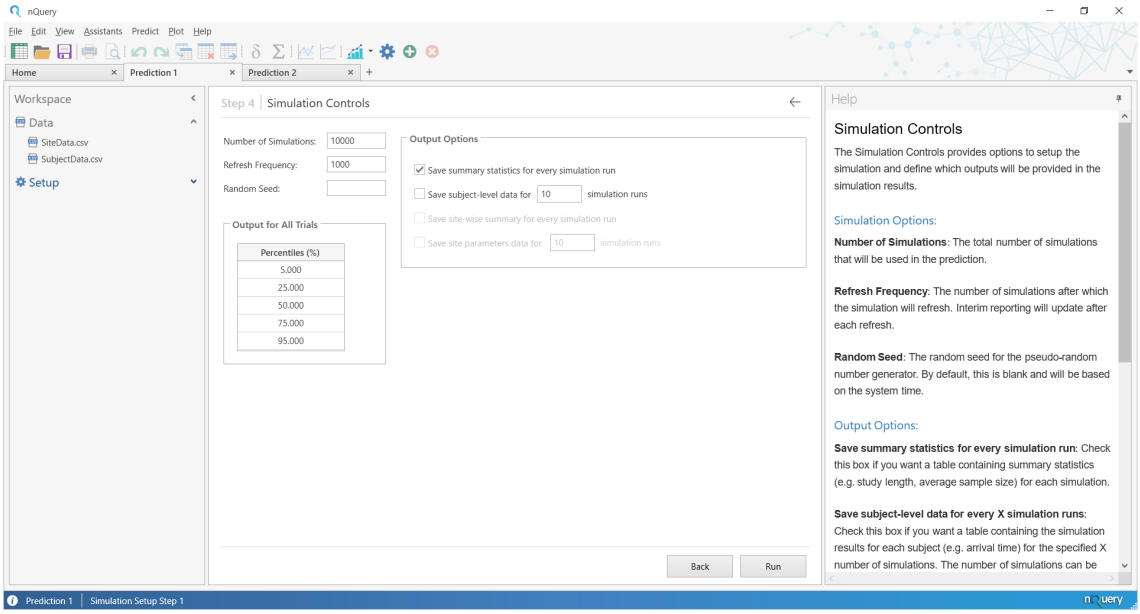



**Figure 6.71:** Simulation Controls Setup Step for Blinded Events Prediction using Subject-level Data

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.72.

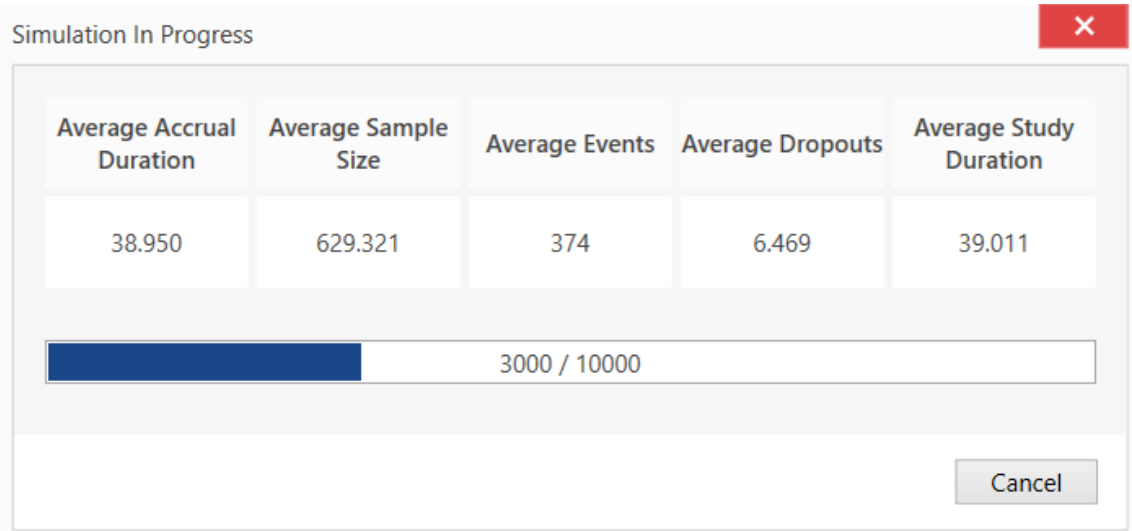



**Figure 6.72:** Simulation in Progress Window for Blinded Events Prediction using Subject-level Data

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.73.

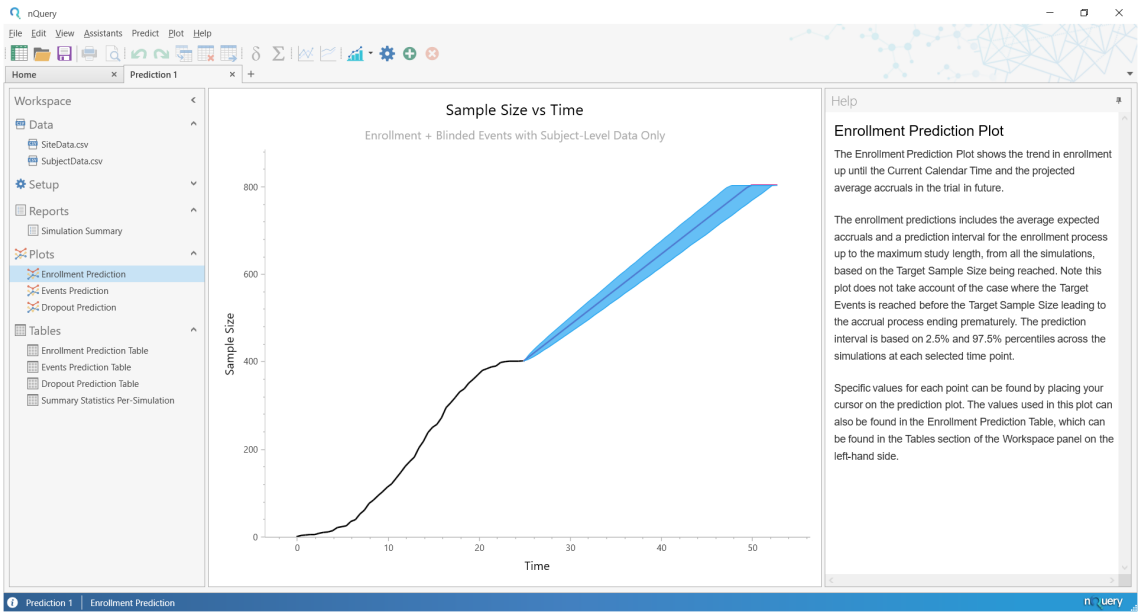



**Figure 6.73:** Enrollment Prediction Plot for Blinded Events Prediction using Subject-level Data

The results of a nQuery Predict milestone prediction can be divided into three

categories: **Reports, Plots** and **Tables**. A high-level summary of the results in nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table, Events Prediction Table, Dropout Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.74.

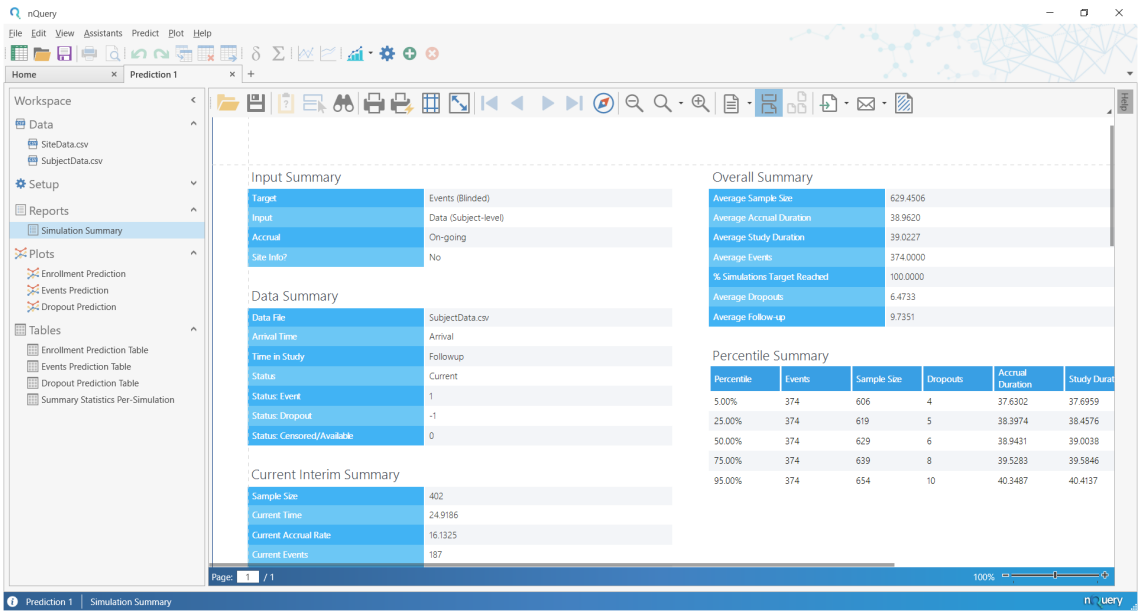



**Figure 6.74:** Simulation Summary Report for Blinded Events Prediction using Subject-level Data

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.75. For this demonstration, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 39.0227. The **Percentile Summary** shows that 90% of predictions had

a **Study Duration** between 37.6959 (5% Percentile) and 40.4137 (95% Percentile), 50% a duration between 38.4576 (25% Percentile) and 39.5846 (75% Percentile) and a median (50% Percentile) duration of 39.0038. The **Average Dropouts** was 6.4733 (Percentiles: 4, 5, 6, 8, 10) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 9.7351.

The accrual process in this study ended early for all simulations with an **Average Sample Size** of 629.4506 (Percentiles: 606, 619, 629, 639, 654) enrolled before the **Target Number of Events** of 804 was reached. This meant the **Average Accrual Duration** of 38.9620 (Percentiles: 37.6302, 38.3974, 38.9431, 39.5283, 40.3487) is effectively equal to the **Average Study Duration** of 39.0227 with the accrual duration effectively the time of the last enrollment before study end.

Overall Summary

| | |
|---|---|
| Average Sample Size | 629.4506 |
| Average Accrual Duration | 38.9620 |
| Average Study Duration | 39.0227 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.4733 |
| Average Follow-up | 9.7351 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 606 | 4 | 37.6302 | 37.6959 |
| 25.00% | 374 | 619 | 5 | 38.3974 | 38.4576 |
| 50.00% | 374 | 629 | 6 | 38.9431 | 39.0038 |
| 75.00% | 374 | 639 | 8 | 39.5283 | 39.5846 |
| 95.00% | 374 | 654 | 10 | 40.3487 | 40.4137 |

**Figure 6.75:** Simulation Summary Report Results for Blinded Events Prediction

To edit the **Simulation Summary** report there a number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment simu-

lation from the study start to when the **Target Sample Size** was reached in all simulations. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.76.

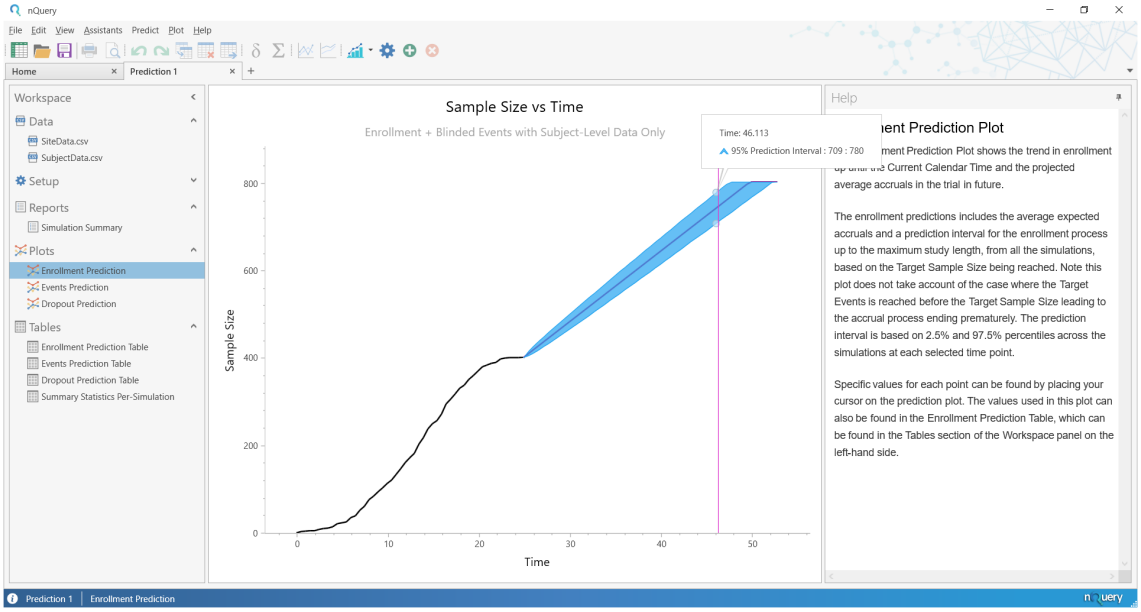



**Figure 6.76:** Enrollment Prediction Plot for Blinded Events Prediction using Subject-level Data

The **Enrollment Prediction** plot consists of two parts: from Time = 0 to **Current Time** and from **Current Time** to the **Maximum Time**. The first part is a black line which plots the accrual pattern from the user inputs provided. The second part is a blue conic section with a grey line where the blue conic section plots the 95% prediction interval for the **Sample Size** enrolled at a given **Time** across all simulations and the grey line is the average **Sample Size** enrolled at a given time. The 95% prediction interval indicates the 95% of predictions had a sample size that fell between the lower and upper limit at a given time.

Note that the **Enrollment Prediction** plot will show the projection for **Target Sample Size** and its associated **Accrual Duration** and therefore will usually have a different **Time** scale than the **Event Prediction** and **Dropout Prediction** plots which will be associated with the **Study Duration.** Note the **Enrollment Prediction** plot ignores the case where **Accrual Duration** is cut short due to the **Target Number of Events** being reached before the enrollment process is complete.

The **Events Prediction** plot is very similar to the **Enrollment Prediction** plot where the first part is a black line which plots the events pattern from the user inputs

provided and the second part is a yellow conic section with a grey line where the yellow conic section plots the 95% prediction interval for the **Number of Events** that have occurred at a given **Time** across all simulations and the grey line is the average **Number of Events** at a given time. The 95% prediction interval indicates the 95% of predictions had a number of events that fell between the lower and upper limit at a given time. The **Events Prediction** plot for this demonstration is shown in Figure 6.77.

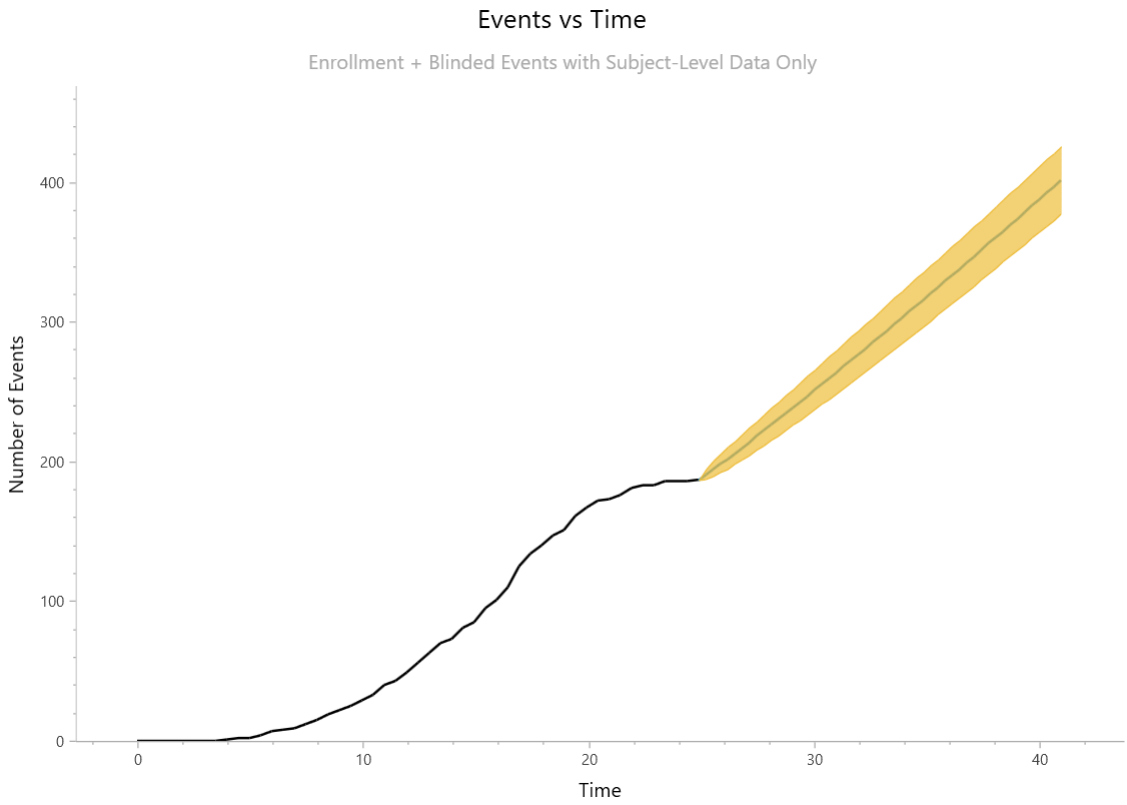



**Figure 6.77:** Events Prediction Plot for Blinded Events Prediction using Subject-level Data

It is important the note that these event times are when these events occurred relative to the start of the study, not how long they happened after follow up i.e. this is not a Kaplan-Meier type plot!

The **Dropout Prediction** plot works effectively the same as the **Events Prediction** plot except for the dropout process. The **Dropout Prediction** plot for this demonstration is shown in Figure 6.78.

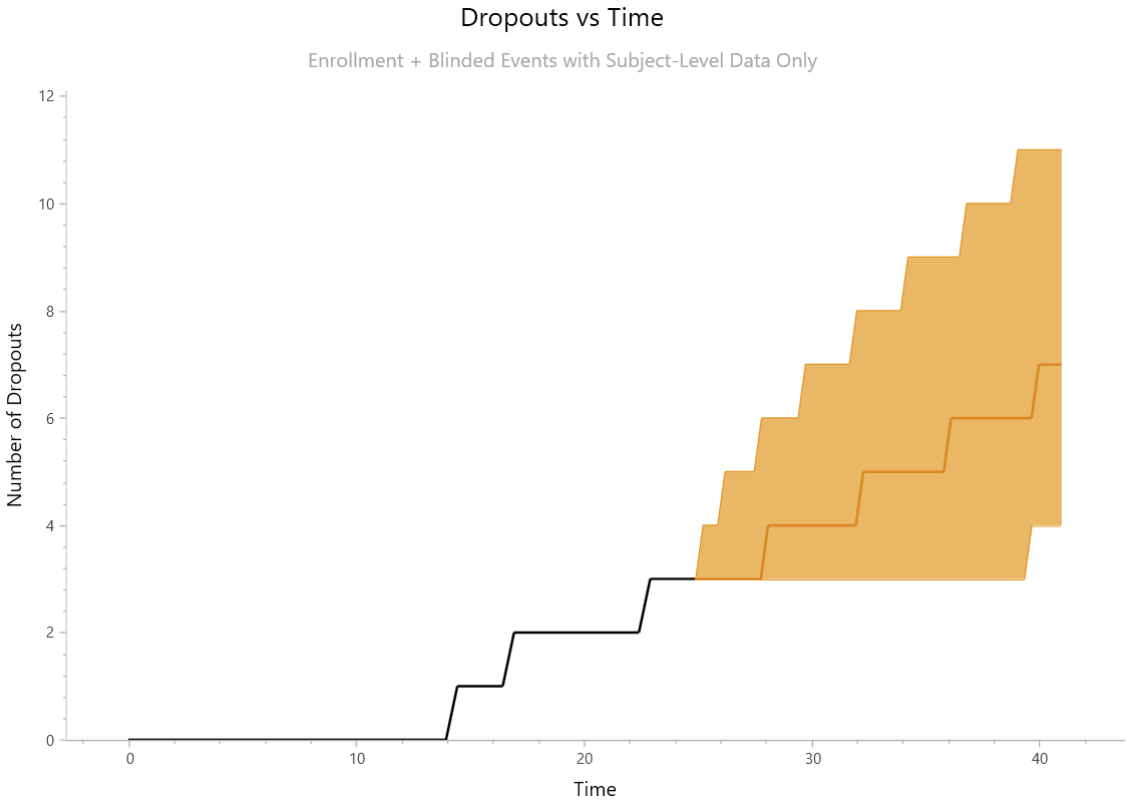
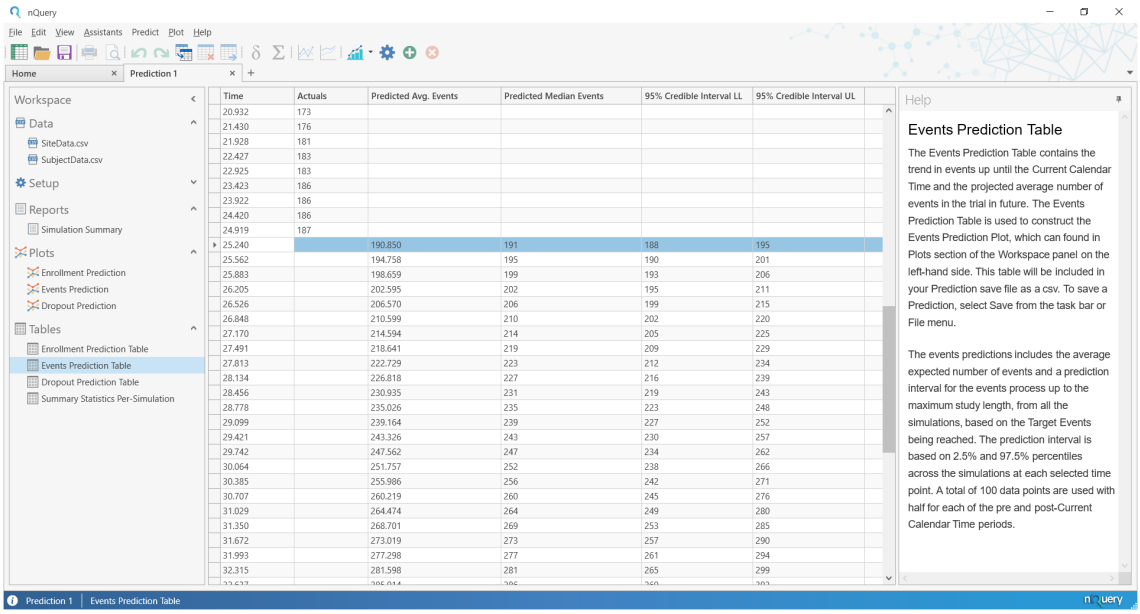



**Figure 6.78:** Dropout Prediction Plot for Blinded Events Prediction using Subject-level Data

The **Enrollment Prediction Table, Events Prediction Table and Dropout Prediction Tables** are the tables used to create the respective **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plots, which by default all consists of ~100 rows (50 pre-simulation and 50 post-simulation). Each consists of five columns: **Time, Actual, Predicted Avg. Sample Size/Events/Dropouts, Predicted Median Sample Size/Events/Dropouts, 95% Prediction Interval LL, 95% Prediction Interval UL**. One prediction table (the **Events Prediction Table**) for this demonstration at the start of the simulated period is given in Figure 6.79.




| Time | Actuals | Predicted Avg. Events | Predicted Median Events | 95% Credible Interval LL | 95% Credible Interval UL |
|---|---|---|---|---|---|
| 20.932 | 173 | | | | |
| 21.430 | 176 | | | | |
| 21.928 | 181 | | | | |
| 22.427 | 183 | | | | |
| 22.925 | 183 | | | | |
| 23.423 | 186 | | | | |
| 23.922 | 186 | | | | |
| 24.420 | 186 | | | | |
| 24.919 | 187 | | | | |
| 25.240 | | 190.850 | 191 | 188 | 195 |
| 25.562 | | 194.758 | 195 | 190 | 201 |
| 25.883 | | 198.659 | 199 | 193 | 206 |
| 26.205 | | 202.595 | 202 | 195 | 211 |
| 26.526 | | 206.570 | 206 | 199 | 215 |
| 26.848 | | 210.599 | 210 | 202 | 220 |
| 27.170 | | 214.594 | 214 | 205 | 225 |
| 27.491 | | 218.641 | 219 | 209 | 229 |
| 27.813 | | 222.729 | 223 | 212 | 234 |
| 28.134 | | 226.818 | 227 | 216 | 239 |
| 28.456 | | 230.935 | 231 | 219 | 243 |
| 28.778 | | 235.026 | 235 | 223 | 248 |
| 29.099 | | 239.164 | 239 | 227 | 252 |
| 29.421 | | 243.326 | 243 | 230 | 257 |
| 29.742 | | 247.562 | 247 | 234 | 262 |
| 30.064 | | 251.757 | 252 | 238 | 266 |
| 30.385 | | 255.986 | 256 | 242 | 271 |
| 30.707 | | 260.219 | 260 | 245 | 276 |
| 31.029 | | 264.474 | 264 | 249 | 280 |
| 31.350 | | 268.701 | 269 | 253 | 285 |
| 31.672 | | 273.019 | 273 | 257 | 290 |
| 31.993 | | 277.298 | 277 | 261 | 294 |
| 32.315 | | 281.598 | 281 | 265 | 299 |

**Figure 6.79:** Events Prediction Table for Blinded Events Prediction using Subject-level Data

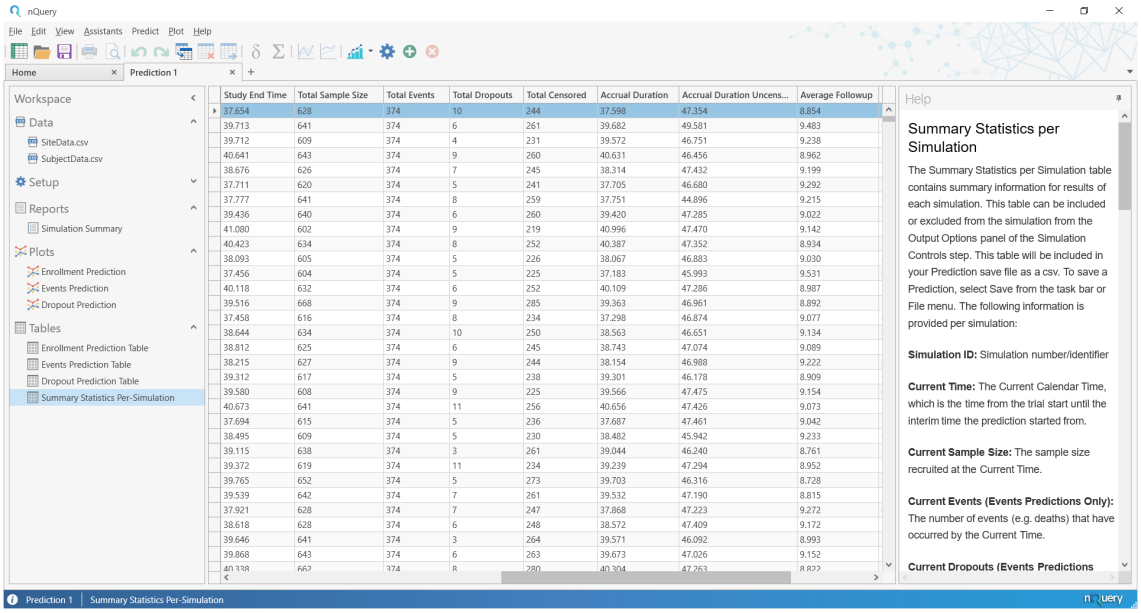
**Time** is the time value relative to the study start time of zero. **Actual** is the actual number of events achieved at a pre-simulation **Time** row. **Predicted Avg. Events** is the predicted average (mean) number of events at a post-simulation **Time** row. **Predicted Median Events** is the predicted median number of events at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the number of events at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the number of events at a post-simulation **Time** row.

Note that the **Enrollment Prediction Table** will show the projection for **Target Sample Size** and its associated **Accrual Duration** and therefore will usually have a different **Time** scale than the **Event Prediction Table** and **Dropout Prediction Table** which will be associated with the **Study Duration.** Note the **Enrollment Prediction Table** ignores the case where **Accrual Duration** is cut short due to the **Target Number of Events** being reached before the enrollment process is complete.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Current Events, Current Dropout, Current Available, Target Sample Size, Target Events, Study End Time, Total Sample Size, Total Events, Total Dropouts, Total Censored, Accrual Duration, Accrual Duration Uncensored, Average Followup, Median Followup, Target Events Reached.**

The **Summary Statistics Per-Simulation Table** for this demonstration, focused

on the simulation results, is shown in Figure 6.80.




| Study End Time | Total Sample Size | Total Events | Total Dropouts | Total Censored | Accrual Duration | Accrual Duration Uncens... | Average Followup |
|---|---|---|---|---|---|---|---|
| 37.654 | 628 | 374 | 10 | 244 | 37.598 | 47.354 | 8.854 |
| 39.713 | 641 | 374 | 6 | 261 | 39.682 | 49.581 | 9.483 |
| 39.712 | 609 | 374 | 4 | 231 | 39.572 | 46.751 | 9.238 |
| 40.641 | 643 | 374 | 9 | 260 | 40.631 | 46.456 | 8.962 |
| 38.676 | 626 | 374 | 7 | 245 | 38.314 | 47.432 | 9.199 |
| 37.711 | 620 | 374 | 5 | 241 | 37.705 | 46.680 | 9.292 |
| 37.777 | 641 | 374 | 8 | 259 | 37.751 | 44.896 | 9.215 |
| 39.436 | 640 | 374 | 6 | 260 | 39.420 | 47.285 | 9.022 |
| 41.080 | 602 | 374 | 9 | 219 | 40.996 | 47.470 | 9.142 |
| 40.423 | 634 | 374 | 8 | 252 | 40.387 | 47.352 | 8.934 |
| 38.093 | 605 | 374 | 5 | 226 | 38.067 | 46.883 | 9.030 |
| 37.456 | 604 | 374 | 5 | 225 | 37.183 | 45.993 | 9.531 |
| 40.118 | 632 | 374 | 6 | 252 | 40.109 | 47.286 | 8.987 |
| 39.516 | 668 | 374 | 9 | 285 | 39.363 | 46.961 | 8.892 |
| 37.458 | 616 | 374 | 8 | 234 | 37.298 | 46.874 | 9.077 |
| 38.644 | 634 | 374 | 10 | 250 | 38.563 | 46.651 | 9.134 |
| 38.812 | 625 | 374 | 6 | 245 | 38.743 | 47.074 | 9.089 |
| 38.215 | 627 | 374 | 9 | 244 | 38.154 | 46.988 | 9.222 |
| 39.312 | 617 | 374 | 5 | 238 | 39.301 | 46.178 | 8.909 |
| 39.580 | 608 | 374 | 9 | 225 | 39.566 | 47.475 | 9.154 |
| 40.673 | 641 | 374 | 11 | 256 | 40.656 | 47.426 | 9.073 |
| 37.694 | 615 | 374 | 5 | 236 | 37.687 | 47.461 | 9.042 |
| 38.495 | 609 | 374 | 5 | 230 | 38.482 | 45.942 | 9.233 |
| 39.115 | 638 | 374 | 3 | 261 | 39.044 | 46.240 | 8.761 |
| 39.372 | 619 | 374 | 11 | 234 | 39.239 | 47.294 | 8.952 |
| 39.765 | 652 | 374 | 5 | 273 | 39.703 | 46.316 | 8.728 |
| 39.539 | 642 | 374 | 7 | 261 | 39.532 | 47.190 | 8.815 |
| 37.921 | 628 | 374 | 7 | 247 | 37.868 | 47.223 | 9.272 |
| 38.618 | 628 | 374 | 6 | 248 | 38.572 | 47.409 | 9.172 |
| 39.646 | 641 | 374 | 3 | 264 | 39.571 | 46.092 | 8.993 |
| 39.868 | 643 | 374 | 6 | 263 | 39.673 | 47.026 | 9.152 |

**Figure 6.80:** Summary Statistics Per-Simulation Table for Blinded Events Prediction using Subject-level Data

Details on each field are provided in the Help window on the right but broadly the "Current" fields described the information provided up to the **Current Time**, the "Target" fields describe the study targets and the remainder give the projection achieved for key parameters in the simulation with that **Simulation ID**.

**Alternative Scenarios** The demonstration has focused on the default inputs using **SubjectData.csv** up to this point. However, as noted in the data summary given in subsection 6.2.4 the study from which the example data was simulated had an original sample size target of 460 (402 enrolled in **SubjectData.csv**). Additionally, in the **Enrollment Prediction Plot** there is noticeable "kink" where the accrual rate was slowing down for the final 4-5 months before the current time and the best fitting **Weibull Shape Parameter** was equal to 0.831 which indicates some deviation from the exponential assumption of a constant hazard rate (Weibull and exponential are equivalent if **Shape Parameter = 1)**

To explore this additional information, we will investigate two additional scenarios

- Scenario A: **Target Sample Size = 460** using default **Accrual Rate**
- Scenario B: **Target Sample Size = 460** using slower "updated" **Accrual Rate** and **Weibull Response Distribution** (using nQuery Predict default fit) model

- Scenario C: **Target Sample Size = 460** using default **Accrual Rate** with **Fixed Followup** = 12 months

**Scenario A** The only changes required compared to the first demonstration is to change the **Target Sample Size** to 460 at the **Accrual Options** step. The simplest way to do this is to select the **Setup** header in the Workspace Navigation Bar on the left and then select the **Accrual Options** drop down option. Then edit the **Target Sample Size** 460 in the **Current and Target Sample Sizes** section in the top-left and select the Next or → button. At the next **Simulation Controls** step, we will add a single simulation run by selecting the checkbox beside the **Save subject-level data for (x) simulation runs** and replace the default of 10 runs with 1 runs at the "(x)" spot. An edited down summary of the **Setup** changes is provided in Figure 6.81.

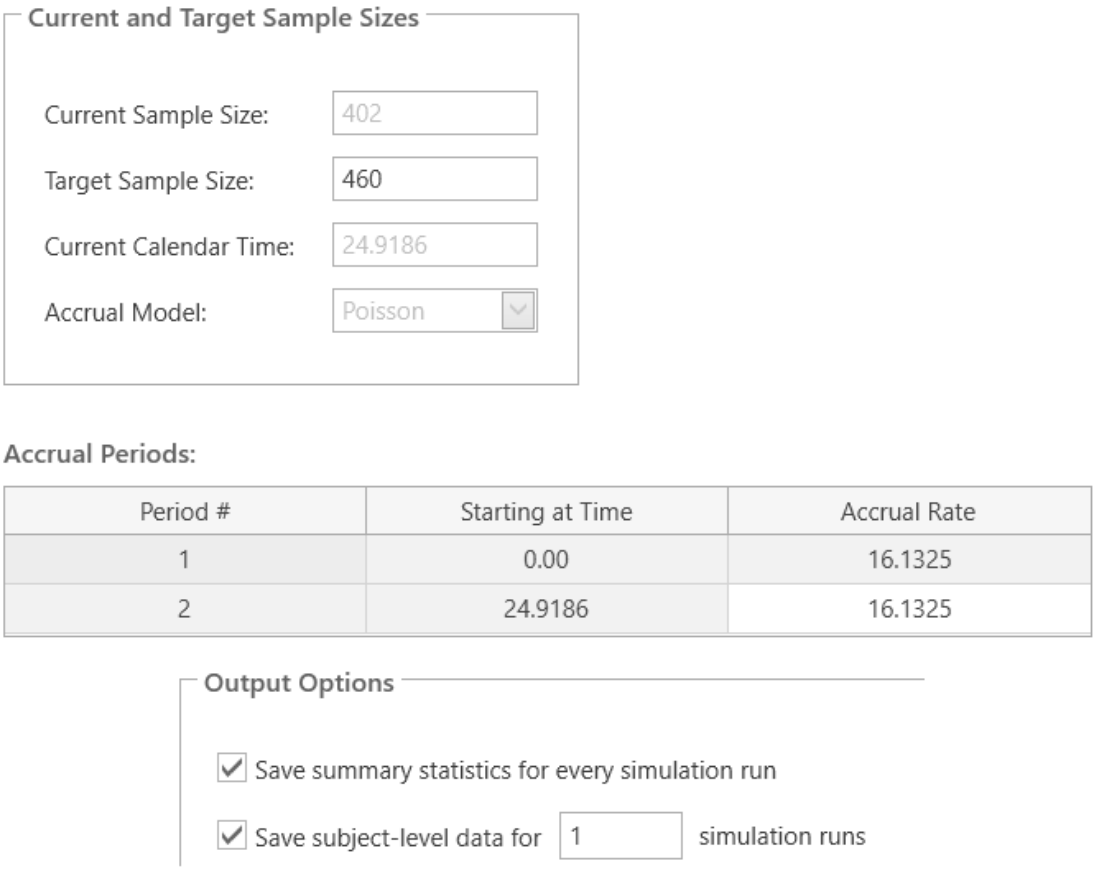



**Figure 6.81:** Scenario A Setup Changes for Blinded Events Prediction using Subject-level Data

To run the simulation again with the new Scenario A, select the Run/→ button.

The main results of this Scenario A simulation are shown in Figure 6.82.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 28.5145 |
| Average Study Duration | 43.4715 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.0007 |
| Average Follow-up | 12.3579 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 3 | 27.7722 | 41.1954 |
| 25.00% | 374 | 460 | 5 | 28.1872 | 42.4874 |
| 50.00% | 374 | 460 | 6 | 28.4944 | 43.4284 |
| 75.00% | 374 | 460 | 7 | 28.8220 | 44.4065 |
| 95.00% | 374 | 460 | 9 | 29.3153 | 45.9318 |

**Figure 6.82:** Scenario A Results for Blinded Events Prediction using Subject-level Data

For Scenario A, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 43.4715. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 41.1954 (5% Percentile) and 45.9318 (95% Percentile), 50% a duration between 42.4874 (25% Percentile) and 44.4065 (75% Percentile) and a median (50% Percentile) duration of 43.4284. The **Average Dropouts** was 6.0007 (Percentiles: 3, 5, 6, 7, 9) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 12.3579.

The original target study duration was 40 months and therefore this study is expected to end 3.5 months later than expected. There are multiple possible causes of delays in a survival trial including, but not limited to, a slower than expected accrual rate, a slower than expected event hazard rates and a higher than expected dropout (or other competing process) rate.

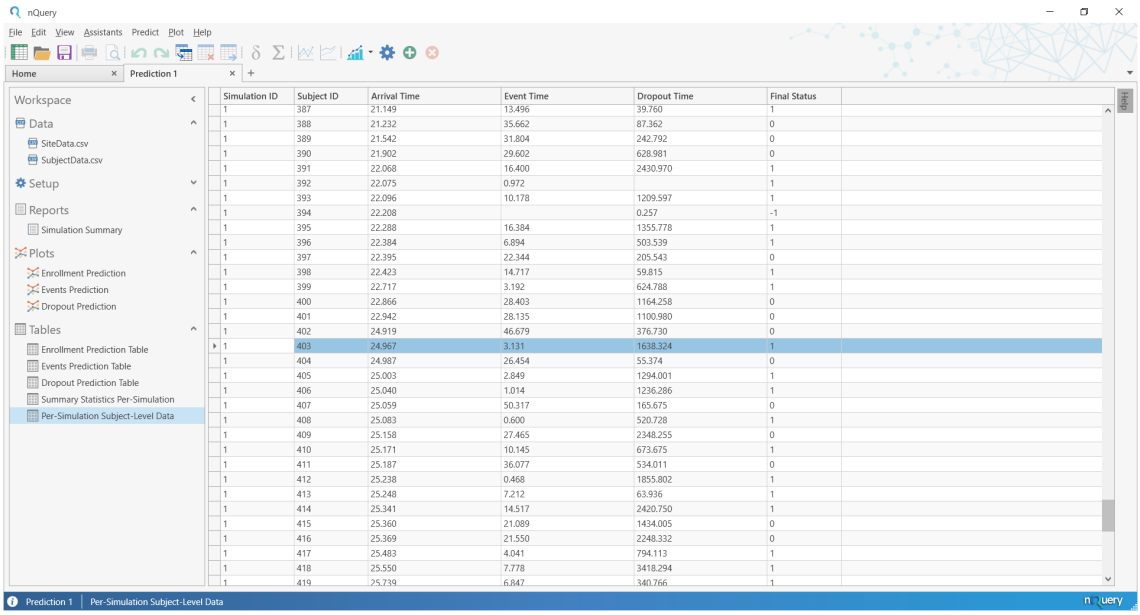
The **Target Sample Size** of 460 was reached in all simulations with an **Average**

**Accrual Duration** of 28.5154 (Percentiles: 27.7722, 28.1872, 28.4944, 28.8220, 29.3153).

Given that the original target accrual period length was 30 months, this indicates that if the average accrual rate over the entire period up to the current time was being maintained that the accrual period would around 1.5 months less than expected and the accrual pattern before the current time does not indicate any significant delay in starting accrual (see Figure 6.76). Since accrual was ahead of target, this indicates the longer than expected **Study Duration** was due to a lower than expected hazard rate since the dropout rate was quite low (6.0007 subjects dropped out on average).

In the original demonstration with a **Target Sample Size** of 804, the **Average Study Duration** was 39.0277. This would indicate a large increase in sample size would be needed to reduce the study duration sufficiently and would likely be infeasible.

The **Per-Simulation Subject-level Data Table** was also generated here and is shown for Scenario A in Figure 6.83 with the first simulated subject (and their **Arrival Time**) in the first simulation (**Simulation ID = 1**) highlighted. The tables consists of the following fields: **Simulation ID**, **Subject ID**, **Arrival Time**, **Event Time**, **Dropout Time** and **Final Status**. The order of the Subject ID is based on **Arrival Time** order.




| Simulation ID | Subject ID | Arrival Time | Event Time | Dropout Time | Final Status |
|---|---|---|---|---|---|
| 1 | 387 | 21.149 | 13.496 | 39.760 | 1 |
| 1 | 388 | 21.232 | 35.662 | 87.362 | 0 |
| 1 | 389 | 21.542 | 31.804 | 242.792 | 0 |
| 1 | 390 | 21.902 | 29.602 | 628.981 | 0 |
| 1 | 391 | 22.068 | 16.400 | 2430.970 | 1 |
| 1 | 392 | 22.075 | 0.972 | | 1 |
| 1 | 393 | 22.096 | 10.178 | 1209.597 | 1 |
| 1 | 394 | 22.208 | | 0.257 | -1 |
| 1 | 395 | 22.288 | 16.384 | 1355.778 | 1 |
| 1 | 396 | 22.384 | 6.894 | 503.539 | 1 |
| 1 | 397 | 22.395 | 22.344 | 205.543 | 0 |
| 1 | 398 | 22.423 | 14.717 | 59.815 | 1 |
| 1 | 399 | 22.717 | 3.192 | 624.788 | 1 |
| 1 | 400 | 22.866 | 28.403 | 1164.258 | 0 |
| 1 | 401 | 22.942 | 28.135 | 1100.980 | 0 |
| 1 | 402 | 24.919 | 46.679 | 376.730 | 0 |
| 1 | 403 | 24.967 | 3.131 | 1638.324 | 1 |
| 1 | 404 | 24.987 | 26.454 | 55.374 | 0 |
| 1 | 405 | 25.003 | 2.849 | 1294.001 | 1 |
| 1 | 406 | 25.040 | 1.014 | 1236.286 | 1 |
| 1 | 407 | 25.059 | 50.317 | 165.675 | 0 |
| 1 | 408 | 25.083 | 0.600 | 520.728 | 1 |
| 1 | 409 | 25.158 | 27.465 | 2348.255 | 0 |
| 1 | 410 | 25.171 | 10.145 | 673.675 | 1 |
| 1 | 411 | 25.187 | 36.077 | 534.011 | 0 |
| 1 | 412 | 25.238 | 0.468 | 1855.802 | 1 |
| 1 | 413 | 25.248 | 7.212 | 63.936 | 1 |
| 1 | 414 | 25.341 | 14.517 | 2420.750 | 1 |
| 1 | 415 | 25.360 | 21.089 | 1434.005 | 0 |
| 1 | 416 | 25.369 | 21.550 | 2248.332 | 0 |
| 1 | 417 | 25.483 | 4.041 | 794.113 | 1 |
| 1 | 418 | 25.550 | 7.778 | 3418.294 | 1 |
| 1 | 419 | 25.739 | 6.847 | 340.766 | 1 |

**Figure 6.83:** Per-Simulation Subject-level Data Table for Blinded Events Subject-level Data

**Arrival Time** is generated for all simulated subjects (i.e. **Subject ID > 402**). **Event Time** and **Dropout Time** are generated for all simulated subjects and all subjects which were **Censored Status** (all status values from **Status Indicator**

field in **Subject-level Data Setup** step) in the subject-level data (for example **Subject ID = 402** in this example). Blank fields for **Event Time** or **Dropout Time** indicate cases where the subject already had the event/dropout and therefore the time for the other status does not exist and does not need to be simulated.

The **Final Status** is from the subject-level data for subjects with **Event Status** or **Dropout Status** and based on the **Event Time** and **Dropout Time** for simulated subjects and **Censored Status** subjects in the subject-level data. **Final Status** will equal **Event Status** if **Event Time** is less than **Dropout Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.80 for per-simulation end times). **Final Status** will equal **Dropout Status** if **Dropout Time** is less than **Event Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.80 for per-simulation end times). **Final Status** will equal **Censored Status** if **Event Time** and **Dropout Time** are greater than the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.80 for per-simulation end times)

**Scenario B** Scenario A assumed the **Accrual Rate** would continue at average rate over the entirety of the period up until this simulation was made and the constant **Exponential Response Distribution** model was appropriate. However, there is evidence that in the last five months there was a significant reduction in the **Accrual Rate**, perhaps reflecting a wind down of the enrollment process and that the **Weibull Response Distribution** fit indicated some deviation from the exponential model (where Weibull **Shape Parameter = 1**).

To evaluate this scenario, note that 31 subjects where recruited after 20 months of the accrual period is passed. The remaining accrual period after 20 months to the current time equals 4.9186 (24.9816 (Current Time) - 20). This gives an average accrual rate over this period of ($31/4.9186 \approx 6.3$). This is significantly lower than the average 16.13 **Accrual Rate** over the entire period up to now. The default **Weibull** model will also be used.

To evaluate this scenario, select **Accrual Options** from under the **Setup** header in the Workspace Navigation Bar on the left then change the **Accrual Rate** in the **Period # = 2** row to 6.3 at the **Accrual Options** step by and then select the Next or → button.

At the next **Event and Dropout Information** step in the **Events Model** tab, change the **Response Distribution** field to **Weibull.** Select the Run or → button to move to the next step.

At the next **Simulation Controls** step, select the Run or → button to simulate Scenario B. The changes and results for Scenario B are summarized in Figure 6.84.

Accrual Periods:

| Period # | Starting at Time | Accrual Rate |
|---|---|---|
| 1 | 0.00 | 16.1325 |
| 2 | 24.9186 | 6.30 |

Events Model | Dropout Model

| | |
|---|---|
| Current Number of Events: | 187 |
| Current Censored: | 212 |
| Target Sample Size: | 460 |
| Target Number of Events: | 374 |
| Response Distribution: | Weibull |
| Scale Parameter: | 0.059 |
| Shape Parameter: | 0.831 |

## Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 34.1232 |
| Average Study Duration | 50.3707 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.8860 |
| Average Follow-up | 14.2191 |

## Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 4 | 32.2322 | 46.9750 |
| 25.00% | 374 | 460 | 5 | 33.2865 | 48.8669 |
| 50.00% | 374 | 460 | 7 | 34.0656 | 50.2908 |
| 75.00% | 374 | 460 | 8 | 34.8904 | 51.7489 |
| 95.00% | 374 | 460 | 10 | 36.2181 | 54.0922 |

**Figure 6.84:** Scenario B for Blinded Event Prediction with Subject-level Data

For Scenario B, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 50.3707 (Percentiles: 46.9750, 48.8669, 50.2908, 51.7489, 54.0922). The **Average Dropouts** was 6.8860 (Percentiles: 4, 5, 7, 8, 10) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 14.2191. The original target study duration was 40 months and therefore this study is expected to end 10.4 months later than expected.

The **Target Sample Size** of 460 was reached in all simulations with an **Average Accrual Duration** of 34.1232 (Percentiles: 32.2322, 33.2865, 34.0656, 34.8904, 36.2181) which is 4.1 months longer than expected. This likely contributed to the higher duration expected compared with **Scenario A (Average Study Duration =43.4715)**. In addition, the Weibull fit **Shape Parameter** of 0.8310 indicates a lower hazard rate the longer subjects were on the study which may also have contributed to the slower event rate in a study that was well underway. Given the larger increase in the expected study length, the case for a larger change (e.g. request ad-hoc sample size increase) may be greater.

**Scenario C** The prior examples have assumed the study was an event-driven (i.e. variable followup) design where all subjects who have not had the event or dropped out are followed until the **Target Number of Events** is reached and then administratively censored. nQuery Predict gives the option to do a fixed followup design where subjects are instead followed for a pre-specified maximum period of time after which are administratively censored.

In Scenario C, we explore what effect setting a **Fixed Followup** of 12 months would have on the event milestone prediction compared to **Scenario A** above.

To evaluate this scenario, select **Accrual Options** from under the **Setup** header in the Workspace Navigation Bar on the left then change the **Accrual Rate** in the **Period # = 2** row back to its default of 16.1325. Then select **For Fixed Period** from the **Followup Option** dropdown and enter 12 in the new **Followup Time** field. Then select the Next or → button.

At the next **Event and Dropout Information** step in the **Events Model** tab, change the **Response Distribution** field back to **Exponential.** Select the Run or → button to move to the next step.

At the next **Simulation Controls** step, select the Run or → button to simulate Scenario C. The changes and results for Scenario C are summarized in Figure 6.85.

Step 3 | Accrual Options

Sample Size & Followup Option

| | |
|---|---|
| Current Sample Size: | 402 |
| Current Censored: | 212 |
| Target Sample Size: | 460 |
| Current Calendar Time: | 24.9186 |
| Followup Option: | For Fixed Period |
| Followup Time: | 12.00 |
| Accrual Model: | Poisson |

Step 4 | Event and Dropout Information

Events Model | Dropout Model

| | |
|---|---|
| Current Number of Events: | 187 |
| Current Censored: | 212 |
| Target Sample Size: | 460 |
| Target Number of Events: | 374 |
| Response Distribution: | Exponential |
| Number of Hazard Pieces: | 1 |
| Hazard Rate: | 0.0658 |

Accrual Periods:

| Period # | Starting at Time | Accrual Rate |
|---|---|---|
| 1 | 0.00 | 16.1325 |
| 2 | 24.9186 | 16.1325 |

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 28.5132 |
| Average Study Duration | 40.4375 |
| Average Events | 253.4656 |
| % Simulations Target Reached | 0.0000 |
| Average Dropouts | 3.0546 |
| Average Follow-up | 7.8001 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 243 | 460 | 2 | 27.7999 | 39.7111 |
| 25.00% | 249 | 460 | 2 | 28.1806 | 40.1057 |
| 50.00% | 253 | 460 | 3 | 28.4874 | 40.4094 |
| 75.00% | 258 | 460 | 4 | 28.8193 | 40.7501 |
| 95.00% | 264 | 460 | 5 | 29.3221 | 41.2540 |

**Figure 6.85:** Scenario C for Blinded Event Prediction with Subject-level Data

For Scenario B, the **Overall Summary** shows that the **Target Number of Events** of 374 was not reached in any simulation with an **Average Events** of 253.4656 (Percentiles: 243, 249, 253, 258, 264) and an **Average Study Duration** over all the simulations of 40.4375 (Percentiles: 39.7111, 40.1507, 40.4094, 40.7501, 41.2540).

From this example it is obvious that the length of **Fixed Followup** would require a significantly higher **Target Sample Size** as many of the event times from the Scenario A would be greater than the 12 month cutoff used here and therefore these subjects are censored here instead of having the event of interest. This highlights the relative inefficiency of the fixed followup design and why it is generally only used over an event-driven design when there is a strong clinical reason to value that fixed followup period e.g. 12 month progression free survival (PFS).

Note that unlike in the other demonstrations, the **Study Duration** achieved here is the time at which the final subject was administratively censored due to reaching their **Fixed Followup** time (since **% Simulation Target Reached = 0**). In the other examples, the **Study Duration** was the time taken to reach the **Target Number of Events**. This definition of **Study Time** is also true for the **Fixed Followup** if **% Simulation Target Reached = 100%** (as for other demonstrations) or it can be a hybrid of these two outcomes if **% Simulation Target Reached** was between 0 and 100%.

The **Average Dropouts** was 3.0546 (Percentiles: 2, 2, 3, 4, 5) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 7.8001. The **Target Sample Size** of 460 was reached in all simulations with an **Average Accrual Duration** of 28.5132 (Percentiles: 27.9999, 28.1806, 28.4874, 28.8193, 29.3221) which replicates Scenario A.

#### 6.2.7.2 Unblinded Events Prediction with Subject + Site Data with Enrollment Ongoing

**Setup** **Unblinded Events** milestone prediction using **Subject + Site-level Data** means that the prediction will project the expected length of the accrual period and study needed to achieve a specified **Target Number of Events** using subject-level and site-level data, though note the site-level data is only used for the enrollment process.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step select **Unblinded Events** from the **Target** field, **Subject + Site Data** from the **Input** field and **Ongoing** from the **Enrollment Status** field.

These selections are shown in Figure 6.86. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

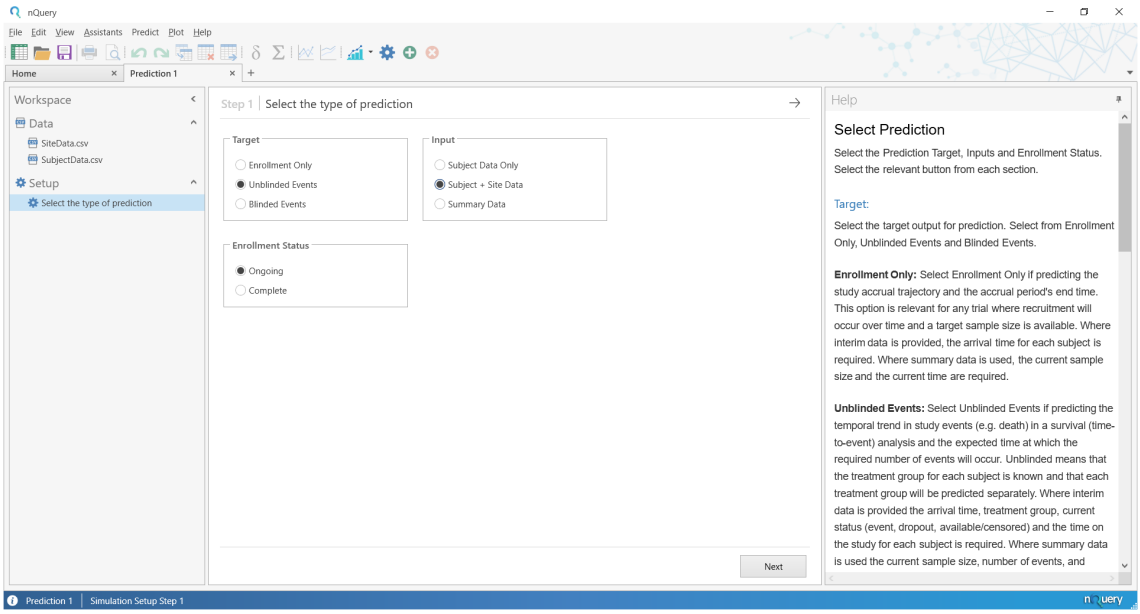



**Figure 6.86:** Select the Type of Prediction Setup Step for Unblinded Events using Subject + Site-level Data

The next step is the **Subject-level Data** step. The **Subject-level Data** step is a Data Field Selection step. The **Subject-level Data** is the dataset where each row corresponds to a specific subject enrolled in the study.

Data Field Selection steps are steps in which the required dataset in the current Workspace is selected and columns in the selected dataset are assigned to the required fields for that prediction problem. The options in a given Data Field Selection field are opened by selecting the button on the right of each field which is found to the right of the field name e.g. **Select Data**. Each field will have specific requirements (e.g. numeric only, two distinct values only) and nQuery Predict only displays options which are consistent with those fields.

In this prediction there are six fields which are as follows:

- Select Dataset: Select the dataset that is the Subject-level Data from the datasets in the current workspace
- Arrival Time: The arrival time column contains the length of time since the study start (Time = 0) until a subject entered the study. Each subject's Arrival Time should be greater than or equal to zero
- Treatment ID: The treatment code associated with each treatment group. The Treatment ID should have two unique values for the following two sub-fields:
  - Control Group: Treatment ID value associated with subject being assigned to control group
  - Treatment Group: Treatment ID value associated with subject being assigned to treatment group

- Status Indicator: The status (event, dropout, censored/active) for each subject. This Status Indicator column should have a minimum of one alphanumeric unique value and a maximum of three unique values. It has the following sub-fields
  - Event Status: Enter value in Status Indicator column that corresponds to subject having had an event
  - Dropout Status: Enter value in Status Indicator column that corresponds to a subject having dropped out
  - Censored Status: Enter value in Status Indicator column that corresponds to a subject having neither had the event or dropped out
- Time on Study: The length of time after Arrival Time that a subject was in the study. Must be numeric value greater than or equal to zero. The definition depends on Status Indicator (see below)
- Site ID: Site ID of site from which a subject was recruited. Site IDs here must correspond to Site ID contained in the Site-level Data

When the **Status Indicator** field is selected, the **Event Status**, **Dropout Status** and **Censored Status** rows become active. These sub-fields are used to select which values in the **Status Indicator** field correspond to each subject status of interest. The **Censored Status** is also referred to as the **Available Status** as the subject would have been administratively censored if the study ended at the **Current Time** but will be available to have the event of interest in the future during the event milestone prediction.

In the **Treatment ID** sub-fields, the lower alphanumeric input will be assigned to the **Control Group** and the other to the **Treatment Group.** To swap the sub-field values, manually enter the required value in the **Control Group** sub-field by either inputting that value directly into the field or selecting the button on the right of the sub-field and selecting from the two unique values in the selected **Treatment ID** field.

If the nQuery Predict default values of 1 = Event, -1 = Dropout, 0 = Censored is used in the **Status Indicator** field in the **Subject-level Data** then the **Status** sub-fields will be filled automatically with those assignments. Otherwise, manually enter values by either inputting a value directly into the field or selecting the button on the right of each **Status** sub-field and selecting from the list of unique values in the selected **Status Indicator** field. If there are less than three unique values in the **Status Indicator** field (for example if there have been no dropouts i.e. zero rows with **Dropout Status**) then the user must manually enter the value they would like to have assigned for that **Status** outcome. Note that appropriate defaults will be provided for the rate of previously unused **Status** category in subsequent steps e.g. Dropout Hazard Rate(s) = 0.

The definition of **Time on Study** will depend on the **Status Indicator** for that subject. If the **Status Indicator** is equal to **Event Status** or **Dropout Status**,

the **Time on Study** is the length of time a subject was followed for after their **Arrival Time** before they had an event/dropped out. If the **Status Indicator** is equal to **Censored Status** , the **Time on Study** is the length of time a subject was followed for after their **Arrival Time** until the **Current Time.**

For this prediction, in the **Select Dataset** field select **SubjectData.csv**. Then in the **Arrival Time** field select **Arrival**, in the **Status Indicator** field select **Current** (assign 1 to **Event Status**, -1 to **Dropout Status**, 0 to **Censored Status**), in the **Time on Study** field select **Followup** and in the **Site ID** field select **SID.** These selections are shown in Figure 6.87. Select the Next or → button to move to the next **Setup** step.

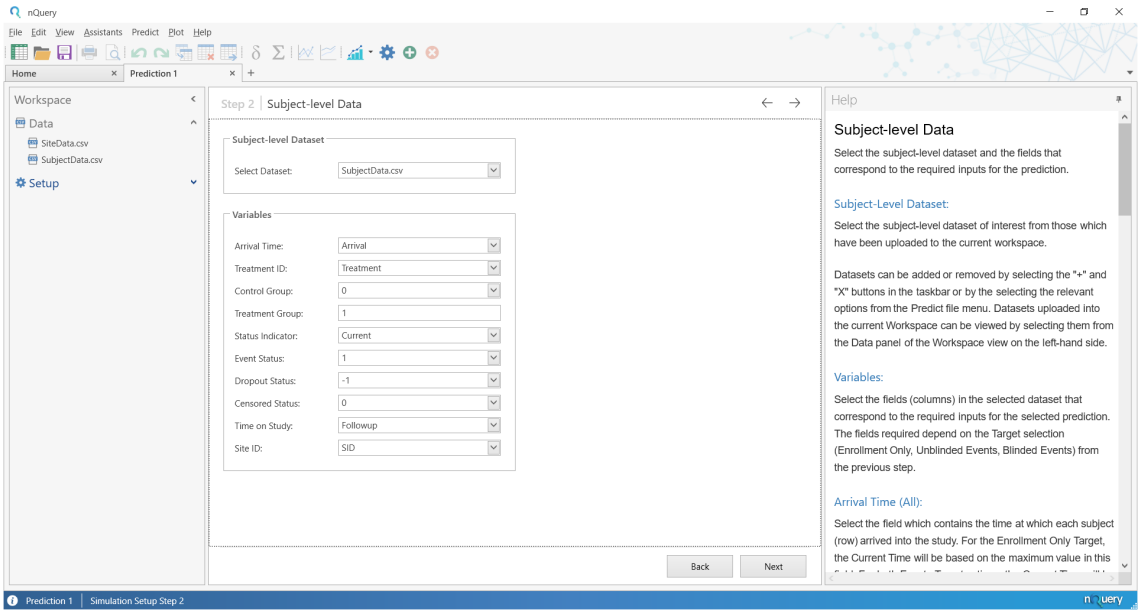



**Figure 6.87:** Subject-level Data Setup Step for Unblinded Events Prediction using Subject + Site-level Data

The next step is the **Site-level Data** step. The **Site-level Data** step is a Data Field Selection step as per the **Subject-level Data** step. The **Site-level Data** is the dataset where each row corresponds to a specific site.

The **Site-level Data** step fields are the same for all predictions. These fields are:

- Select Dataset: Select the dataset that is the Site-level Data from the datasets in the current workspace
- Site ID: Site ID for each site. Each row should have a distinct Site ID and Site IDs from Subject-level Data must be included in the list of Site IDs selected here
- Accrual Rate/Site: The planned accrual rate (subjects recruited per unit time) in a site. Each site's Accrual Rate/Site should be a numeric value greater than

zero

- Enrollment Cap: The maximum number of subjects that can be recruited from that site. Each site's Enrollment Cap should be an integer greater than the higher of zero and the No. of Accruals (see below)
- Site Initiation Time (Opened Sites): The time at which a site was open and able to start enrolling subjects. Each site's Site Initiation Time should be a numeric value less than the Current Time
- Start Time (Unopened Sites): The start time of the window during which an unopened site could open. Optional field for unopened sites, should be numeric value greater than or equal to the Current Time
- End Time (Unopened Sites): The start time of the window during which an unopened site could open. Optional field for unopened sites, should be numeric value greater than or equal to the Start Time (Unopened Sites) for that site

For this prediction, in the **Select Dataset** field select **SiteData.csv,** in the **Site ID** field select **SID**, in the **Accrual Rate/Site** field select **Rate**, in the **Enrollment Cap** field select **Cap**, in the **Site Initiation Time** field select **OpenTime**, in the **Start Time** field select **Start** and in the **End Time** field select **End.** These selections are shown in Figure 6.30. Select the Next or → button to move to the next **Setup** step.

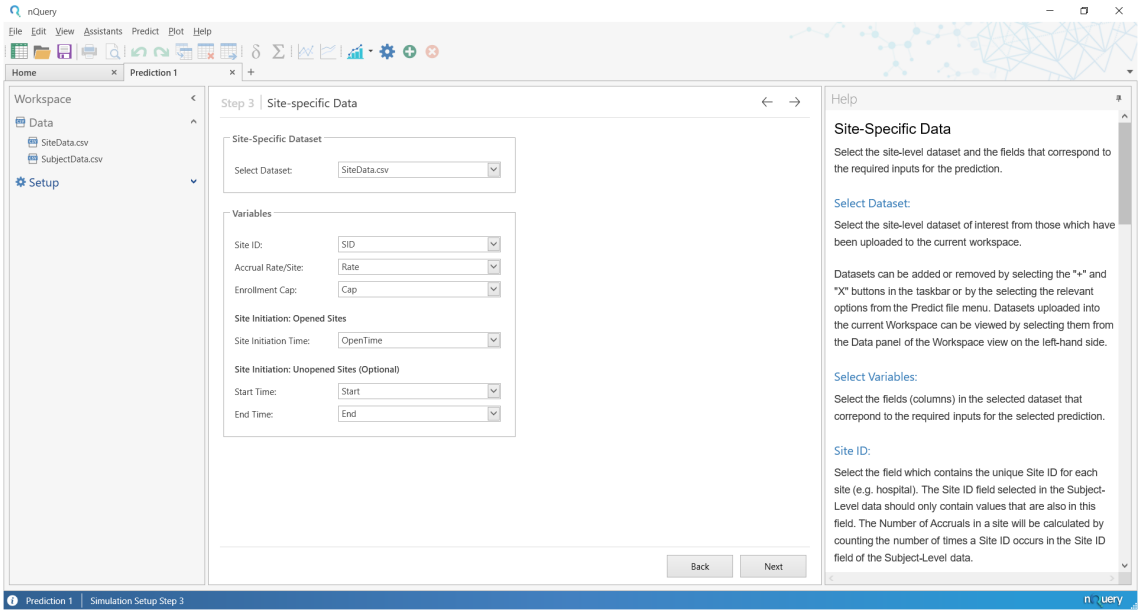



**Figure 6.88:** Site-level Data Setup Step

The next step is the **Accrual Options** step. When **Site-level Data** is used, the **Accrual Options** consists of two tabs: **Sample Size & Followup Options** and

**Accrual Options/Site**. A given tab can be opened by selecting the grey tab name at the top of the main window, just below the **Step 4 | Accrual Options** name. These tabs are highlighted in Figure 6.89.

The **Sample Size & Followup Options** tab provides overall information about the trial up until the current time alongside the **Target Sample Size** and **Accrual Model.** The **Sample Size & Followup Options** tab for this demonstration is shown in Figure 6.89.

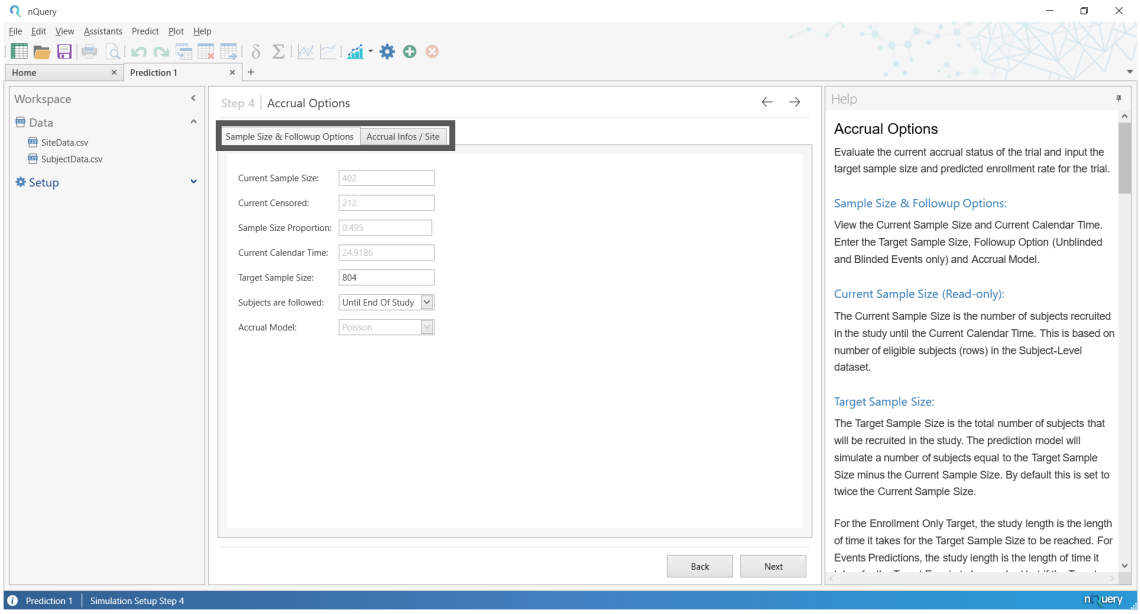



**Figure 6.89:** Accrual Options Step Sample Size & Followup Options tab for Site-level Data

The **Sample Size & Followup Options** fields are:

- Current Sample Size (Read-only): Sample Size recruited into a study so far. Equals the number of rows in the Subject-level Data
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Sample Size Proportion (Read-only): Proportion of subjects in the Control Group. Based on number of Control Group subjects in Treatment ID field. Controls the allocation proportion in simulated data
- Current Time (Read-only): The length of time that has passed since the study started. Equals the maximum Arrival Time value
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study. The number of simulated subjects equals the Total Sample Size minus the Current Sample Size.

- Accrual Model (Editable): The statistical model that will be used to generate simulated enrollments

In nQuery Predict, the **Target Sample Size** defaults to twice the **Current Sample Size** and the **Accrual Model** defaults to **Poisson.** In this demonstration, the **Current Sample Size** equals 402 and therefore the default **Target Sample Size** is $402\,(2) = 804$.

To open the **Accrual Options/Site** tab select the tab at the top of the main window with that name. This is shown in Figure 6.90

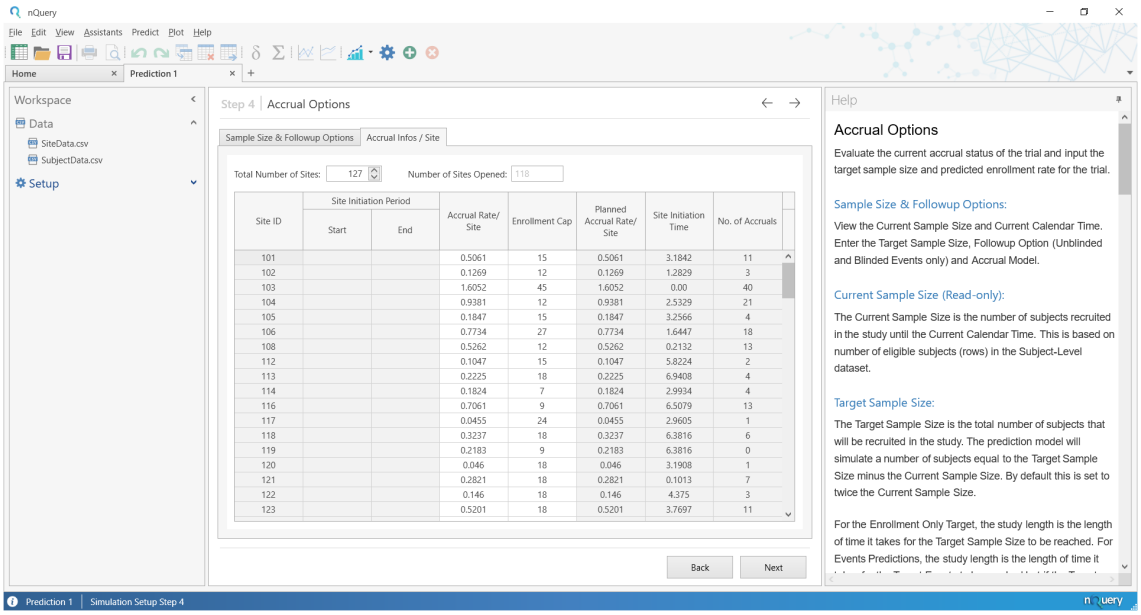



**Figure 6.90:** Accrual Options Step Accrual Options/Site tab for Site-level Data

The **Accrual Options/Site** tab has three main elements: the **Total Number of Sites** field, the **Number of Sites Open** field and the **Site Accrual Table**.

The **Total Number of Sites** and **Number of Sites Open** field are found at the top of tab window. The **Total Number of Sites** field sets the total number of sites that will be used to simulate subjects in this milestone prediction and by default is equal to the number of rows/sites in the **Site-level Data**.

The **Number of Sites Open** field is the number sites that are already open at the time the simulation is being performed and is the number of sites which had a **Site Initiation Time** in the site-level data. The **Total Number of Sites** must always be greater than or equal to the **Number of Sites Open**.

The **Site Accrual Table** consists of eight columns with each row corresponding to required information for each site. The columns are defined as follows:

- Site ID (Read-only for Open Sites, Editable for Unopened Sites): Site ID for a site. Taken directly from Site-level Data. All Site ID must be unique and include all Site ID in the subject-level Site ID field
- Site Initiation Period - Start Time (Editable for Unopened Sites, Numeric $\geq$ Current Time): Start time for window during which unopened site can open. Not used for open sites
- Site Initiation Period - End Time (Editable for Unopened Sites, Numeric $\geq$ Start Time): End time for window during which unopened site can open. Not used for open sites
- Accrual Rate/Site (Editable, Numeric $> 0$): The accrual rate (subjects recruited per unit time) that will be used to simulate subjects for that site
- Enrollment Cap (Editable, Numeric $\geq$ No. of Accruals) The maximum number of subjects that can be recruited in a site. This must be greater than the No. of Accruals in a site
- Planned Accrual Rate/Site (Read-only): The original planned accrual rate in a site. Taken from Site-level Data
- Site Initiation Time (Read-only): The site initiation time for opened sites. Taken from Site-level Data. Not used for unopened sites
- No. of Accruals (Read-only): The number of subjects that have been enrolled in an open site at the current time. Not used for unopened sites

**Site ID, Site Initiation Period (Start Time, End Time), Enrollment Cap, Planned Accrual Rate/Site** and **Site Initiation Time** are taken directly from the **Site-level Data**.

For sites which have already recruited subjects (i.e. **No. of Accruals** $> 0$), the default **Accrual Rate/Site** is calculated using the **No. of Accruals** and **Site Initiation Time** using the following relationship:

$$R = \frac{\#A}{(CT - SIT)}$$

where $R$ is the **Accrual Rate/Site**, $\#A$ is the **No. of Accruals** in a site, $CT$ is the **Current Time** and $SIT$ is the **Site Initiation Time** for that site.

For sites which are unopened sites or which have not recruited a subject (i.e. **No. of Accruals** $= 0$), the **Planned Accrual Rate/Site** from the **Site-level Data's Accrual Rate** field is used.

The **No. of Accruals** is calculated by counting the number of times each **Site ID** value occurred in the **Subject-level Data's Site ID** column.

As referenced above, the inputs for the **Site Accrual Table** for each site depend on whether a site is open or unopened. Unopened sites are automatically placed at the bottom of the **Site Accrual Table**.

The user can also add additional sites (rows) by increasing the **Total Number of Sites** field above the original value derived from the number of rows in the **Site-level Data**. This will add a blank row at the bottom of the **Site Accrual Table**. Note **Setup** cannot continue until all new rows are filled appropriately. An example of **Site Accrual Table** with the unopened sites from the **SiteLevel.csv** data alongside one blank new site by increasing **Total Number of Sites** to 128 is provided in Figure 6.91.

Sample Size & Followup Options | Accrual Infos / Site

Total Number of Sites: 128 Number of Sites Opened: 118

| Site ID | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap | Planned Accrual Rate/ Site | Site Initiation Time | No. of Accruals |
|---|---|---|---|---|---|---|---|
| | Start | End | | | | | |
| 607 | | | 0.3517 | 10 | 0.3517 | 7.8592 | 6 |
| 608 | | | 0.2815 | 9 | 0.2815 | 10.0658 | 0 |
| 609 | | | 2.2244 | 12 | 2.2244 | 19.375 | 0 |
| 610 | | | 0.1848 | 16 | 0.1848 | 8.6842 | 3 |
| 611 | | | 0.1739 | 9 | 0.1739 | 7.6645 | 3 |
| 620 | | | 0.1307 | 9 | 0.1307 | 9.2434 | 0 |
| 621 | | | 0.0567 | 18 | 0.0567 | 7.2697 | 1 |
| 622 | | | 0.052 | 18 | 0.052 | 5.6908 | 1 |
| 623 | | | 0.0517 | 10 | 0.0517 | 5.5592 | 1 |
| 144 | 25.51 | 26.11 | 0.50 | 14 | 0.50 | | |
| 145 | 25.51 | 26.11 | 0.50 | 14 | 0.50 | | |
| 347 | 25.00 | 25.30 | 0.10 | 6 | 0.10 | | |
| 368 | 25.00 | 25.75 | 0.20 | 12 | 0.20 | | |
| 369 | 25.00 | 25.75 | 0.20 | 12 | 0.20 | | |
| 420 | 25.10 | 25.20 | 0.30 | 16 | 0.30 | | |
| 423 | 25.10 | 25.20 | 0.15 | 8 | 0.15 | | |
| 443 | 25.00 | 25.50 | 0.12 | 8 | 0.12 | | |
| 543 | 25.00 | 26.00 | 0.50 | 24 | 0.50 | | |

**Figure 6.91:** Unopened and New Sites in Site Accrual Table

For this demonstration, assume the **Total Number of Sites** is reset to its default of 127. Select the Next or → button to move to the next **Setup** step.

The next step is the **Event and Dropout Information** step which specifies the **Target Number of Events,** event model (and associated statistical parameters) and dropout model (and associated statistical parameters). The **Event and Dropout Information** step on the default **Events Model** tab is shown in Figure 6.92.

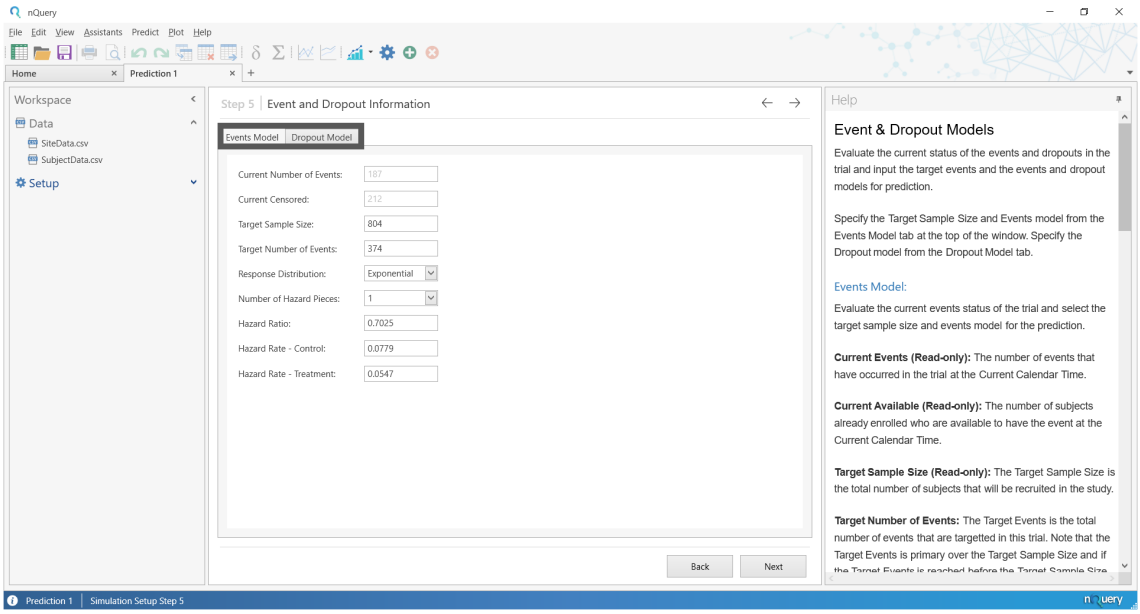



**Figure 6.92:** Event and Dropout Information Setup Step for Unblinded Events Prediction using Subject + Site-level Data

The **Event and Dropout Information** step consists of two tabs: the **Events Model** tab and the **Dropout Model** tab. These tabs have been highlighted in Figure 6.92 and are opened by selecting the required tab name at the top of the main window, below the step title. For **Unblinded Events**, it is assumed that survival and dropout times will be simulated individually for each treatment group using their own user specified model.

The **Events Model** tab provides fields which provide information on the current number of subject who have had an event, could have an event going forward and allows the input of the **Target Number of Events, Target Sample Size, Response Distribution** for the events model and additional parameters needed for that specific model. The following fields are provided in this prediction:

- Current Events (Read-only): The number of subjects who had the event of interest at the current time. Equals the number of subjects with the Event Status in the Status Indicator field
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study. The number of simulated subjects equals the Total Sample Size minus the Current Sample Size.
- Target Number of Events (Editable): The required number of subjects needed to have the event of interest for the study to end

- Response Distribution: The statistical model that will be used to generate survival times for subjects available to have the event
- Number of Hazard Pieces (Response Distribution = Exponential only): Number of hazard pieces used to specify the piecewise exponential model. Select value between 1 and 10
- Hazard Ratio (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The hazard ratio between the Treatment and Control Hazard Rates
- Hazard Rate - Control (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The control group's exponential hazard rate for the constant exponential survival model specified above
- Hazard Rate - Treatment (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The treatment group's exponential hazard rate for the constant exponential survival model specified above

In nQuery Predict, the **Target Sample Size** at this step is inherited from the prior **Accrual Options** step, the **Target Number of Events** defaults to twice the **Current Events**, the default **Response Distribution** is **Exponential** with the **Number of Hazard Pieces** set to 1 and the default **Hazard Ratio** and **Hazard Rate - Control/Treatment** set using the provided information (see subsubsection 6.1.2.2 for details on the survival distributions used in nQuery Predict, see section 3.5 for details on how to calculate a exponential **Hazard Rate** using a median survival or survival probability)**.** When one of the **Hazard Ratio, Hazard Rate - Control** and **Hazard Rate - Treatment** is changed (both for **Number of Hazard Pieces = 1** and **>1** cases, for latter see below) one of the other parameters will be auto-calculated to ensure consistency (**Hazard Rate - Treatment** changes if **Hazard Ratio** or **Hazard Rate - Control** edited, **Hazard Rate - Control** changes if **Hazard Rate - Treatment** edited).

In this demonstration, the **Current Number of Events** is 187 and therefore the default **Target Number of Events** is $187\,(2) = 374$ giving an estimated exponential **Hazard Ratio** of 0.7025, **Hazard Rate - Control** of 0.0779 and **Hazard Rate - Treatment** of 0.0547. Note that if the **Target Number of Events** is reached before the **Target Sample Size** has been enrolled then the it is assumed the enrollment process will stop early and the total sample size achieved may be less than the **Target Sample Size** specified.

In nQuery, there are effectively three primary survival distributions that survival times can be generated from:

- Constant Exponential (Response Distribution = Exponential, Number of Hazard Pieces =1): Input Hazard Rate
- Piecewise Exponential (Response Distribution = Exponential, Number of Hazard Pieces > 1): Input table of Starting at Time and Hazard Rate columns

- Weibull (Response Distribution = Weibull): Input Scale Parameter and Shape Parameter

Details on how survival times are generated from each of these distributions is provided in subsubsection 6.1.2.2, alongside how the default values provided in nQuery Predict are estimated from the provided information. The change in the **Events Model** tab for each of these choices is shown in Figure 6.93.

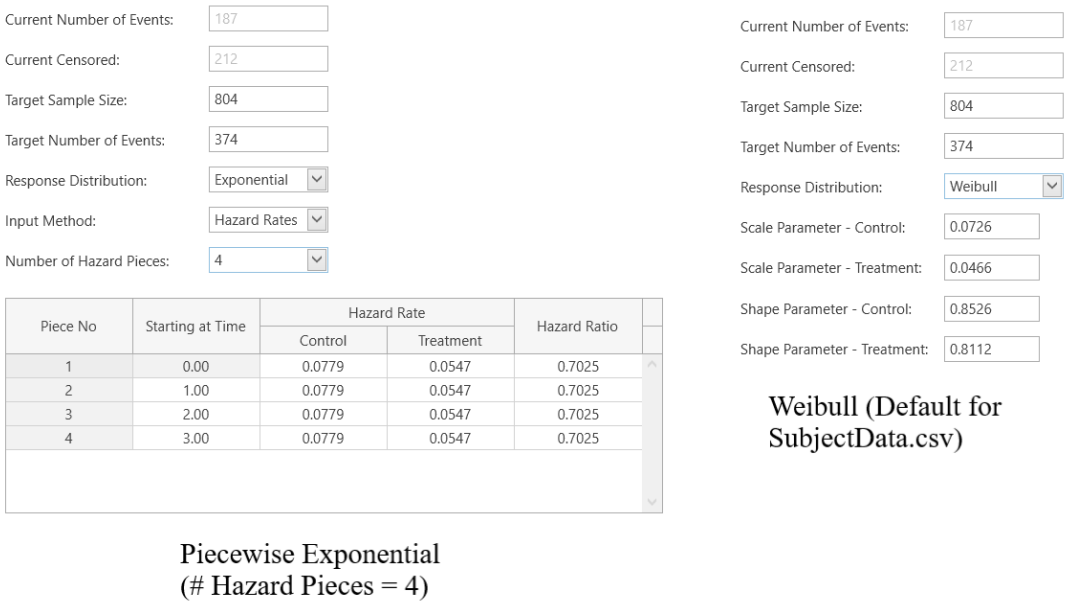



**Figure 6.93:** Piecewise Exponential (4 Pieces) and Weibull Defaults for SubjectData.csv

The piecewise exponential table will have a number of rows equal to **Number of Hazard Pieces.** The **Piece No** column indicates the hazard piece number of a row. The **Starting at Time** column sets the start time for when the **Hazard Rate** value in that row takes effect. The **Hazard Rate** columns sets the exponential hazard rate for the **Control** group and **Treatment** group in effect from the start time of the current hazard piece to the start time in the next row. The **Hazard Ratio** row is also available for each period and changing either **Hazard Rate** or the **Hazard Ratio** will automatically update a cell in that row to ensure consistency as for the **Number of Hazard Pieces = 1** case.

See section 3.5 for details on how to calculate a exponential **Hazard Rate** using a median survival or survival probability and the Compute Effect Size side table for STT3 may also be useful. By default the piecewise exponential assumes constant control and treatment exponential hazard rates (same **Hazard Rate** as for the **Number of Hazard Pieces** = 1 case) with **Starting at Time** being values increasing by 1 up to the final row.

The Weibull distribution has four parameters: the Weibull **Scale Parameter - Control,** Weibull **Shape Parameter - Treatment**, the Weibull **Shape Parameter - Control** and Weibull **Shape Parameter - Treatment**. **Control/Treatment** indicates the treatment group the Weibull parameter is associated with. The default for these is found by fitting the survival data from the subject-level data (see subsubsection 6.1.2.2 for details).

The **Dropout Model** tab is opened by selecting the **Dropout Model** tab name at the top of the main window. For this demonstration, the **Dropout Model** tab is shown in Figure 6.94.

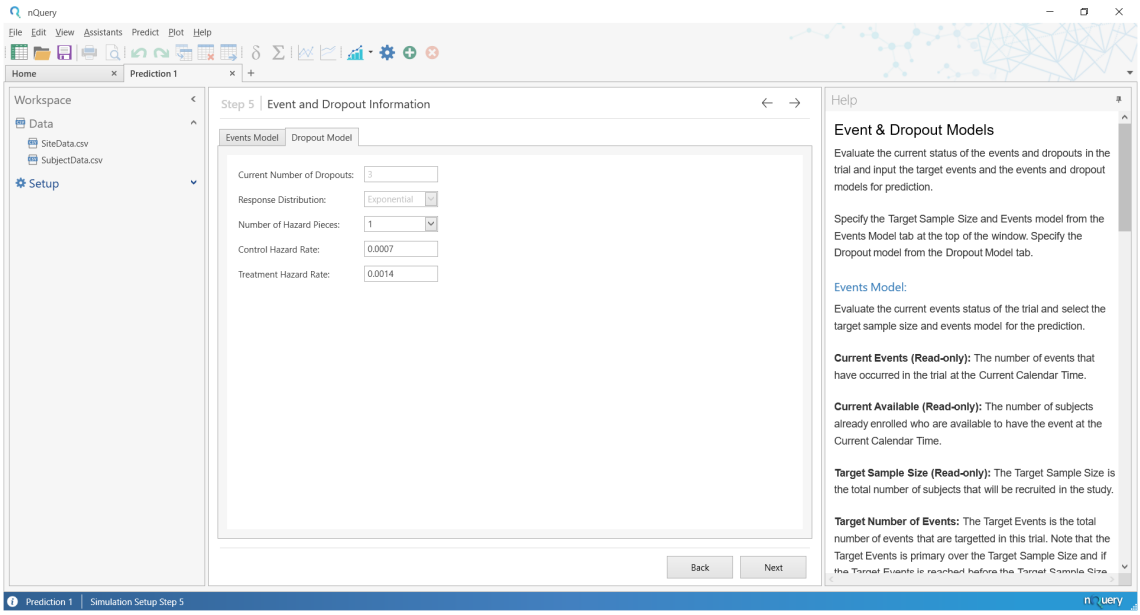



**Figure 6.94:** Dropout Model tab for Unblinded Events with Subject + Site-level Data

For **Dropout Model**, the **Current Number of Dropouts** (number of subjects who have dropped out up to the **Current Time**), **Response Distribution** and **Number of Hazard Pieces** are available. For dropout, the only **Response Distribution** option is **Exponential**. The inputs and defaults for dropout **Exponential Response Distribution**, for both the case where **Hazard Pieces = 1** (Constant Exponential) or **Hazard Pieces > 1** (Piecewise Exponential), are the same as from the **Events Model** described above (with the exception of removing the **Hazard Ratio** field) and therefore will not be replicated here. In this demonstration, the constant exponential **Control Hazard Rate** and **Treatment Hazard Rate** for dropout calculated from the information provided equals 0.0007 and 0.0014 respectively.

In this demonstration the default for all inputs (**Target Sample Size = 804, Target Number of Events = 374, Events Model Exponential Control Hazard Rate = 0.0779, Events Model Exponential Treatment Hazard Rate**

**= 0.0547, Dropout Model Exponential Control Hazard Rate = 0.0007, Dropout Model Exponential Treatment Hazard Rate = 0.0014**) are used. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.95. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

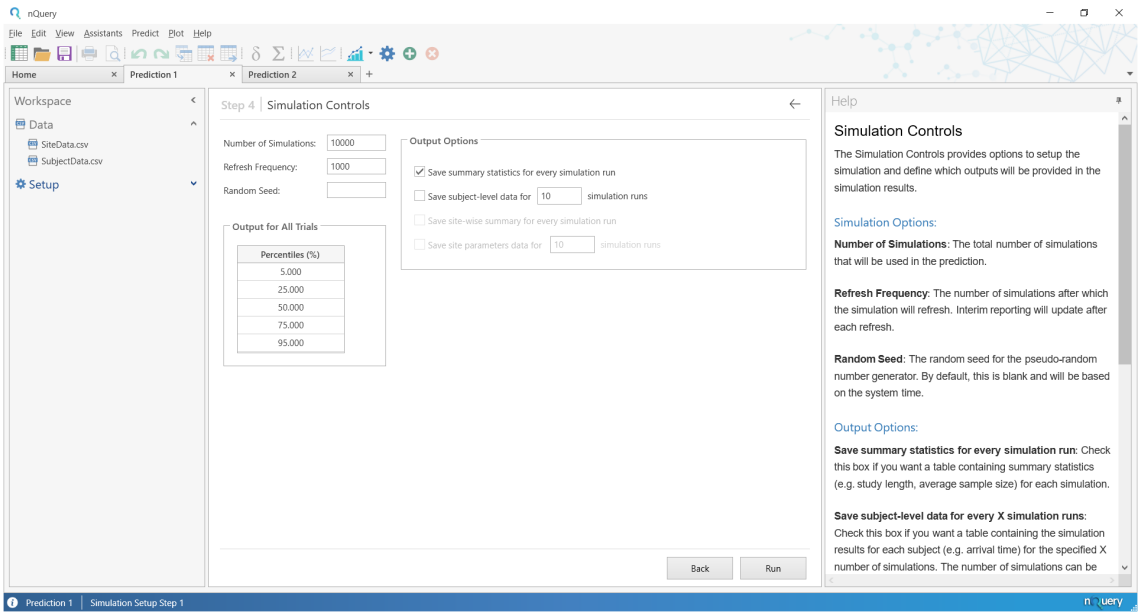



**Figure 6.95:** Simulation Controls Setup Step for Unblinded Events Prediction using Subject + Site-level Data

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.96.

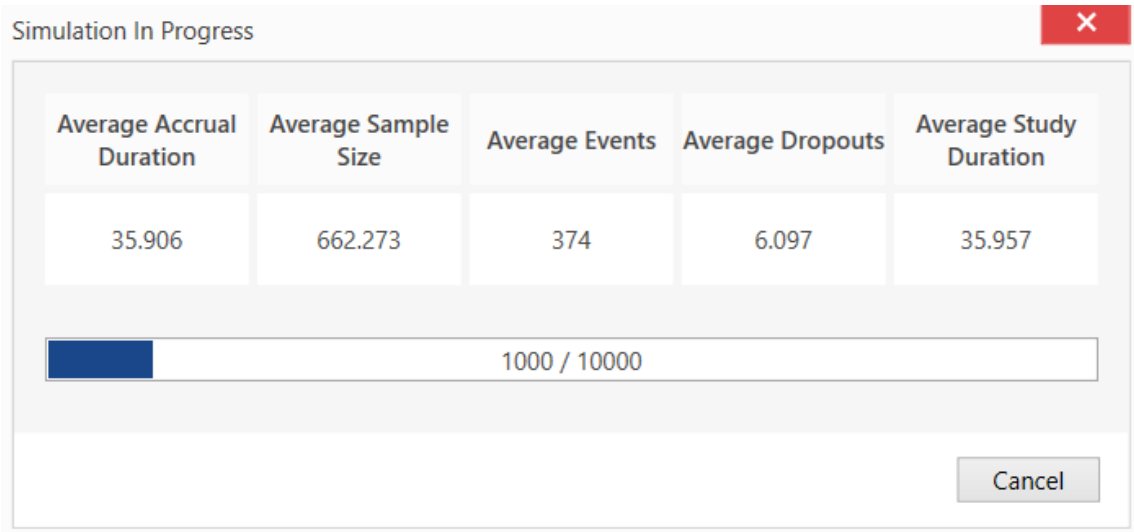



**Figure 6.96:** Simulation in Progress Window for Unblinded Events Prediction using Subject + Site-level Data

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.97.

**Figure 6.97:** Enrollment Prediction Plot for Unblinded Events Prediction using Subject + Site-level Data

The results of a nQuery Predict milestone prediction can be divided into three categories: **Reports, Plots** and **Tables**. A high-level summary of the results in nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table, Events Prediction Table, Dropout Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.98.

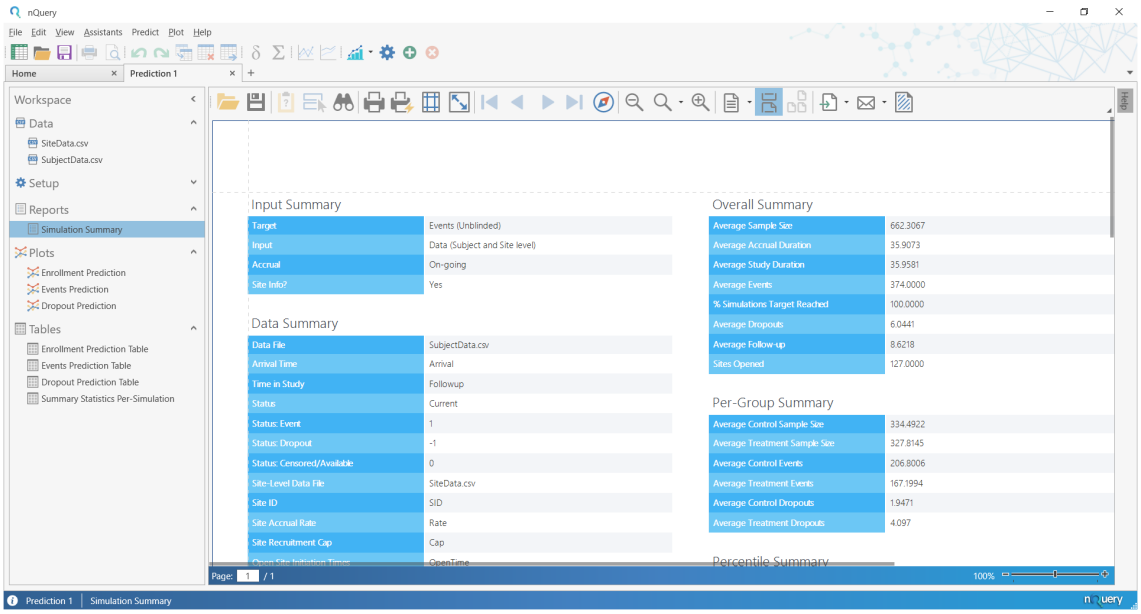



**Figure 6.98:** Simulation Summary Report for Unblinded Events Prediction using Subject + Site-level Data

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.99. For this demonstration, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 35.9581. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 34.9694 (5% Percentile) and 36.9366 (95% Percentile), 50% a duration between 35.4936 (25% Percentile) and 36.3094 (75% Percentile) and a median (50% Percentile) duration of 35.8942 The **Average Dropouts** was 6.0441 (Percentiles: 4, 5, 6, 7, 9) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 8.6218.

The accrual process in this study ended early for all simulations with an **Average Sample Size** of 662.3067 (Percentiles: 638, 652, 662, 672, 687) enrolled before the **Target Number of Events** of 804 was reached. This meant the **Average Accrual Duration** of 38.9620 (Percentiles: 34.9164, 35.4936, 35.8942, 36.3094, 36.9366) is effectively equal to the **Average Study Duration** of 35.9073 with the accrual duration effectively the time of the last enrollment before study end.

For **Unblinded Events,** a **Per-Group Summary** which provides the average sample size, events and dropouts for each group. The **Control Group** had an **Average Sample Size** 334.4922, **Average Events** of 206.8006 and **Dropouts** of

4.097. The **Treatment Group** had an **Average Sample Size** 327.8145, **Average Events** of 167.1994 and **Dropouts** of 4.097.

Overall Summary

| | |
|---|---|
| Average Sample Size | 662.3067 |
| Average Accrual Duration | 35.9073 |
| Average Study Duration | 35.9581 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.0441 |
| Average Follow-up | 8.6218 |
| Sites Opened | 127.0000 |

Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 334.4922 |
| Average Treatment Sample Size | 327.8145 |
| Average Control Events | 206.8006 |
| Average Treatment Events | 167.1994 |
| Average Control Dropouts | 1.9471 |
| Average Treatment Dropouts | 4.097 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration | Sites Opened |
|---|---|---|---|---|---|---|
| 5.00% | 374 | 638 | 4 | 34.9164 | 34.9694 | 127 |
| 25.00% | 374 | 652 | 5 | 35.4936 | 35.5429 | 127 |
| 50.00% | 374 | 662 | 6 | 35.8942 | 35.9449 | 127 |
| 75.00% | 374 | 672 | 7 | 36.3094 | 36.3630 | 127 |
| 95.00% | 374 | 687 | 9 | 36.9366 | 36.9843 | 127 |

**Figure 6.99:** Simulation Summary Report Results for Unblinded Events Prediction

To edit the **Simulation Summary report there a** number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment simulation from the study start to when the **Target Sample Size** was reached in all

simulations. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.100.

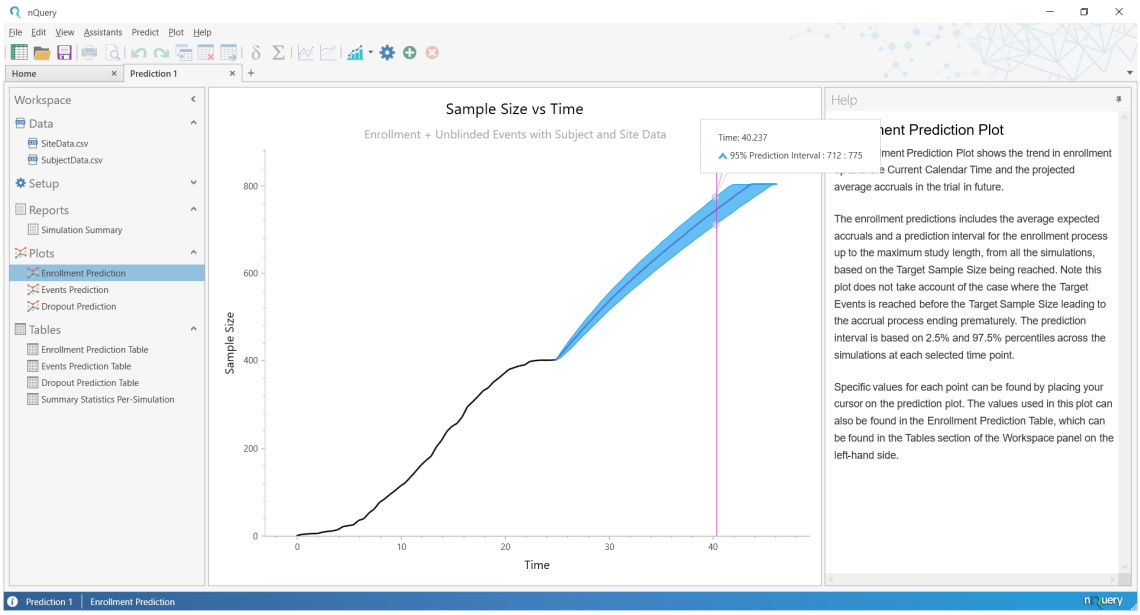



**Figure 6.100:** Enrollment Prediction Plot for Unblinded Events Prediction using Subject + Site-level Data

The **Enrollment Prediction** plot consists of two parts: from Time = 0 to **Current Time** and from **Current Time** to the **Maximum Time**. The first part is a black line which plots the accrual pattern from the user inputs provided. The second part is a blue conic section with a grey line where the blue conic section plots the 95% prediction interval for the **Sample Size** enrolled at a given **Time** across all simulations and the grey line is the average **Sample Size** enrolled at a given time. The 95% prediction interval indicates the 95% of predictions had a sample size that fell between the lower and upper limit at a given time.

Note that the **Enrollment Prediction** plot will show the projection for **Target Sample Size** and its associated **Accrual Duration** and therefore will usually have a different **Time** scale than the **Event Prediction** and **Dropout Prediction** plots which will be associated with the **Study Duration.** Note the **Enrollment Prediction** plot ignores the case where **Accrual Duration** is cut short due to the **Target Number of Events** being reached before the enrollment process is complete.

The **Events Prediction** plot is very similar to the **Enrollment Prediction** plot where the first part is a black line which plots the events pattern from the user inputs provided and the second part is a yellow conic section with a grey line where the

yellow conic section plots the 95% prediction interval for the **Number of Events** that have occurred at a given **Time** across all simulations and the grey line is the average **Number of Events** at a given time. The 95% prediction interval indicates the 95% of predictions had a number of events that fell between the lower and upper limit at a given time. The **Events Prediction** plot for this demonstration is shown in Figure 6.101.

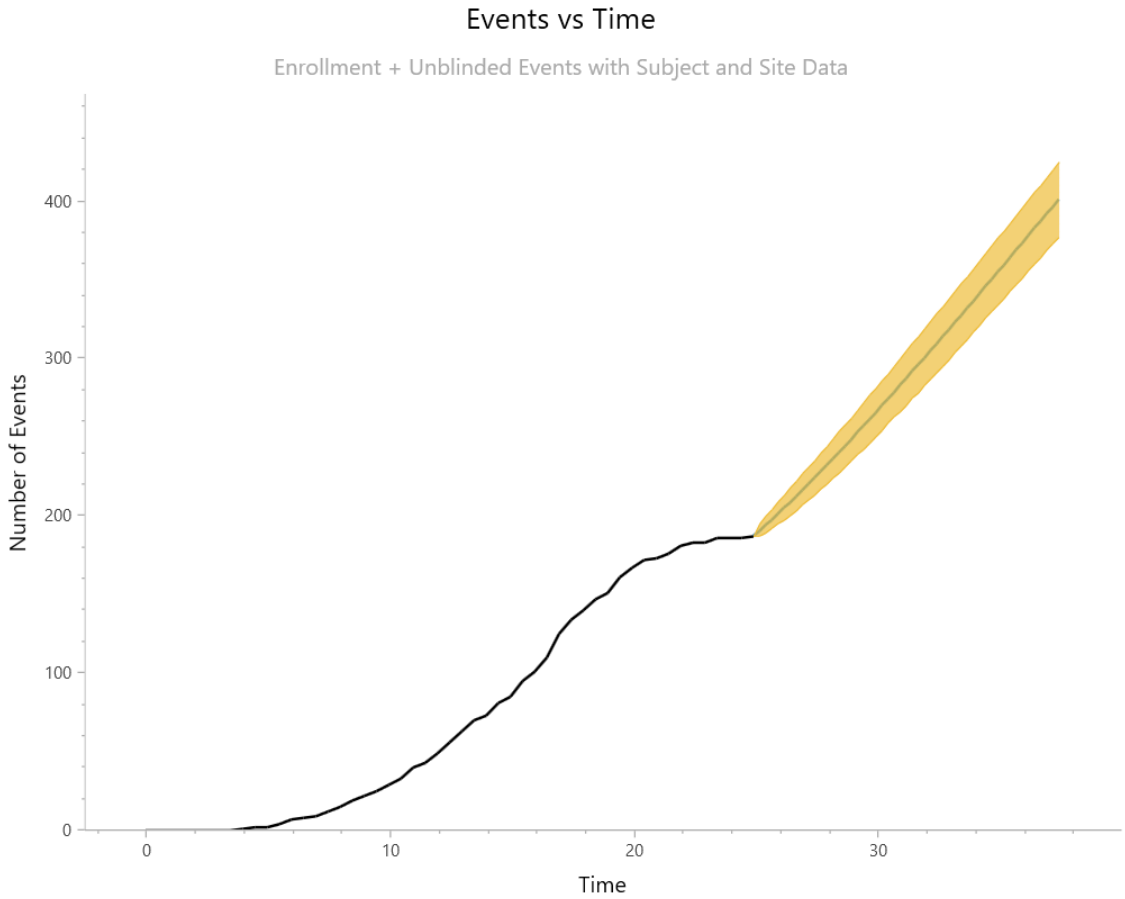



**Figure 6.101:** Events Prediction Plot for Unblinded Events Prediction using Subject + Site-level Data

It is important the note that these event times are when these events occurred relative to the start of the study, not how long they happened after follow up i.e. this is not a Kaplan-Meier type plot!

The **Dropout Prediction** plot works effectively the same as the **Events Prediction** plot except for the dropout process. The **Dropout Prediction** plot for this demonstration is shown in Figure 6.102.

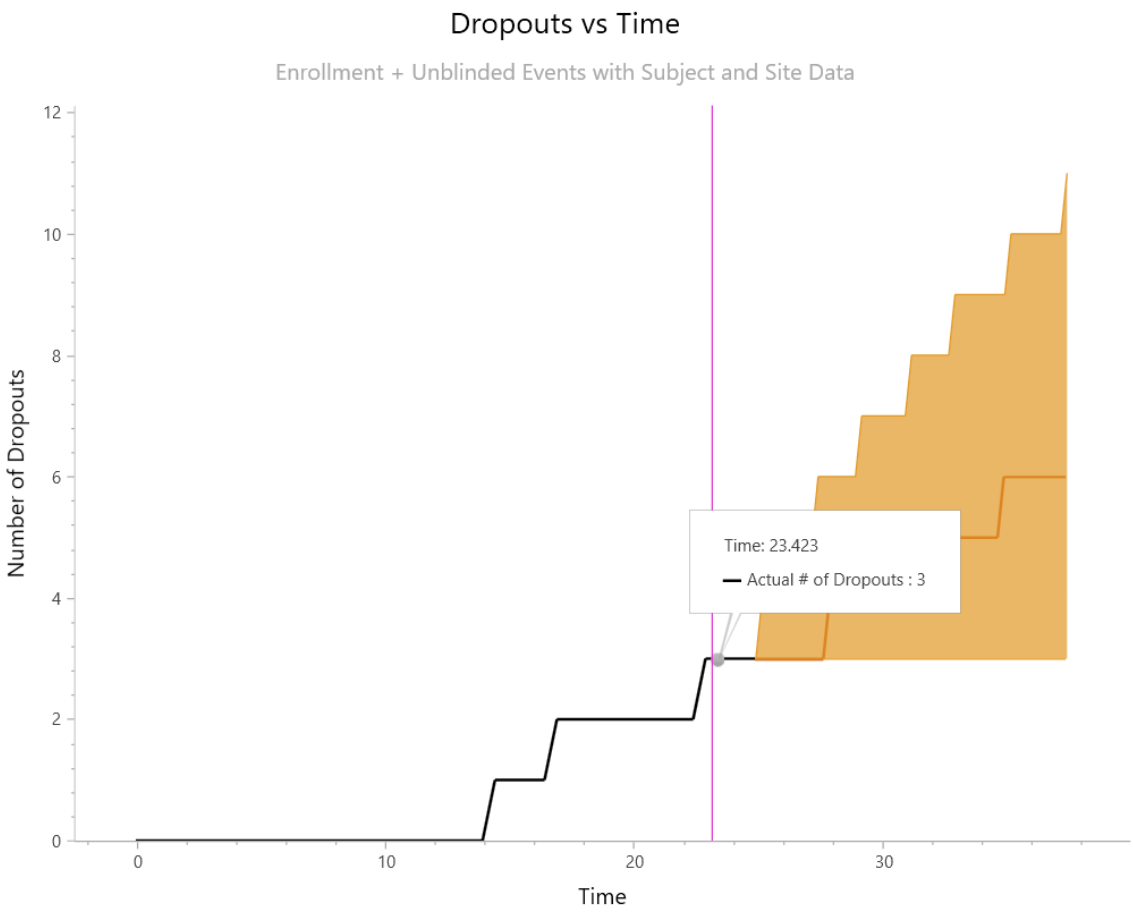
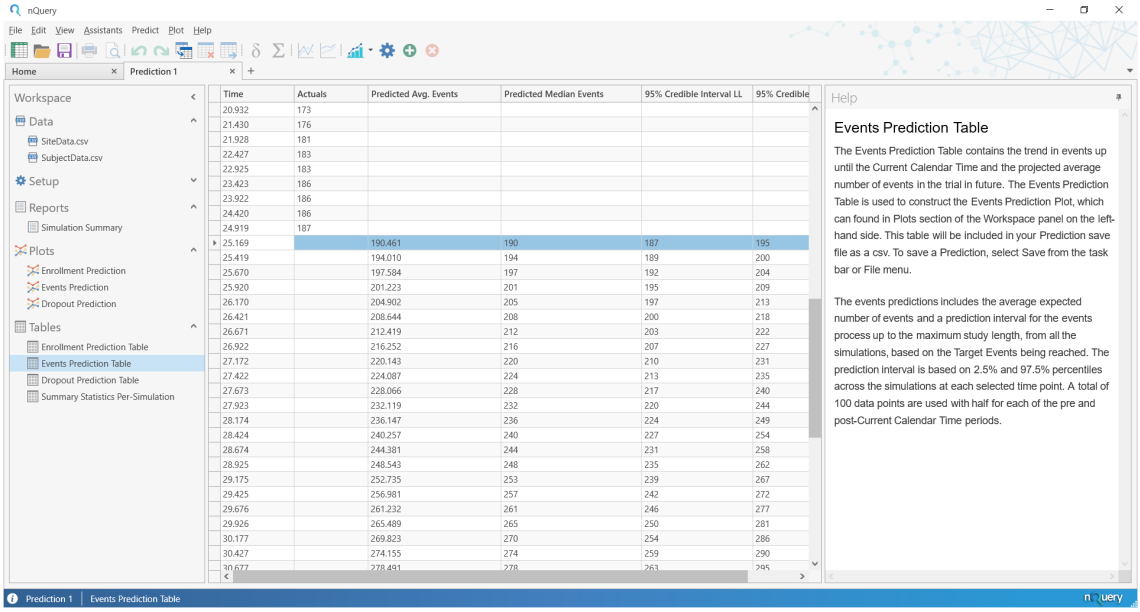



**Figure 6.102:** Dropout Prediction Plot for Unblinded Events Prediction using Subject + Site-level Data

The **Enrollment Prediction Table, Events Prediction Table and Dropout Prediction Tables** are the tables used to create the respective **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plots, which by default all consists of ~100 rows (50 pre-simulation and 50 post-simulation). Each consists of five columns: **Time, Actual, Predicted Avg. Sample Size/Events/Dropouts, Predicted Median Sample Size/Events/Dropouts, 95% Prediction Interval LL, 95% Prediction Interval UL**. One prediction table (the **Events Prediction Table**) for this demonstration at the start of the simulated period is given in Figure 6.103.




| Time | Actuals | Predicted Avg. Events | Predicted Median Events | 95% Credible Interval LL | 95% Credible |
|---|---|---|---|---|---|
| 20.932 | 173 | | | | |
| 21.430 | 176 | | | | |
| 21.928 | 181 | | | | |
| 22.427 | 183 | | | | |
| 22.925 | 183 | | | | |
| 23.423 | 186 | | | | |
| 23.922 | 186 | | | | |
| 24.420 | 186 | | | | |
| 24.919 | 187 | | | | |
| 25.169 | | 190.461 | 190 | 187 | 195 |
| 25.419 | | 194.010 | 194 | 189 | 200 |
| 25.670 | | 197.584 | 197 | 192 | 204 |
| 25.920 | | 201.223 | 201 | 195 | 209 |
| 26.170 | | 204.902 | 205 | 197 | 213 |
| 26.421 | | 208.644 | 208 | 200 | 218 |
| 26.671 | | 212.419 | 212 | 203 | 222 |
| 26.922 | | 216.252 | 216 | 207 | 227 |
| 27.172 | | 220.143 | 220 | 210 | 231 |
| 27.422 | | 224.087 | 224 | 213 | 235 |
| 27.673 | | 228.066 | 228 | 217 | 240 |
| 27.923 | | 232.119 | 232 | 220 | 244 |
| 28.174 | | 236.147 | 236 | 224 | 249 |
| 28.424 | | 240.257 | 240 | 227 | 254 |
| 28.674 | | 244.381 | 244 | 231 | 258 |
| 28.925 | | 248.543 | 248 | 235 | 262 |
| 29.175 | | 252.735 | 253 | 239 | 267 |
| 29.425 | | 256.981 | 257 | 242 | 272 |
| 29.676 | | 261.232 | 261 | 246 | 277 |
| 29.926 | | 265.489 | 265 | 250 | 281 |
| 30.177 | | 269.823 | 270 | 254 | 286 |
| 30.427 | | 274.155 | 274 | 259 | 290 |
| 30.677 | | 278.491 | 278 | 263 | 295 |

**Figure 6.103:** Events Prediction Table for Unblinded Events Prediction using Subject + Site-level Data

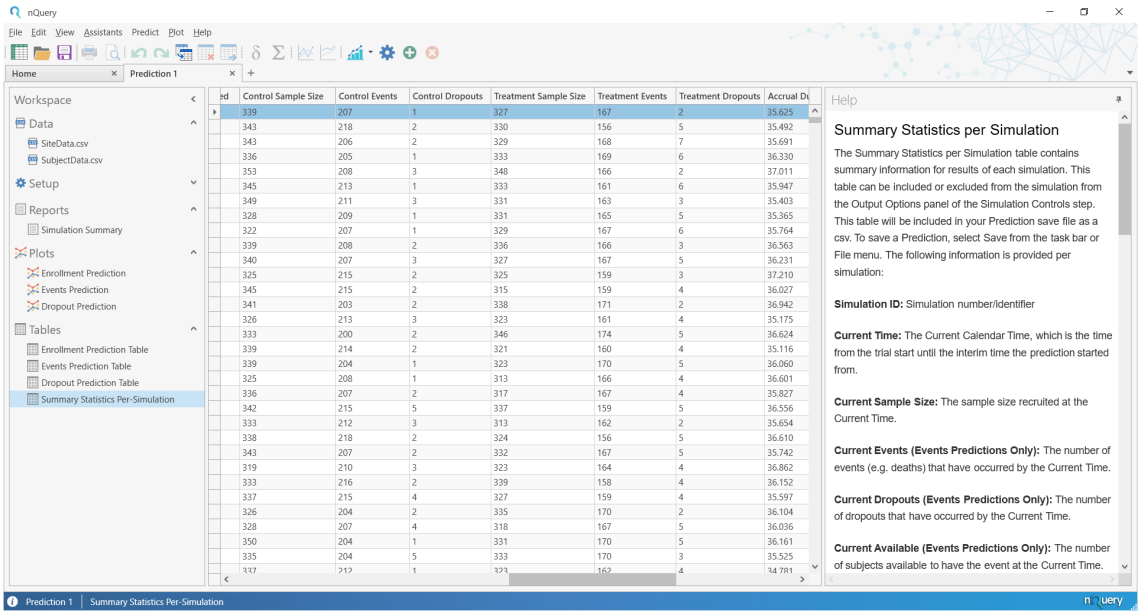
**Time** is the time value relative to the study start time of zero. **Actual** is the actual number of events achieved at a pre-simulation **Time** row. **Predicted Avg. Events** is the predicted average (mean) number of events at a post-simulation **Time** row. **Predicted Median Events** is the predicted median number of events at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the number of events at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the number of events at a post-simulation **Time** row.

Note that the **Enrollment Prediction Table** will show the projection for **Target Sample Size** and its associated **Accrual Duration** and therefore will usually have a different **Time** scale than the **Event Prediction Table** and **Dropout Prediction Table** which will be associated with the **Study Duration.** Note the **Enrollment Prediction Table** ignores the case where **Accrual Duration** is cut short due to the **Target Number of Events** being reached before the enrollment process is complete.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Current Events, Current Dropout, Current Available, Target Sample Size, Target Events, Study End Time, Total Sample Size, Total Events, Total Dropouts, Total Censored, Control Sample Size, Control Events, Control Dropouts, Treatment Sample Size, Treatment Events, Treatment Dropouts, Accrual Duration, Accrual Duration Uncensored, Average Fol-**

**lowup, Median Followup, Target Events Reached.**

The **Summary Statistics Per-Simulation Table** for this demonstration, focused on the per-group simulation results, is shown in Figure 6.104.




| Control Sample Size | Control Events | Control Dropouts | Treatment Sample Size | Treatment Events | Treatment Dropouts | Accrual D |
|---|---|---|---|---|---|---|
| 339 | 207 | 1 | 327 | 167 | 2 | 35.625 |
| 343 | 218 | 2 | 330 | 156 | 5 | 35.492 |
| 343 | 206 | 2 | 329 | 168 | 7 | 35.691 |
| 336 | 205 | 1 | 333 | 169 | 6 | 36.330 |
| 353 | 208 | 3 | 348 | 166 | 2 | 37.011 |
| 345 | 213 | 1 | 333 | 161 | 6 | 35.947 |
| 349 | 211 | 3 | 331 | 163 | 3 | 35.403 |
| 328 | 209 | 1 | 331 | 165 | 5 | 35.365 |
| 322 | 207 | 1 | 329 | 167 | 6 | 35.764 |
| 339 | 208 | 2 | 336 | 166 | 3 | 36.563 |
| 340 | 207 | 3 | 327 | 167 | 5 | 36.231 |
| 325 | 215 | 2 | 325 | 159 | 3 | 37.210 |
| 345 | 215 | 2 | 315 | 159 | 4 | 36.027 |
| 341 | 203 | 2 | 338 | 171 | 2 | 36.942 |
| 326 | 213 | 3 | 323 | 161 | 4 | 35.175 |
| 333 | 200 | 2 | 346 | 174 | 5 | 36.624 |
| 339 | 214 | 2 | 321 | 160 | 4 | 35.116 |
| 339 | 204 | 1 | 323 | 170 | 5 | 36.060 |
| 325 | 208 | 1 | 313 | 166 | 4 | 36.601 |
| 336 | 207 | 2 | 317 | 167 | 4 | 35.827 |
| 342 | 215 | 5 | 337 | 159 | 5 | 36.556 |
| 333 | 212 | 3 | 313 | 162 | 2 | 35.654 |
| 338 | 218 | 2 | 324 | 156 | 5 | 36.610 |
| 343 | 207 | 2 | 332 | 167 | 5 | 35.742 |
| 319 | 210 | 3 | 323 | 164 | 4 | 36.862 |
| 333 | 216 | 2 | 339 | 158 | 4 | 36.152 |
| 337 | 215 | 4 | 327 | 159 | 4 | 35.597 |
| 326 | 204 | 2 | 335 | 170 | 2 | 36.104 |
| 328 | 207 | 4 | 318 | 167 | 5 | 36.036 |
| 350 | 204 | 1 | 331 | 170 | 5 | 36.161 |
| 335 | 204 | 5 | 333 | 170 | 3 | 35.525 |

**Figure 6.104:** Summary Statistics Per-Simulation Table for Unblinded Events Prediction using Subject + Site-level Data

Details on each field are provided in the Help window on the right but broadly the "Current" fields described the information provided up to the **Current Time**, the "Target" fields describe the study targets, the "Total" fields (and final Accrual, Followup and Target Event Reached columns) give the projection achieved for key parameters in total, the "Control" fields give the key results in the **Control Group** and the "Treatment" fields give the results in the **Treatment Group.**

**Alternative Scenarios** The demonstration has focused on the default inputs using **SubjectData.csv** up to this point. However, as noted in the data summary given in subsection 6.2.4 the study from which the example data was simulated had an original sample size target of 460 (402 enrolled in **SubjectData.csv**).

However, note that the percentage of the **Target Sample Size** enrolled $(100(402/406) = 87.5)$ at the **Current Time** of 24.9186 months is higher than the percentage of the targetted accrual period length (30 months) used so far $(100\,(24.9186/30) = 83.1\%)$. Based on this, it could be decided to discontinue the opening of new sites and to close the "worst" performing site in terms of recruitment. Additionally, there is some evidence of deviations from the constant exponential model and the more flexible nQuery Predict default Weibull model is used.

To explore this additional information, we will investigate two additional scenarios

- Scenario A: **Target Sample Size = 460** using defaults for each site
- Scenario B: **Target Sample Size = 460** with no unopened sites added, "worst" site closed and **Weibull** model

**Scenario A** The change required compared to the first demonstration is to change the **Target Sample Size** to 460 in the **Sample Size & Followup Options** tab of the **Accrual Options** step. The simplest way to do this is to select the **Setup** header in the Workspace Navigation Bar on the left and then select the **Accrual Options** drop down option. Then edit the **Target Sample Size** 460 in the **Sample Size & Followup Options tab** in the top-left and select the Next or → button.

At the **Simulation Controls** step, we will add the three additional tables under the **Output Options** field by selecting the checkbox beside each. We will replace the default of 10 with 1 (at "(x)" spot) for **Save subject-level data for (x) simulation runs** and **Save site parameters data for (x) simulation runs**. An edited down summary of the **Setup** changes is provided in Figure 6.105.

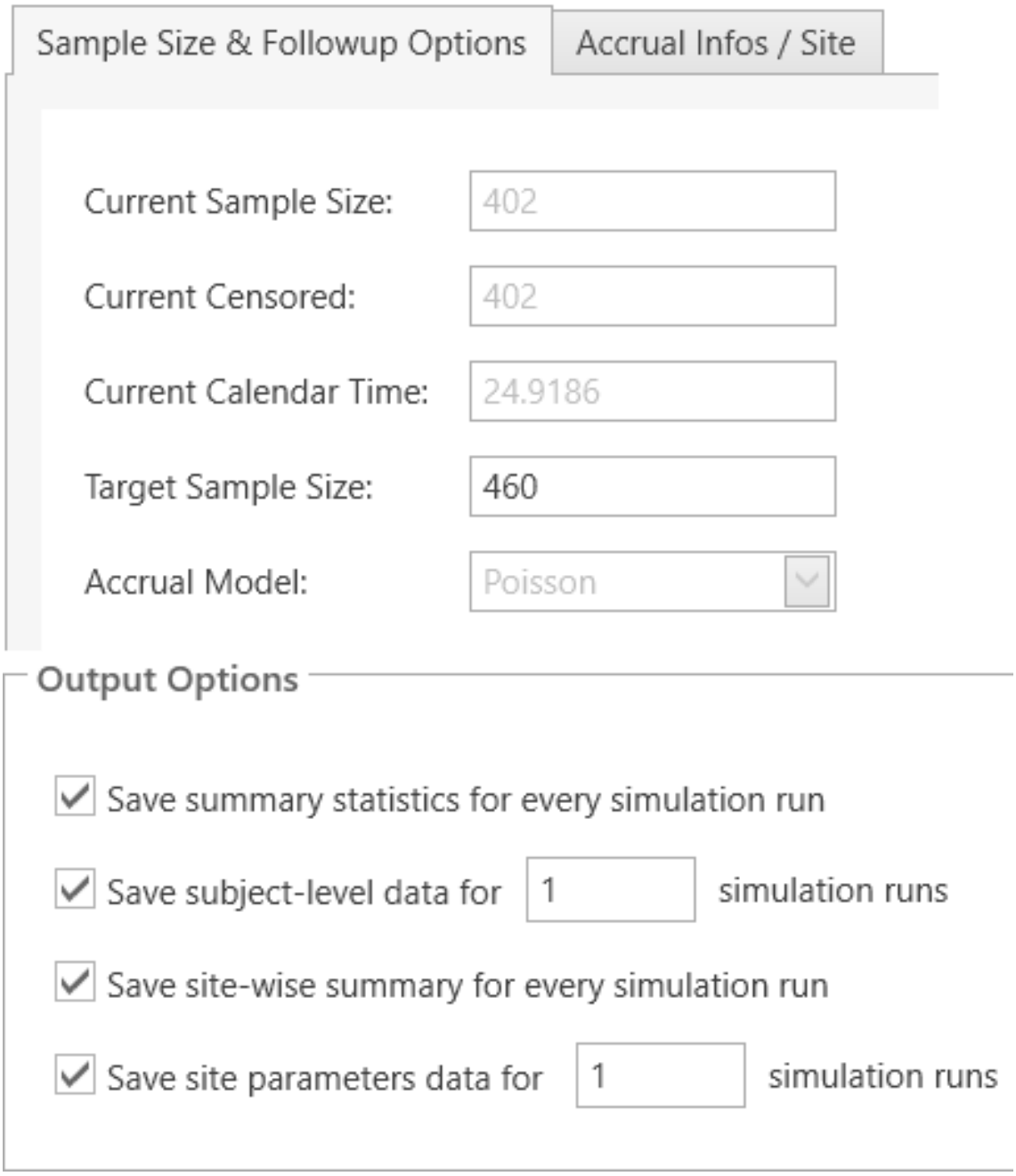



**Figure 6.105:** Scenario A Setup Changes for Unblinded Events Prediction using Subject + Site-level Data

To run the simulation again with the new Scenario A, select the Run/→ button. The main results of this Scenario A simulation are shown in Figure 6.106.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 27.0596 |
| Average Study Duration | 43.7440 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.1895 |
| Average Follow-up | 12.4849 |
| Sites Opened | 127.0000 |

Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 232 |
| Average Treatment Sample Size | 228 |
| Average Control Events | 201.1749 |
| Average Treatment Events | 172.8251 |
| Average Control Dropouts | 1.9175 |
| Average Treatment Dropouts | 4.272 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration | Sites Opened |
|---|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 4 | 26.6271 | 41.3850 | 127 |
| 25.00% | 374 | 460 | 5 | 26.8709 | 42.7187 | 127 |
| 50.00% | 374 | 460 | 6 | 27.0490 | 43.6915 | 127 |
| 75.00% | 374 | 460 | 7 | 27.2382 | 44.7063 | 127 |
| 95.00% | 374 | 460 | 9 | 27.5262 | 46.2982 | 127 |

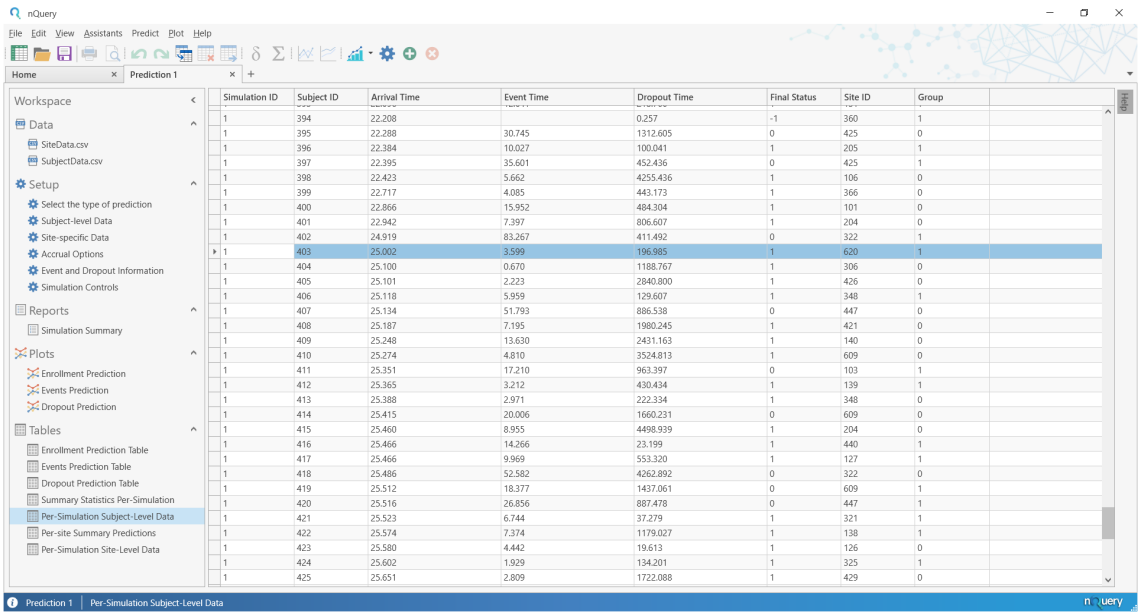
**Figure 6.106:** Scenario A Results for Unblinded Events Prediction using Subject + Site-level Data

For Scenario A, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 43.7440. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 41.3850 (5% Percentile) and 46.2982 (95% Percentile), 50% a duration between 42.7187 (25% Percentile) and 44.7063 (75% Percentile) and a median (50% Percentile) duration of 43.9615. The **Target**

**Sample Size** of 460 was reached in all simulations with an **Average Accrual Duration** of 27.0596.

With an expected **Study Duration** of 40 months, this indicates the study is currently behind schedule by around 3.7 months. However, the enrollment process is expected to be completed after 27.0596 months which is around 3 months earlier than the expected 30 months accrual. This implies the most likely reason for the delay is a lower than expected event hazard rate and this is consistent with a 1-2 month increase in median survival compared to the sample size determination used during the trial's design (see Figure 6.168).

The **Per-Simulation Subject-level Data Table** was also generated here and is shown for Scenario A in Figure 6.107 with the first simulated subject in the first simulation (**Simulation ID = 1**) highlighted. The tables consists of the following fields: **Simulation ID, Subject ID, Arrival Time, Event Time, Dropout Time, Final Status, Site ID, Group**. Note that the order of the Subject ID is based on **Arrival Time** order.




| Simulation ID | Subject ID | Arrival Time | Event Time | Dropout Time | Final Status | Site ID | Group |
|---|---|---|---|---|---|---|---|
| 1 | 394 | 22.208 | | 0.257 | -1 | 360 | 1 |
| 1 | 395 | 22.288 | 30.745 | 1312.605 | 0 | 425 | 0 |
| 1 | 396 | 22.384 | 10.027 | 100.041 | 1 | 205 | 1 |
| 1 | 397 | 22.395 | 35.601 | 452.436 | 0 | 425 | 1 |
| 1 | 398 | 22.423 | 5.662 | 4255.436 | 1 | 106 | 0 |
| 1 | 399 | 22.717 | 4.085 | 443.173 | 1 | 366 | 0 |
| 1 | 400 | 22.866 | 15.952 | 484.304 | 1 | 101 | 0 |
| 1 | 401 | 22.942 | 7.397 | 806.607 | 1 | 204 | 0 |
| 1 | 402 | 24.919 | 83.267 | 411.492 | 0 | 322 | 1 |
| 1 | 403 | 25.002 | 3.599 | 196.985 | 1 | 620 | 1 |
| 1 | 404 | 25.100 | 0.670 | 1188.767 | 1 | 306 | 0 |
| 1 | 405 | 25.101 | 2.223 | 2840.800 | 1 | 426 | 0 |
| 1 | 406 | 25.118 | 5.959 | 129.607 | 1 | 348 | 1 |
| 1 | 407 | 25.134 | 51.793 | 886.538 | 0 | 447 | 0 |
| 1 | 408 | 25.187 | 7.195 | 1980.245 | 1 | 421 | 0 |
| 1 | 409 | 25.248 | 13.630 | 2431.163 | 1 | 140 | 0 |
| 1 | 410 | 25.274 | 4.810 | 3524.813 | 1 | 609 | 0 |
| 1 | 411 | 25.351 | 17.210 | 963.397 | 0 | 103 | 1 |
| 1 | 412 | 25.365 | 3.212 | 430.434 | 1 | 139 | 1 |
| 1 | 413 | 25.388 | 2.971 | 222.334 | 1 | 348 | 0 |
| 1 | 414 | 25.415 | 20.006 | 1660.231 | 0 | 609 | 0 |
| 1 | 415 | 25.460 | 8.955 | 4498.939 | 1 | 204 | 0 |
| 1 | 416 | 25.466 | 14.266 | 23.199 | 1 | 440 | 1 |
| 1 | 417 | 25.466 | 9.969 | 553.320 | 1 | 127 | 1 |
| 1 | 418 | 25.486 | 52.582 | 4262.892 | 0 | 322 | 0 |
| 1 | 419 | 25.512 | 18.377 | 1437.061 | 0 | 609 | 1 |
| 1 | 420 | 25.516 | 26.856 | 887.478 | 0 | 447 | 1 |
| 1 | 421 | 25.523 | 6.744 | 37.279 | 1 | 321 | 1 |
| 1 | 422 | 25.574 | 7.374 | 1179.027 | 1 | 138 | 1 |
| 1 | 423 | 25.580 | 4.442 | 19.613 | 1 | 126 | 0 |
| 1 | 424 | 25.602 | 1.929 | 134.201 | 1 | 325 | 1 |
| 1 | 425 | 25.651 | 2.809 | 1722.088 | 1 | 429 | 0 |

**Figure 6.107:** Per-Simulation Subject-level Data Table for Subject + Site-level Data

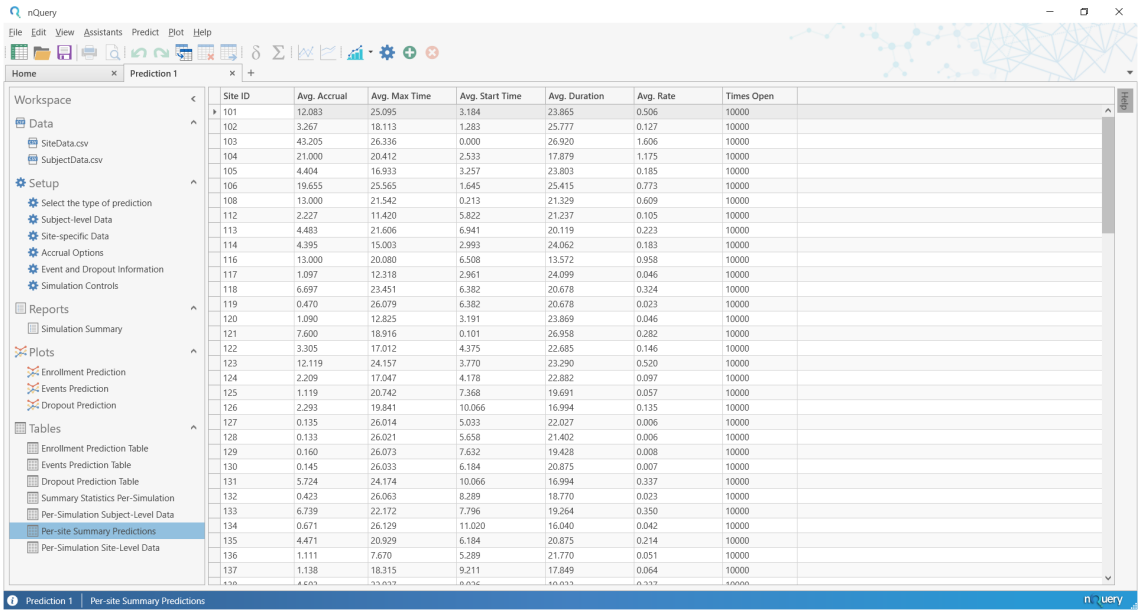
**Arrival Time** is generated for all simulated subjects (i.e. **Subject ID > 402**). **Event Time** and **Dropout Time** are generated for all simulated subjects and all subjects which were **Censored Status** (all status values from **Status Indicator** field in **Subject-level Data Setup** step) in the subject-level data (for example **Subject ID = 402** in this example). **Site ID** and treatment **Group** are also generated for all simulated subjects. Blank fields for **Event Time** or **Dropout Time** indicate cases where the subject already had the event/dropout and therefore the time for the other status does not exist and does not need to be simulated.

The **Final Status** is from the subject-level data for subjects with **Event Status** or **Dropout Status** and based on the **Event Time** and **Dropout Time** for simulated subjects and **Censored Status** subjects in the subject-level data. **Final Status** will equal **Event Status** if **Event Time** is less than **Dropout Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.80 for per-simulation end times). **Final Status** will equal **Dropout Status** if **Dropout Time** is less than **Event Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.80 for per-simulation end times). **Final Status** will equal **Censored Status** if **Event Time** and **Dropout Time** are greater than the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.80 for per-simulation end times)

The **Per-Site Summary Predictions Table** was also generated here and is shown for Scenario A in Figure 6.108. The **Per-Site Summary Predictions Table** contains the following columns:

- Site ID: Site ID for a site
- Avg. Accrual: The average number of subjects enrolled from this site across all simulations
- Avg. Max Time: The average time of the final enrollment in a site
- Avg. Start Time: The average site initiation time for a site. This is from the Site-level Data for open sites and generated randomly for unopened sites
- Avg. Duration: The average length of time a site was open and enrolling subjects
- Avg. Rate: The average accrual rate for a site during the period it was open
- Times Open: The number of simulations in which the site opened before enrollment was complete

Note that **Times Opens** will equals the number of simulations except for unopened sites with any generated start times higher than a given simulation's **Accrual Duration**.

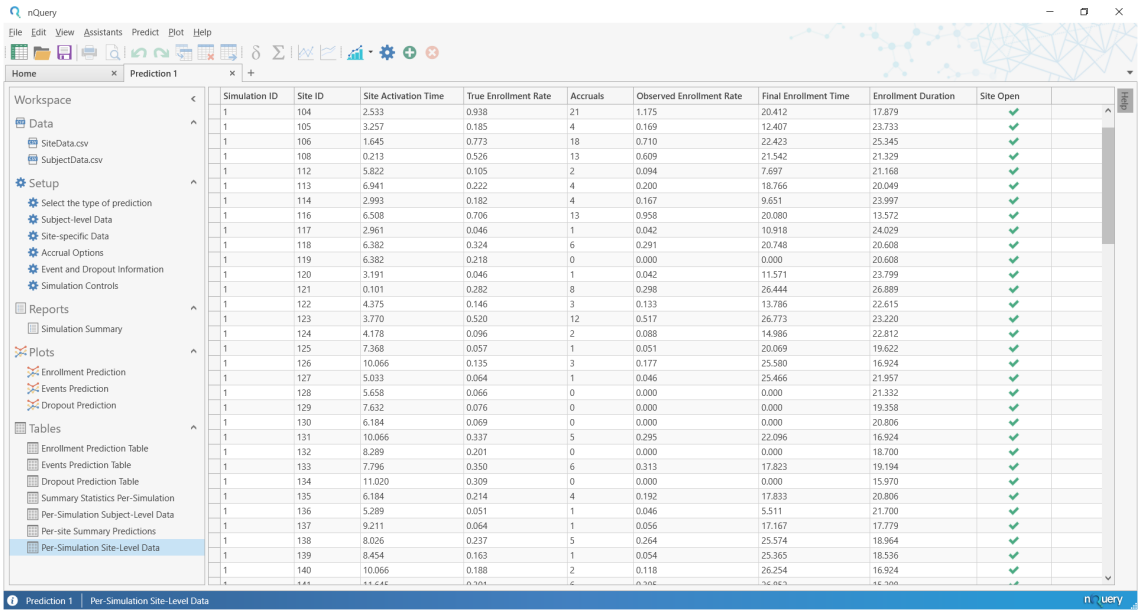



| Site ID | Avg. Accrual | Avg. Max Time | Avg. Start Time | Avg. Duration | Avg. Rate | Times Open |
|---|---|---|---|---|---|---|
| 101 | 12.083 | 25.095 | 3.184 | 23.865 | 0.506 | 10000 |
| 102 | 3.267 | 18.113 | 1.283 | 25.777 | 0.127 | 10000 |
| 103 | 43.205 | 26.336 | 0.000 | 26.920 | 1.606 | 10000 |
| 104 | 21.000 | 20.412 | 2.533 | 17.879 | 1.175 | 10000 |
| 105 | 4.404 | 16.933 | 3.257 | 23.803 | 0.185 | 10000 |
| 106 | 19.655 | 25.565 | 1.645 | 25.415 | 0.773 | 10000 |
| 108 | 13.000 | 21.542 | 0.213 | 21.329 | 0.609 | 10000 |
| 112 | 2.227 | 11.420 | 5.822 | 21.237 | 0.105 | 10000 |
| 113 | 4.483 | 21.606 | 6.941 | 20.119 | 0.223 | 10000 |
| 114 | 4.395 | 15.003 | 2.993 | 24.062 | 0.183 | 10000 |
| 116 | 13.000 | 20.080 | 6.508 | 13.572 | 0.958 | 10000 |
| 117 | 1.097 | 12.318 | 2.961 | 24.099 | 0.046 | 10000 |
| 118 | 6.697 | 23.451 | 6.382 | 20.678 | 0.324 | 10000 |
| 119 | 0.470 | 26.079 | 6.382 | 20.678 | 0.023 | 10000 |
| 120 | 1.090 | 12.825 | 3.191 | 23.869 | 0.046 | 10000 |
| 121 | 7.600 | 18.916 | 0.101 | 26.958 | 0.282 | 10000 |
| 122 | 3.305 | 17.012 | 4.375 | 22.685 | 0.146 | 10000 |
| 123 | 12.119 | 24.157 | 3.770 | 23.290 | 0.520 | 10000 |
| 124 | 2.209 | 17.047 | 4.178 | 22.882 | 0.097 | 10000 |
| 125 | 1.119 | 20.742 | 7.368 | 19.691 | 0.057 | 10000 |
| 126 | 2.293 | 19.841 | 10.066 | 16.994 | 0.135 | 10000 |
| 127 | 0.135 | 26.014 | 5.033 | 22.027 | 0.006 | 10000 |
| 128 | 0.133 | 26.021 | 5.658 | 21.402 | 0.006 | 10000 |
| 129 | 0.160 | 26.073 | 7.632 | 19.428 | 0.008 | 10000 |
| 130 | 0.145 | 26.033 | 6.184 | 20.875 | 0.007 | 10000 |
| 131 | 5.724 | 24.174 | 10.066 | 16.994 | 0.337 | 10000 |
| 132 | 0.423 | 26.063 | 8.289 | 18.770 | 0.023 | 10000 |
| 133 | 6.739 | 22.172 | 7.796 | 19.264 | 0.350 | 10000 |
| 134 | 0.671 | 26.129 | 11.020 | 16.040 | 0.042 | 10000 |
| 135 | 4.471 | 20.929 | 6.184 | 20.875 | 0.214 | 10000 |
| 136 | 1.111 | 7.670 | 5.289 | 21.770 | 0.051 | 10000 |
| 137 | 1.138 | 18.315 | 9.211 | 17.849 | 0.064 | 10000 |

**Figure 6.108:** Per-Site Summary Predictions for Scenario A

The **Per-Simulation Site-level Data Table** was also generated here and is shown for Scenario A in Figure 6.109. The **Per-Simulation Site-level Data Table** contains the following columns:

- Simulation ID: The simulation number
- Site ID: Site ID for a site
- Site Activation Time: Time after study start when site become open to enroll subjects. Taken from Site-level Data for open sites, generated for unopened sites
- True Enrollment Rate: The user specified enrollment rate for a site
- Accruals: Total number of subjects enrolled from a site in a simulation
- Observed Enrollment Rate: The actual enrollment rate (subjects enrolled per unit time) achieved in a site in a simulation
- Final Enrollment Time: Time of the final enrollment in a site in a simulation
- Enrollment Duration: The length of time a site was open in a simulation
- Site Open: Indicator of whether site opened in a simulation. Tick = Open, X = Did not open




| Simulation ID | Site ID | Site Activation Time | True Enrollment Rate | Accruals | Observed Enrollment Rate | Final Enrollment Time | Enrollment Duration | Site Open |
|---|---|---|---|---|---|---|---|---|
| 1 | 104 | 2.533 | 0.938 | 21 | 1.175 | 20.412 | 17.879 | ✓ |
| 1 | 105 | 3.257 | 0.185 | 4 | 0.169 | 12.407 | 23.733 | ✓ |
| 1 | 106 | 1.645 | 0.773 | 18 | 0.710 | 22.423 | 25.345 | ✓ |
| 1 | 108 | 0.213 | 0.526 | 13 | 0.609 | 21.542 | 21.329 | ✓ |
| 1 | 112 | 5.822 | 0.105 | 2 | 0.094 | 7.697 | 21.168 | ✓ |
| 1 | 113 | 6.941 | 0.222 | 4 | 0.200 | 18.766 | 20.049 | ✓ |
| 1 | 114 | 2.993 | 0.182 | 4 | 0.167 | 9.651 | 23.997 | ✓ |
| 1 | 116 | 6.508 | 0.706 | 13 | 0.958 | 20.080 | 13.572 | ✓ |
| 1 | 117 | 2.961 | 0.046 | 1 | 0.042 | 10.918 | 24.029 | ✓ |
| 1 | 118 | 6.382 | 0.324 | 6 | 0.291 | 20.748 | 20.608 | ✓ |
| 1 | 119 | 6.382 | 0.218 | 0 | 0.000 | 0.000 | 20.608 | ✓ |
| 1 | 120 | 3.191 | 0.046 | 1 | 0.042 | 11.571 | 23.799 | ✓ |
| 1 | 121 | 0.101 | 0.282 | 8 | 0.298 | 26.444 | 26.889 | ✓ |
| 1 | 122 | 4.375 | 0.146 | 3 | 0.133 | 13.786 | 22.615 | ✓ |
| 1 | 123 | 3.770 | 0.520 | 12 | 0.517 | 26.773 | 23.220 | ✓ |
| 1 | 124 | 4.178 | 0.096 | 2 | 0.088 | 14.986 | 22.812 | ✓ |
| 1 | 125 | 7.368 | 0.057 | 1 | 0.051 | 20.069 | 19.622 | ✓ |
| 1 | 126 | 10.066 | 0.135 | 3 | 0.177 | 25.580 | 16.924 | ✓ |
| 1 | 127 | 5.033 | 0.064 | 1 | 0.046 | 25.466 | 21.957 | ✓ |
| 1 | 128 | 5.658 | 0.066 | 0 | 0.000 | 0.000 | 21.332 | ✓ |
| 1 | 129 | 7.632 | 0.076 | 0 | 0.000 | 0.000 | 19.358 | ✓ |
| 1 | 130 | 6.184 | 0.069 | 0 | 0.000 | 0.000 | 20.806 | ✓ |
| 1 | 131 | 10.066 | 0.337 | 5 | 0.295 | 22.096 | 16.924 | ✓ |
| 1 | 132 | 8.289 | 0.201 | 0 | 0.000 | 0.000 | 18.700 | ✓ |
| 1 | 133 | 7.796 | 0.350 | 6 | 0.313 | 17.823 | 19.194 | ✓ |
| 1 | 134 | 11.020 | 0.309 | 0 | 0.000 | 0.000 | 15.970 | ✓ |
| 1 | 135 | 6.184 | 0.214 | 4 | 0.192 | 17.833 | 20.806 | ✓ |
| 1 | 136 | 5.289 | 0.051 | 1 | 0.046 | 5.511 | 21.700 | ✓ |
| 1 | 137 | 9.211 | 0.064 | 1 | 0.056 | 17.167 | 17.779 | ✓ |
| 1 | 138 | 8.026 | 0.237 | 5 | 0.264 | 25.574 | 18.964 | ✓ |
| 1 | 139 | 8.454 | 0.163 | 1 | 0.054 | 25.365 | 18.536 | ✓ |
| 1 | 140 | 10.066 | 0.188 | 2 | 0.118 | 26.254 | 16.924 | ✓ |

**Figure 6.109:** Per-Simulation Site-level Data Table for Scenario A

**Scenario B** Scenario A assumed that an additional 9 sites would open after the **Current Time**, as per the original study plan. Given that the enrollment is ahead of schedule in Scenario A, it was decided to remove all unopened sites from the study plan and that the site with the lowest enrollment rate up to the **Current Time** will be closed.

To evaluate this scenario, select **Accrual Options** from under the **Setup** header in the Workspace Navigation Bar. In **Accrual Options**, open the **Accrual Infos/Sites** tab. Change the **Total Number of Sites** to 118 (equal to the **Number of Sites Opened**). The site with the lowest accrual rate was Site 117. To close this site, set the **Enrollment Cap** equal to 1 (i.e. the current **No. of Accruals**) in the **Site ID = 117** row.

The updated **Accrual Infos/Sites** tab for Scenario B is shown in Figure 6.110.

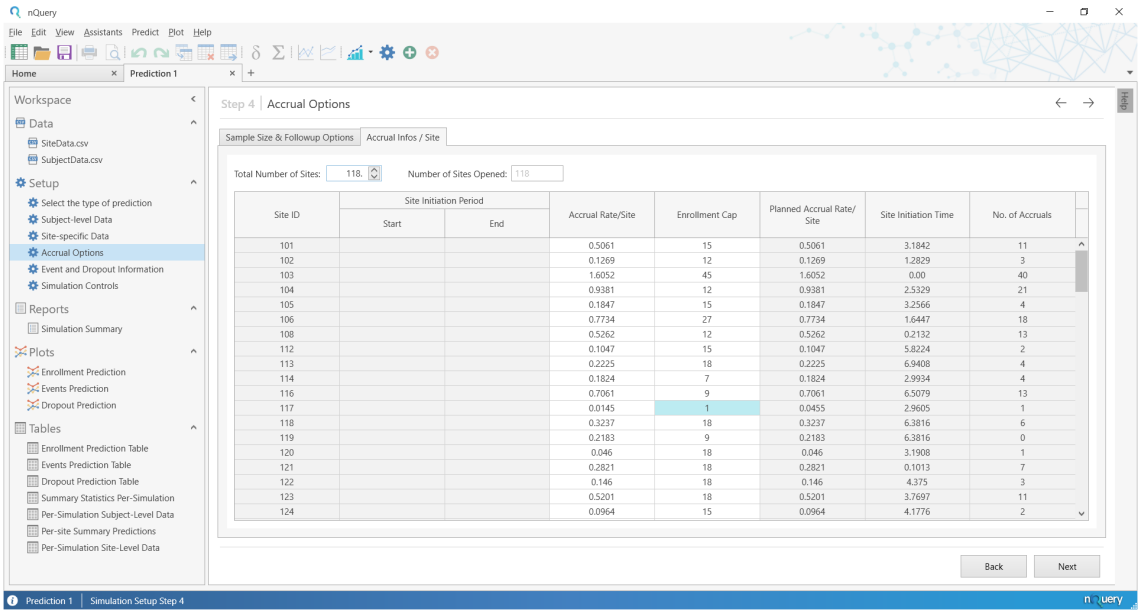



**Figure 6.110:** Scenario B Accrual Infos/Sites Tab

Next or → button. At the next **Simulation Controls** step, select the Run or → button to simulate Scenario B. The results for Scenario B are summarized in Figure 6.111.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 27.2258 |
| Average Study Duration | 43.7791 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.1701 |
| Average Follow-up | 12.4911 |
| Sites Opened | 118.0000 |

Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 232 |
| Average Treatment Sample Size | 228 |
| Average Control Events | 201.1793 |
| Average Treatment Events | 172.8207 |
| Average Control Dropouts | 1.9108 |
| Average Treatment Dropouts | 4.2593 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration | Sites Opened |
|---|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 4 | 26.7442 | 41.4105 | 118 |
| 25.00% | 374 | 460 | 5 | 27.0171 | 42.7222 | 118 |
| 50.00% | 374 | 460 | 6 | 27.2126 | 43.7471 | 118 |
| 75.00% | 374 | 460 | 7 | 27.4201 | 44.7514 | 118 |
| 95.00% | 374 | 460 | 9 | 27.7463 | 46.2956 | 118 |

**Figure 6.111:** Scenario B for Enrollment Prediction with Subject + Site-level Data

For Scenario B, the **Overall Summary** shows that the **Target Sample Size** of 460 was reached in every simulation with an **Average Accrual Duration** over all the simulations of 27.2258 with the **Percentile Summary** showing that 95% of predictions had a **Study Duration** between 26.7442 and 27.7463, 50% a duration between 27.0171 and 27.4201 with a median duration of 27.2126 with 118 **Sites Opened** in all cases as expected. This was approximately a 0.16 months increase compared to **Scenario A** which is, as expected, a minimal change due to choosing to

not open any additional sites and closing Site 117. The **Average Study Duration** of 43.7991 was also a minimal increase over **Scenario A's** 43.7440.

#### 6.2.7.3 Blinded Events Prediction with Subject Data Only with Enrollment Complete

**Setup** **Blinded Events** milestone prediction using **Subject-level Data Only** means that the prediction will project the expected length of study needed to achieve a specified **Target Number of Events** using subject-level interim data alone. In this demonstration, the enrollment process is assumed to be complete and therefore the total sample size available will equal the number of rows in **SubjectData.csv**, which is 402.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step select **Blinded Events** from the **Target** field, **Subject Data Only** from the **Input** field and **Complete** from the **Enrollment Status** field.

These selections are shown in Figure 6.112. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

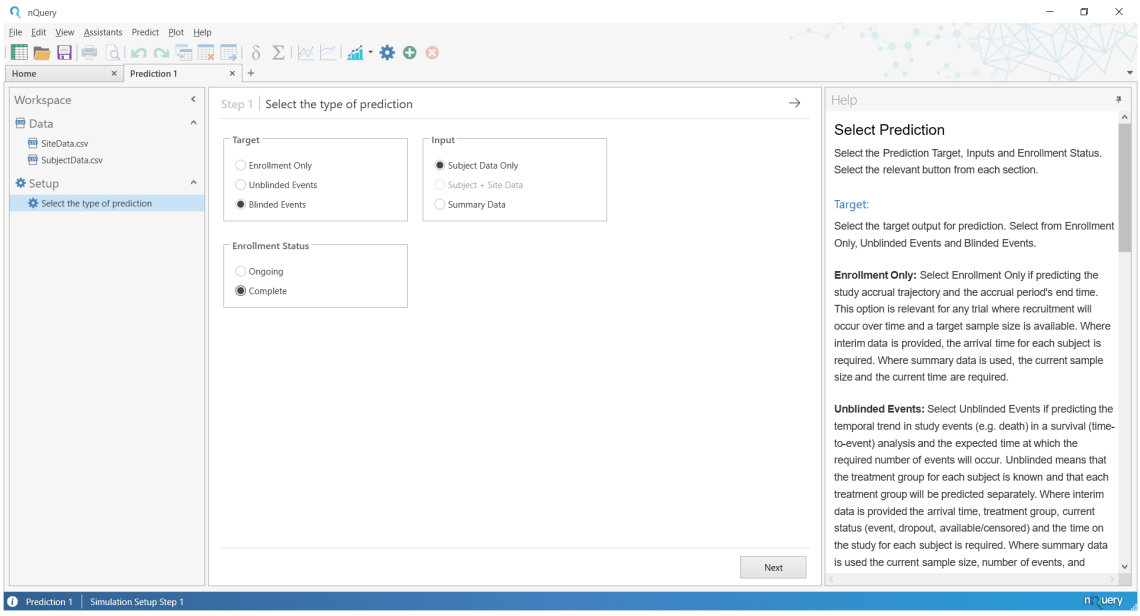



**Figure 6.112:** Select the Type of Prediction Setup Step for Blinded Events when Enrollment Complete

The next step is the **Subject-level Data** step. The **Subject-level Data** step is a Data Field Selection step. The **Subject-level Data** is the dataset where each row corresponds to a specific subject enrolled in the study.

Data Field Selection steps are steps in which the required dataset in the current Workspace is selected and columns in the selected dataset are assigned to the required fields for that prediction problem. The options in a given Data Field Selection field are opened by selecting the button on the right of each field which is found to the right of the field name e.g. **Select Data**. Each field will have specific requirements (e.g. numeric only, two distinct values only) and nQuery Predict only displays options which are consistent with those fields.

In this prediction there are two fields which are as follows:

- Select Dataset: Select the dataset that is the Subject-level Data from the datasets in the current workspace
- Arrival Time: The arrival time column contains the length of time since the study start (Time = 0) until a subject entered the study. Each subject's Arrival Time should be greater than or equal to zero
- Status Indicator: The status (event, dropout, censored/active) for each subject. This Status Indicator column should have a minimum of one alphanumeric unique value and a maximum of three unique values. It has the following sub-fields
  - Event Status: Enter value in Status Indicator column that corresponds to subject having had an event
  - Dropout Status: Enter value in Status Indicator column that corresponds to a subject having dropped out
  - Censored Status: Enter value in Status Indicator column that corresponds to a subject having neither had the event or dropped out
- Time on Study: The length of time after Arrival Time that a subject was in the study. Must be numeric value greater than or equal to zero. The definition depends on Status Indicator (see below)

When the **Status Indicator** field is selected, the **Event Status**, **Dropout Status** and **Censored Status** rows become active. These sub-fields are used to select which values in the **Status Indicator** field correspond to each subject status of interest. The **Censored Status** is also referred to as the **Available Status** as the subject would have been administratively censored if the study ended at the **Current Time** but will be available to have the event of interest in the future during the event milestone prediction.

If the nQuery Predict default values of 1 = Event, -1 = Dropout, 0 = Censored is used in the **Status Indicator** field in the **Subject-level Data** then the **Status** sub-fields will be filled automatically with those assignments. Otherwise, manually enter values by either inputting a value directly into the field or selecting the button on the right of each **Status** sub-field and selecting from the list of unique values in the selected **Status Indicator** field. If there are less than three unique values in the **Status Indicator** field (for example if there have been no dropouts i.e.

zero rows with **Dropout Status**) then the user must manually enter the value they would like to have assigned for that **Status** outcome. Note that appropriate defaults will be provided for the rate of previously unused **Status** category in subsequent steps e.g. Dropout Hazard Rate(s) = 0.

The definition of **Time on Study** will depend on the **Status Indicator** for that subject. If the **Status Indicator** is equal to **Event Status** or **Dropout Status**, the **Time on Study** is the length of time a subject was followed for after their **Arrival Time** before they had an event/dropped out. If the **Status Indicator** is equal to **Censored Status** , the **Time on Study** is the length of time a subject was followed for after their **Arrival Time** until the **Current Time.**

For this prediction, in the **Select Dataset** field select **SubjectData.csv**. Then in the **Arrival Time** field select **Arrival**, in the **Status Indicator** field select **Current** (assign 1 to **Event Status**, -1 to **Dropout Status**, 0 to **Censored Status**) and in the **Time on Study** field select **Followup.** These selections are shown in Figure 6.113. Select the Next or → button to move to the next **Setup** step.

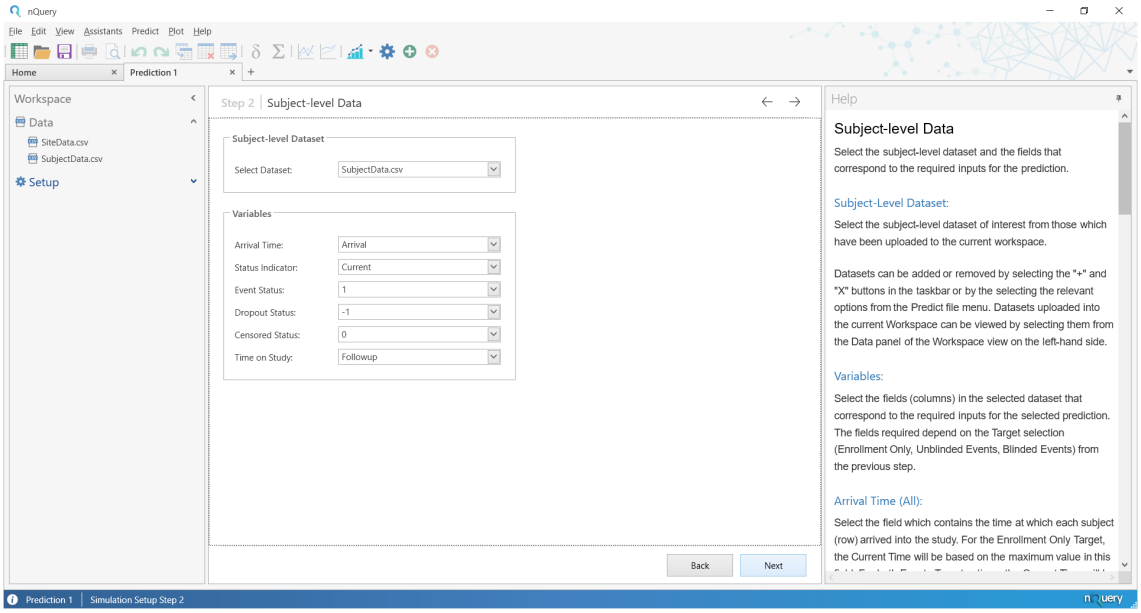



**Figure 6.113:** Subject-level Data Setup Step for Blinded Events when Enrollment Complete

The next step is the **Accrual Options** step which specifies the **Target Sample Size** and **Followup Option**. The **Accrual Options** step, and its defaults, for this demonstration are shown in Figure 6.114.

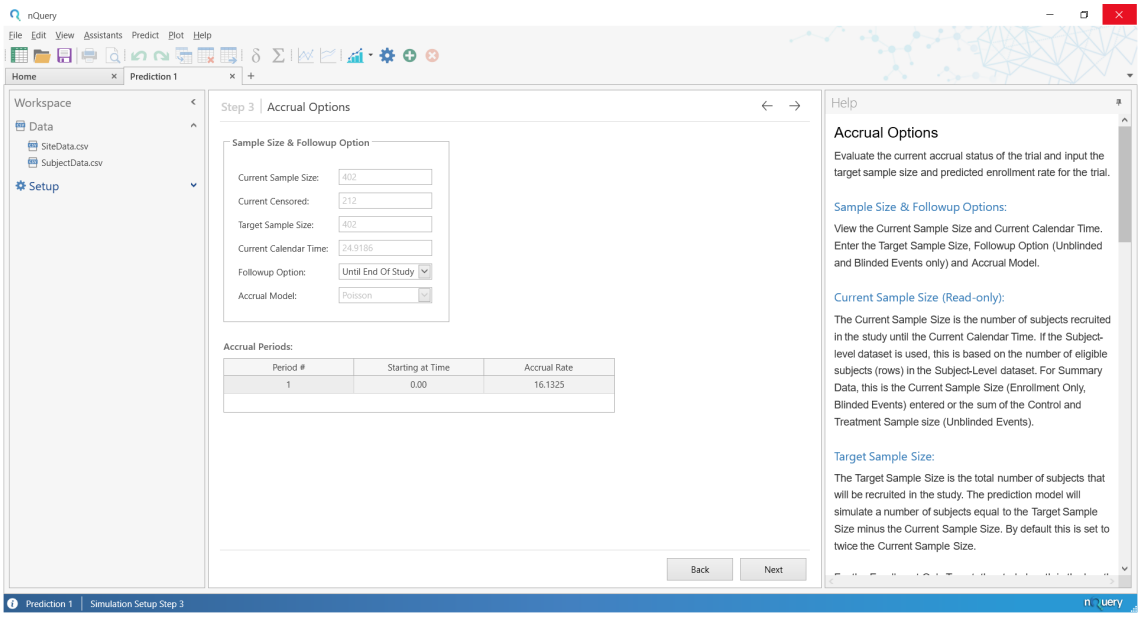



**Figure 6.114:** Accrual Options Setup Step for Blinded Events Prediction when Enrollment Complete

The **Accrual Options** step consists of two main elements: the **Sample Size & Followup** input fields and the **Accrual Periods** table.

The **Sample Size & Followup** input fields provides information on enrollment process, censored status and current time based on the subject-level data and allows the user to edit the **Target Sample Size** and select the **Followup Option** and **Accrual Model**. The following fields are provided in this prediction:

- Current Sample Size (Read-only): Sample Size recruited into a study so far. Equals the number of rows in the Subject-level Data. When Enrollment Status = Complete, Current and Target Sample Size are the same
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Target Sample Size (Read-only): Total number of subjects that will be recruited in this study. When Enrollment Status = Complete, Current and Target Sample Size are the same
- Current Time (Read-only): The length of time that has passed since the study started. Equals the maximum of the sum of the Arrival Time + Time on Study values for each subject
- Followup Option: Specifies whether subjects who do not have the event/dropout will be administratively censored at the end of the study or after a fixed follow up period

  - Followup Time (Followup Option = For Fixed Period only): Length of time a subject is followed for until administratively censored if they do not have the event/dropout

- Accrual Model (Read-only): The statistical model that will be used to generate simulated enrollments

When **Enrollment Status = Complete**, the **Accrual Options** step is available to see information on ernollment up to the **Current Time** but is not used to set the **Target Sample Size** (see read-only and equal to **Current Sample Size**) and therefore the **Accrual Model** is not needed.

The only editable option is **Followup Option** which by default is set to **Until End of Study**.

If **Followup Option** is set to **Until End of Study** then subjects who do not have the event or dropout will be followed until the end of the study when the **Target Events** is reached, at which point they will be administratively right-censored.

If **Followup Option** is set to **For Fixed Period** then subjects who do not have the event or dropout will be followed until they have a **Time on Study** equal to the value specified in the **Followup Time** field, which appears automatically below the **Followup Option** field when **For Fixed Period** is selected, at which point a subject will be administratively right-censored. Note that for subjects whose **Followup Time** is less than length of the time a subject was on study before the study ended when the **Target Events** was reached, they will be administratively censored at the study end rather than after their **Followup Time**.

The **Accrual Periods** table contains information on the **Accrual Rate** up the **Current Time.** When **Enrollment Status = Complete**, the **Accrual Periods** table will consist of one row which will contain zero in the **Starting at Time** column and the accrual rate from the information provided in the **Accrual Rate** column.

In this demonstration, the default **Followup Option = Until End of Study** is used as per Figure 6.114. Select the Next or → button to move to the next **Setup** step.

The next step is the **Event and Dropout Information** step which specifies the **Target Number of Events,** event model (and associated statistical parameters) and dropout model (and associated statistical parameters). The **Event and Dropout Information** step on the default **Events Model** tab is shown in Figure 6.115.

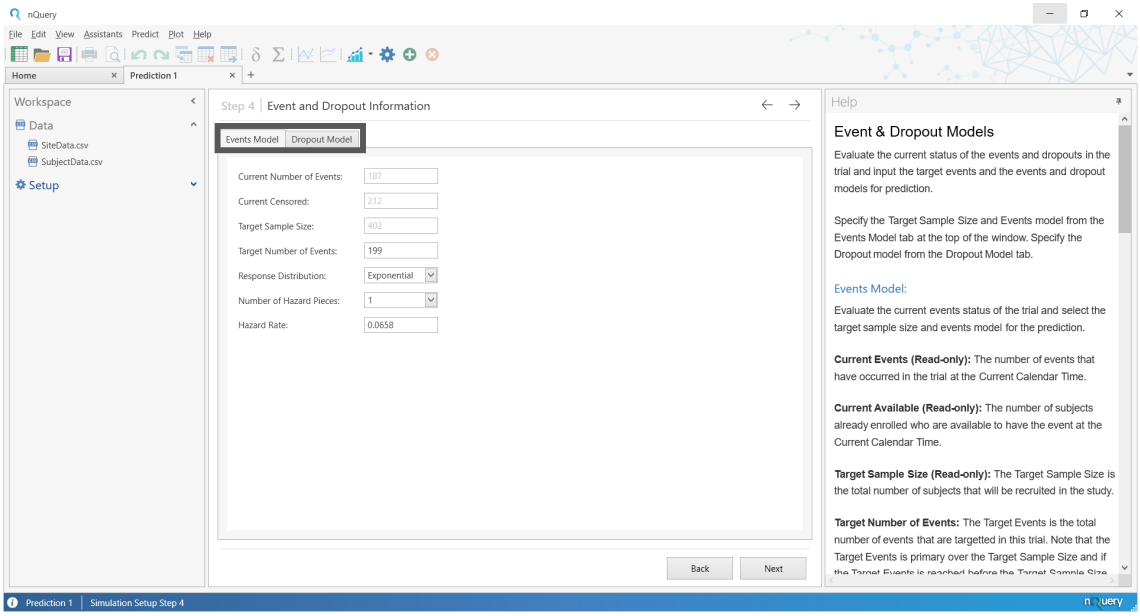



**Figure 6.115:** Event and Dropout Information Setup Step for Blinded Events Prediction when Enrollment Complete

The **Event and Dropout Information** step consists of two tabs: the **Events Model** tab and the **Dropout Model** tab. These tabs have been highlighted in Figure 6.115 and are opened by selecting the required tab name at the top of the main window, below the step title. For **Blinded Events**, it is assumed that survival and dropout times will be simulated using a single "global" process which is specified by the user.

The **Events Model** tab provides fields which provide information on the current number of subject who have had an event, could have an event going forward and allows the input of the **Target Number of Events** and **Response Distribution** for the events model and additional parameters needed for that specific model. The following fields are provided in this prediction:

- Current Events (Read-only): The number of subjects who had the event of interest at the current time. Equals the number of subjects with the Event Status in the Status Indicator field
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Target Sample Size (Read-only): Total number of subjects that will be recruited in this study. When Enrollment Status = Complete, this is the Current Sample Size
- Target Number of Events (Editable): The required number of subjects needed to have the event of interest for the study to end

- Response Distribution: The statistical model that will be used to generate survival times for subjects available to have the event
- Number of Hazard Pieces (Response Distribution = Exponential only): Number of hazard pieces used to specify the piecewise exponential model. Select value between 1 and 10
- Hazard Rate (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The exponential hazard rate for the constant exponential survival model specified above

In nQuery Predict, the default **Response Distribution** is **Exponential** with the **Number of Hazard Pieces** set to 1 and the default **Hazard Rate** set the expected event rate from the provided information (see subsubsection 6.1.2.2 for details on the survival distributions used in nQuery Predict, see section 3.5 for details on how to calculate a exponential **Hazard Rate** using a median survival or survival probability)**.** The **Target Number of Events** for this example is 199 and the default estimated exponential **Hazard Rate** is 0.0658.

In nQuery, there are effectively three primary survival distributions that survival times can be generated from:

- Constant Exponential (Response Distribution = Exponential, Number of Hazard Pieces =1): Input Hazard Rate
- Piecewise Exponential (Response Distribution = Exponential, Number of Hazard Pieces > 1): Input table of Starting at Time and Hazard Rate columns
- Weibull (Response Distribution = Weibull): Input Scale Parameter and Shape Parameter

Details on how survival times are generated from each of these distributions is provided in subsubsection 6.1.2.2, alongside how the default values provided in nQuery Predict are estimated from the provided information. The change in the **Events Model** tab for each of these choices is shown in Figure 6.116.

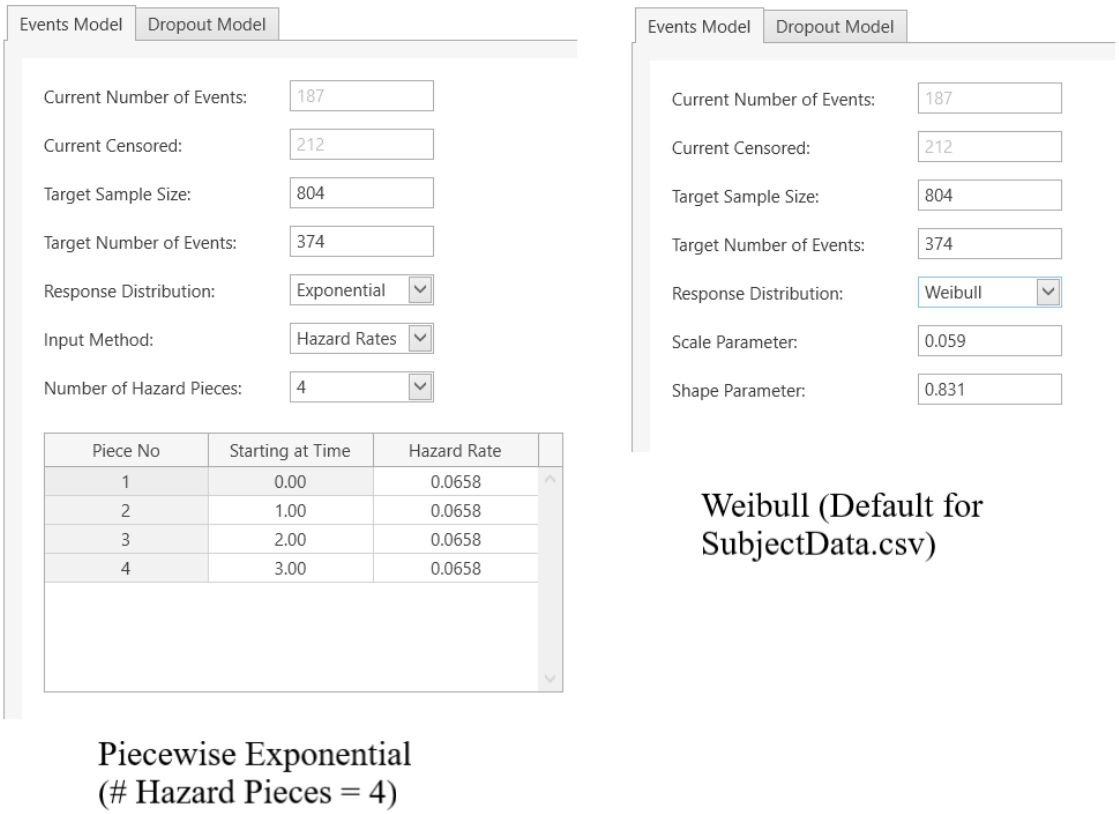



**Figure 6.116:** Piecewise Exponential (4 Pieces) and Weibull Defaults for SubjectData.csv

The piecewise exponential table will have a number of rows equal to **Number of Hazard Pieces.** The **Piece No** column indicates the hazard piece number of a row. The **Starting at Time** column sets the start time for when the **Hazard Rate** value in that row takes effect. The **Hazard Rate** column sets the exponential hazard rate in effect from the start time of the current hazard piece to the start time in the next row (see section 3.5 for details on how to calculate a exponential **Hazard Rate** using a median survival or survival probability and the Compute Effect Size side table for STT3 may also be useful). By default the piecewise exponential assumes a constant exponential model **Hazard Rate** (same **Hazard Rate** as for the **Number of Hazard Pieces** = 1 case) with **Starting at Time** being values increasing by 1 up to the final row.

The Weibull distribution has two parameters: the Weibull **Scale Parameter** and Weibull **Shape Parameter**. The default for these is found by fitting the survival data from the subject-level data (see subsubsection 6.1.2.2 for details).

The **Dropout Model** tab is opened by selecting the **Dropout Model** tab name at the top of the main window. For this demonstration, the **Dropout Model** tab is shown in Figure 6.117.

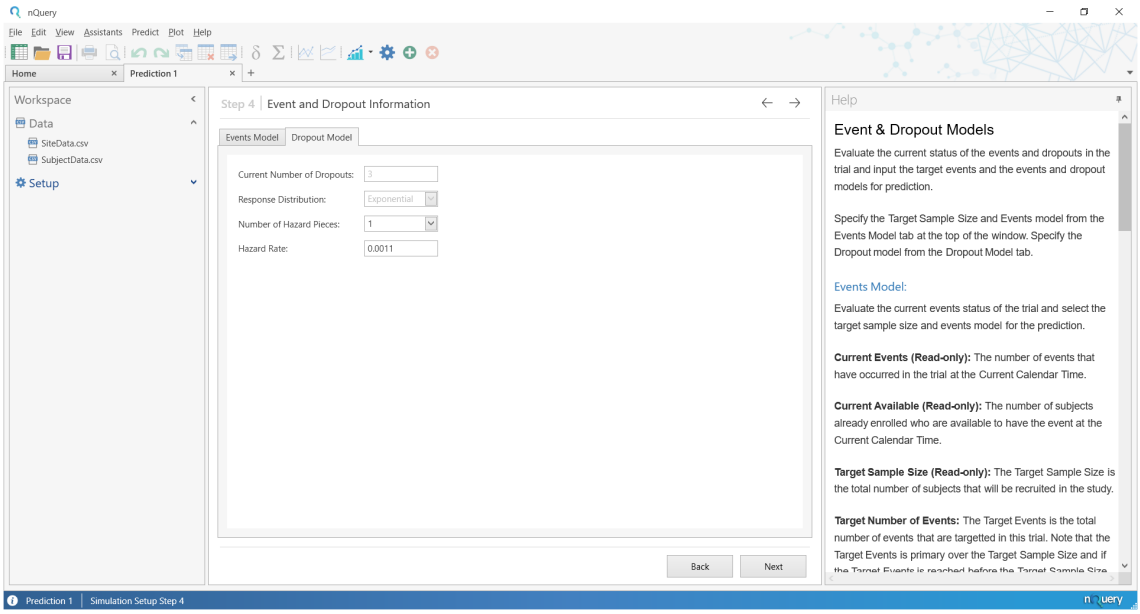



**Figure 6.117:** Dropout Model tab for Blinded Events with Subject-level Data

For **Dropout Model**, the **Current Number of Dropouts** (number of subjects who have dropped out up to the **Current Time**), **Response Distribution** and **Number of Hazard Pieces** are available. For dropout, the only **Response Distribution** option is **Exponential**. The inputs and defaults for dropout **Exponential Response Distribution**, for both the case where **Hazard Pieces = 1** (Constant Exponential) or **Hazard Pieces > 1** (Piecewise Exponential), are the same as from the **Events Model** described above and therefore will not be replicated here. In this demonstration, the constant exponential **Hazard Rate** for dropout calculated from the information provided equals 0.0011.

In this demonstration the default for all inputs (**Current/Target Sample Size = 402, Target Number of Events = 199, Events Model Exponential Response Distribution (1 Hazard Piece) Hazard Rate = 0.0658, Dropout Model Exponential Response Distribution (1 Hazard Piece) Hazard Rate = 0.0011**) are used. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.118. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

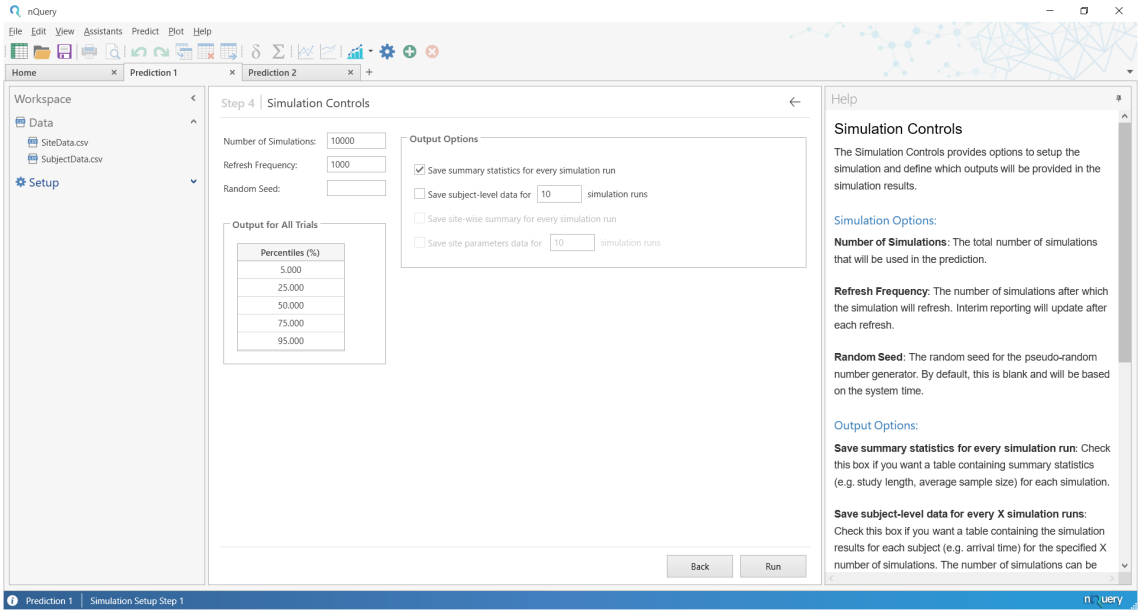



**Figure 6.118:** Simulation Controls Setup Step for Blinded Events Prediction when Enrollment Complete

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.119.

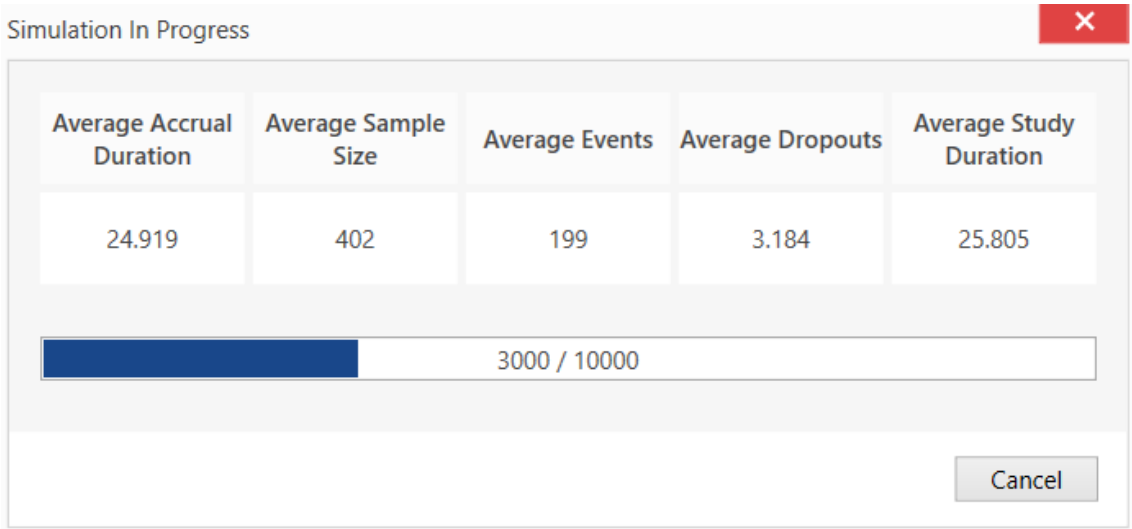



**Figure 6.119:** Simulation in Progress Window for Blinded Events Prediction when Enrollment Complete

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.120.

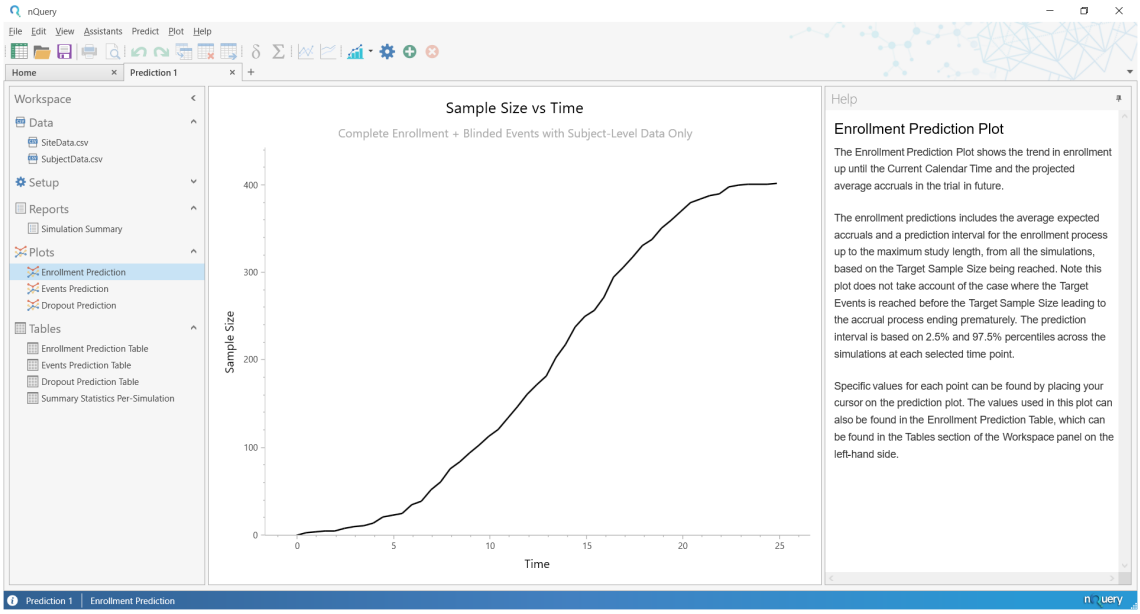



**Figure 6.120:** Enrollment Prediction Plot for Blinded Events Prediction when Enrollment Complete

The results of a nQuery Predict milestone prediction can be divided into three

categories: **Reports, Plots** and **Tables**. A high-level summary of the results in nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table, Events Prediction Table, Dropout Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.121.

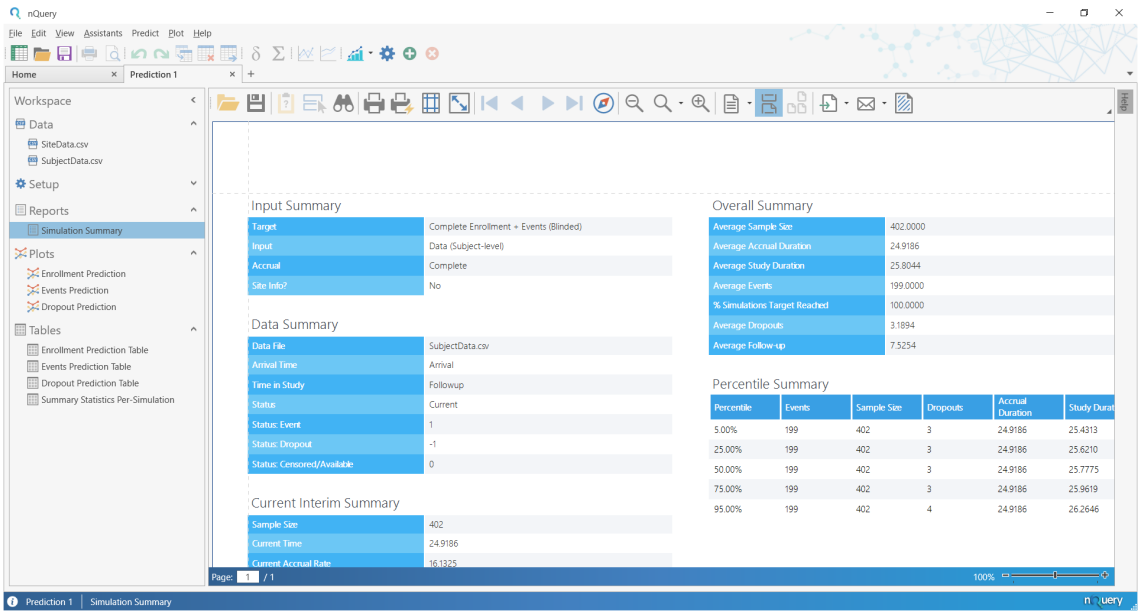



**Figure 6.121:** Simulation Summary Report for Blinded Events Prediction when Enrollment Complete

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.122. For this demonstration, the **Overall Summary** shows that the **Target Number of Events** of 199 was reached in every simulation with an **Average Study Duration** over all the simulations of 25.8044. The **Percentile Summary** shows that 90% of predictions had

a **Study Duration** between 25.4313 (5% Percentile) and 26.2646 (95% Percentile), 50% a duration between 25.6210 (25% Percentile) and 25.9619 (75% Percentile) and a median (50% Percentile) duration of 25.7775. The **Average Dropouts** was 3.1894 (Percentiles:3, 3, 3, 3, 4) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 7.5254. As enrollment was complete, the **Accrual Duration** and **Sample Size** are constant.

Overall Summary

| | |
|---|---|
| Average Sample Size | 402.0000 |
| Average Accrual Duration | 24.9186 |
| Average Study Duration | 25.8044 |
| Average Events | 199.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 3.1894 |
| Average Follow-up | 7.5254 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 199 | 402 | 3 | 24.9186 | 25.4313 |
| 25.00% | 199 | 402 | 3 | 24.9186 | 25.6210 |
| 50.00% | 199 | 402 | 3 | 24.9186 | 25.7775 |
| 75.00% | 199 | 402 | 3 | 24.9186 | 25.9619 |
| 95.00% | 199 | 402 | 4 | 24.9186 | 26.2646 |

**Figure 6.122:** Simulation Summary Report Results for Blinded Events Prediction when Enrollment Complete

To edit the **Simulation Summary report there a** number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.123.

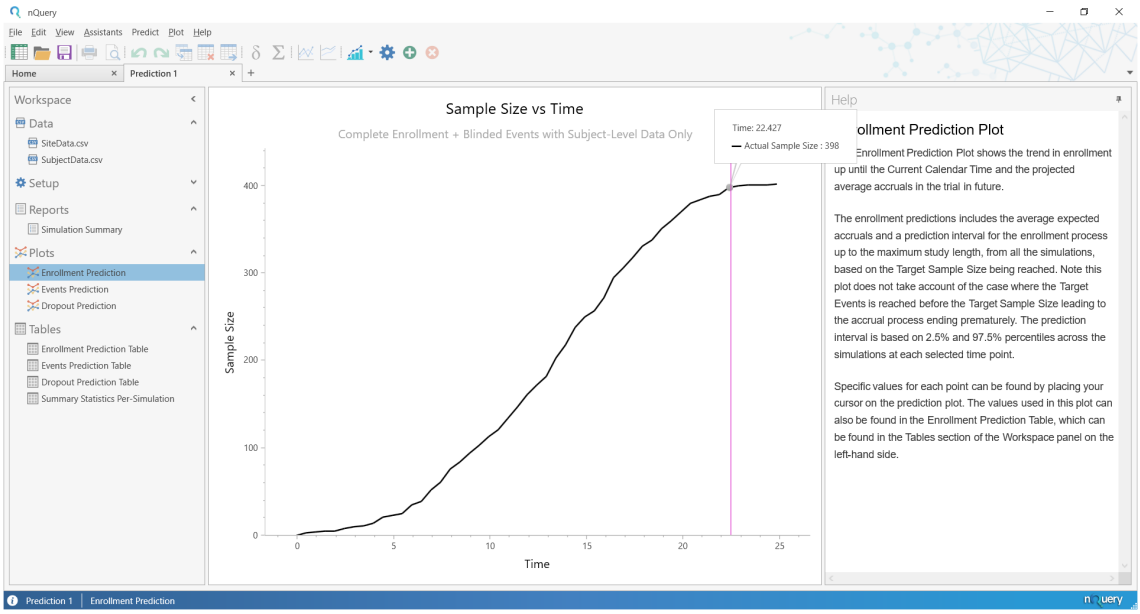



**Figure 6.123:** Enrollment Prediction Plot for Blinded Events Prediction when Enrollment Complete

The **Enrollment Prediction** plot when **Enrollment Status = Complete** will show the accrual pattern up to the **Current Time**, with the sample size at a given prior time given by placing the cursor over the plot. Note this means a different **Time** scale than the **Event Prediction** and **Dropout Prediction** plots which will be associated with the **Study Duration.**

The **Events Prediction** plot consists of two parts: from Time = 0 to **Current Time** and from **Current Time** to the **Maximum Time**. The first part is a black line which plots the accrual pattern from the user inputs provided. The second part is a yellow conic section with a grey line where the yellow conic section plots the 95% prediction interval for the **Events** achieved at a given **Time** across all simulations and the grey line is the average **Events** at a given time. The 95% prediction interval indicates the 95% of predictions had a number of events that fell between the lower and upper limit at a given time. The **Events Prediction** plot for this demonstration is shown in Figure 6.124.

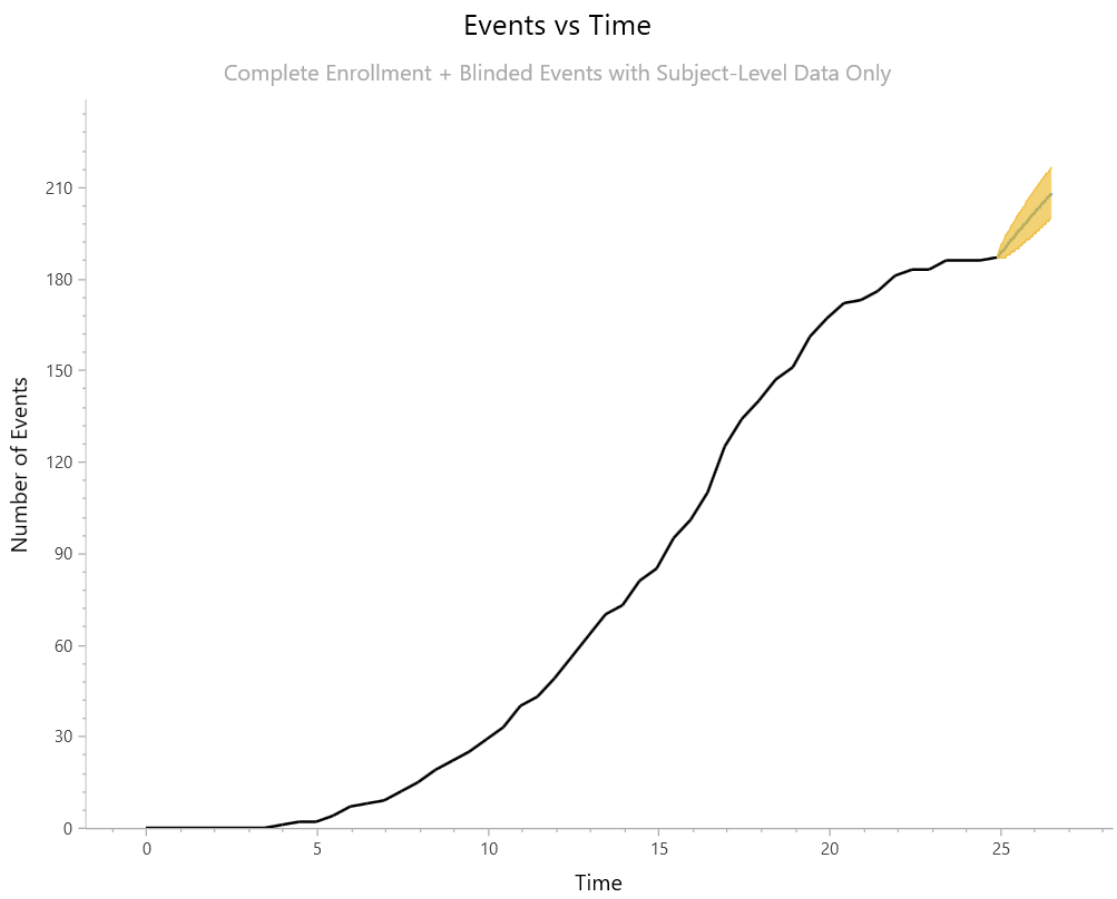



**Figure 6.124:** Events Prediction Plot for Blinded Events Prediction when Enrollment Complete

It is important the note that these event times are when these events occurred relative to the start of the study, not how long they happened after follow up i.e. this is not a Kaplan-Meier type plot!

The **Dropout Prediction** plot works effectively the same as the **Events Prediction** plot except for the dropout process. The **Dropout Prediction** plot for this demonstration is shown in Figure 6.125.

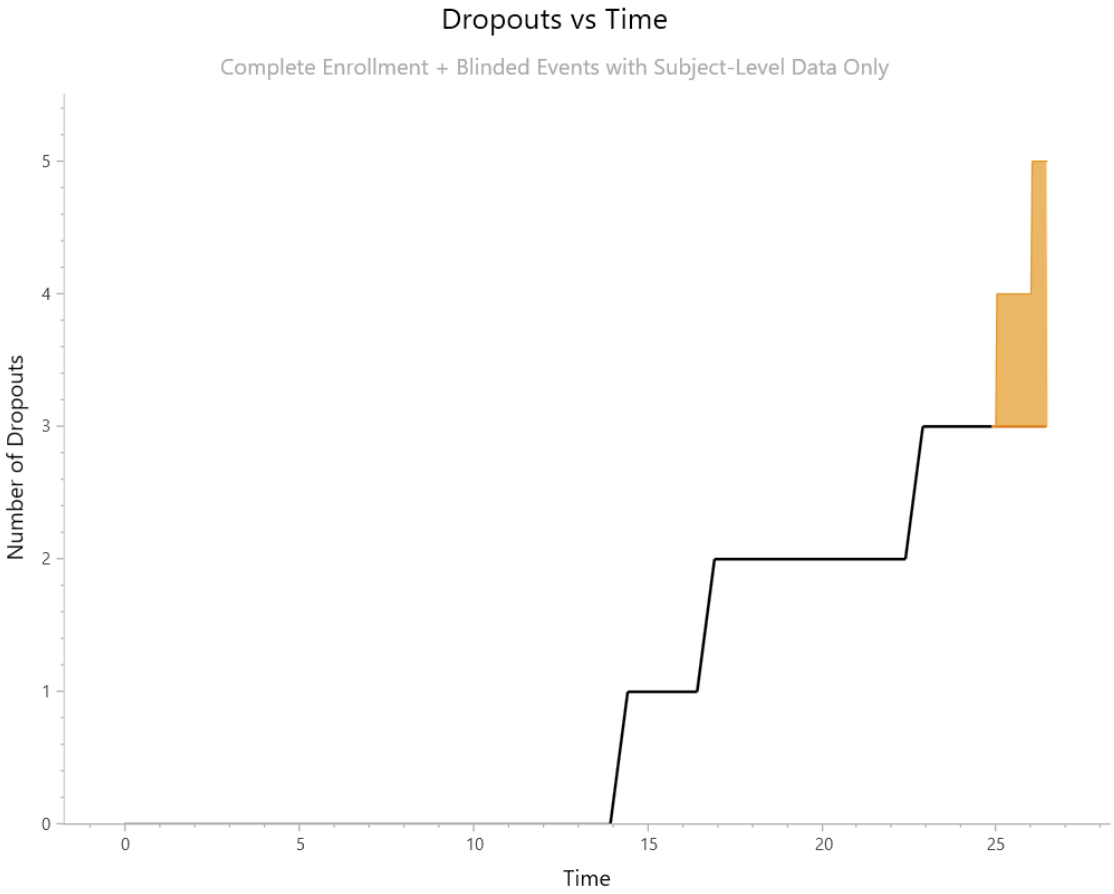
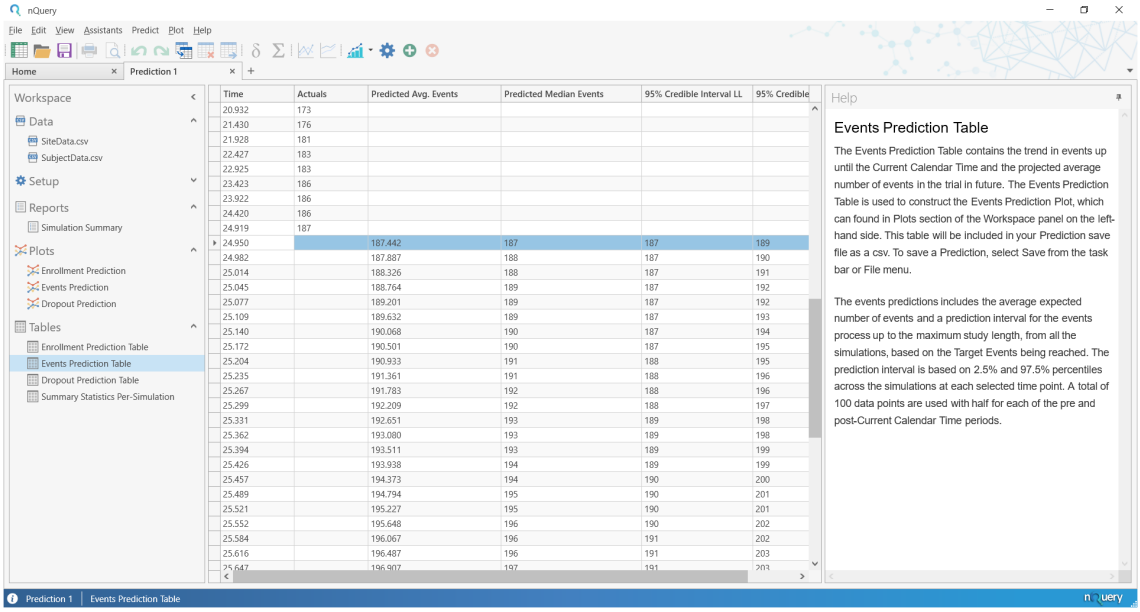



**Figure 6.125:** Dropout Prediction Plot for Blinded Events Prediction when Enrollment Complete

The **Enrollment Prediction Table, Events Prediction Table and Dropout Prediction Tables** are the tables used to create the respective **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plots, which by default all consists of ~100 rows (50 pre-simulation and 50 post-simulation). Each consists of five columns: **Time, Actual, Predicted Avg. Sample Size/Events/Dropouts, Predicted Median Sample Size/Events/Dropouts, 95% Prediction Interval LL, 95% Prediction Interval UL**. One prediction table (the **Events Prediction Table**) for this demonstration at the start of the simulated period is given in Figure 6.126.




| Time | Actuals | Predicted Avg. Events | Predicted Median Events | 95% Credible Interval LL | 95% Credible |
|---|---|---|---|---|---|
| 20.932 | 173 | | | | |
| 21.430 | 176 | | | | |
| 21.928 | 181 | | | | |
| 22.427 | 183 | | | | |
| 22.925 | 183 | | | | |
| 23.423 | 186 | | | | |
| 23.922 | 186 | | | | |
| 24.420 | 186 | | | | |
| 24.919 | 187 | | | | |
| 24.950 | | 187.442 | 187 | 187 | 189 |
| 24.982 | | 187.887 | 188 | 187 | 190 |
| 25.014 | | 188.326 | 188 | 187 | 191 |
| 25.045 | | 188.764 | 189 | 187 | 192 |
| 25.077 | | 189.201 | 189 | 187 | 192 |
| 25.109 | | 189.632 | 189 | 187 | 193 |
| 25.140 | | 190.068 | 190 | 187 | 194 |
| 25.172 | | 190.501 | 190 | 187 | 195 |
| 25.204 | | 190.933 | 191 | 188 | 195 |
| 25.235 | | 191.361 | 191 | 188 | 196 |
| 25.267 | | 191.783 | 192 | 188 | 196 |
| 25.299 | | 192.209 | 192 | 188 | 197 |
| 25.331 | | 192.651 | 193 | 189 | 198 |
| 25.362 | | 193.080 | 193 | 189 | 198 |
| 25.394 | | 193.511 | 193 | 189 | 199 |
| 25.426 | | 193.938 | 194 | 189 | 199 |
| 25.457 | | 194.373 | 194 | 190 | 200 |
| 25.489 | | 194.794 | 195 | 190 | 201 |
| 25.521 | | 195.227 | 195 | 190 | 201 |
| 25.552 | | 195.648 | 196 | 190 | 202 |
| 25.584 | | 196.067 | 196 | 191 | 202 |
| 25.616 | | 196.487 | 196 | 191 | 203 |
| 25.647 | | 196.907 | 197 | 191 | 203 |

**Figure 6.126:** Events Prediction Table for Blinded Events Prediction when Enrollment Complete

**Time** is the time value relative to the study start time of zero. **Actual** is the actual number of events achieved at a pre-simulation **Time** row. **Predicted Avg. Events** is the predicted average (mean) number of events at a post-simulation **Time** row. **Predicted Median Events** is the predicted median number of events at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the number of events at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the number of events at a post-simulation **Time** row.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Current Events, Current Dropout, Current Available, Target Sample Size, Target Events, Study End Time, Total Sample Size, Total Events, Total Dropouts, Total Censored, Accrual Duration, Accrual Duration Uncensored, Average Followup, Median Followup, Target Events Reached.**

The **Summary Statistics Per-Simulation Table** for this demonstration, focused on the simulation results, is shown in Figure 6.127.

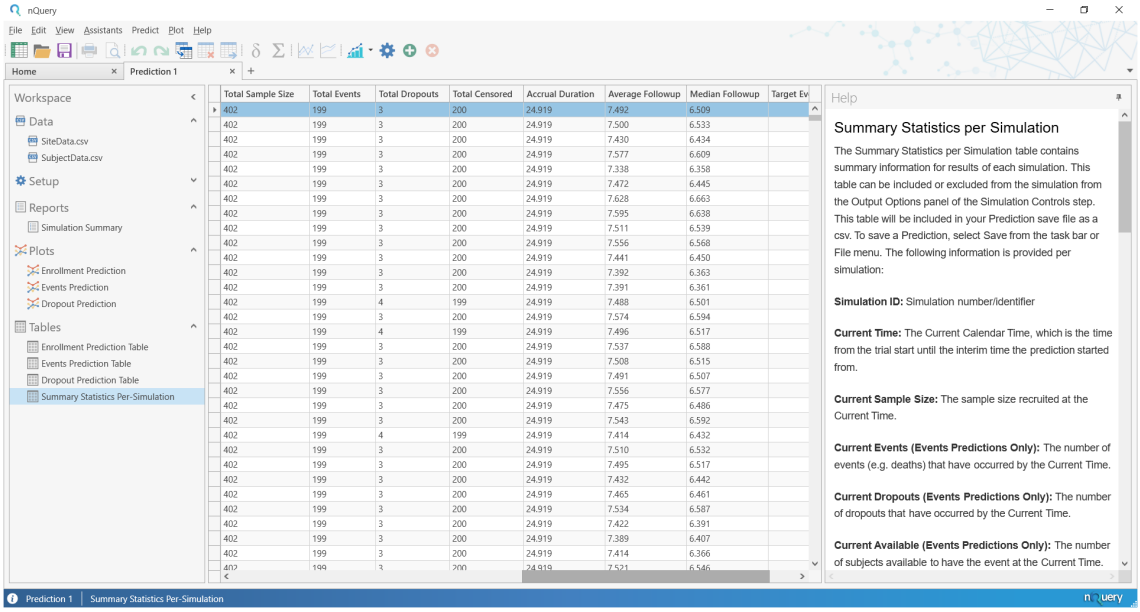



**Figure 6.127:** Summary Statistics Per-Simulation Table for Blinded Events Prediction when Enrollment Complete

Details on each field are provided in the Help window on the right but broadly the "Current" fields described the information provided up to the **Current Time**, the "Target" fields describe the study targets and the remainder give the projection achieved for key parameters in the simulation with that **Simulation ID**.

**Alternative Scenarios** The demonstration has focused on the default inputs using **SubjectData.csv** up to this point. However, as noted in the data summary given in subsection 6.2.4 the study from which the example data was simulated had an original sample size target of 460 (402 enrolled in **SubjectData.csv**) and target number of events of 374. Given that **Enrollment Status = Complete** in this demonstration, the following scenarios will evaluate the outcome if the **Target Number of Events** is set to its original value of 374. A piecewise exponential model will also be investigated.

To explore this additional information, we will investigate two additional scenarios

- Scenario A: **Target Number of Events = 374** using all other defaults from above
- Scenario B: **Target Number of Events = 374** using Piecewise Exponential **Response Distribution** (i.e. Exponential with **Number of Hazard Pieces > 1**)

**Scenario A** The only changes required compared to the first demonstration is to change the **Target Number of Events** to 374 in the **Events Model** tab at

the **Event and Dropout Information** step. The simplest way to do this is to select the **Setup** header in the Workspace Navigation Bar on the left and then select the **Event and Dropout Information** drop down option. Then edit the **Target Number of Events** to 374 and select the Next or → button. At the next **Simulation Controls** step, we will add a single simulation run by selecting the checkbox beside the **Save subject-level data for (x) simulation runs** and replace the default of 10 runs with 1 runs at the "(x)" spot. An edited down summary of the **Setup** changes is provided in Figure 6.128.

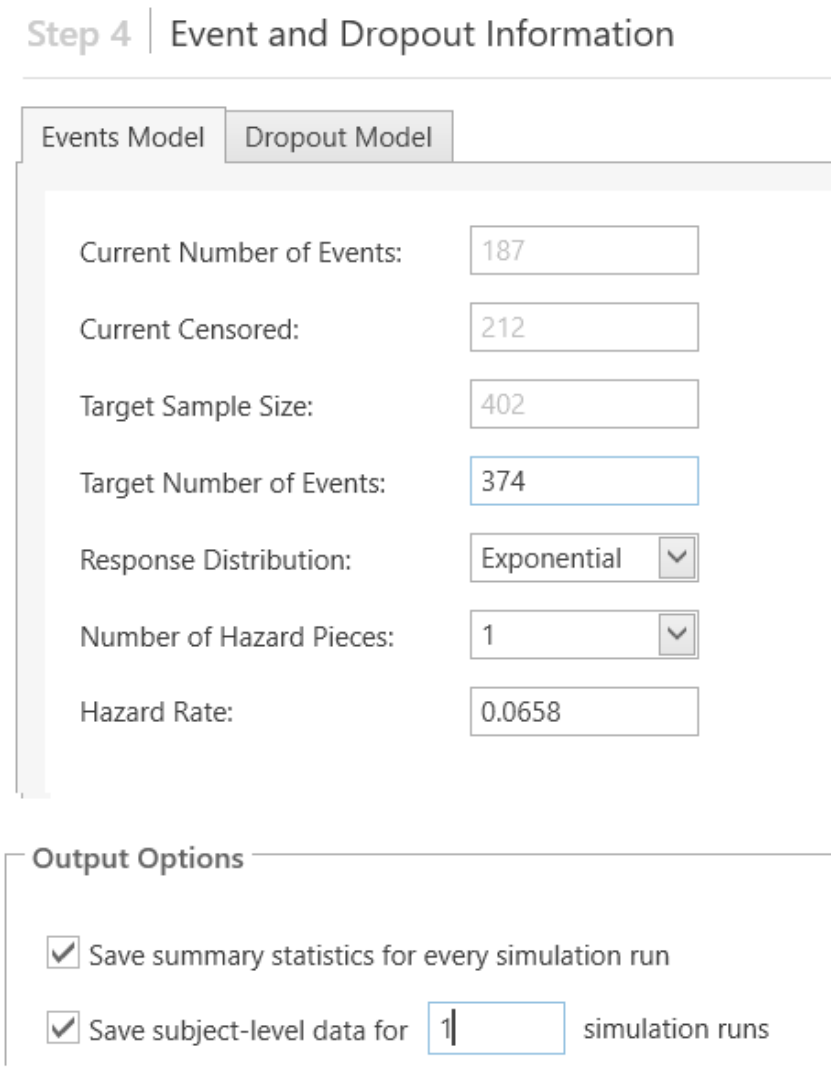



**Figure 6.128:** Scenario A Setup Changes for Blinded Events Prediction when Enrollment Complete

To run the simulation again with the new Scenario A, select the Run/→ button.

The main results of this Scenario A simulation are shown in Figure 6.129.

Overall Summary

| | |
|---|---|
| Average Sample Size | 402.0000 |
| Average Accrual Duration | 24.9186 |
| Average Study Duration | 58.5459 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.0327 |
| Average Follow-up | 14.1409 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 402 | 4 | 24.9186 | 53.5240 |
| 25.00% | 374 | 402 | 5 | 24.9186 | 56.2714 |
| 50.00% | 374 | 402 | 6 | 24.9186 | 58.3829 |
| 75.00% | 374 | 402 | 7 | 24.9186 | 60.6105 |
| 95.00% | 374 | 402 | 9 | 24.9186 | 64.1180 |

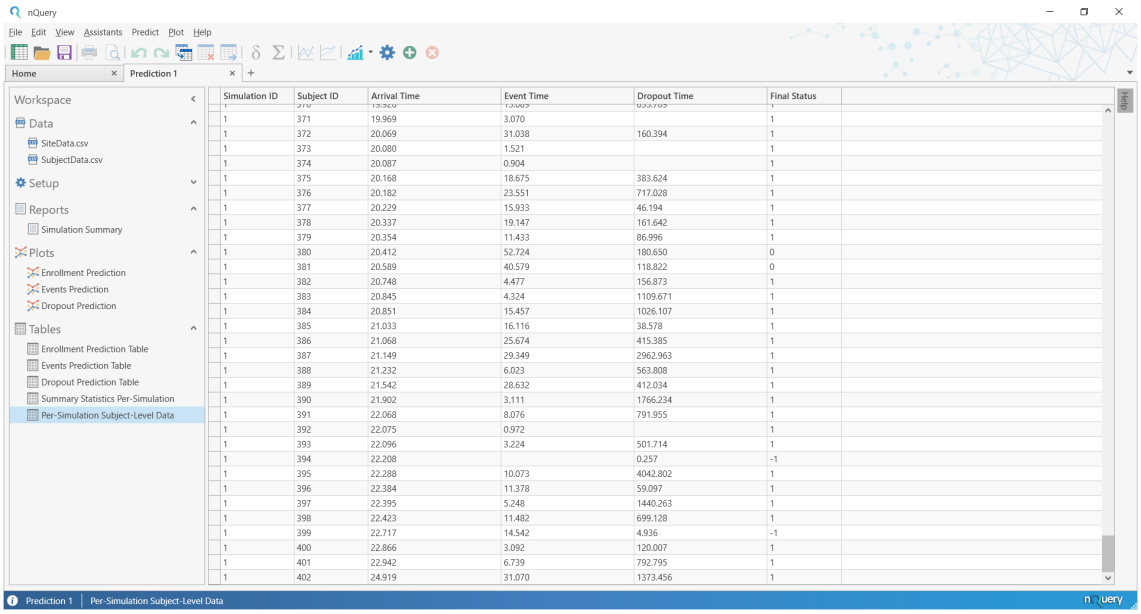
**Figure 6.129:** Scenario A Results for Blinded Events Prediction when Enrollment Complete

For Scenario A, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 58.5459. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 53.5240 (5% Percentile) and 64.1180 (95% Percentile), 50% a duration between 56.2714 (25% Percentile) and 60.6150 (75% Percentile) and a median (50% Percentile) duration of 58.3829. The **Average Dropouts** was 6.0327 (Percentiles: 4, 5, 6, 7, 9) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 14.1409.

The original target study duration was 40 months and therefore this study is expected to end 18.5 months later than expected. As expected, stopping the enrollment process prematurely causes a significant delay compared to the original plan and the expected duration based on **SubjectData.csv** if the original **Target Sample Size** of 460 had been recruited (see Scenario A for subsubsection 6.2.1.1)

The **Per-Simulation Subject-level Data Table** was also generated here and is shown for Scenario A in Figure 6.130 with the first simulation (**Simulation ID**

= **1**). The tables consists of the following fields: **Simulation ID**, **Subject ID**, **Arrival Time**, **Event Time**, **Dropout Time** and **Final Status**. The order of the Subject ID is based on **Arrival Time** order.




| Simulation ID | Subject ID | Arrival Time | Event Time | Dropout Time | Final Status |
|---|---|---|---|---|---|
| 1 | 371 | 19.969 | 3.070 | | 1 |
| 1 | 372 | 20.069 | 31.038 | 160.394 | 1 |
| 1 | 373 | 20.080 | 1.521 | | 1 |
| 1 | 374 | 20.087 | 0.904 | | 1 |
| 1 | 375 | 20.168 | 18.675 | 383.624 | 1 |
| 1 | 376 | 20.182 | 23.551 | 717.028 | 1 |
| 1 | 377 | 20.229 | 15.933 | 46.194 | 1 |
| 1 | 378 | 20.337 | 19.147 | 161.642 | 1 |
| 1 | 379 | 20.354 | 11.433 | 86.996 | 1 |
| 1 | 380 | 20.412 | 52.724 | 180.650 | 0 |
| 1 | 381 | 20.589 | 40.579 | 118.822 | 0 |
| 1 | 382 | 20.748 | 4.477 | 156.873 | 1 |
| 1 | 383 | 20.845 | 4.324 | 1109.671 | 1 |
| 1 | 384 | 20.851 | 15.457 | 1026.107 | 1 |
| 1 | 385 | 21.033 | 16.116 | 38.578 | 1 |
| 1 | 386 | 21.068 | 25.674 | 415.385 | 1 |
| 1 | 387 | 21.149 | 29.349 | 2962.963 | 1 |
| 1 | 388 | 21.232 | 6.023 | 563.808 | 1 |
| 1 | 389 | 21.542 | 28.632 | 412.034 | 1 |
| 1 | 390 | 21.902 | 3.111 | 1766.234 | 1 |
| 1 | 391 | 22.068 | 8.076 | 791.955 | 1 |
| 1 | 392 | 22.075 | 0.972 | | 1 |
| 1 | 393 | 22.096 | 3.224 | 501.714 | 1 |
| 1 | 394 | 22.208 | | 0.257 | -1 |
| 1 | 395 | 22.288 | 10.073 | 4042.802 | 1 |
| 1 | 396 | 22.384 | 11.378 | 59.097 | 1 |
| 1 | 397 | 22.395 | 5.248 | 1440.263 | 1 |
| 1 | 398 | 22.423 | 11.482 | 699.128 | 1 |
| 1 | 399 | 22.717 | 14.542 | 4.936 | -1 |
| 1 | 400 | 22.866 | 3.092 | 120.007 | 1 |
| 1 | 401 | 22.942 | 6.739 | 792.795 | 1 |
| 1 | 402 | 24.919 | 31.070 | 1373.456 | 1 |

**Figure 6.130:** Per-Simulation Subject-level Data Table for Blinded Events when Enrollment Complete

**Event Time** and **Dropout Time** are generated for all subjects which were **Censored Status** (all status values from **Status Indicator** field in **Subject-level Data Setup** step) in the subject-level data (for example **Subject ID = 402** in this example). Blank fields for **Event Time** or **Dropout Time** indicate cases where the subject already had the event/dropout and therefore the time for the other status does not exist and does not need to be simulated.

The **Final Status** is from the subject-level data for subjects with **Event Status** or **Dropout Status** and based on the **Event Time** and **Dropout Time** for simulated subjects and **Censored Status** subjects in the subject-level data. **Final Status** will equal **Event Status** if **Event Time** is less than **Dropout Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.127 for per-simulation end times). **Final Status** will equal **Dropout Status** if **Dropout Time** is less than **Event Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.127 for per-simulation end times). **Final Status** will equal **Censored Status** if **Event Time** and **Dropout Time** are greater than the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.127 for per-simulation end times)

**Scenario B** Scenario A assumed an **Exponential Response Distribution** which had a constant hazard rate (i.e. **Number of Hazard Pieces = 1**). However, a piecewise **Exponential** model can also be used where the hazard rate changes over time. There are methods available for the automatic fitting of piecewise exponential curves [Fang and Su, 2011] but here a simple case where the hazard rate halves after 5 months of follow up is used.

To evaluate this scenario, from the **Setup** header in the Workspace Navigation Bar on the left select the **Event and Dropout Information** drop down option. Set the **Response Distribution** to **Exponential** and the **Number of Hazard Pieces** to 3. The piecewise exponential model for this case will equal the following:

| Piece # | Starting at Time | Hazard Rate |
|---|---|---|
| 1 | 0 | 0.065789598 |
| 2 | 5 | 0.032894799 |

**Table 6.1:** Piecewise Exponential Fit for Scenario B

In nQuery Predict, this will be as per Figure 6.131.

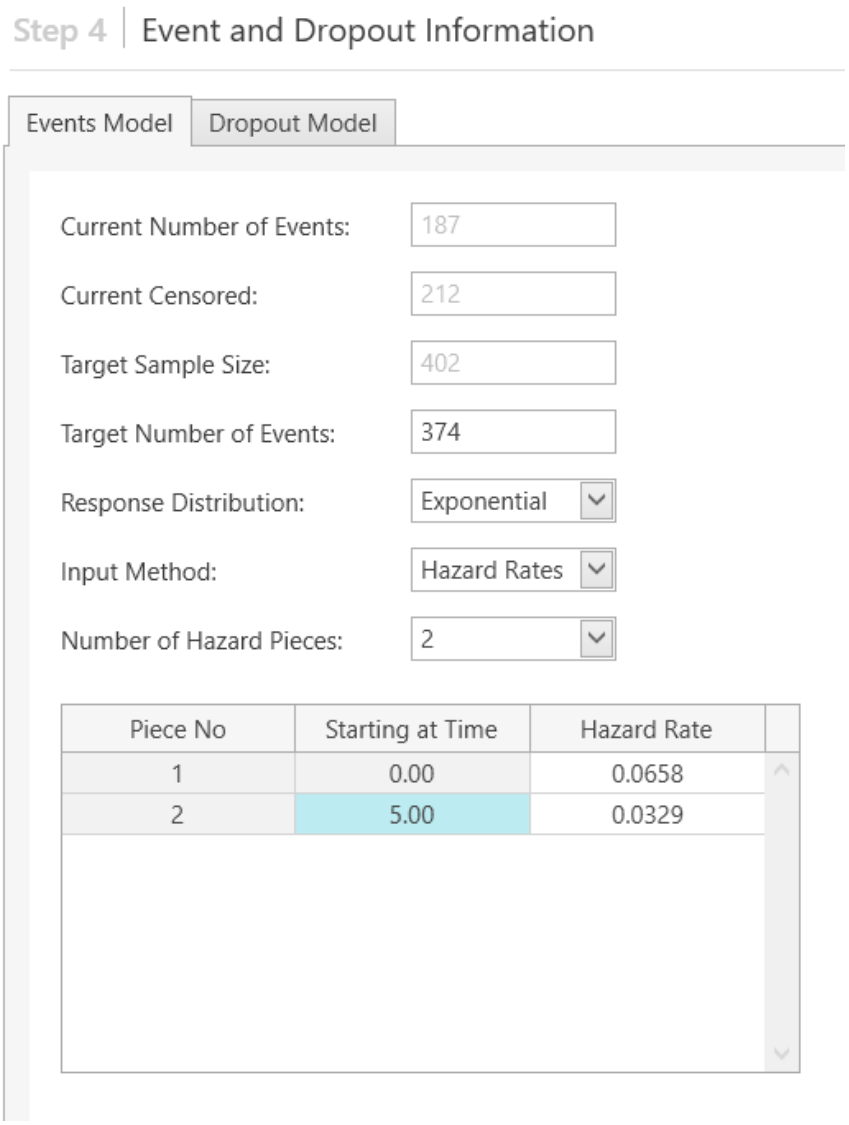



**Figure 6.131:** Scenario B for Blinded Event Prediction when Enrollment Complete

Select the Run or → button to move to the next step. At the next **Simulation Controls** step, select the Run or → button to simulate Scenario B. The changes and results for Scenario B are summarized in

Overall Summary

| | |
|---|---|
| Average Sample Size | 402.0000 |
| Average Accrual Duration | 24.9186 |
| Average Study Duration | 89.1947 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 8.0357 |
| Average Follow-up | 18.9834 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 402 | 5 | 24.9186 | 78.2691 |
| 25.00% | 374 | 402 | 6 | 24.9186 | 84.1662 |
| 50.00% | 374 | 402 | 8 | 24.9186 | 88.8058 |
| 75.00% | 374 | 402 | 9 | 24.9186 | 93.6197 |
| 95.00% | 374 | 402 | 12 | 24.9186 | 101.4556 |

**Figure 6.132:** Results for Scenario B for Blinded Event Prediction when Enrollment Complete

For Scenario B, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 89.1947 (Percentiles: 78.2691, 84.1662, 88.8058. 93.6197, 101.4556). The **Average Dropouts** was 8.0357 (Percentiles: 5, 6, 8, 9, 12) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 18.9834. This gives an increase of around 32 months over the constant exponential case which shows that the study length is sensitive to the assumed survival model.

#### 6.2.7.4 Unblinded Events Prediction with Summary Data with Enrollment Ongoing

**Setup** **Unblinded Events** milestone prediction using **Summary Data** means that the prediction will project the expected length of study needed to achieve a specified **Target Number of Events** using summary data fixed parameters from each treatment group.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step

select **Unblinded Events** from the **Target** field, **Summary Data** from the **Input** field and **Ongoing** from the **Enrollment Status** field.

These selections are shown in Figure 6.133. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

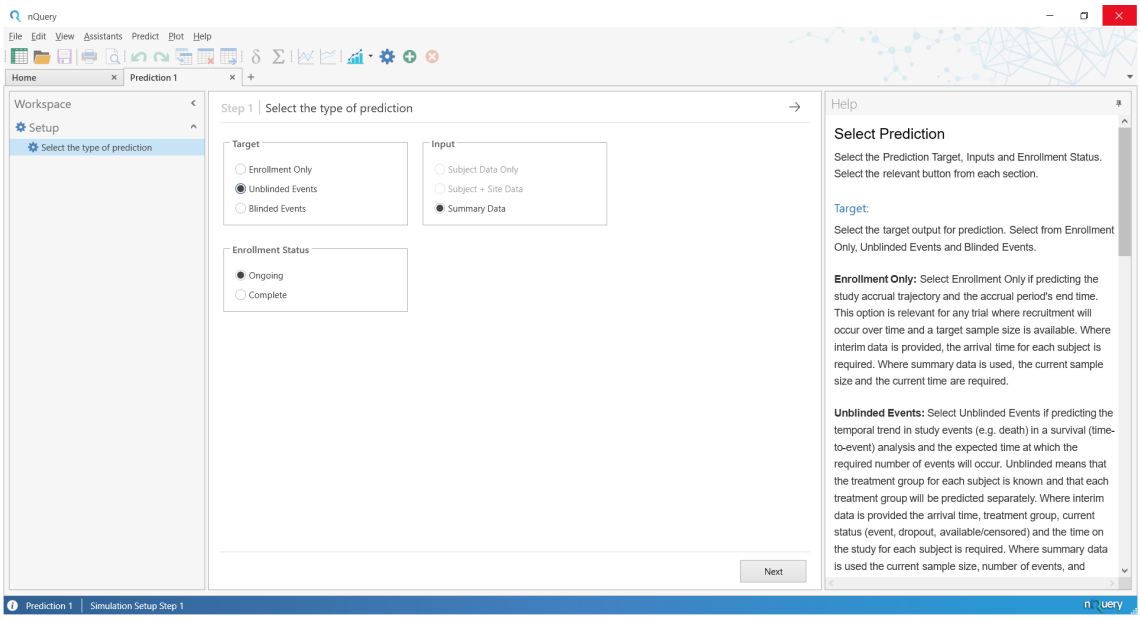



**Figure 6.133:** Select the Type of Prediction Setup Step for Unblinded Events using Summary Data

The next step is the **Fixed Parameters** step. For this demonstration, there is a **Sample Sizes, Events & Dropouts** section which contains nine fields which are as follows:

- Current Sample Size (Control): The sample size enrolled into the control group at the Current Time
- Current Sample Size (Treatment): The sample size enrolled into the treatment group at the Current Time
- Current No. of Events (Control): The number of events that have occurred in the control group at the Current Time
- Current No. of Events (Treatment): The number of events that have occurred in the treatment group at the Current Time
- Current No. of Dropouts (Control): The number of dropouts that have occurred in the control group at the Current Time
- Current No. of Dropouts (Treatment): The number of dropouts that have occurred in the treatment group at the Current Time

- Current No. Censored (Control) (Read-only): The subjects censored (not had event or dropout) in the control group if the study ended at the Current Time
- Current No. Censored (Treatment) (Read-only): The subjects censored (not had event or dropout) in the treatment group if the study ended at the Current Time
- Current Time: The length of time the study has been on-going since the study began

For this demonstration, inputs are selected which mirror those from Scenario A of the **Unblinded Events Prediction with Subject + Site Data with Enrollment Ongoing (subsubsection 6.2.7.2)** example using **SubjectData.csv.**

Based on **SubjectData.csv,** the fixed parameters for this example give a **Current Sample Size (Control)** of 203, **Current Sample Size (Treatment)** of 199, **Current No. of Events (Control)** of 106, **Current No. of Events (Treatment)** of 81, **Current No. of Dropouts (Control)** of 1, **Current No. of Dropouts (Treatment)** of 2 and a **Current Time** of 24.9186. The **Current No. Censored (Control)** is automatically calculated as 96 and the **Current No. Censored (Treatment)** is automatically calculated as 116.

The **Fixed Parameters** step with these values inputted is shown in Figure 6.134. After all fields all filled, select the Next or → button to move to the next **Setup** step.

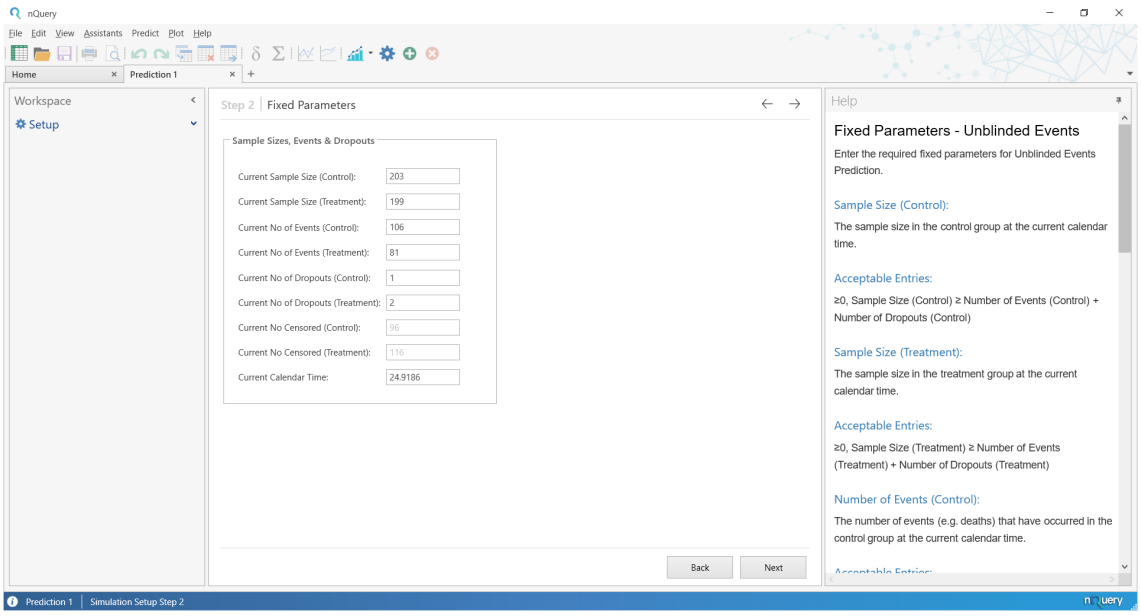



**Figure 6.134:** Fixed Parameters Setup Step for Unblinded Events using Summary Data

The next step is the **Accrual Options** step which specifies the **Target Sample**

**Size**. The **Accrual Options** step, and its defaults, for this demonstration are shown in Figure 6.135.

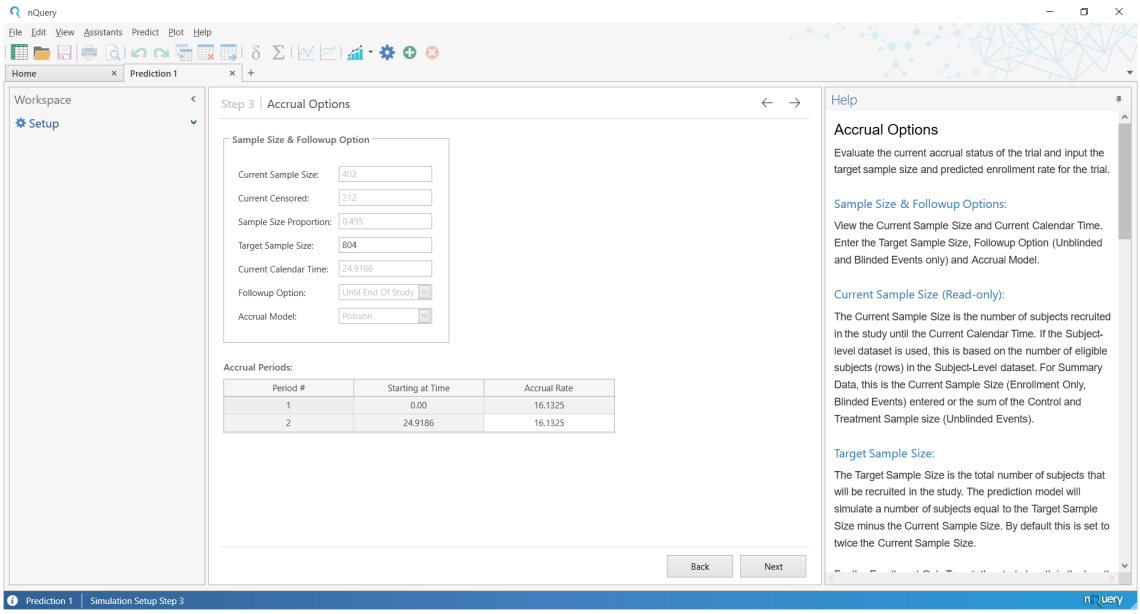



**Figure 6.135:** Accrual Options Setup Step Defaults for Unblinded Events Prediction using Summary Data

The **Accrual Options** step consists of two main elements: the **Sample Size & Followup** input fields and the **Accrual Periods** table.

The **Sample Size & Followup** input fields provides information on enrollment process, censored status and current time based on the subject-level data and allows the user to edit the **Target Sample Size** and select the **Followup Option** and **Accrual Model**. The following fields are provided in this prediction:

- Current Sample Size (Read-only): Sample Size recruited into a study so far. Equals the number of rows in the Subject-level Data
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Sample Size Proportion (Read-only): Proportion of subjects in the Control Group. Based on number of Control Group subjects in Treatment ID field. Controls the allocation proportion in simulated data
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study
- Current Time (Read-only): The length of time that has passed since the study started. Equals the maximum of the sum of the Arrival Time + Time on Study values for each subject

- Followup Option (Read-only): Specifies whether subjects who do not have the event/dropout will be administratively censored at the end of the study or after a fixed follow up period. For Summary Data, this is set to Until End of Study
- Accrual Model (Editable): The statistical model that will be used to generate simulated enrollments

In nQuery Predict, the **Target Sample Size** defaults to twice the **Current Sample Size** and the **Accrual Model** defaults to **Poisson.** In this demonstration, the **Current Sample Size** equals 402 and therefore the default **Target Sample Size** is 402 (2) = 804. However, in this demonstration the **Target Sample Size** is set to 460.

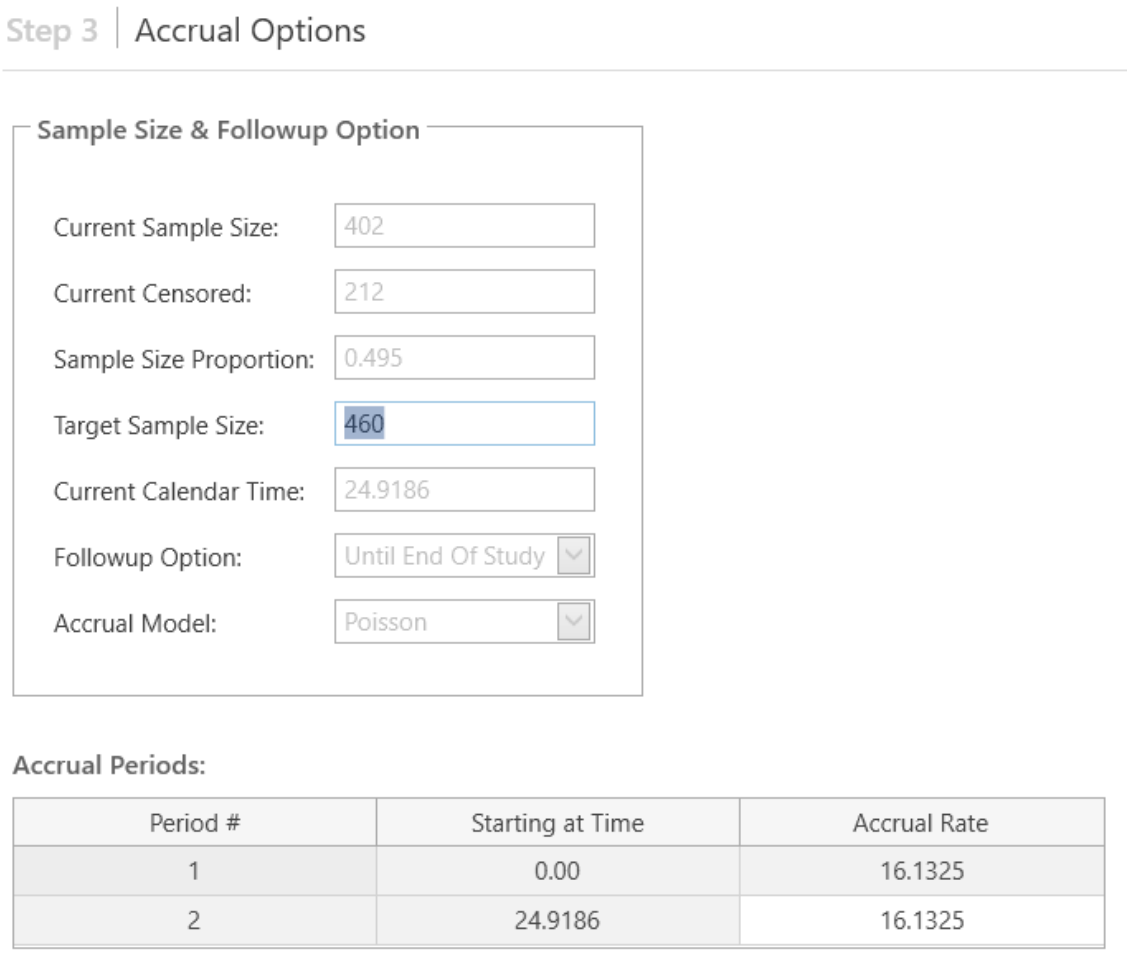
When **Input = Summary Data,** the **Followup Option** must be set to **Until End of Study.**

The **Accrual Periods** table contains information on the **Accrual Rate** up the **Current Time** and allows the user to edit the future **Accrual Rate** used to generate enrollment simulations. The **Accrual Periods** has three columns ,with each row corresponding to a time period, with the columns defined as follows:

- Period #: A numeric ID for the current time period. Increases in increments of one for each subsequent row
- Starting at Time: The starting time for the current time period row. The first row will equal 0 (rate from study start to Current Time) and the second row will equal the Current Time
- Accrual Rate: The accrual rate is the average number of subjects recruited per unit time (months). The accrual rate in the first row will equal the accrual rate up to the Current Time.

For the **Poisson Accrual Model**, the **Accrual Periods** table will consist of two time period rows. The first row will correspond to the information provided by the user regarding the study up until the **Current Time.** The second row will correspond to the inputs used to generate future enrollments. By default the **Accrual Rate** in the future second row will equal the **Accrual Rate** from the first row i.e. the accrual rate up until the **Current Time.** For the inputs used in this demonstration, this **Accrual Rate** column values in each row equal 16.13254936. The user can edit the **Accrual Rate** for generating future enrollments by editing the second cell of the **Accrual Rate** column.

In this demonstration, **Target Sample Size** = 460 and **Accrual Rate** =16.13254936 as per Figure 6.136. Select the Next or → button to move to the next **Setup** step.




| Period # | Starting at Time | Accrual Rate |
|---|---|---|
| 1 | 0.00 | 16.1325 |
| 2 | 24.9186 | 16.1325 |

**Figure 6.136:** Accrual Options Setup Step Final Inputs for Unblinded Events Prediction using Summary Data

The next step is the **Event and Dropout Information** step which specifies the **Target Number of Events,** event model (and associated statistical parameters) and dropout model (and associated statistical parameters). The **Event and Dropout Information** step on the default **Events Model** tab is shown in Figure 6.137.

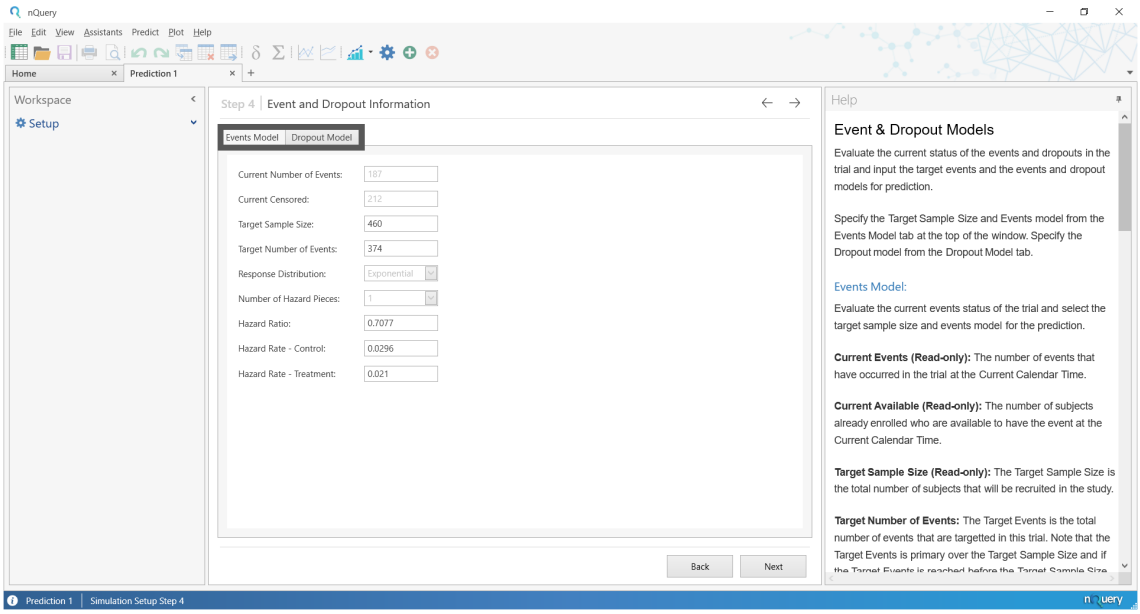



**Figure 6.137:** Event and Dropout Information Setup Step for Unblinded Events Prediction using Summary Data

The **Event and Dropout Information** step consists of two tabs: the **Events Model** tab and the **Dropout Model** tab. These tabs have been highlighted in Figure 6.137 and are opened by selecting the required tab name at the top of the main window, below the step title. For **Unblinded Events**, it is assumed that survival and dropout times will be simulated individually for each treatment group using their own user specified model.

The **Events Model** tab provides fields which provide information on the current number of subject who have had an event, could have an event going forward and allows the input of the **Target Number of Events** and **Response Distribution** for the events model and additional parameters needed for that specific model. The following fields are provided in this prediction:

- Current Events (Read-only): The number of subjects who had the event of interest at the current time. Equals the number of subjects with the Event Status in the Status Indicator field
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Target Sample Size (Read-only): Total number of subjects that will be recruited in this study
- Target Number of Events (Editable): The required number of subjects needed to have the event of interest for the study to end

- Response Distribution: The statistical model that will be used to generate survival times for subjects available to have the event
- Number of Hazard Pieces (Response Distribution = Exponential only): Number of hazard pieces used to specify the piecewise exponential model. Set to 1 for Input = Summary Data
- Hazard Ratio (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The hazard ratio between the Treatment and Control Hazard Rates
- Hazard Rate - Control (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The control group's exponential hazard rate for the constant exponential survival model specified above
- Hazard Rate - Treatment (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The treatment group's exponential hazard rate for the constant exponential survival model specified above

When **Input = Summary Data**, the only **Response Distribution** is **Exponential** (**Number of Hazard Pieces** = 1) and the default **Hazard Rates** set at the expected event rate for each group from the provided **Fixed Parameters**.

In nQuery Predict when **Input = Summary Data**, the default **Hazard Rates** are calculated based on the assumption that all recruitment in a group occurred at the start of the trial and using the implied survival rate from the **Current Sample Size** (N) and **Current No. of Events** (E) i.e. 1 - E/N. The **Survival Parameter Converter** tool from the Assistants menu can be used to show this as per Figure 6.138.

STT00-1 / Conversion to alternate rates for exponential survival curves

| | Control | Treatment | HR |
|---|---|---|---|
| **Time t** | 24.91860000 | 24.91860000 | |
| **Group 1 Proportion $\pi_1$ at Time t** | 0.47800000 | 0.59300000 | |
| **Group 1 Median Survival, $med_1$** | 23.39955963 | 33.05310060 | 33.05310060 |
| **Group 1 Exponential Parameter, $\lambda_1$** | 0.02962223 | 0.02097072 | 0.02097072 |
| **Group 2 Proportion $\pi_2$ at Time t** | | | |
| **Group 2 Median Survival, $med_2$** | | | 23.39955963 |
| **Group 2 Exponential Parameter, $\lambda_2$** | | | 0.02962223 |
| **Hazard Ratio, $h=\ln(\pi_1)/\ln(\pi_2)=med_2/med$** | | | 0.70793841 |

Notes

Calibri 11 B I

User Notes for STT00-1

Control Sample Size = 203
Control Events = 106
=> Control P(Survival) = 1 - 106/203 = 0.478

Treatment Sample Size = 199
Treatment Events =81
=> Trmt P(Survival) = 1 - 106/203 = 0.593

**Figure 6.138:** Default Hazard Rate Calculation using Summary Data

The constant **Exponential** model (i.e. **Number of Hazards Pieces** = 1) is the only model provided as the other (conditional) models available in nQuery Predict rely on subject-level survival times being available, which are obviously not available for hypothetical subjects assumed to have had the event using fixed parameters inputs.

The default **Target Number of Events** is double the **Current Events**, which for this example is 374, with default estimated exponential **Hazard Rate - Control** of 0.0296, **Hazard Rate - Treatment** of 0.021 (**Hazard Ratio = 0.7077**).

The **Dropout Model** tab is opened by selecting the **Dropout Model** tab name at the top of the main window. For this demonstration, the **Dropout Model** tab is shown in Figure 6.139.

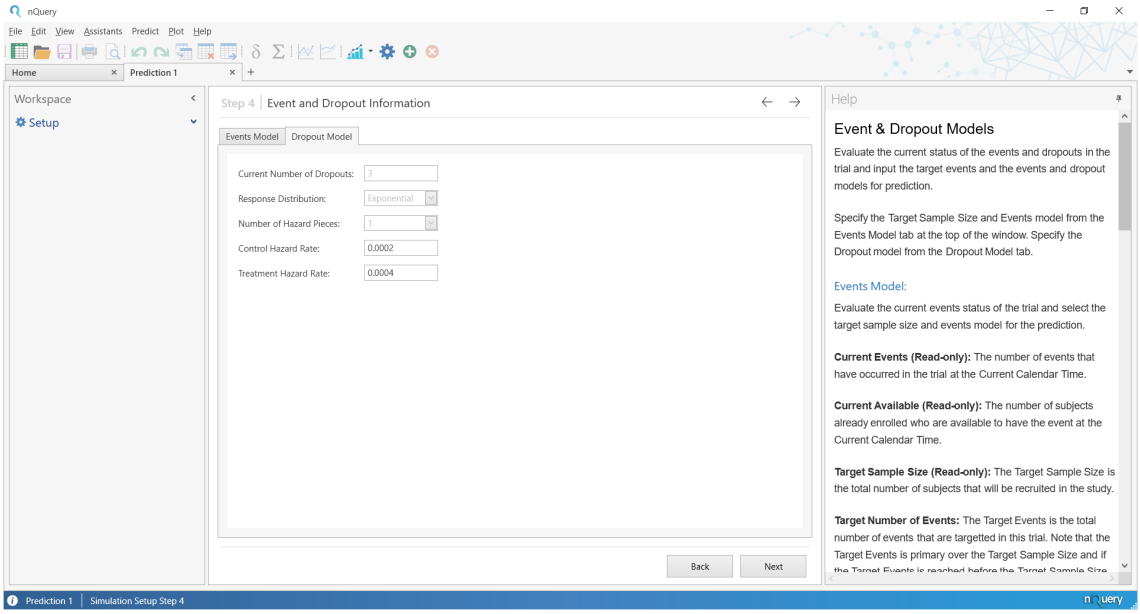



**Figure 6.139:** Dropout Model tab for Unblinded Events using Summary Data

For **Dropout Model**, the **Current Number of Dropouts** (number of subjects who have dropped out up to the **Current Time**), **Response Distribution** (**Number of Hazard Pieces** must equal 1 as per **Event Model** tab). For dropout, the only **Response Distribution** option is **Exponential**. The inputs and defaults for dropout **Exponential Response Distribution** are the same as from the **Events Model** described above and therefore will not be replicated here. In this demonstration, the constant exponential **Hazard Rate - Control** and **Hazard Rate - Treatment** for dropout calculated from the information provided equal 0.0002 and 0.0004 respectively. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.140. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

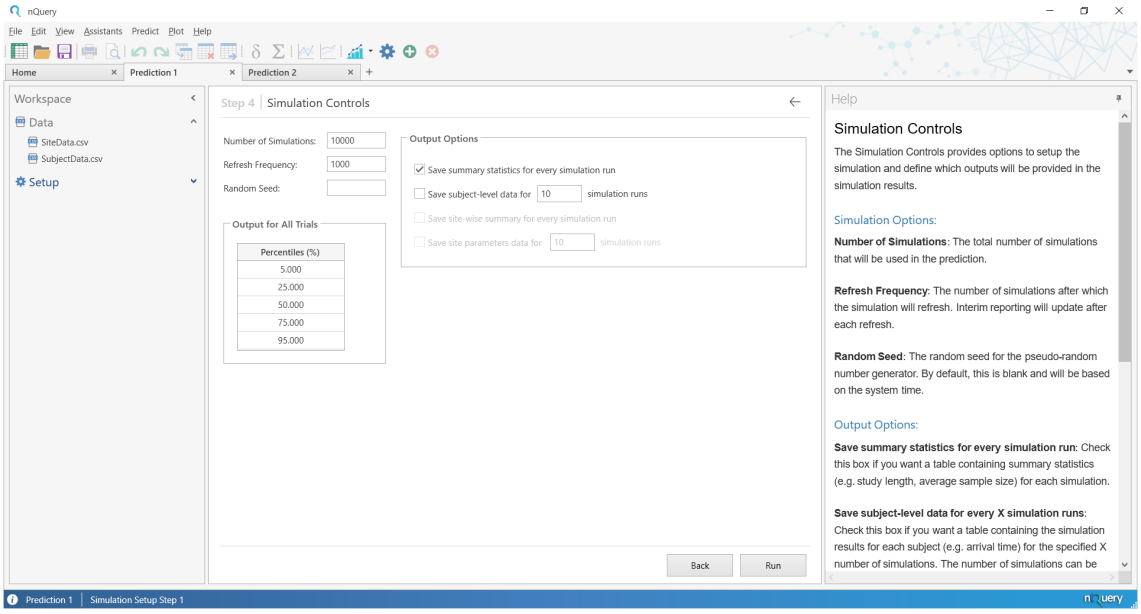



**Figure 6.140:** Simulation Controls Setup Step for Unblinded Events Prediction using Summary Data

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.141.

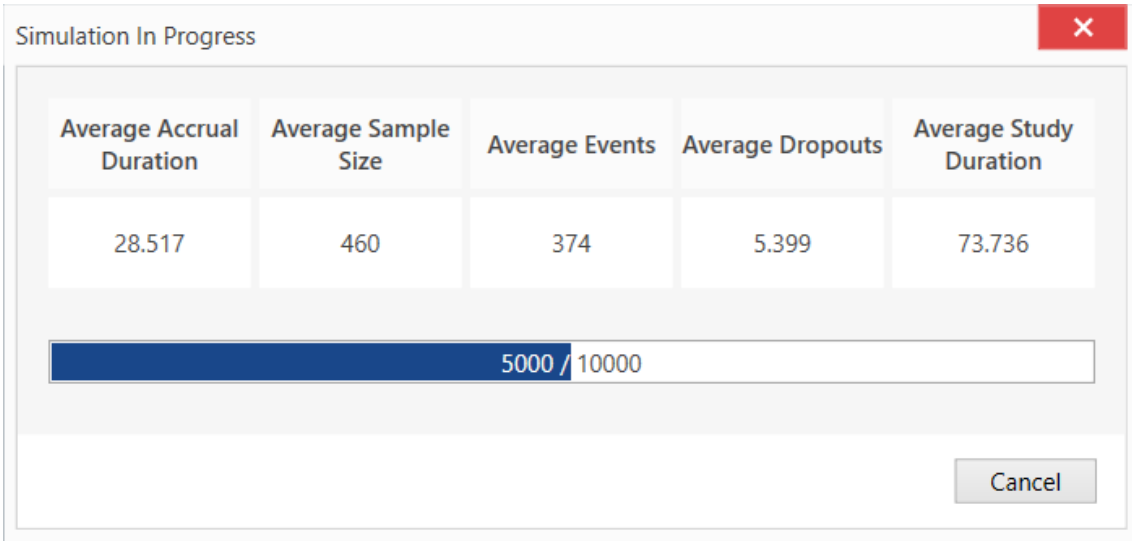



**Figure 6.141:** Simulation in Progress Window for Unblinded Events Prediction using Summary Data

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.142.

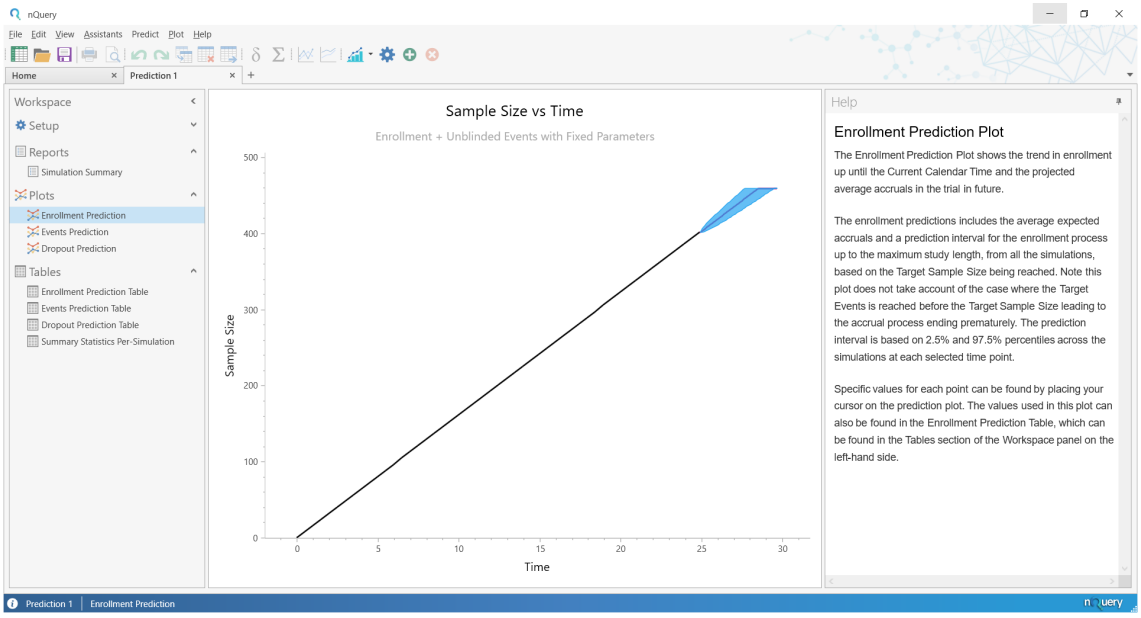



**Figure 6.142:** Enrollment Prediction Plot for Unblinded Events Prediction using Summary Data

The results of a nQuery Predict milestone prediction can be divided into three

categories: **Reports, Plots** and **Tables**. A high-level summary of the results in nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table, Events Prediction Table, Dropout Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.143.

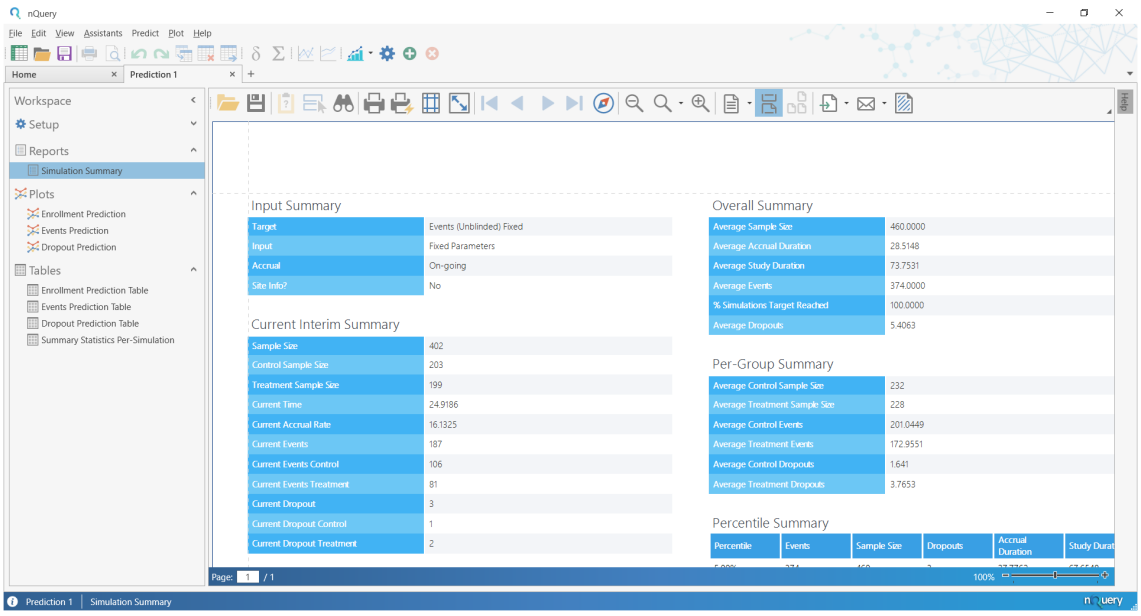



**Figure 6.143:** Simulation Summary Report for Unblinded Events Prediction using Summary Data

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.144. For this demonstration, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 73.7531. The **Percentile Summary** shows that 90% of predictions had

a **Study Duration** between 67.6548 (5% Percentile) and 80.2711 (95% Percentile), 50% a duration between 71.1336 (25% Percentile) and 76.2677 (75% Percentile) and a median (50% Percentile) duration of 73.6307.

The **Average Dropouts** was 5.4063 (Percentiles: 3, 4, 5, 6, 8) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 5.4063. For the enrollment process, the **Average Accrual Duration** was 28.5148 (Percentiles: 27.7762, 28.1781, 28.4940, 28.8244, 29.3288) with the **Target Sample Size** of 460 reached in all simulations. The **Control Group** had a **Average Sample Size** 232, **Average Events** of 201.0449 and **Dropouts** of 1.641. The **Treatment Group** had a **Average Sample Size** 228, **Average Events** of 172.9551 and **Dropouts** of 3.7653.

### Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 28.5148 |
| Average Study Duration | 73.7531 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 5.4063 |

### Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 232 |
| Average Treatment Sample Size | 228 |
| Average Control Events | 201.0449 |
| Average Treatment Events | 172.9551 |
| Average Control Dropouts | 1.641 |
| Average Treatment Dropouts | 3.7653 |

### Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 3 | 27.7762 | 67.6548 |
| 25.00% | 374 | 460 | 4 | 28.1781 | 71.1336 |
| 50.00% | 374 | 460 | 5 | 28.4940 | 73.6307 |
| 75.00% | 374 | 460 | 6 | 28.8244 | 76.2677 |
| 95.00% | 374 | 460 | 8 | 29.3288 | 80.2711 |

**Figure 6.144:** Simulation Summary Report Results for Unblinded Events Prediction using Summary Data

To edit the **Simulation Summary report there a** number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.145.

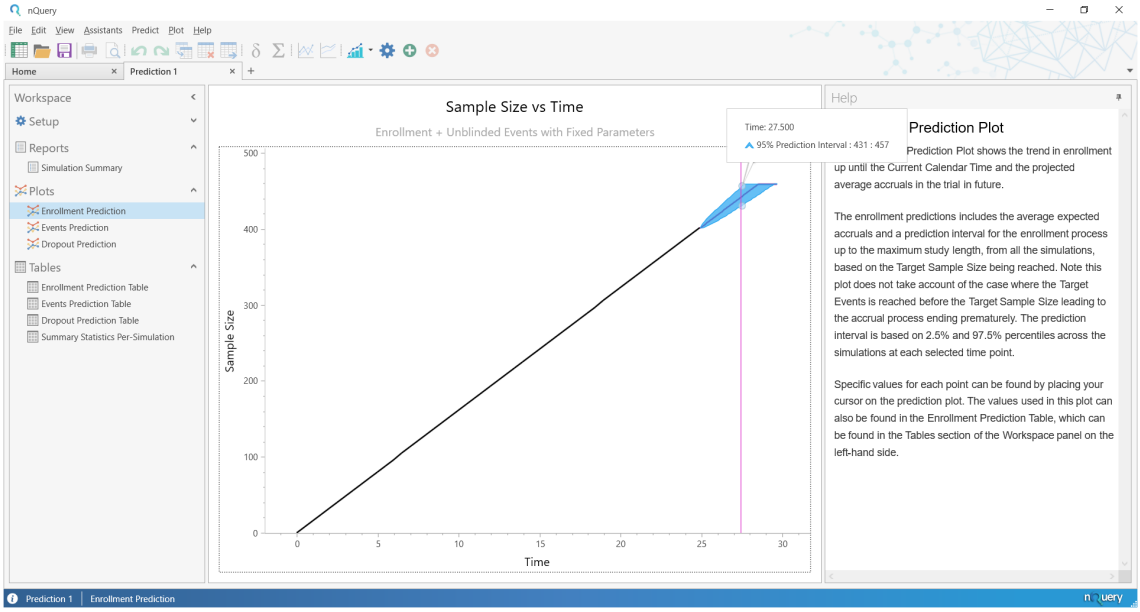



**Figure 6.145:** Enrollment Prediction Plot for Unblinded Events Prediction using Summary Data

The **Enrollment Prediction** plot when **Input = Summary Data** will assume a uniform (linear) accrual pattern up to the **Current Time**, with the assumed sample size at a given prior time given by placing the cursor over the plot. After the **Current Time,** a blue conic section with a grey line where the blue conic section plots the 95% prediction interval for the **Events** achieved at a given **Time** across all simulations and the grey line is the average **Sample Size** enrolled at a given time. Note the **Enrollment Prediction** plot **Time** scale will be related to the **Accrual Duration** rather than the **Event Prediction** and **Dropout Prediction** plots which will be associated with the **Study Duration.**

The **Events Prediction** plot consists of two parts: from Time = 0 to **Current Time** and from **Current Time** to the **Maximum Time**. The first part is a black line which plots the events pattern from the user inputs provided, which for **Input = Summary Data** will be a linear constant event rate. The second part is a yellow conic section with a grey line where the yellow conic section plots the 95% prediction interval for the **Events** achieved at a given **Time** across all simulations and the grey line is the average **Events** at a given time. The 95% prediction interval indicates the 95% of predictions had a number of events that fell between the lower and upper

limit at a given time. The **Events Prediction** plot for this demonstration is shown in Figure 6.146.

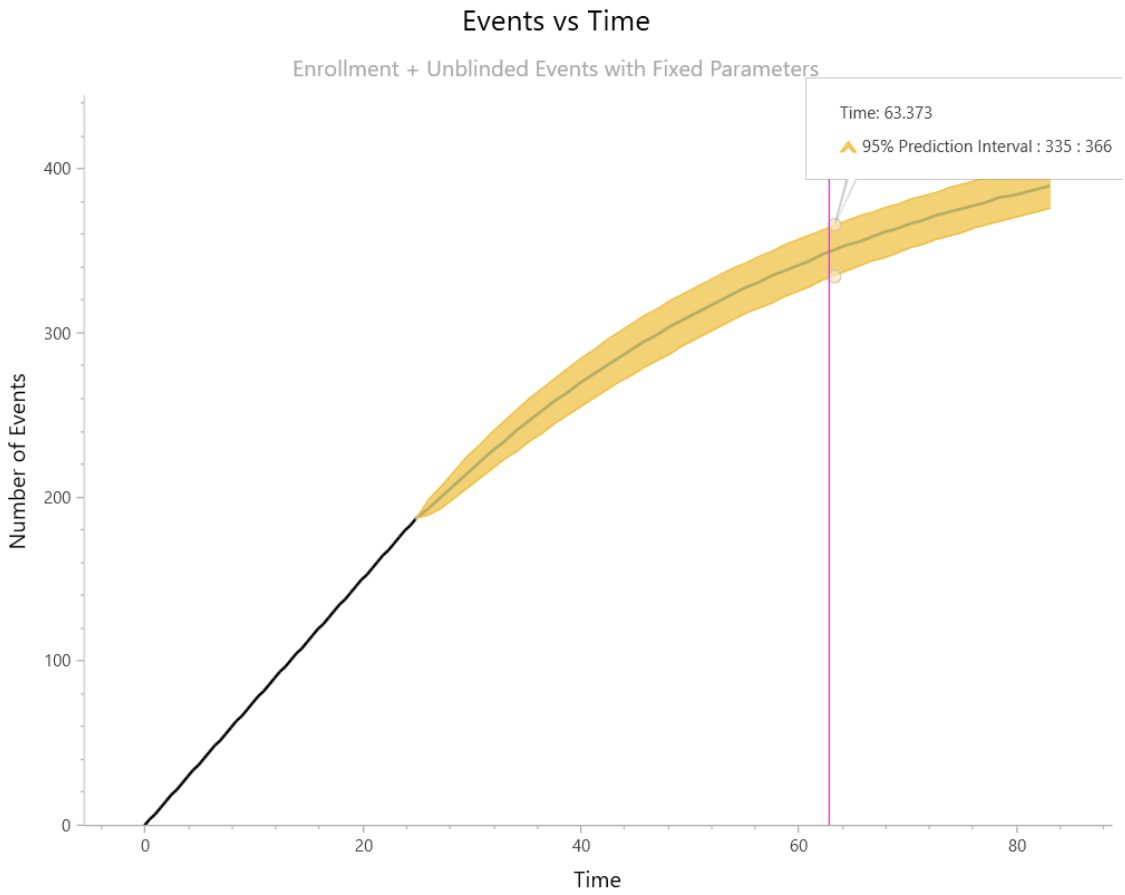



**Figure 6.146:** Events Prediction Plot for Unblinded Events Prediction using Summary Data

The **Dropout Prediction** plot works effectively the same as the **Events Prediction** plot except for the dropout process. The **Dropout Prediction** plot for this demonstration is shown in Figure 6.147.

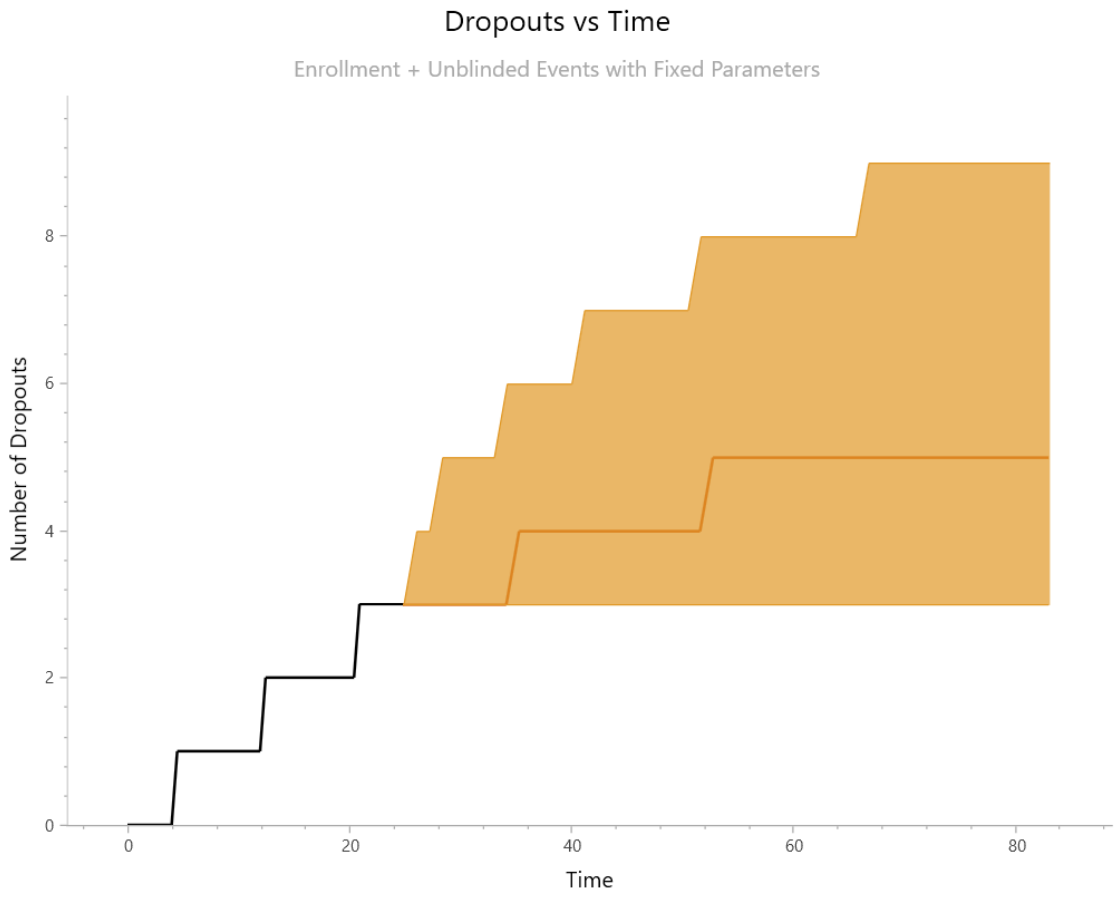
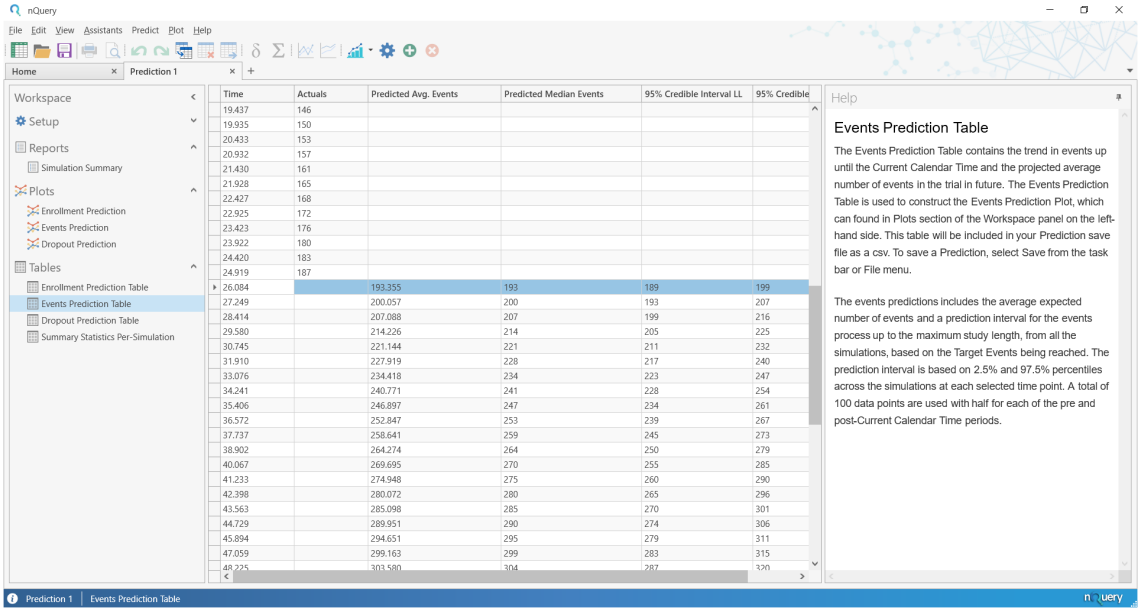



**Figure 6.147:** Dropout Prediction Plot for Unblinded Events Prediction using Summary Data

The **Enrollment Prediction Table, Events Prediction Table and Dropout Prediction Tables** are the tables used to create the respective **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plots, which by default all consists of ~100 rows (50 pre-simulation and 50 post-simulation). Each consists of five columns: **Time, Actual, Predicted Avg. Sample Size/Events/Dropouts, Predicted Median Sample Size/Events/Dropouts, 95% Prediction Interval LL, 95% Prediction Interval UL**. One prediction table (the **Events Prediction Table**) for this demonstration at the start of the simulated period is given in Figure 6.148.




| Time | Actuals | Predicted Avg. Events | Predicted Median Events | 95% Credible Interval LL | 95% Credible |
|---|---|---|---|---|---|
| 19.437 | 146 | | | | |
| 19.935 | 150 | | | | |
| 20.433 | 153 | | | | |
| 20.932 | 157 | | | | |
| 21.430 | 161 | | | | |
| 21.928 | 165 | | | | |
| 22.427 | 168 | | | | |
| 22.925 | 172 | | | | |
| 23.423 | 176 | | | | |
| 23.922 | 180 | | | | |
| 24.420 | 183 | | | | |
| 24.919 | 187 | | | | |
| 26.084 | | 193.355 | 193 | 189 | 199 |
| 27.249 | | 200.057 | 200 | 193 | 207 |
| 28.414 | | 207.088 | 207 | 199 | 216 |
| 29.580 | | 214.226 | 214 | 205 | 225 |
| 30.745 | | 221.144 | 221 | 211 | 232 |
| 31.910 | | 227.919 | 228 | 217 | 240 |
| 33.076 | | 234.418 | 234 | 223 | 247 |
| 34.241 | | 240.771 | 241 | 228 | 254 |
| 35.406 | | 246.897 | 247 | 234 | 261 |
| 36.572 | | 252.847 | 253 | 239 | 267 |
| 37.737 | | 258.641 | 259 | 245 | 273 |
| 38.902 | | 264.274 | 264 | 250 | 279 |
| 40.067 | | 269.695 | 270 | 255 | 285 |
| 41.233 | | 274.948 | 275 | 260 | 290 |
| 42.398 | | 280.072 | 280 | 265 | 296 |
| 43.563 | | 285.098 | 285 | 270 | 301 |
| 44.729 | | 289.951 | 290 | 274 | 306 |
| 45.894 | | 294.651 | 295 | 279 | 311 |
| 47.059 | | 299.163 | 299 | 283 | 315 |
| 48.225 | | 303.580 | 304 | 287 | 320 |

**Figure 6.148:** Events Prediction Table for Unblinded Events Prediction using Summary Data

**Time** is the time value relative to the study start time of zero. **Actual** is the actual number of events achieved at a pre-simulation **Time** row. **Predicted Avg. Events** is the predicted average (mean) number of events at a post-simulation **Time** row. **Predicted Median Events** is the predicted median number of events at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the number of events at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the number of events at a post-simulation **Time** row.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Current Events, Current Dropout, Current Available, Target Sample Size, Target Events, Study End Time, Total Sample Size, Total Events, Total Dropouts, Total Censored, Control Sample Size, Control Events, Control Dropouts, Treatment Sample Size, Treatment Events, Treatment Dropouts, Accrual Duration, Accrual Duration Uncensored, Average Followup, Median Followup, Target Events Reached.**

The **Summary Statistics Per-Simulation Table** for this demonstration, focused on the simulation results, is shown in Figure 6.149.

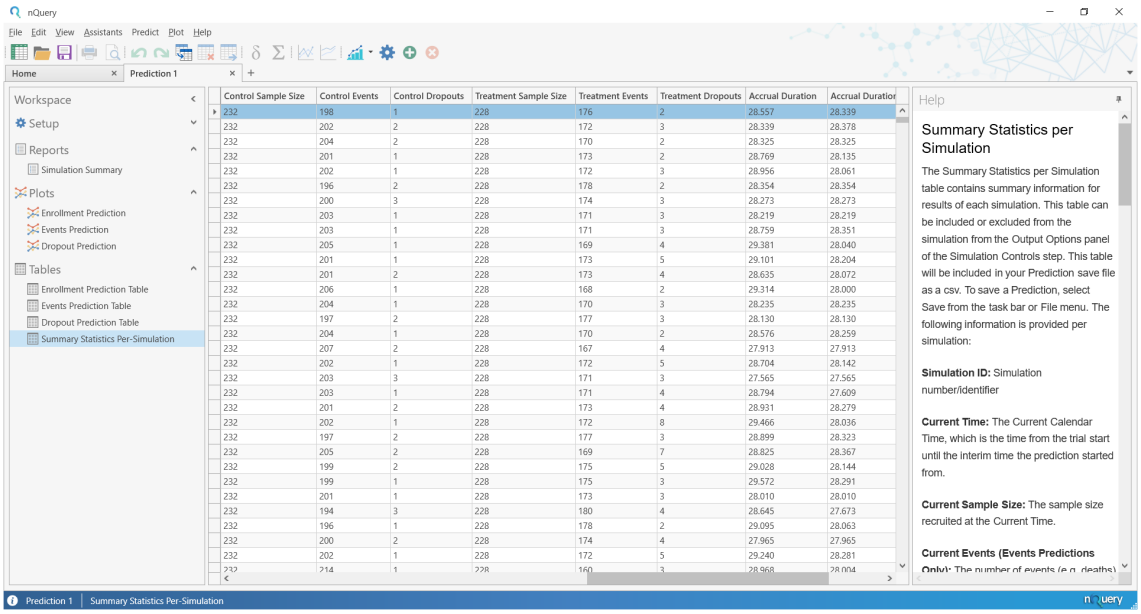



**Figure 6.149:** Summary Statistics Per-Simulation Table for Unblinded Events Prediction using Summary Data

Details on each field are provided in the Help window on the right but broadly the "Current" fields described the information provided up to the **Current Time**, the "Target" fields describe the study targets, the "Total" fields (and final Accrual, Followup and Target Event Reached columns) give the projection achieved for key parameters in total, the "Control" fields give the key results in the **Control Group** and the "Control" fields give the results in the **Treatment Group.**

**Alternative Scenarios** The demonstration used inputs based on **SubjectData.csv** and its original design. However, the "real" prediction using **Target = Unblinded Events** with **SubjectData.csv** (subsubsection 6.2.7.2) gave significantly different estimates for event and dropout hazard rates. As noted above, when **Input = Summary Data** the default hazard rates are based on all subjects being recruited at the start of the study but this is effectively a lower bound for hazard rate where every subject's maximum potential follow up time was at its highest possible value.

- Scenario A: **Hazard Rates** set to "true" values based on **SubjectData.csv**

**Scenario A** For **Scenario A**, the **Event Rates** and **Hazard Rates** will be replaced with the values derived from the demonstration using **SubjectData.csv** (see subsubsection 6.2.7.2). The defaults based on **SubjectData.csv** for the **Event Model Hazard Rates** are 0.0779 and 0.0547 in the **Control** and **Treatment Group** respectively. The default **Control** and **Treatment Group** dropout hazard rates are 0.0007 and 0.0014.

To make the required changes, edit the **Hazard Rate - Control** field to 0.0779 and **Hazard Rate - Treatment** to 0.0547 in the **Events Model** tab at the **Event and Dropout Information** step. The simplest way to do this is to select the **Setup** header in the Workspace Navigation Bar on the left and then select the **Event and Dropout Information** drop down option.

Then select the **Dropout Model** tab and edit **Hazard Rate - Control** field to 0.0007 and **Hazard Rate - Treatment** to 0.0014. Then select the Next or → button.

At the next **Simulation Controls** step, we will add a single simulation run by selecting the checkbox beside the **Save subject-level data for (x) simulation runs** and replace the default of 10 runs with 1 runs at the "(x)" spot. An edited down summary of the **Setup** changes is provided in Figure 6.150.

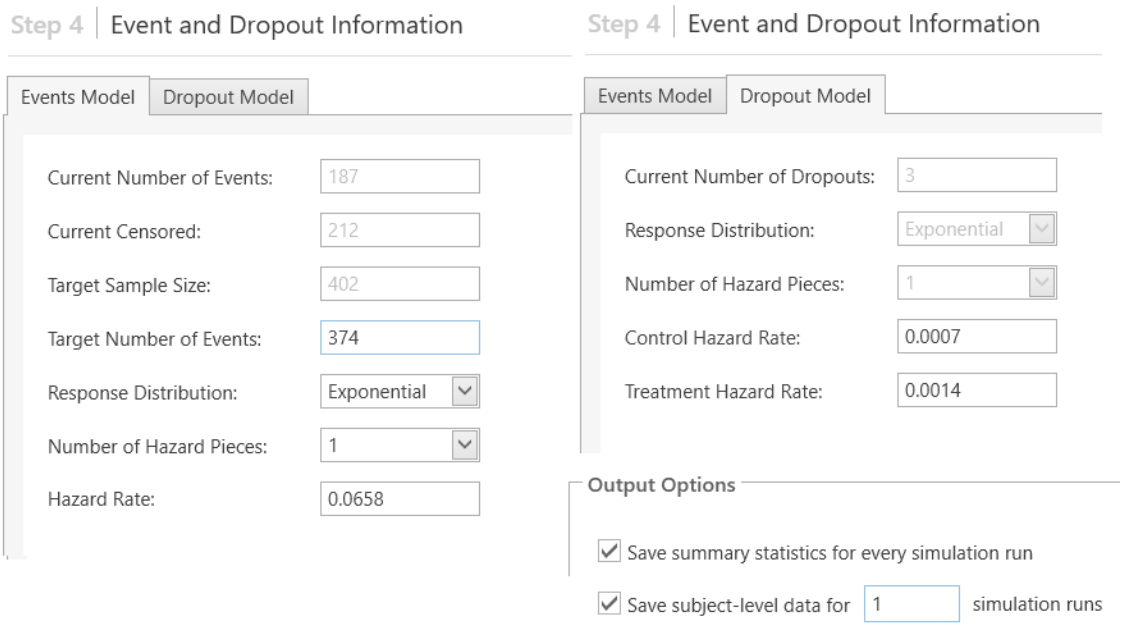



**Figure 6.150:** Scenario A Setup Changes for Unblinded Events Prediction using Summary Data

To run the simulation again with the new Scenario A, select the Run/→ button. The main results of this Scenario A simulation are shown in Figure 6.151.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 28.5066 |
| Average Study Duration | 43.9197 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 6.1932 |

Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 232 |
| Average Treatment Sample Size | 228 |
| Average Control Events | 201.16 |
| Average Treatment Events | 172.84 |
| Average Control Dropouts | 1.8957 |
| Average Treatment Dropouts | 4.2975 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 4 | 27.7784 | 41.5365 |
| 25.00% | 374 | 460 | 5 | 28.1814 | 42.9178 |
| 50.00% | 374 | 460 | 6 | 28.4869 | 43.8695 |
| 75.00% | 374 | 460 | 7 | 28.8076 | 44.8857 |
| 95.00% | 374 | 460 | 9 | 29.2999 | 46.4096 |

**Figure 6.151:** Scenario A Results for Unblinded Events Prediction using Summary Data

For Scenario A, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 43.9197. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 41.5365 (5% Percentile) and 46.4096 (95% Percentile), 50% a duration between 42.9178 (25% Percentile) and 44.8857 (75% Percentile) and a median (50% Percentile) duration of 43.8695. The **Average Dropouts** was 6.1932 (Percentiles: 4, 5, 6, 7, 9) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of

6.1932.

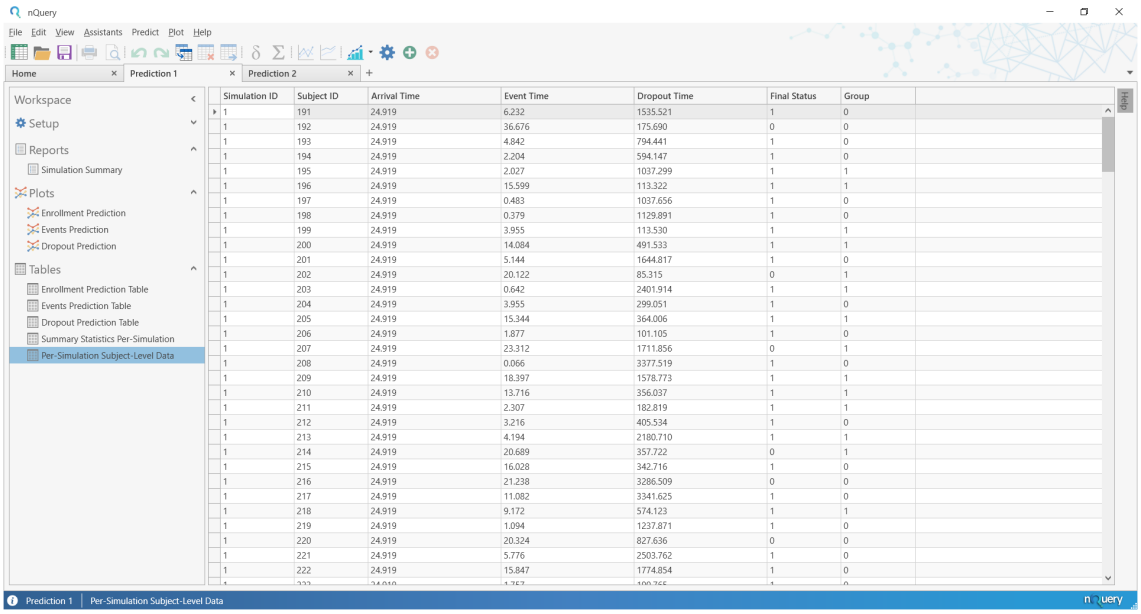
The original target study duration was 40 months and therefore this study is expected to end 3.9 months later than expected. However, this is significantly lower than the **Average Study Duration** of 73.7531 found above using the "worst case" **Summary Data** default and is effectively the same as the results generated using the actual **SubjectData.csv** (see subsubsection 6.2.7.2 Scenario A for example).

The enrollment process led to the **Target Sample Size** of 460 being achieved in all cases with an **Average Accrual Duration** of 28.5066 (Percentiles: 27.7784, 28.1814, 28.4869, 28.8076, 29.2999) which is effectively the same as before.

The per-group estimates were broadly the same as expected as the hazard ratio was effectively the same for **Scenario A** and the above example. The **Control Group** had a **Average Sample Size** 232, **Average Events** of 201.16 and **Dropouts** of 1.8957. The **Treatment Group** had a **Average Sample Size** 228, **Average Events** of 172.84 and **Dropouts** of 4.2975.

For "better" **Hazard Rate** estimates in the absence of real data, the user may wish to simulate an artificial dataset(s) using pre-trial assumptions (see subsubsection 6.2.7.5 for example) and use this artificial data (by selecting **Save subject-level data for (x) simulation runs** at the **Simulation Controls** step and using the **Per-Simulation Subject-level Data Table**, see below) as an input to a **Input = Subject Data Only** simulation to generate "new" defaults at the **Event & Dropout Information** step.

The **Per-Simulation Subject-level Data Table** was also generated here and is shown for Scenario A in Figure 6.152 with the first simulation (**Simulation ID = 1**). The tables consists of the following fields: **Simulation ID**, **Subject ID**, **Arrival Time**, **Event Time**, **Dropout Time, Final Status** and **Group.** The order of the Subject ID is based on **Arrival Time** order.




| Simulation ID | Subject ID | Arrival Time | Event Time | Dropout Time | Final Status | Group |
|---|---|---|---|---|---|---|
| 1 | 191 | 24.919 | 6.232 | 1535.521 | 1 | 0 |
| 1 | 192 | 24.919 | 36.676 | 175.690 | 0 | 0 |
| 1 | 193 | 24.919 | 4.842 | 794.441 | 1 | 0 |
| 1 | 194 | 24.919 | 2.204 | 594.147 | 1 | 0 |
| 1 | 195 | 24.919 | 2.027 | 1037.299 | 1 | 1 |
| 1 | 196 | 24.919 | 15.599 | 113.322 | 1 | 1 |
| 1 | 197 | 24.919 | 0.483 | 1037.656 | 1 | 0 |
| 1 | 198 | 24.919 | 0.379 | 1129.891 | 1 | 0 |
| 1 | 199 | 24.919 | 3.955 | 113.530 | 1 | 1 |
| 1 | 200 | 24.919 | 14.084 | 491.533 | 1 | 1 |
| 1 | 201 | 24.919 | 5.144 | 1644.817 | 1 | 0 |
| 1 | 202 | 24.919 | 20.122 | 85.315 | 0 | 1 |
| 1 | 203 | 24.919 | 0.642 | 2401.914 | 1 | 1 |
| 1 | 204 | 24.919 | 3.955 | 299.051 | 1 | 0 |
| 1 | 205 | 24.919 | 15.344 | 364.006 | 1 | 1 |
| 1 | 206 | 24.919 | 1.877 | 101.105 | 1 | 0 |
| 1 | 207 | 24.919 | 23.312 | 1711.856 | 0 | 1 |
| 1 | 208 | 24.919 | 0.066 | 3377.519 | 1 | 0 |
| 1 | 209 | 24.919 | 18.397 | 1578.773 | 1 | 1 |
| 1 | 210 | 24.919 | 13.716 | 356.037 | 1 | 1 |
| 1 | 211 | 24.919 | 2.307 | 182.819 | 1 | 1 |
| 1 | 212 | 24.919 | 3.216 | 405.534 | 1 | 0 |
| 1 | 213 | 24.919 | 4.194 | 2180.710 | 1 | 1 |
| 1 | 214 | 24.919 | 20.689 | 357.722 | 0 | 1 |
| 1 | 215 | 24.919 | 16.028 | 342.716 | 1 | 0 |
| 1 | 216 | 24.919 | 21.238 | 3286.509 | 0 | 0 |
| 1 | 217 | 24.919 | 11.082 | 3341.625 | 1 | 0 |
| 1 | 218 | 24.919 | 9.172 | 574.123 | 1 | 1 |
| 1 | 219 | 24.919 | 1.094 | 1237.871 | 1 | 0 |
| 1 | 220 | 24.919 | 20.324 | 827.636 | 0 | 0 |
| 1 | 221 | 24.919 | 5.776 | 2503.762 | 1 | 0 |
| 1 | 222 | 24.919 | 15.847 | 1774.854 | 1 | 0 |

**Figure 6.152:** Per-Simulation Subject-level Data Table for Unblinded Events using Summary Data

When **Input = Summary Data**, **Event Time** and **Dropout Time** are generated for all subjects. This set of subjects includes the **No. of Censored** subjects at the **Current Time** from the **Fixed Parameters** step and the new subjects simulated after the **Current Time**. For **Censored Subject**, the **Arrival Time** is set to the **Current Time** (since no information exists for an actual **Arrival Time**). New subjects have an **Arrival Time** generated after the **Current Time** as per the normal. Note that the implied subjects who had **Event/Dropout** before the **Current Time** based on the **Fixed Parameters** are not simulated and therefore are not shown here.

**Final Status** will equal **Dropout Status** if **Dropout Time** is less than **Event Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.167 for per-simulation end times). **Final Status** will equal **Censored Status** if **Event Time** and **Dropout Time** are greater than the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.167 for per-simulation end times).

#### 6.2.7.5 Blinded Events Prediction with Summary Data Pre-Trial

**Setup** **Blinded Events** milestone prediction using **Summary Data** means that the prediction will project the expected length of study needed to achieve a specified **Target Number of Events** using summary data fixed parameters for the whole study as per-group parameters are unknown. In this demonstration, the **Fixed**

**Parameters** are based on pre-trial estimates rather than estimates from an on-going trial.

In the **Setup** for this prediction, first in the **Select the type of Prediction** step select **Blinded Events** from the **Target** field, **Summary Data** from the **Input** field and **Ongoing** from the **Enrollment Status** field.

These selections are shown in Figure 6.153. Select the Next button in the bottom-right or → button in the top-right of the Main Window to move to the next step of the prediction **Setup** stage.

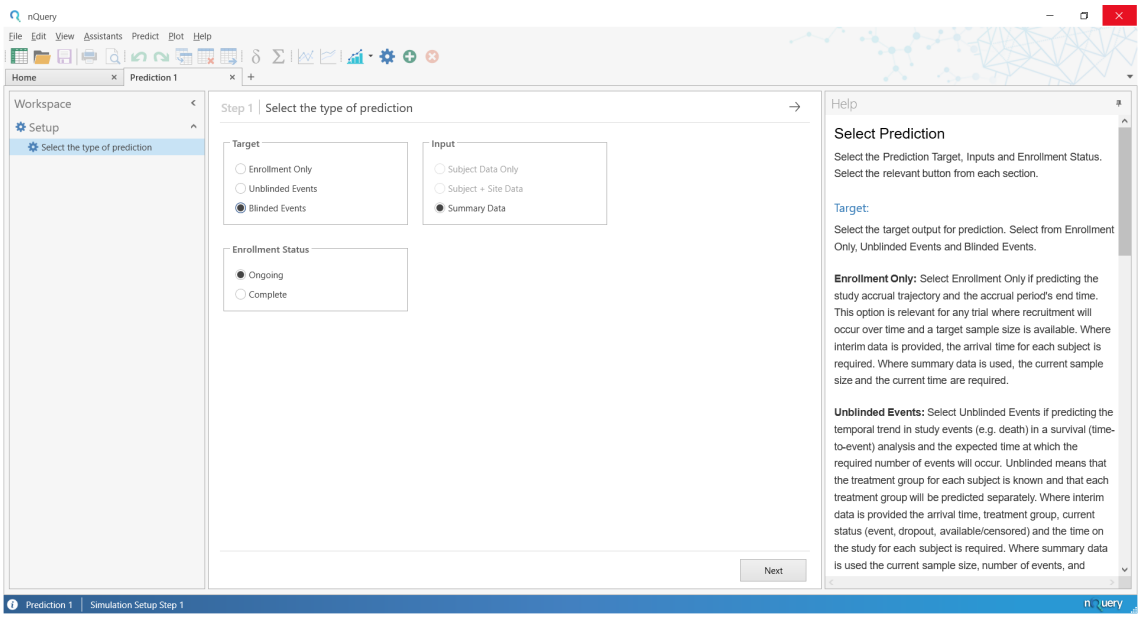



**Figure 6.153:** Select the Type of Prediction Setup Step for Blinded Events Pre-Trial

The next step is the **Fixed Parameters** step. For this demonstration, there is a **Sample Sizes, Events & Dropouts** section which contains five fields which are as follows:

- Current Sample Size: The total sample size enrolled at the Current Time
- Current No. of Events: The total number of events at the Current Time
- Current No. of Dropouts: The total number of dropouts that have occurred at the Current Time
- Current No. Censored (Read-only): The subjects censored (not had event or dropout) if the study ended at the Current Time
- Current Time: The length of time the study has been on-going since the study began

For this demonstration, it assumed the prediction is being made pre-trial. To make a pre-trial prediction set all of the fields above equal to zero as per Figure 6.154. After all fields all filled, select the Next or → button to move to the next **Setup** step.

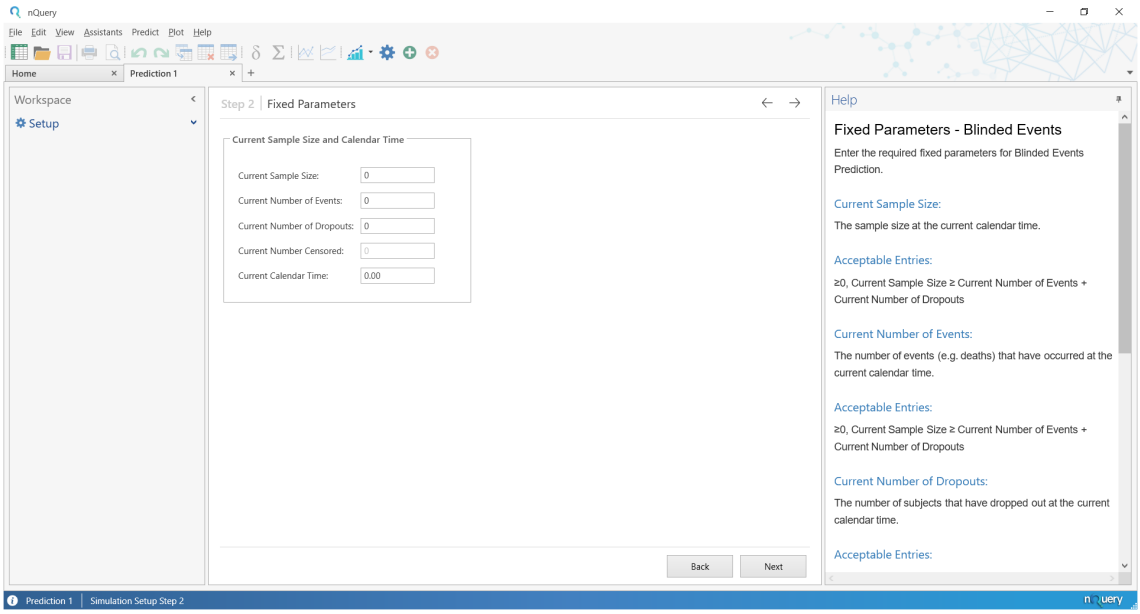



**Figure 6.154:** Fixed Parameters Setup Step for Blinded Events Pre-Trial

The next step is the **Accrual Options** step which specifies the **Target Sample Size**. The **Accrual Options** step, and its defaults, for this demonstration are shown in Figure 6.155.

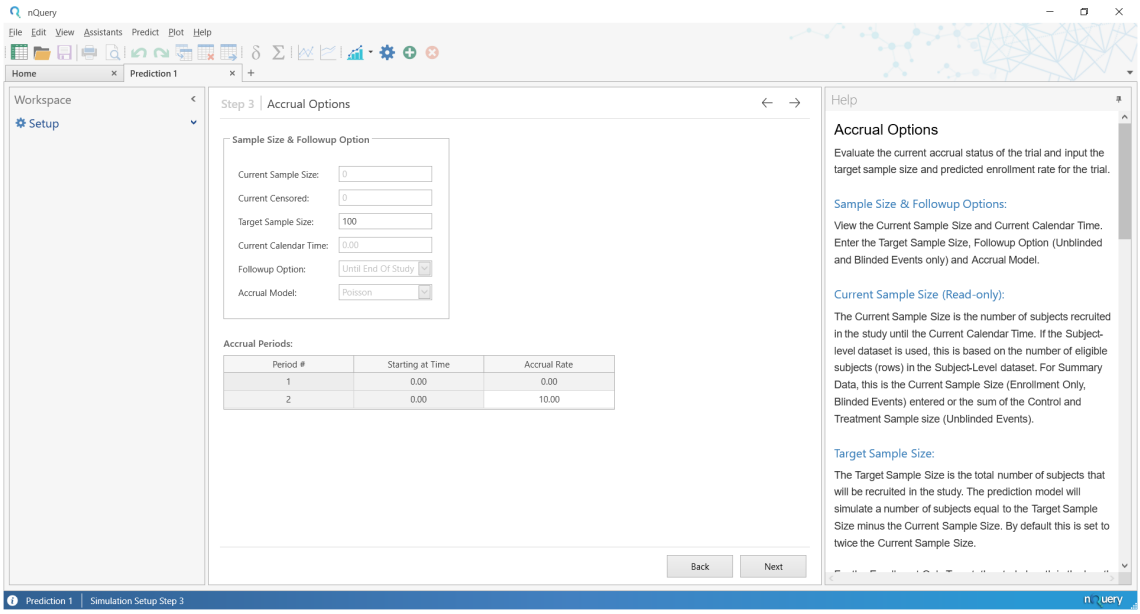



**Figure 6.155:** Accrual Options Setup Step Defaults for Blinded Events Pre-Trial

The **Accrual Options** step consists of two main elements: the **Sample Size & Followup** input fields and the **Accrual Periods** table.

The **Sample Size & Followup** input fields provides information on enrollment process, censored status and current time based on the subject-level data and allows the user to edit the **Target Sample Size** and select the **Followup Option** and **Accrual Model**. The following fields are provided in this prediction:

- Current Sample Size (Read-only): Sample Size recruited into a study so far
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Sample Size Proportion (Read-only): Proportion of subjects in the Control Group. Based on number of Control Group subjects in Treatment ID field. Controls the allocation proportion in simulated data
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study
- Current Time (Read-only): The length of time that has passed since the study started. Equals the maximum of the sum of the Arrival Time + Time on Study values for each subject
- Followup Option (Read-only): Specifies whether subjects who do not have the event/dropout will be administratively censored at the end of the study or after a fixed follow up period. For Summary Data, this is set to Until End of Study

- Accrual Model (Editable): The statistical model that will be used to generate simulated enrollments

When **Input = Summary Data** and a **Pre-Trial** prediction is being conducted, nQuery Predict defaults to a **Target Sample Size** of 100 and a **Poisson Accrual Model**.

When **Input = Summary Data,** the **Followup Option** must be set to **Until End of Study.**

The **Accrual Periods** table contains information on the **Accrual Rate** up the **Current Time** and allows the user to edit the future **Accrual Rate** used to generate enrollment simulations. The **Accrual Periods** has three columns ,with each row corresponding to a time period, with the columns defined as follows:

- Period #: A numeric ID for the current time period. Increases in increments of one for each subsequent row
- Starting at Time: The starting time for the current time period row. The first row will equal 0 (rate from study start to Current Time) and the second row will equal 0 for pre-trial prediction
- Accrual Rate: The accrual rate is the average number of subjects recruited per unit time (months). The accrual rate in the first row will equal the accrual rate up to the Current Time.

For the **Poisson Accrual Model**, the **Accrual Periods** table will consist of two time period rows. For pre-trial prediction, the first row will just indicate that no information has been provided since **Starting at Time** equals zero for both rows. The second row will correspond to the inputs used to generate future enrollments. By default the **Accrual Rate** for **Pre-Trial** prediction in the future second row will equal 10. The user can edit the **Accrual Rate** for generating future enrollments by editing the second cell of the **Accrual Rate** column. Select the Next or → button to move to the next **Setup** step.

The next step is the **Event and Dropout Information** step which specifies the **Target Number of Events,** event model (and associated statistical parameters) and dropout model (and associated statistical parameters). The **Event and Dropout Information** step on the default **Events Model** tab is shown in Figure 6.156.

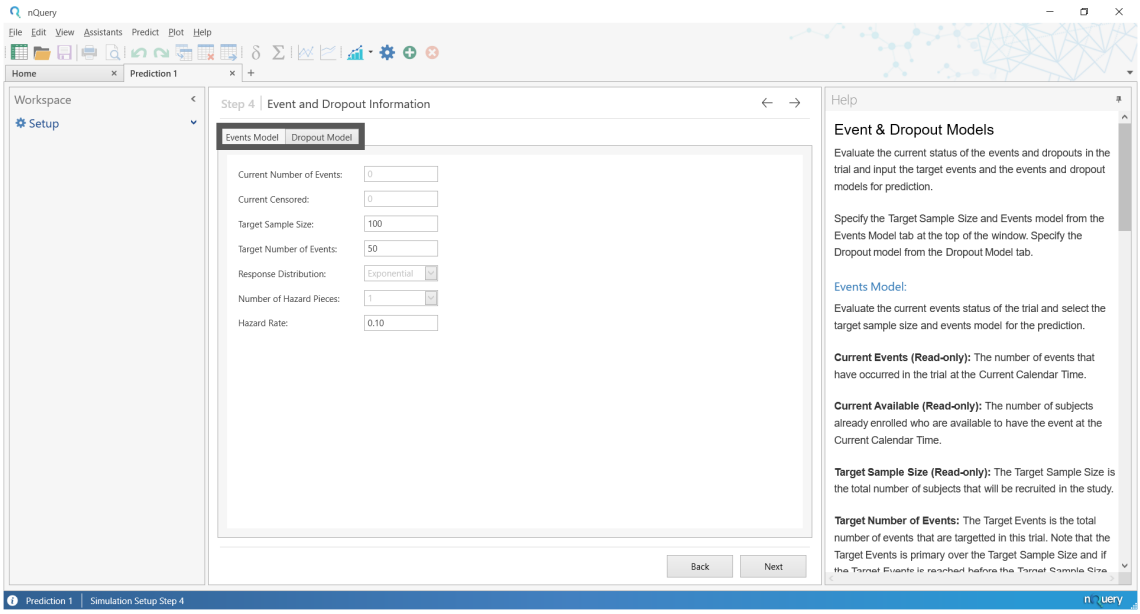



**Figure 6.156:** Event and Dropout Information Setup Step for Blinded Events Pre-Trial

The **Event and Dropout Information** step consists of two tabs: the **Events Model** tab and the **Dropout Model** tab. These tabs have been highlighted in Figure 6.156 and are opened by selecting the required tab name at the top of the main window, below the step title. For **Blinded Events**, it is assumed that survival and dropout times will be simulated using a single "global" process which is specified by the user.

The **Events Model** tab provides fields which provide information on the current number of subject who have had an event, could have an event going forward and allows the input of the **Target Number of Events** and **Response Distribution** for the events model and additional parameters needed for that specific model. The following fields are provided in this prediction:

- Current Events (Read-only): The number of subjects who had the event of interest at the current time. Equals the number of subjects with the Event Status in the Status Indicator field
- Current Censored (Read-only): The number of subjects who have not had the event or dropped out at the current time. Equals the number of subjects with the Censored Status in the Status Indicator field
- Target Sample Size (Editable): Total number of subjects that will be recruited in this study
- Target Number of Events: The required number of subjects needed to have the event of interest for the study to end

- Response Distribution: The statistical model that will be used to generate survival times for subjects available to have the event
- Number of Hazard Pieces (Response Distribution = Exponential only): Number of hazard pieces used to specify the piecewise exponential model. Select value between 1 and 10
- Hazard Rate (Response Distribution = Exponential only, Number of Hazard Pieces =1, Editable): The exponential hazard rate for the constant exponential survival model specified above

When **Input = Summary Data** pre-trial the only **Response Distribution** is **Exponential** (**Number of Hazard Pieces** = 1) and the default **Hazard Rate** equals 0.1 with a default **Target Number of Events** of 50.

The **Dropout Model** tab is opened by selecting the **Dropout Model** tab name at the top of the main window. For this demonstration, the **Dropout Model** tab is shown in Figure 6.157.

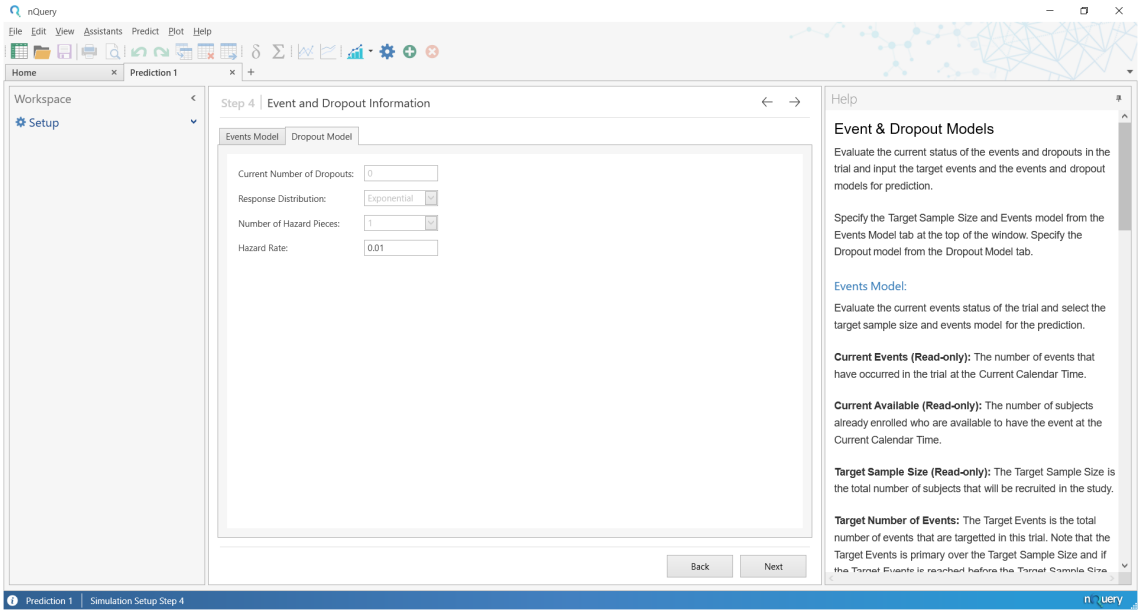



**Figure 6.157:** Dropout Model tab for Blinded Events Pre-Trial

For dropout, the only **Response Distribution** option is exponential with one hazard piece with a **Hazard Rate** equal to 0.01. The inputs for dropout **Exponential Response Distribution** are the same as from the **Events Model** described above and therefore will not be replicated here. Select the Next or → button to move to the next **Setup** step.

The final step is the **Simulation Controls** step. The **Simulation Controls** step consists of three sections: Simulation Controls, Output for All Trials and Output Options.

**Simulation Controls** (top-left) sets the **Number of Simulations**, **Refresh Frequency** and **Random Seed**

**Output For All Trials** (bottom-left) sets the Percentiles desired in the **Percentile Summary** table in the **Simulation Summary** report for key simulation outputs such as study length and sample size.

**Output Options** (right) selects which additional datasets the user wants in the **Tables** field of the simulation results. Additional datasets will be summary statistics for each simulation and outputs from a set number of individual of simulations.

The **Simulation Controls** step is described in detail in subsubsection 6.2.9.4. In this demonstration, the defaults will be used which corresponds to 1000 simulations with the default percentiles (5%, 25%, 50%, 75%, 95%) and with the "Save summary statistics for every simulation run" **Output Options** table active. The **Simulation Controls** defaults for this demonstration are shown in Figure 6.158. To start the milestone prediction, Select the Run (where Next button was in previous steps) or → button.

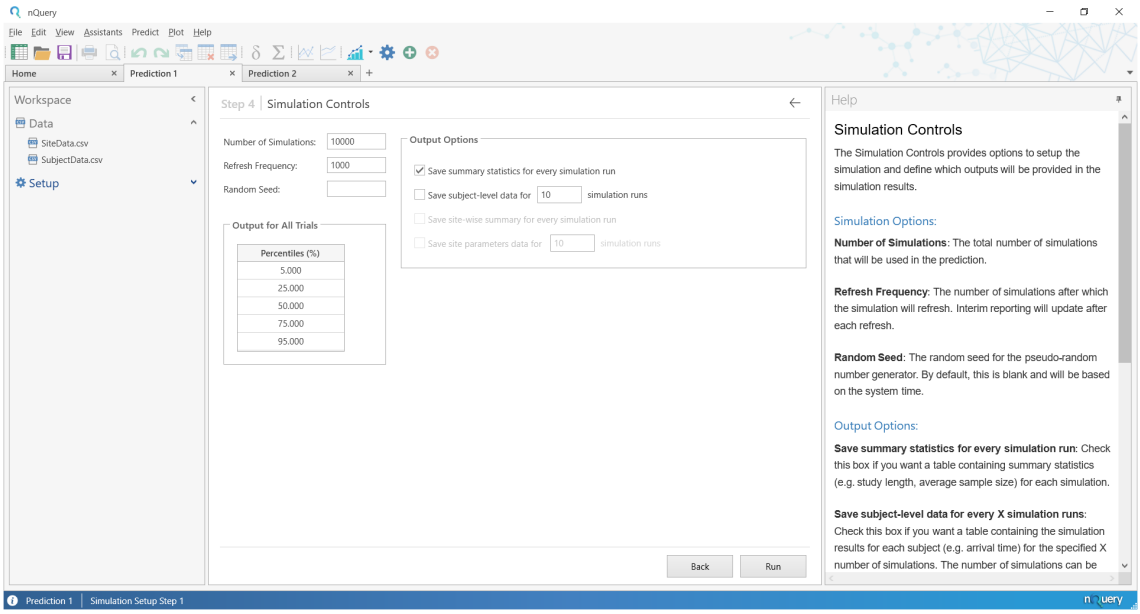



**Figure 6.158:** Simulation Controls Setup Step for Blinded Events Pre-Trial

**Results** While a milestone prediction is being run, a **Simulation in Progress** window will be displayed. The **Simulation in Progress** window provides updated information on key simulation metrics while the simulation is ongoing. An example for this demonstration is given in Figure 6.159.

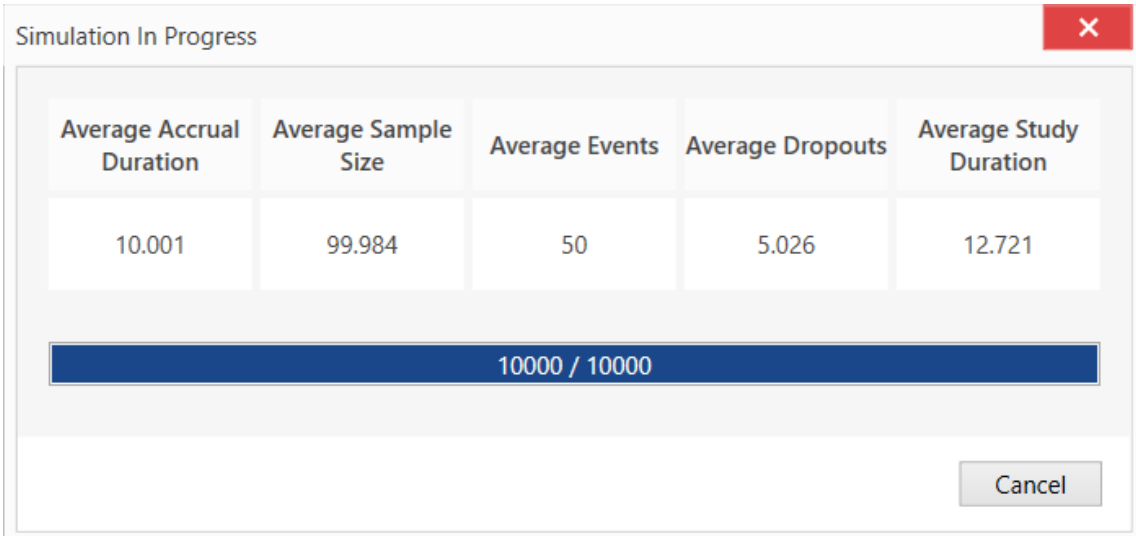



**Figure 6.159:** Simulation in Progress Window for Blinded Events Pre-Trial

Once the simulation is complete, the results will automatically be added to the Workspace Navigation Bar on the left and the first element under the **Plots** header will be displayed. In this demonstration, the **Enrollment Prediction Plot** will be displayed. This is shown in Figure 6.160.

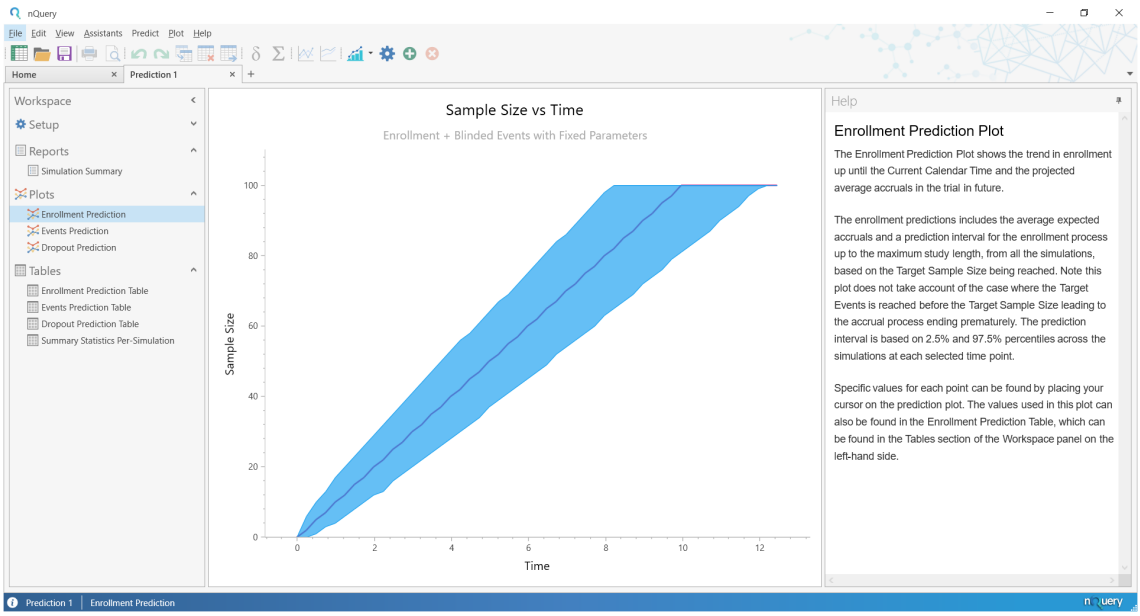



**Figure 6.160:** Enrollment Prediction Plot for Blinded Events Pre-Trial

The results of a nQuery Predict milestone prediction can be divided into three categories: **Reports, Plots** and **Tables**. A high-level summary of the results in

nQuery Predict is provided in subsection 6.2.8 with demonstration specific elements highlighted here.

For this demonstration the **Reports** header contains the **Simulation Summary** report, the **Plots** header contains the **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plot and the **Tables** header contains the **Enrollment Prediction Table, Events Prediction Table, Dropout Prediction Table** and **Summary Statistics Per-Simulation**. To select any result element, select it from the appropriate header in the Workspace Navigation Bar on the left (see subsection 6.2.2 for details on the Navigation Bar).

The **Simulation Summary** provides a tabular summary of the inputs and outputs from the current milestone prediction simulation. The **Simulation Summary** for this demonstration is shown in Figure 6.161.

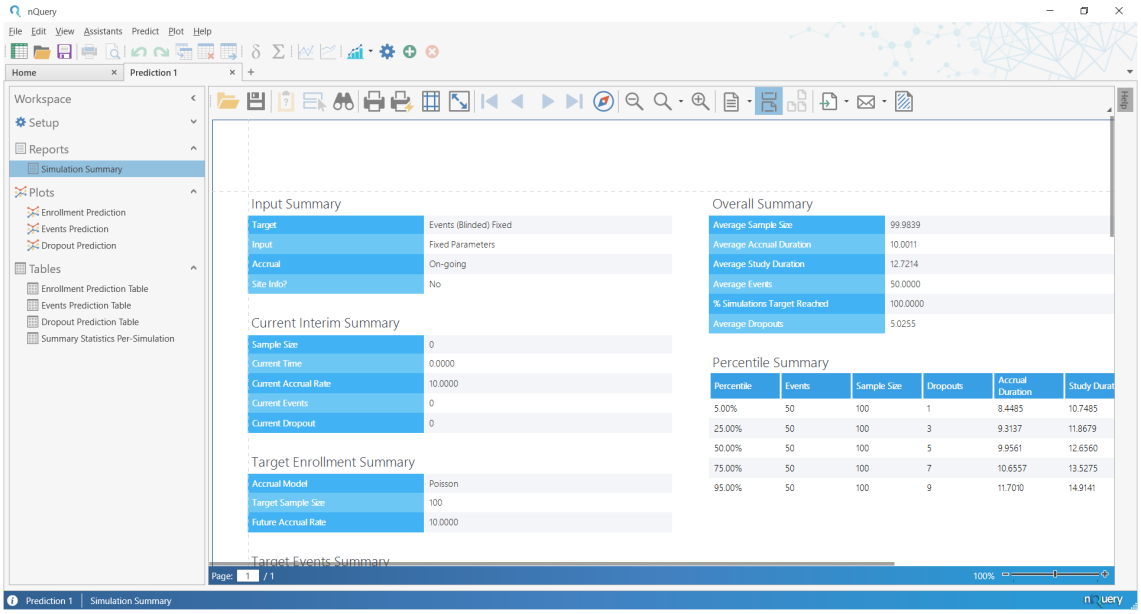



**Figure 6.161:** Simulation Summary Report for Blinded Events Pre-Trial

In the **Simulation Summary** report, the left-hand column provides a summary of inputs used to generate the current simulation and the right-hand column provides the results of the current simulation. In this demonstration, we will skip the left-hand Inputs column and focus on the main results in the right-hand column and the main report options.

The results column (on the right) is shown in Figure 6.162. For this demonstration, the **Overall Summary** shows that the **Target Number of Events** of 50 was reached in every simulation with an **Average Study Duration** over all the simulations of 12.7214. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 10.7485 (5% Percentile) and 14.9141 (95% Percentile),

50% a duration between 11.8679 (25% Percentile) and 13.5275 (75% Percentile) and a median (50% Percentile) duration of 12.656.

The **Average Dropouts** was 5.0255 (Percentiles: 1, 3, 5, 7, 9) and subjects had an **Average Followup** (until either event/dropout or until study end for censored subjects) of 5.0255. For the enrollment process, the **Average Accrual Duration** was 10.0011 (Percentiles: 8.4485, 9.3137, 9.9561, 10.6557, 11.7010) with the **Target Sample Size** of 100 reached in all simulations.

Overall Summary

| | |
|---|---|
| Average Sample Size | 99.9839 |
| Average Accrual Duration | 10.0011 |
| Average Study Duration | 12.7214 |
| Average Events | 50.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 5.0255 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 50 | 100 | 1 | 8.4485 | 10.7485 |
| 25.00% | 50 | 100 | 3 | 9.3137 | 11.8679 |
| 50.00% | 50 | 100 | 5 | 9.9561 | 12.6560 |
| 75.00% | 50 | 100 | 7 | 10.6557 | 13.5275 |
| 95.00% | 50 | 100 | 9 | 11.7010 | 14.9141 |

**Figure 6.162:** Simulation Summary Report Results for Blinded Events Pre-Trial

To edit the **Simulation Summary report there a** number of buttons available at the top of the report. For full detail on the **Simulation Summary** report and options to edit the report see subsubsection 6.2.8.2. However the Zoom options , Page Layout and Setup options and Export option will be the more commonly used options.

The **Enrollment Prediction** plot provides a visual summary of enrollment. The X-axis is **Time** (relative to study start) and the Y-axis is **Sample Size** (enrolled at Time X-axis value). When the cursor is placed over the **Enrollment Prediction** plot, information on the **Sample Size** enrolled at a given **Time** is shown as per Figure 6.163.

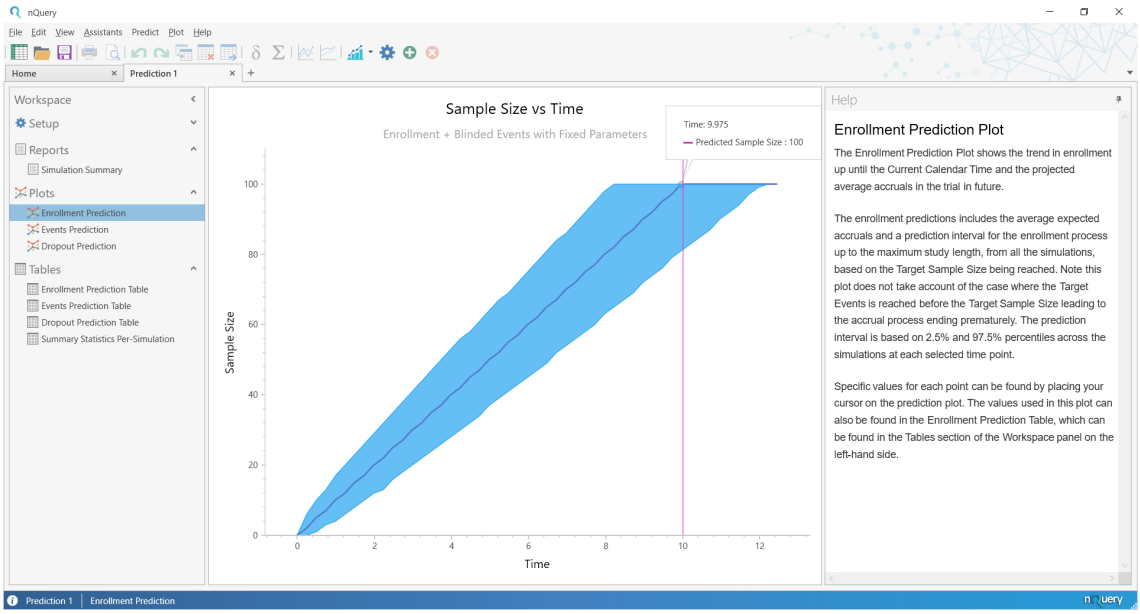



**Figure 6.163:** Enrollment Prediction Plot for Blinded Events Pre-Trial

The **Enrollment Prediction** plot for pre-trial prediction will have a blue conic section with a grey line where the blue conic section plots the 95% prediction interval for the **Sample Size** achieved at a given **Time** across all simulations and the grey line is the average **Sample Size** enrolled at a given time. The 95% prediction interval indicates the 95% of predictions had a sample size that fell between the lower and upper limit at a given time. Note the **Enrollment Prediction** plot **Time** scale will be related to the **Accrual Duration** rather than the **Event Prediction** and **Dropout Prediction** plots which will be associated with the **Study Duration.**

The **Event Prediction** plot for pre-trial prediction will have a yellow conic section with a grey line where the yellow conic section plots the 95% prediction interval for the **Events** achieved at a given **Time** across all simulations and the grey line is the average **Events** enrolled at a given time. The 95% prediction interval indicates the 95% of predictions had an event that fell between the lower and upper limit at a given time. The **Events Prediction** plot for this demonstration is shown in Figure 6.164.

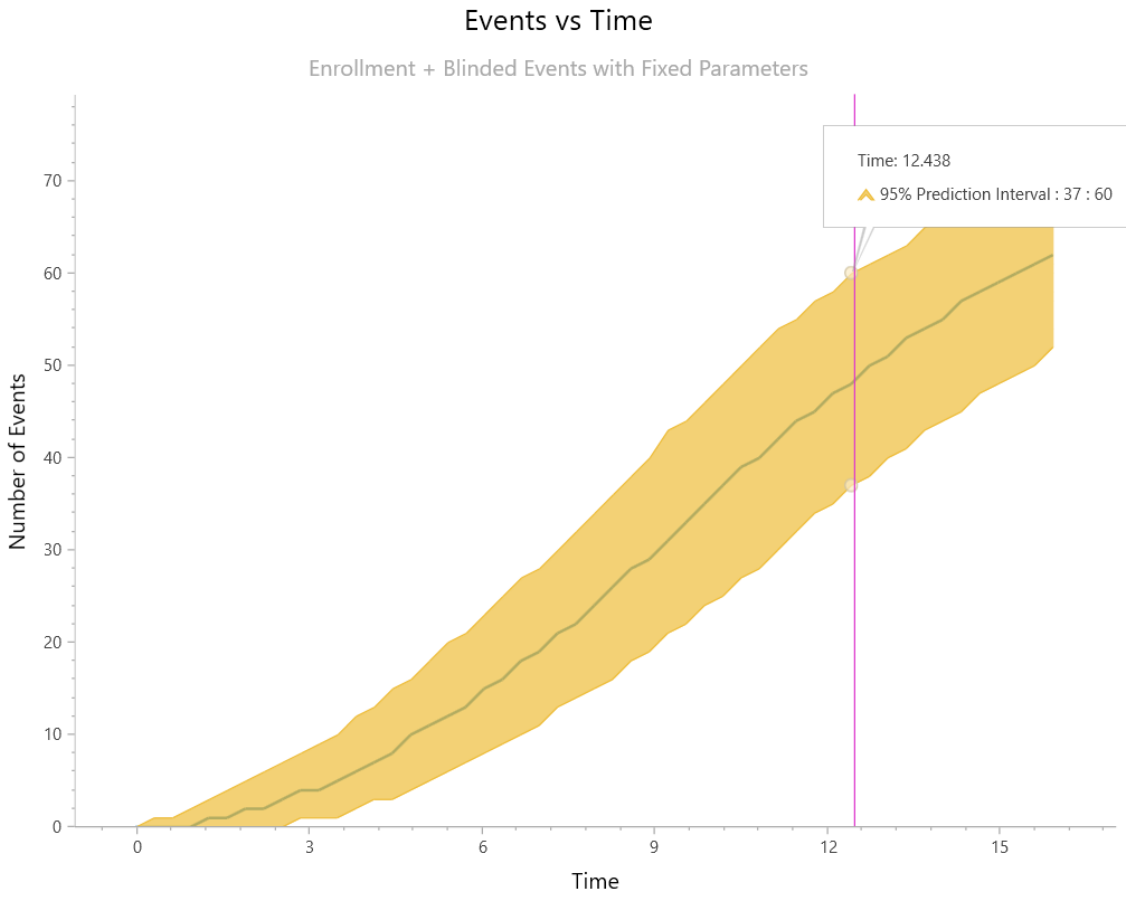



**Figure 6.164:** Events Prediction Plot for Blinded Events Pre-Trial

The **Dropout Prediction** plot works effectively the same as the **Events Prediction** plot except for the dropout process. The **Dropout Prediction** plot for this demonstration is shown in Figure 6.165.

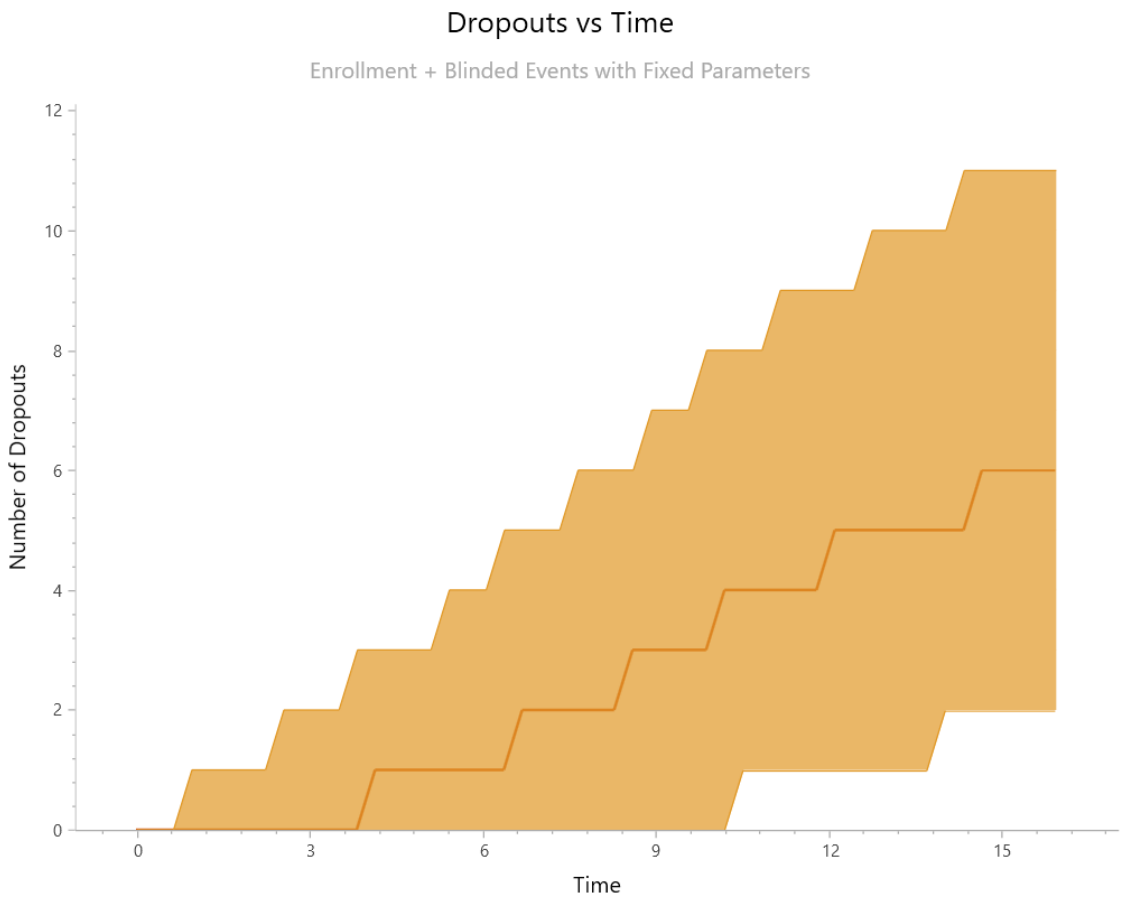
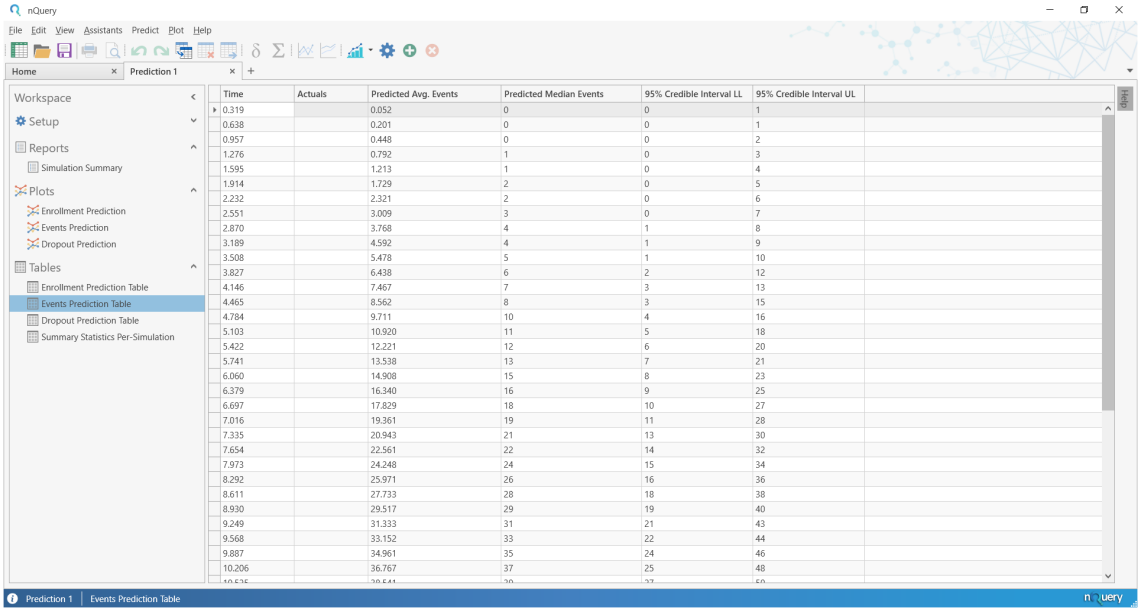



**Figure 6.165:** Dropout Prediction Plot for Blinded Events Pre-Trial

The **Enrollment Prediction Table, Events Prediction Table and Dropout Prediction Tables** are the tables used to create the respective **Enrollment Prediction, Events Prediction** and **Dropout Prediction** plots, which by default all consists of ~100 rows (50 pre-simulation and 50 post-simulation). Each consists of five columns: **Time, Actual, Predicted Avg. Sample Size/Events/Dropouts, Predicted Median Sample Size/Events/Dropouts, 95% Prediction Interval LL, 95% Prediction Interval UL**. One prediction table (the **Events Prediction Table**) for this demonstration at the start is given in Figure 6.166.



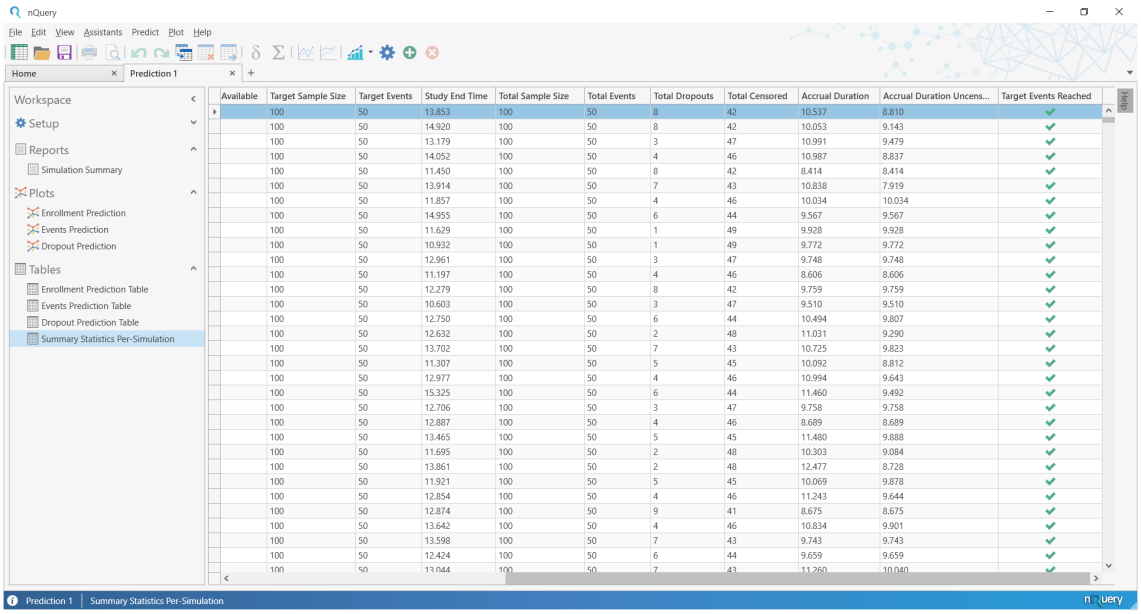
| Time | Actuals | Predicted Avg. Events | Predicted Median Events | 95% Credible Interval LL | 95% Credible Interval UL |
|---|---|---|---|---|---|
| 0.319 | | 0.052 | 0 | 0 | 1 |
| 0.638 | | 0.201 | 0 | 0 | 1 |
| 0.957 | | 0.448 | 0 | 0 | 2 |
| 1.276 | | 0.792 | 1 | 0 | 3 |
| 1.595 | | 1.213 | 1 | 0 | 4 |
| 1.914 | | 1.729 | 2 | 0 | 5 |
| 2.232 | | 2.321 | 2 | 0 | 6 |
| 2.551 | | 3.009 | 3 | 0 | 7 |
| 2.870 | | 3.768 | 4 | 1 | 8 |
| 3.189 | | 4.592 | 4 | 1 | 9 |
| 3.508 | | 5.478 | 5 | 1 | 10 |
| 3.827 | | 6.438 | 6 | 2 | 12 |
| 4.146 | | 7.467 | 7 | 3 | 13 |
| 4.465 | | 8.562 | 8 | 3 | 15 |
| 4.784 | | 9.711 | 10 | 4 | 16 |
| 5.103 | | 10.920 | 11 | 5 | 18 |
| 5.422 | | 12.221 | 12 | 6 | 20 |
| 5.741 | | 13.538 | 13 | 7 | 21 |
| 6.060 | | 14.908 | 15 | 8 | 23 |
| 6.379 | | 16.340 | 16 | 9 | 25 |
| 6.697 | | 17.829 | 18 | 10 | 27 |
| 7.016 | | 19.361 | 19 | 11 | 28 |
| 7.335 | | 20.943 | 21 | 13 | 30 |
| 7.654 | | 22.561 | 22 | 14 | 32 |
| 7.973 | | 24.248 | 24 | 15 | 34 |
| 8.292 | | 25.971 | 26 | 16 | 36 |
| 8.611 | | 27.733 | 28 | 18 | 38 |
| 8.930 | | 29.517 | 29 | 19 | 40 |
| 9.249 | | 31.333 | 31 | 21 | 43 |
| 9.568 | | 33.152 | 33 | 22 | 44 |
| 9.887 | | 34.961 | 35 | 24 | 46 |
| 10.206 | | 36.767 | 37 | 25 | 48 |

**Figure 6.166:** Events Prediction Table for Blinded Events Pre-Trial

**Time** is the time value relative to the study start time of zero. **Actual** is the actual number of events achieved at a pre-simulation **Time** row but will be empty for pre-trial prediction. **Predicted Avg. Events** is the predicted average (mean) number of events at a post-simulation **Time** row. **Predicted Median Events** is the predicted median number of events at a post-simulation **Time** row. **95% Prediction Interval LL** is the lower limit for the 95% prediction interval for the number of events at a post-simulation **Time** row. **95% Prediction Interval UL** is the upper limit for the 95% prediction interval for the number of events at a post-simulation **Time** row.

The **Summary Statistics Per-Simulation Table** provides the summary statistics achieved for primary outputs in each simulation. For this demonstration this includes the **Simulation ID, Current Time, Current Sample Size, Current Events, Current Dropout, Current Available, Target Sample Size, Target Events, Study End Time, Total Sample Size, Total Events, Total Dropouts, Total Censored, Accrual Duration, Accrual Duration Uncensored, Average Followup, Median Followup, Target Events Reached.**

The **Summary Statistics Per-Simulation Table** for this demonstration, focused on the simulation results, is shown in Figure 6.167.




| Available | Target Sample Size | Target Events | Study End Time | Total Sample Size | Total Events | Total Dropouts | Total Censored | Accrual Duration | Accrual Duration Uncens... | Target Events Reached |
|---|---|---|---|---|---|---|---|---|---|---|
| | 100 | 50 | 13.853 | 100 | 50 | 8 | 42 | 10.537 | 8.810 | ✔ |
| | 100 | 50 | 14.920 | 100 | 50 | 8 | 42 | 10.053 | 9.143 | ✔ |
| | 100 | 50 | 13.179 | 100 | 50 | 3 | 47 | 10.991 | 9.479 | ✔ |
| | 100 | 50 | 14.052 | 100 | 50 | 4 | 46 | 10.987 | 8.837 | ✔ |
| | 100 | 50 | 11.450 | 100 | 50 | 8 | 42 | 8.414 | 8.414 | ✔ |
| | 100 | 50 | 13.914 | 100 | 50 | 7 | 43 | 10.838 | 7.919 | ✔ |
| | 100 | 50 | 11.857 | 100 | 50 | 4 | 46 | 10.034 | 10.034 | ✔ |
| | 100 | 50 | 14.955 | 100 | 50 | 6 | 44 | 9.567 | 9.567 | ✔ |
| | 100 | 50 | 11.629 | 100 | 50 | 1 | 49 | 9.928 | 9.928 | ✔ |
| | 100 | 50 | 10.932 | 100 | 50 | 1 | 49 | 9.772 | 9.772 | ✔ |
| | 100 | 50 | 12.961 | 100 | 50 | 3 | 47 | 9.748 | 9.748 | ✔ |
| | 100 | 50 | 11.197 | 100 | 50 | 4 | 46 | 8.606 | 8.606 | ✔ |
| | 100 | 50 | 12.279 | 100 | 50 | 8 | 42 | 9.759 | 9.759 | ✔ |
| | 100 | 50 | 10.603 | 100 | 50 | 3 | 47 | 9.510 | 9.510 | ✔ |
| | 100 | 50 | 12.750 | 100 | 50 | 6 | 44 | 10.494 | 9.807 | ✔ |
| | 100 | 50 | 12.632 | 100 | 50 | 2 | 48 | 11.031 | 9.290 | ✔ |
| | 100 | 50 | 13.702 | 100 | 50 | 7 | 43 | 10.725 | 9.823 | ✔ |
| | 100 | 50 | 11.307 | 100 | 50 | 5 | 45 | 10.092 | 8.812 | ✔ |
| | 100 | 50 | 12.977 | 100 | 50 | 4 | 46 | 10.994 | 9.643 | ✔ |
| | 100 | 50 | 15.325 | 100 | 50 | 6 | 44 | 11.460 | 9.492 | ✔ |
| | 100 | 50 | 12.706 | 100 | 50 | 3 | 47 | 9.758 | 9.758 | ✔ |
| | 100 | 50 | 12.887 | 100 | 50 | 4 | 46 | 8.689 | 8.689 | ✔ |
| | 100 | 50 | 13.465 | 100 | 50 | 5 | 45 | 11.480 | 9.888 | ✔ |
| | 100 | 50 | 11.695 | 100 | 50 | 2 | 48 | 10.303 | 9.084 | ✔ |
| | 100 | 50 | 13.861 | 100 | 50 | 2 | 48 | 12.477 | 8.728 | ✔ |
| | 100 | 50 | 11.921 | 100 | 50 | 5 | 45 | 10.069 | 9.878 | ✔ |
| | 100 | 50 | 12.854 | 100 | 50 | 4 | 46 | 11.243 | 9.644 | ✔ |
| | 100 | 50 | 12.874 | 100 | 50 | 9 | 41 | 8.675 | 8.675 | ✔ |
| | 100 | 50 | 13.642 | 100 | 50 | 4 | 46 | 10.834 | 9.901 | ✔ |
| | 100 | 50 | 13.598 | 100 | 50 | 7 | 43 | 9.743 | 9.743 | ✔ |
| | 100 | 50 | 12.424 | 100 | 50 | 6 | 44 | 9.659 | 9.659 | ✔ |
| | 100 | 50 | 13.044 | 100 | 50 | 7 | 43 | 11.260 | 10.040 | ✔ |

**Figure 6.167:** Summary Statistics Per-Simulation Table for Blinded Events Pre-Trial

Details on each field are provided in the Help window on the right but broadly the "Current" fields described the information provided up to the **Current Time**, the "Target" fields describe the study targets and the "Total" fields (and final Accrual, Followup and Target Event Reached columns) give the projection achieved for key parameters in total.

**Alternative Scenarios** The prior demonstration used the default values for pre-trial prediction. In the following scenarios, pre-trial prediction will be created for original study design from which the **SubjectData.csv** interim results were generated.

The original study plan assumed a **Target Number of Events** of 374, **Target Sample Size** of 460, **Accrual Period** of 30 and **Study Duration** of 40. The sample size determination (0.025 1-sided significance level for 90% power) assumed a **Hazard Ratio** of ~0.715, **Hazard Rate - Control** of ~0.091 (median survival time of ~10.7 months) and zero dropout. A replication of this sample size calculation in STT2 is provided in Figure 6.168.

STT2-2 / Two Sample Log-Rank Test of Exponential Survival with Exponential Dropout

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **Test Significance Level, α** | 0.02500000 | | | |
| **1 or 2 Sided Test?** | 1 | 2 | 2 | 2 |
| **Length of Accrual Period, a** | 30.00000000 | | | |
| **Maximum Length of Followup, f** | 40.00000000 | | | |
| **Common Exponential Dropout Rate, d** | 0.00000000 | | | |
| **Group 1 Exponential Parameter, $\lambda_1$** | 0.06500000 | | | |
| **Group 2 Exponential Parameter, $\lambda_2$** | 0.09090000 | | | |
| **Hazard Ratio, $h=\lambda_1/\lambda_2$** | 0.71507151 | | | |
| **Power (%)** | 90 | | | |
| **Sample Size per Group, n** | **230** | | | |
| **Total number of events required, E** | 374 | | | |

**Figure 6.168:** Original Sample Size Determination for Example Study

- Scenario A: Unblinded Pre-trial Prediction using original study plan
- Scenario B: Unblinded Events Pre-Trial Prediction using Study Plan with updated hazard rate based on **SubjectData.csv**

**Scenario A** For **Scenario A**, create a new empty prediction workspace (see subsection 6.2.1).

Select **Target = Unblinded Events**, **Input = Summary Data** and **Enrollment Status = Complete** at the **Select the Type of Prediction** step. Then select the Next or → button.

At the **Fixed Parameters** step, enter zero into all fields. Then select the Next or → button.

At the **Accrual Options** step, enter a **Target Sample Size** of 460 and an **Accrual Rate** of 15.3333 (460/30) in the second row of the **Accrual Periods** table. Then select the Next or → button.

At the **Event & Dropout Information** step, in the **Events Model** tab input 374 in the **Target Sample Size** field, 0.715 in the **Hazard Ratio** field and 0.091 in the **Hazard Rate - Control** field giving an automatically calculated value for **Hazard Rate - Treatment** of 0.0651. In the **Dropout Model** tab, enter zero for both **Hazard Rate** fields. Then select the Next or → button.

At the next **Simulation Controls** step, we will add a single simulation run by selecting the checkbox beside the **Save subject-level data for (x) simulation runs** and replace the default of 10 runs with 1 runs at the "(x)" spot. An edited down summary of the **Setup** changes is provided in Figure 6.169.

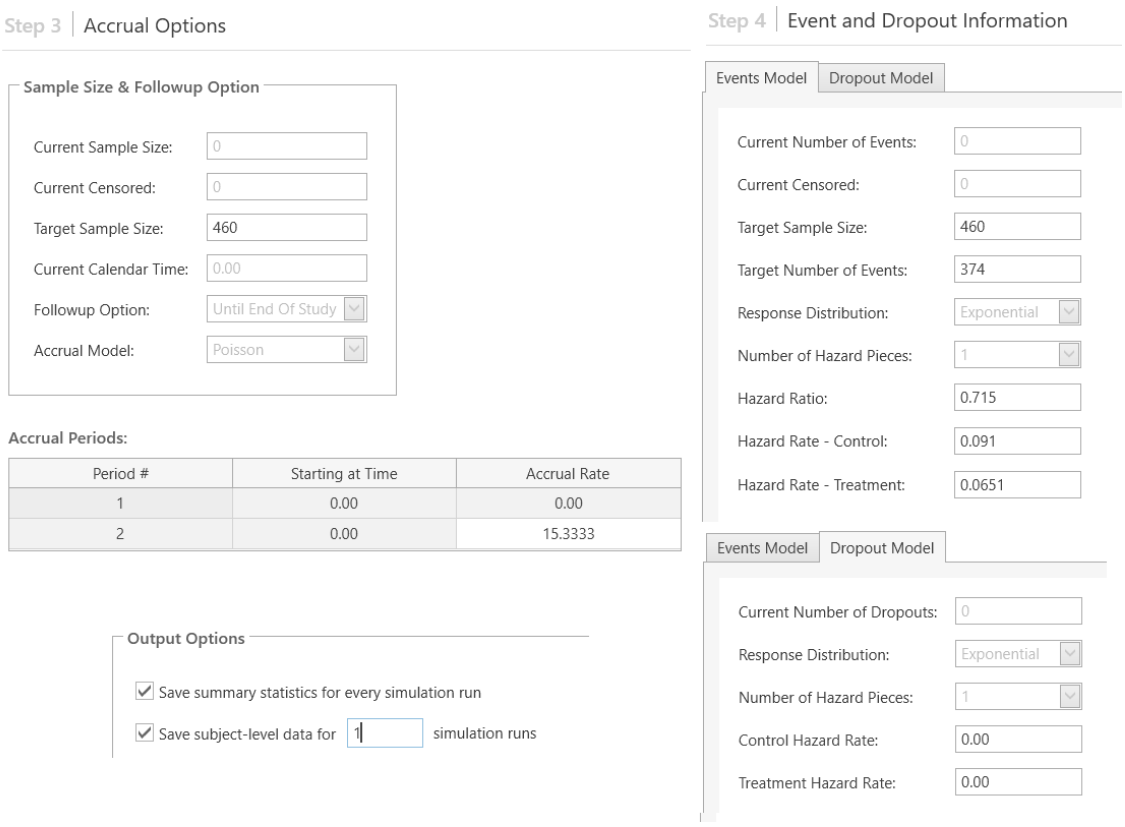



**Figure 6.169:** Scenario A Setup Changes for Unblinded Events Prediction Pre-Trial

To run the simulation again with the new Scenario A, select the Run/$^{\rightarrow}$ button. The main results of this Scenario A simulation are shown in Figure 6.170.

### Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 30.0179 |
| Average Study Duration | 39.7490 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 0.0000 |

### Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 230 |
| Average Treatment Sample Size | 230 |
| Average Control Events | 197.5927 |
| Average Treatment Events | 176.4073 |
| Average Control Dropouts | 0 |
| Average Treatment Dropouts | 0 |

### Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 0 | 27.7927 | 37.2135 |
| 25.00% | 374 | 460 | 0 | 29.0644 | 38.6711 |
| 50.00% | 374 | 460 | 0 | 29.9905 | 39.7116 |
| 75.00% | 374 | 460 | 0 | 30.9549 | 40.7874 |
| 95.00% | 374 | 460 | 0 | 32.3704 | 42.3793 |

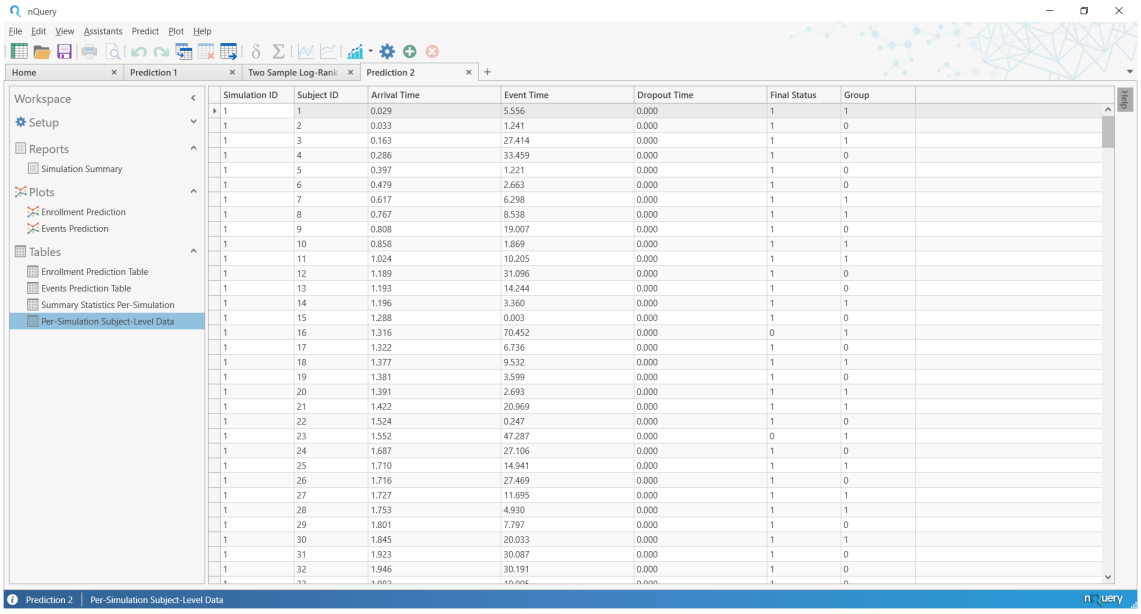
**Figure 6.170:** Scenario A Results for Unblinded Events Prediction Pre-Trial

For Scenario A, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 39.7490. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 37.2135 (5% Percentile) and 42.3793 (95% Percentile), 50% a duration between 38.6711 (25% Percentile) and 40.7874 (75% Percentile) and a median (50% Percentile) duration of 39.7116.

As expected, the **Average Study Duration** is very close to the pre-trial target of 40 months.

The enrollment process led to the **Target Sample Size** of 460 being achieved in all cases with an **Average Accrual Duration** of 30.0179 (Percentiles: 27.7927, 29.0644, 29.9905, 30.9549, 32.3704) which is effectively the same as the target accrual period of 30. This verifies that this prediction based on the original study plan and sample size determination is consistent with that plan.

The **Per-Simulation Subject-level Data Table** was also generated here and is shown for Scenario A in Figure 6.171 with the first simulation (**Simulation ID = 1**). The tables consists of the following fields: **Simulation ID**, **Subject ID**, **Arrival Time**, **Event Time**, **Dropout Time, Final Status** and **Group.** The order of the Subject ID is based on **Arrival Time** order.




| Simulation ID | Subject ID | Arrival Time | Event Time | Dropout Time | Final Status | Group |
|---|---|---|---|---|---|---|
| 1 | 1 | 0.029 | 5.556 | 0.000 | 1 | 1 |
| 1 | 2 | 0.033 | 1.241 | 0.000 | 1 | 0 |
| 1 | 3 | 0.163 | 27.414 | 0.000 | 1 | 1 |
| 1 | 4 | 0.286 | 33.459 | 0.000 | 1 | 0 |
| 1 | 5 | 0.397 | 1.221 | 0.000 | 1 | 0 |
| 1 | 6 | 0.479 | 2.663 | 0.000 | 1 | 0 |
| 1 | 7 | 0.617 | 6.298 | 0.000 | 1 | 1 |
| 1 | 8 | 0.767 | 8.538 | 0.000 | 1 | 1 |
| 1 | 9 | 0.808 | 19.007 | 0.000 | 1 | 0 |
| 1 | 10 | 0.858 | 1.869 | 0.000 | 1 | 1 |
| 1 | 11 | 1.024 | 10.205 | 0.000 | 1 | 1 |
| 1 | 12 | 1.189 | 31.096 | 0.000 | 1 | 0 |
| 1 | 13 | 1.193 | 14.244 | 0.000 | 1 | 0 |
| 1 | 14 | 1.196 | 3.360 | 0.000 | 1 | 1 |
| 1 | 15 | 1.288 | 0.003 | 0.000 | 1 | 0 |
| 1 | 16 | 1.316 | 70.452 | 0.000 | 0 | 1 |
| 1 | 17 | 1.322 | 6.736 | 0.000 | 1 | 0 |
| 1 | 18 | 1.377 | 9.532 | 0.000 | 1 | 1 |
| 1 | 19 | 1.381 | 3.599 | 0.000 | 1 | 0 |
| 1 | 20 | 1.391 | 2.693 | 0.000 | 1 | 1 |
| 1 | 21 | 1.422 | 20.969 | 0.000 | 1 | 1 |
| 1 | 22 | 1.524 | 0.247 | 0.000 | 1 | 0 |
| 1 | 23 | 1.552 | 47.287 | 0.000 | 0 | 1 |
| 1 | 24 | 1.687 | 27.106 | 0.000 | 1 | 0 |
| 1 | 25 | 1.710 | 14.941 | 0.000 | 1 | 1 |
| 1 | 26 | 1.716 | 27.469 | 0.000 | 1 | 0 |
| 1 | 27 | 1.727 | 11.695 | 0.000 | 1 | 1 |
| 1 | 28 | 1.753 | 4.930 | 0.000 | 1 | 1 |
| 1 | 29 | 1.801 | 7.797 | 0.000 | 1 | 0 |
| 1 | 30 | 1.845 | 20.033 | 0.000 | 1 | 1 |
| 1 | 31 | 1.923 | 30.087 | 0.000 | 1 | 0 |
| 1 | 32 | 1.946 | 30.191 | 0.000 | 1 | 0 |

**Figure 6.171:** Per-Simulation Subject-level Data Table for Unblinded Events Pre-Trial

For **Pre-Trial** prediction, **Event Time** and **Dropout Time** are generated for all subjects. **Final Status** will equal **Dropout Status** if **Dropout Time** is less than **Event Time** and the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.167 for per-simulation end times). **Final Status** will equal **Censored Status** if **Event Time** and **Dropout Time** are greater than the **Study End Time** (i.e. study duration) for that simulation (see Figure 6.167 for per-simulation end times)

**Scenario B** Scenario B will be the same as Scenario A expect with the **Event Model** tab at the **Event & Dropout Information** step edited to have exponential hazard rates equal to those default values found using the **SubjectData.csv** (see subsubsection 6.2.7.2 for example).

To evaluate this scenario, select **Event & Dropout Information** from the Navigation Bar from the **Setup** heading on the left. On the **Events Model** tab. Then input the **SubjectData.csv** default estimates for the **Hazard Ratio** (0.7025, compared to pre-trial 0.715) and **Hazard Rate - Control** (0.0779, compared to pre-trial 0.091). **Hazard Rate - Treatment** will automatically update to the correct value of 0.0547. Then select the Next or → button at this step and the **Simulation Controls** step. A summary of the changes compared to Scenario A is provided in Figure 6.172.

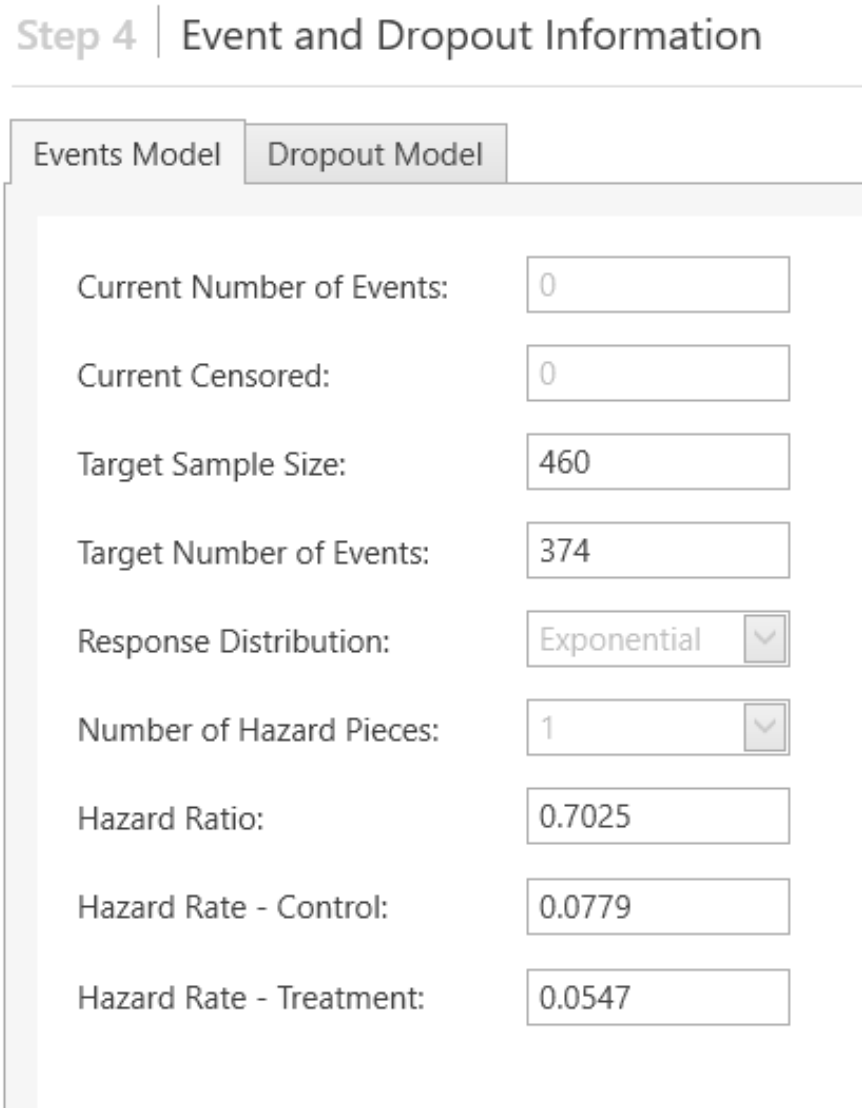



**Figure 6.172:** Scenario B Inputs for Unblinded Events Pre-Trial

The main results for Scenario B are shown in Figure 6.173.

Overall Summary

| | |
|---|---|
| Average Sample Size | 460.0000 |
| Average Accrual Duration | 30.0078 |
| Average Study Duration | 43.3128 |
| Average Events | 374.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 0.0000 |

Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 230 |
| Average Treatment Sample Size | 230 |
| Average Control Events | 198.4642 |
| Average Treatment Events | 175.5358 |
| Average Control Dropouts | 0 |
| Average Treatment Dropouts | 0 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 460 | 0 | 27.7144 | 40.4476 |
| 25.00% | 374 | 460 | 0 | 29.0696 | 42.1066 |
| 50.00% | 374 | 460 | 0 | 30.0010 | 43.2917 |
| 75.00% | 374 | 460 | 0 | 30.9451 | 44.4537 |
| 95.00% | 374 | 460 | 0 | 32.2997 | 46.2845 |

**Figure 6.173:** Scenario B Results for Unblinded Events Prediction Pre-Trial

For Scenario B, the **Overall Summary** shows that the **Target Number of Events** of 374 was reached in every simulation with an **Average Study Duration** over all the simulations of 43.3128. The **Percentile Summary** shows that 90% of predictions had a **Study Duration** between 40.4476 (5% Percentile) and 46.2845 (95% Percentile), 50% a duration between 42.1066 (25% Percentile) and 44.4537 (75% Percentile) and a median (50% Percentile) duration of 43.2917.

With the lower "actual" rates taken from **SubjectData.csv**, the study is expected to around 3.3 months longer than the original plan of 40 months. However, the lower

hazard ratio implies the power would be slightly higher than expected pre-trial all else being equal since it is further from the null hazard ratio of 1.

The enrollment process led to the **Target Sample Size** of 460 being achieved in all cases with an **Average Accrual Duration** of 30.0078 (Percentiles: 27.7927, 29.0644, 29.9905, 30.9549, 32.3704) which is effectively the same as the target accrual period of 30 as per Scenario B.

## 6.2.8 Results Output Summary

### 6.2.8.1 Plots

nQuery Predict provides a number of plots to better explore the results of a milestone prediction. These plots are interactive and editable by the user to maximise the value of these tools. Note that all plots can be saved and edited, see subsubsection 6.2.9.2 for details. To open a given plot, select the required plot from under the **Plots** header in the Workspace Navigation Bar on the left-hand side of the nQuery Predict workspace after a prediction has been completed.

nQuery Predict provides three main plots in the results output. These are:

- Enrollment Prediction Plot (All Cases)
- Events Prediction Plot (Target = Unblinded Events or Blinded Events)
- Dropout Prediction Plot (Target = Unblinded Events or Blinded Events + Current Dropout or Dropout Hazard Rate(s) > 0)

The Enrollment Prediction, Events Prediction and Dropout Prediction plots are all self-similar. Multiple examples of each these plots is provided in the demonstrations above (Enrollment Demonstrations: subsection 6.2.6, Events Prediction: subsection 6.2.7). This section will provide a description of each plot and the tools available to explore and edit these plots.

**Enrollment Prediction Plot** The Enrollment Prediction Plot provides a visual summary of the actual and predicted accrual pattern using a cumulative frequency line graph. An example of an nQuery Prediction workspace with the Enrollment Prediction plot displayed is provided in Figure 6.174.

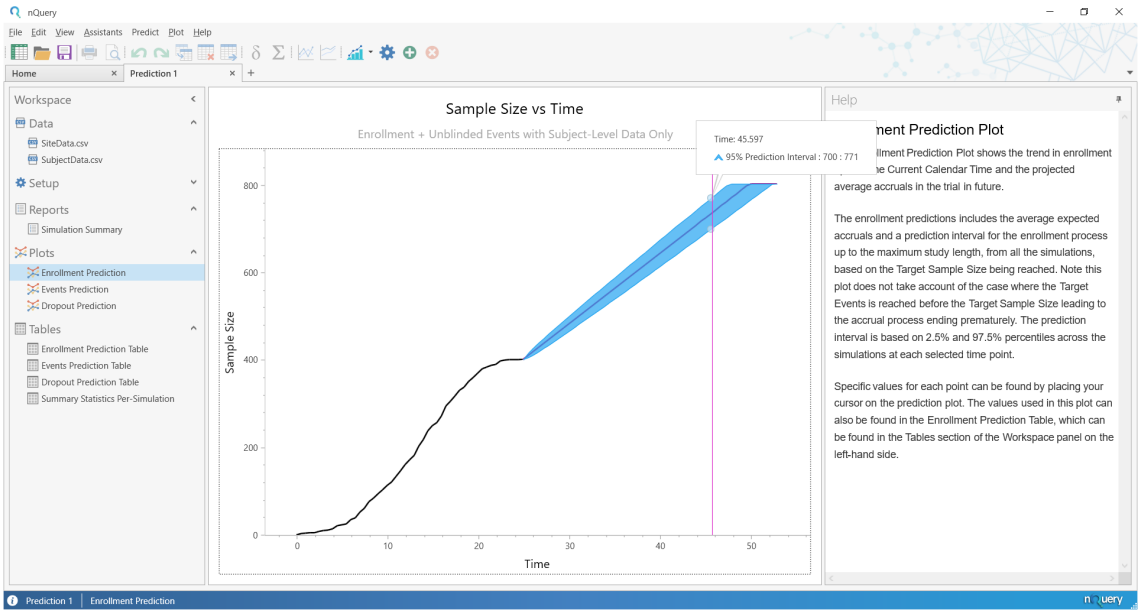



**Figure 6.174:** Enrollment Prediction Plot Example

In the Enrollment Prediction (Sample Size vs Time) plot, the X-axis is **Time** and the Y-axis is **Sample Size**. **Time** is the length of time after study start (**Time = 0**) at which the Y-axis is being evaluated. **Sample Size** is the actual/predicted cumulative number of subjects enrolled at a given value of **Time**. The X-axis length will be related to the **Accrual Duration** (i.e. length of time over which subjected were being enrolled to reach the **Target Sample Size**) over all simulations. The Y-axis height will be related to the **Target Sample Size** specified for that prediction.

In the majority of cases (exceptions discussed below), the Enrollment Prediction plot will consist of two parts:

1. The actual accrual pattern up to the **Current Time**
2. The predicted accrual pattern from the **Current Time** up to the maximum projected **Accrual Duration**

The actual accrual pattern provides a visual summary of when subjects entered the study. The actual accrual pattern will be represented by a solid black line. If the user places the cursor over the plot over this part of the line, a pop-up will be shown on the plot with following text:

Time: A

Actual Sample Size: B

where A and B are the actual values at the currently highlighted **Time** value.

If **Subject-level Data** has been provided, the actual pattern will be based on the **Arrival Time** column values. If **Summary Data** is used, a linear projection is

used from zero to the **Current Sample Size**. If conducting a pre-trial prediction, there will be no actual accrual pattern to show. The plot subtitle will display the input used for the current plot.

The predicted accrual pattern provides a visual summary of the projected number of subjects enrolled at a given future **Time** value after the **Current Time**. The predicted **Average Sample Size** at a given **Time** is represented by a grey line. This grey line is surrounded by a blue area which represents the 95% Prediction Interval for the **Sample Size** at a given time. If the user places the cursor over the plot over the grey part of the line, a pop-up will be shown on the plot with the following text:

Time : A

Predicted Sample Size: B

where A and B are the actual values at the currently highlighted **Time** value.

If the user places the cursor over the blue prediction interval section, a pop-up will be shown on the plot with the following text:

Time = A

95% Prediction Interval: B:C

where A is the **Time** and B:C is the range of the 95% Prediction Interval at the currently highlighted **Time** value. This is shown in Figure 6.174.

Note that this prediction interval section is not shown if **Enrollment Status = Complete** and only the actual accrual pattern up to the **Current Time** is shown. See subsubsection 6.2.7.3 for an example of this.

Note that where the enrollment process is stopped early due the **Target Number of Events** being reached before the **Target Sample Size**, the **Enrollment Prediction Plot** will still include the projection for the original **Target Sample Size** despite the **Simulation Summary Report** showing the sample size at the time the **Target Number of Events** was achieved instead.

Note that values used in this plot can be inspected in the **Enrollment Prediction Table.** See subsubsection 6.2.8.3 for details. This plot can also be edited and saved, see subsubsection 6.2.9.2 for details.

**Event Prediction Plot** The Event Prediction Plot provides a visual summary of the actual and predicted event pattern using a cumulative frequency line graph. An example of an nQuery Prediction workspace with the Event Prediction plot displayed is provided in Figure 6.175.

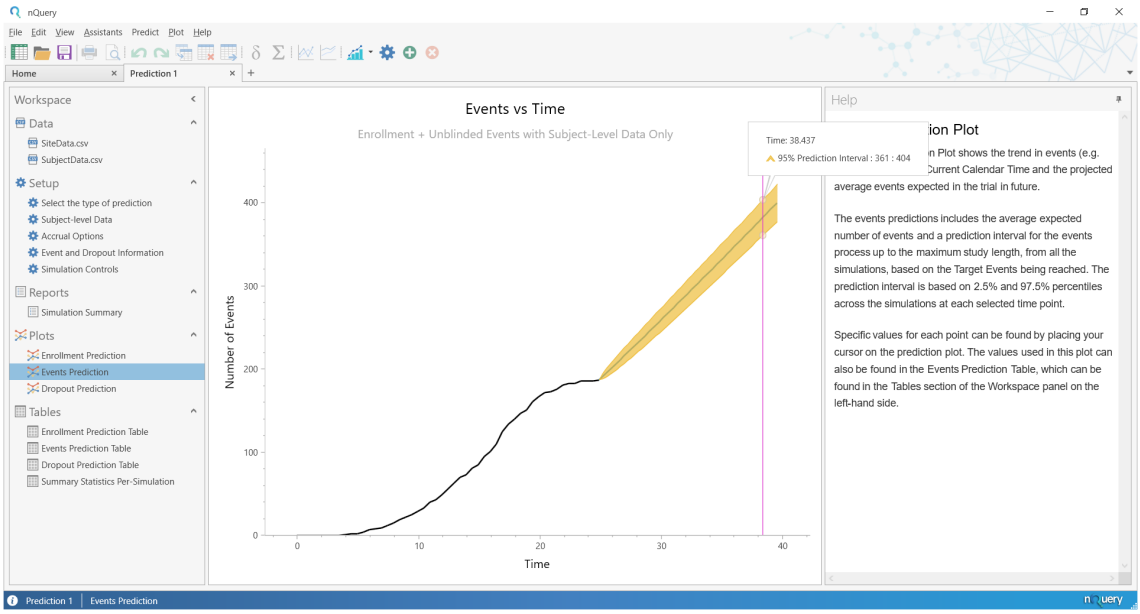



**Figure 6.175:** Event Prediction Plot Example

In the Event Prediction (Events vs Time) plot, the X-axis is **Time** and the Y-axis is **Number of Events**. **Time** is the length of time after study start (**Time = 0**) at which the Y-axis is being evaluated. **Number of Events** is the actual/predicted cumulative number of subjects who have had the event of interest at a given value of **Time**. The X-axis length will be related to the **Study Duration** (i.e. length of time for the **Target Number of Events** to occur) over all simulations. The Y-axis height will be related to the **Target Number of Events** specified for that prediction.

Note that the time of events represented here is when these events occurred relative to the study start (**Time = 0**) NOT when these events occurred relative to subject entry (i.e. follow up time). Therefore, this plot should not be confused with a Kaplan-Meier type plot.

In the majority of cases (exceptions discussed below), the Events Prediction plot will consist of two parts:

1. The actual event pattern up to the **Current Time**
2. The predicted event pattern from the **Current Time** up to the maximum projected **Study Duration**

The actual event pattern provides a visual summary of when subjects had the event. The actual event pattern will be represented by a solid black line. If the user places the cursor over the plot over this part of the line, a pop-up will be shown on the plot with following text:

Time: A

Actual # of Events: B

where A and B are the actual values at the currently highlighted **Time** value.

If **Subject-level Data** has been provided, the event pattern will be based on the sum of the **Arrival Time** and **Time on Study** column values (i.e. length of time after study start at which event occurred) for subjects who had **Status Indicator = Event Status**. If **Summary Data** is used, a linear projection is used from zero to the **Current Number of Events**. If conducting a pre-trial prediction, there will be no actual event pattern to show. The plot subtitle will display the input used for the current plot.

The predicted event pattern provides a visual summary of the projected number of subjects who will have had the event of interest at a given future **Time** value after the **Current Time**. The predicted **Average Number of Events** at a given **Time** is represented by a grey line. This grey line is surrounded by a yellow area which represents the 95% Prediction Interval for the **Number of Events** at a given time. If the user places the cursor over the plot over the grey part of the line, a pop-up will be shown on the plot with the following text:

Time : A

Predicted # of Events: B

where A and B are the actual values at the currently highlighted **Time** value.

If the user places the cursor over the yellow prediction interval section, a pop-up will be shown on the plot with the following text:

Time = A

95% Prediction Interval: B:C

where A is the **Time** and B:C is the range of the 95% Prediction Interval at the currently highlighted **Time** value. This is shown in Figure 6.175.

Note that values used in this plot can be inspected in the **Events Prediction Table.** See subsubsection 6.2.8.3 for details. This plot can also be edited and saved, see subsubsection 6.2.9.2 for details.

**Dropout Prediction Plot** The Dropout Prediction Plot provides a visual summary of the actual and predicted dropout pattern using a cumulative frequency line graph. An example of an nQuery Prediction workspace with the Dropout Prediction plot displayed is provided in Figure 6.176.

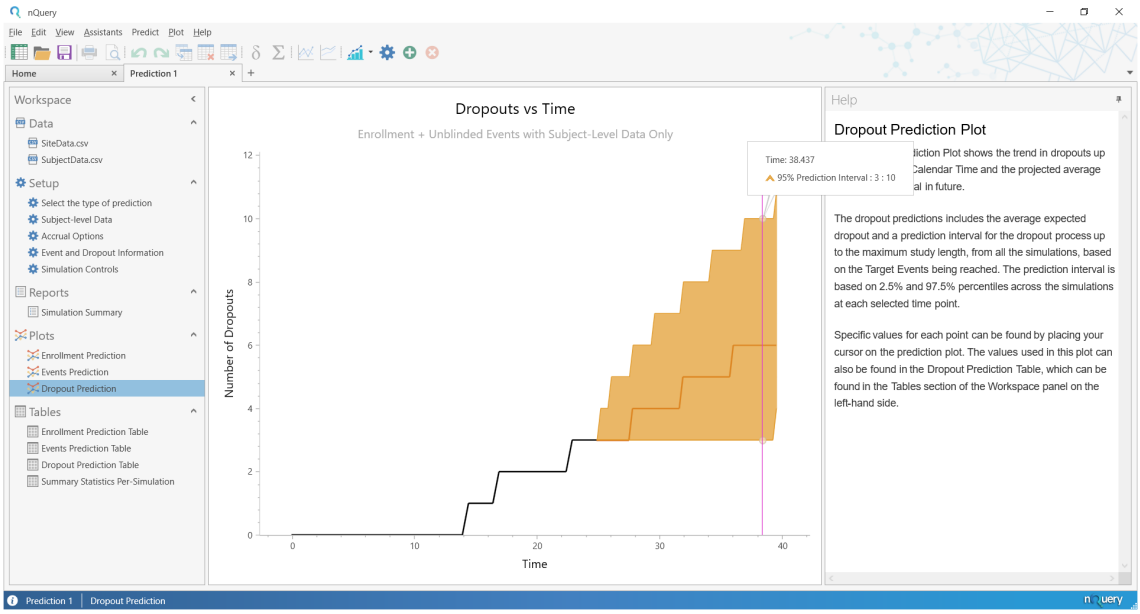



**Figure 6.176:** Dropout Prediction Plot Example

In the Event Prediction (Dropouts vs Time) plot, the X-axis is **Time** and the Y-axis is **Number of Dropouts**. **Time** is the length of time after study start (**Time = 0**) at which the Y-axis is being evaluated. **Number of Dropouts** is the actual/predicted cumulative number of subjects who have dropped out at a given value of **Time**. The X-axis length will be related to the **Study Duration** (i.e. length of time for the **Target Number of Events** to occur) over all simulations. The Y-axis height will be related to the upper 95% Prediction Interval Limit value for the projected **Number of Dropouts** at study end.

In the majority of cases (exceptions discussed below), the Dropout Prediction plot will consist of two parts:

1. The actual dropout pattern up to the **Current Time**
2. The predicted dropout pattern from the **Current Time** up to the maximum projected **Study Duration**

The actual dropout pattern provides a visual summary of when subjects have dropped out. The actual dropout pattern will be represented by a solid black line. If the user places the cursor over the plot over this part of the line, a pop-up will be shown on the plot with following text:

Time: A

Actual # of Dropouts: B

where A and B are the actual values at the currently highlighted **Time** value.

If **Subject-level Data** has been provided, the dropout pattern will be based on the sum of the **Arrival Time** and **Time on Study** column values (i.e. length

of time after study start at which dropout occurred) for subjects who had **Status Indicator = Dropout Status**. If **Summary Data** is used, a linear projection is used from zero to the **Current Number of Dropouts**. If conducting a pre-trial prediction, there will be no actual dropout pattern to show. The plot subtitle will display the input used for the current plot.

The predicted dropout pattern provides a visual summary of the projected number of subjects who will have dropped out at a given future **Time** value after the **Current Time**. The predicted **Average Number of Dropouts** at a given **Time** is represented by a grey line. This grey line is surrounded by an orange area which represents the 95% Prediction Interval for the **Number of Dropouts** at a given time. If the user places the cursor over the plot over the grey part of the line, a pop-up will be shown on the plot with the following text:

Time : A

Predicted # of Dropouts: B

where A and B are the actual values at the currently highlighted **Time** value.

If the user places the cursor over the orange prediction interval section, a pop-up will be shown on the plot with the following text:

Time = A

95% Prediction Interval: B:C

where A is the **Time** and B:C is the range of the 95% Prediction Interval at the currently highlighted **Time** value. This is shown in Figure 6.176.

Note that values used in this plot can be inspected in the **Dropout Prediction Table.** See subsubsection 6.2.8.3 for details. This plot can also be edited and saved, see subsubsection 6.2.9.2 for details.

Note that the **Dropout Prediction** plot and table will not be generated if there are no dropouts in the study. This will occur if up to the **Current Time** no dropouts occur (i.e. no values with **Dropout Status** in **Status Indicator** column if **Subject-level Data** used or **Number of Dropout** field(s) set to zero if **Summary Data** used) **AND** the **Hazard Rate**(s) in the **Dropout Model** tab of the **Event & Dropout Information** step are set to zero.

#### 6.2.8.2 Reports

nQuery Predict provides reports to explore the inputs used to generate a prediction and results of the current predictions. Reports are editable by the user to maximise the functionality of the report. To open a given report, select the required report from under the **Reports** header in the Workspace Navigation Bar on the left-hand side of the nQuery Predict workspace after a prediction has been completed.

nQuery Predict provides one main report in the results output This is the:

- Simulation Summary (All Cases)

Multiple examples the **Simulation Summary** report are provided in the demonstrations above (Enrollment Demonstrations: subsection 6.2.6, Events Prediction: subsection 6.2.7). This section will provide a description of the report and the tools available to explore and edit this report.

The **Simulation Summary** report can be split into the two parts:

1. Inputs (Left-Hand Column)
2. Results (Right-Hand Column)

The **Inputs** column gives a summary of all the inputs used in the **Setup** phase to generate the current prediction. The **Results** column on the right gives a summary of the results of the current prediction. An example **Simulation Summary Report** displayed in an nQuery Predict workspace is shown in

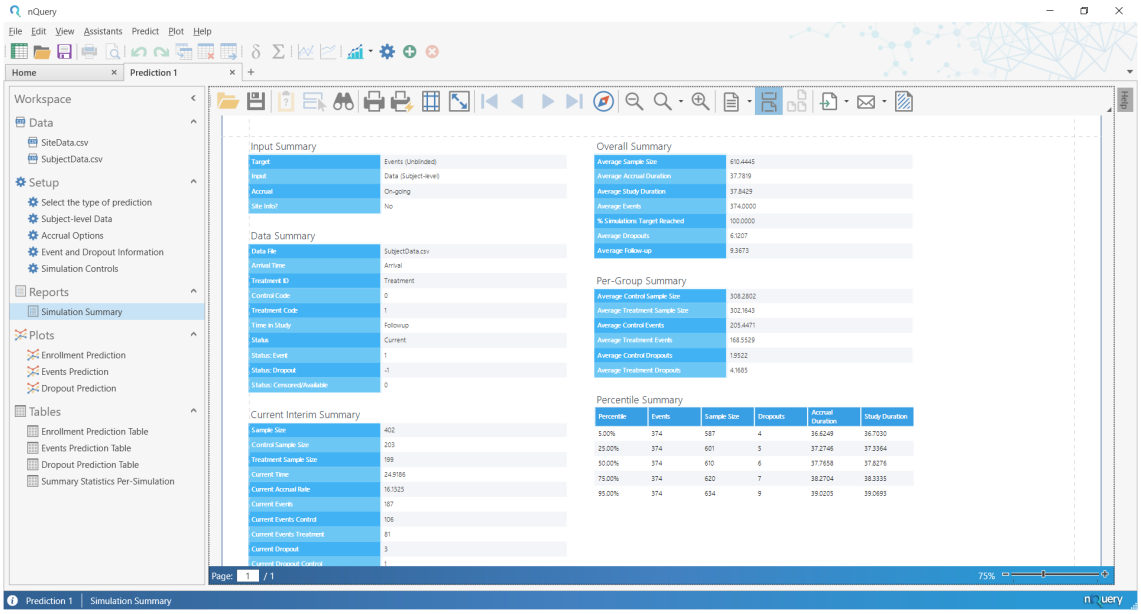



**Figure 6.177:** Simulation Summary Report Example

Each column is made of multiple sections which contain specific information on the setup (e.g. **Input Summary**) or results (e.g. **Overall Summary**). The sections provided in each column will sometimes change depending on the inputs provided during the **Setup** stage of the current prediction. A summary of the **Inputs** and **Results** sections follows.

**Inputs** The **Inputs** sections shown and the information shown in a given section will change to reflect the inputs provided by the user during the **Setup** stage. The following **Input** sections (alongside their **Select the type of Prediction** step

condition of appearance if any) are as follows (in order of appearance within the **Simulation Summary Report**):

- Input Summary (All Predictions)
- Data Summary (Input = Subject Data Only, Subject + Site Data)
- Current Interim Summary (All Predictions)
- Current Site Summary (Input = Subject + Site Data)
- Target Enrollment Summary (Enrollment Status = Ongoing)
- Target Events Summary (Target = Unblinded Events, Blinded Events)
- Dropouts Summary (Target = Unblinded Events, Blinded Events + Dropouts $> 0$ for all simulations)
- Simulation Summary (All Predictions)

The **Input Summary** section is shown for all predictions and provides a high-level summary of the target prediction and its inputs. The **Input Summary** provides a summary of the **Target** (Enrollment Only, Events (Unblinded), Events(Blinded)), **Input** (Subject Data Only, Subject + Site Data, Summary Data), **Accrual** status (Ongoing, Complete) and if site-level data was used (**Site Info?** Yes/No). The **Target** field will also include additional information if Enrollment is complete (**Complete Enrollment**) or summary data was used (**Fixed**). Examples of **Input Summary** sections are provided in Figure 6.178.

Input Summary

| | |
|---|---|
| Target | Complete Enrollment + Events (Unblinded) |
| Input | Data (Subject-level) |
| Accrual | Complete |
| Site Info? | No |

Input Summary

| | |
|---|---|
| Target | Events (Blinded) |
| Input | Data (Subject and Site level) |
| Accrual | On-going |
| Site Info? | Yes |

Input Summary

| | |
|---|---|
| Target | Enrollment Fixed |
| Input | Fixed Parameters |
| Accrual | On-going |
| Site Info? | No |

**Figure 6.178:** Input Summary Section Examples

The **Data Summary** section is shown only for predictions which used **Subject Data Only** or **Subject + Site Data** as an **Input** and provides a summary of which datasets were used and the columns in the selected subject and site data (if used) which were assigned to the required fields for the current prediction. These fields will correspond the choices made at the **Subject-level Data** and **Site-level Data** steps of the **Setup** stage (see Enrollment Demonstrations: subsection 6.2.6, Events Prediction: subsection 6.2.7). Examples of **Data Summary** sections are provided in Figure 6.179.

Data Summary

| | |
|---|---|
| Subject-Level Data File | SubjectData.csv |
| Arrival Time | Arrival |
| Site-Level Data File | SiteData.csv |
| Site ID | SID |
| Site Accrual Rate | Rate |
| Site Recruitment Cap | Cap |
| Open Site Initiation Times | OpenTime |
| Unopened Site Initiation Start Times | Start |
| Unopened Site Initiation End Times | End |

Target = Enrollment Only
Input = Subject + Site Data

Data Summary

| | |
|---|---|
| Data File | SubjectData.csv |
| Arrival Time | Arrival |
| Treatment ID | Treatment |
| Control Code | 0 |
| Treatment Code | 1 |
| Time in Study | Followup |
| Status | Current |
| Status: Event | 1 |
| Status: Dropout | -1 |
| Status: Censored/Available | 0 |

Target =Unblinded Events
Input = Subject Data Only

Data Summary

| | |
|---|---|
| Data File | SubjectData.csv |
| Arrival Time | Arrival |
| Time in Study | Followup |
| Status | Current |
| Status: Event | 1 |
| Status: Dropout | -1 |
| Status: Censored/Available | 0 |

Target = Blinded Events
Input = Subject Data Only

**Figure 6.179:** Data Summary Section Examples

The **Current Interim Summary** section is shown for all prediction and provides a summary of study up to the **Current Time** either based on the **Subject-level Data** provided (**Input = Subject Data Only** or **Subject + Site Data**) or the **Fixed Parameters** provided if **Input = Summary Data**. Note this is referred to as the **Current Status Summary** for **Target = Enrollment** only but its purpose is the same. These fields provide a summary of the current study time, sample size, events, dropouts and follow-up from the subject-level data or summary data. Per-Group summaries are also provided for the **Unblinded Events Target** (see Enrollment Demonstrations: subsection 6.2.6, Events Prediction: subsection 6.2.7). Examples of **Current Interim Summary** sections are provided in Figure 6.180.

Current Status Summary

| | |
|---|---|
| Current Sample Size | 100 |
| Current Time | 10.0000 |
| Current Accrual Rate | 10.0000 |

Target = Enrollment Only

Current Interim Summary

| | |
|---|---|
| Sample Size | 402 |
| Current Time | 24.9186 |
| Current Accrual Rate | 16.1325 |
| Current Events | 187 |
| Current Dropout | 3 |
| Average Follow-up | 9.3352 |

Target = Blinded Events

Current Interim Summary

| | |
|---|---|
| Sample Size | 402 |
| Control Sample Size | 203 |
| Treatment Sample Size | 199 |
| Current Time | 24.9186 |
| Current Accrual Rate | 16.1325 |
| Current Events | 187 |
| Current Events Control | 106 |
| Current Events Treatment | 81 |
| Current Dropout | 3 |
| Current Dropout Control | 1 |
| Current Dropout Treatment | 2 |
| Average Follow-up | 7.5306 |

Target = Unblinded Events

**Figure 6.180:** Current Interim Summary Section Examples

The **Current Site Summary** section is shown for predictions with **Input = Subject + Site Data** and provides a summary of site status at the **Current Time**. It consists of two fields: **Sites Open** and **Sites Available**. These correspond to the number of sites currently open (i.e. have a **Site Initiation Time** value for

that subject) or are available to open after the **Current Time** (i.e. do not have **Site Initiation Time** value for that subject). An example of a **Current Site Summary** section is provided in Figure 6.181.

Current Site Summary

| | |
|---|---|
| Sites Open | 118 |
| Sites Available | 9 |

**Figure 6.181:** Current Site Summary Section Example

The **Target Enrollment Summary** section is shown for when **Enrollment Status = Ongoing** and provides a summary of the **Target Sample Size**, **Accrual** (i.e. Enrollment) **Model** and required parameters for generating enrollments using the selected **Accrual Model**. For example, for the **Poisson Accrual Model**, a **Future Accrual Rate** field is provided to correspond to the bottom-right cell of the **Accrual Periods** table at the **Accrual Options** step (see Enrollment Demonstrations: subsection 6.2.6, Events Prediction: subsection 6.2.7). Note that if **Input = Subject + Site Data**, this will not include the required parameters with input parameters viewable from the **Accrual Infos/Site** tab at the **Accrual Options** step of **Setup**. An example of the **Target Enrollment Summary** section is provided in Figure 6.182.

Target Enrollment Summary

| | |
|---|---|
| Accrual Model | Poisson |
| Target Sample Size | 400 |
| Future Accrual Rate | 20.0000 |

**Figure 6.182:** Target Enrollment Summary Section Example

The **Target Events Summary** section is shown for predictions with **Target = Unblinded Events** or **Blinded Events** and provides a summary of the (Number of) **Target Events**, **Events Model** and required parameters for generating event times using the selected **Events Model**. The **Events Model** and its parameters will depend on the choices made in the **Events Model** tab of the **Event & Dropout Information** step of **Setup** (see Events Prediction Demonstrations: subsection 6.2.7).

Note that for the case where **Response Distribution = Exponential** and **Number of Hazard Pieces > 1** in the **Events Model** tab a **Piecewise Exponential**

**Summary** table is provided at the bottom of the **Results** column on the right. Examples of **Target Events Summary** sections are provided in Figure 6.183.

Target Events Summary

| | |
|---|---|
| Target Events | 51 |
| Events Model | Exponential |
| Event Hazard Rate | 0.0693 |

Input = Blinded Events
Events Model = Exponential
# Hazard Pieces = 1

Target Events Summary

| | |
|---|---|
| Target Events | 199 |
| Events Model | Weibull |
| Control Scale | 0.0726 |
| Treatment Scale | 0.0466 |
| Control Shape | 0.8526 |
| Treatment Shape | 0.8112 |

Input = Unblinded Events
Events Model = Weibull

Target Events Summary

| | |
|---|---|
| Target Events | 374 |
| Events Model | Exponential |

Piecewise Exponential Summary

| Time | Event Hazard Rate |
|---|---|
| 0.0000 | 0.0658 |
| 5.0000 | 0.0500 |
| 10.0000 | 0.0400 |

Input = Blinded Events
Events Model = Exponential
# Hazard Pieces = 3 (Table above)

**Figure 6.183:** Target Events Summary Section Examples

The **Dropouts Summary** section is shown for predictions with **Target = Unblinded Events** or **Blinded Events** and provides a summary of the **Dropout Model** and required parameters for generating dropout times using the selected **Dropout Model**. The **Dropout Model** and its parameters will depend on the choices made in the **Dropouts Model** tab of the **Event & Dropout Information** step of **Setup** (see Events Prediction Demonstrations: subsection 6.2.7). Note that for the case where **Response Distribution = Exponential** and **Number of Hazard Pieces > 1** in the **Dropouts Model** tab a **Piecewise Dropouts Summary** table is provided at the bottom of the **Results** column on the right. Examples of **Dropouts Summary** sections are provided in Figure 6.184.

Dropouts Summary

| | |
|---|---|
| Dropout Model | Exponential |
| Dropout Hazard Rate | 0.0051 |

Target = Blinded Events
Model = Exponential
# Hazard Pieces = 1

Dropouts Summary

| | |
|---|---|
| Dropout Model | Exponential |

Piecewise Dropouts Summary

| Time | Control Dropout Hazard Rate | Treatment Dropout Hazard Rate |
|---|---|---|
| 0.0000 | 0.0007 | 0.0014 |
| 10.0000 | 0.0000 | 0.0020 |

Target = Unblinded Events
Model = Exponential
# Hazard Pieces = 2 (Table above)

**Figure 6.184:** Target Events Summary Section Examples

The **Simulation Summary** section is shown for all predictions and provides a summary of the simulation setup parameters selected at the **Simulation Controls**

step. It consists of three fields: **Seed**, **Simulations** and **Simulation Time**. These correspond to the random seed used in the current simulation, the total number of simulations used in the current simulation and the length of time (in seconds) the current prediction took to simulate. An example of a **Simulation Summary** section is provided in Figure 6.185.

Simulation Summary

| | |
|---|---|
| Seed | 27282703 |
| Simulations | 10000 |
| Simulation Time | 6.3904 |

**Figure 6.185:** Simulation Summary Section Example

**Results Column** The **Results** sections shown and the information shown in a given section will change to reflect the inputs provided by the user during the **Setup** stage. The following sections (alongside their **Select the type of Prediction** step condition of appearance if any) is as follows:

- Overall Summary (All Predictions)
- Per-Group Summary (Target = Unblinded Events)
- Percentile Summary (All Predictions)

**Inputs Column Sections** The **Overall Summary** section is shown for all predictions and provides the average results for the main outputs of the prediction. Depending on the Target (**Enrollment Only** versus **Blinded Events / Unblinded Events**) and Inputs (**Subject Data Only/Summary Data** versus **Subject + Site Data**) the Overall Summary will include results for the average across all simulations for the **Sample Size (All), Accrual Period Duration (All), Study Duration (Blinded Events / Unblinded Events only), Events (Blinded Events / Unblinded Events only), Dropouts (Blinded Events / Unblinded Events only), Follow-up Time (Blinded Events / Unblinded Events only)** and **Number of Sites Opened (Subject + Site Data only)** as well as the **% of Simulations Target Reached** (**Blinded Events / Unblinded Events only**; where "target" is **Target Number of Events** (which may not be reached if **Dropout Hazard Rate(s)** are high enough for example). Examples of **Overall Summary** sections are provided in Figure 6.186.

Overall Summary

| | |
|---|---|
| Average Sample Size | 804.0000 |
| Average Study Duration | 43.7401 |
| % Simulations Target Reached | 100.0000 |
| Sites Opened | 127.0000 |

Target = Enrollment Only
Input = Subject + Site Data

Overall Summary

| | |
|---|---|
| Average Sample Size | 402.0000 |
| Average Accrual Duration | 24.9186 |
| Average Study Duration | 25.9942 |
| Average Events | 199.0000 |
| % Simulations Target Reached | 100.0000 |
| Average Dropouts | 3.2316 |
| Average Follow-up | 7.6226 |

Target = Events (both same)
Input = Subject Data Only/
Summary Data

**Figure 6.186:** Overall Summary Section Examples

The **Per-Group Summary** section is shown for predictions where **Target = Unblinded Events** and provides the average results for the main outputs of the prediction for the **Control Group** and **Treatment Group**. For each group it provides the average **Sample Size, Events** and **Dropouts**. An example of the **Per-Group Summary** section is provided in Figure 6.187.

Per-Group Summary

| | |
|---|---|
| Average Control Sample Size | 203 |
| Average Treatment Sample Size | 199 |
| Average Control Events | 112.5783 |
| Average Treatment Events | 86.4217 |
| Average Control Dropouts | 1.0706 |
| Average Treatment Dropouts | 2.161 |

**Figure 6.187:** Per-Group Summary Section Example

The **Percentile Summary** section is shown for all predictions and provides the percentile results for the main outputs of the prediction. The required percentiles are specified in the **Simulation Controls** step of **Setup**. Depending on the Target (**Enrollment Only** versus **Blinded Events / Unblinded Events**) and Inputs (**Subject Data Only/Summary Data** versus **Subject + Site Data**) the **Percentile Summary** will include percentile results for the **Sample Size (All), Accrual Events (Blinded Events / Unblinded Events only), Dropouts (Blinded Events / Unblinded Events only), Period Duration (All), Study Duration (Blinded Events / Unblinded Events only)** and **Number of Sites Opened (Subject + Site Data only).** Examples of **Percentile Summary** sections (top is **Target = Enrollment Only**, **Input = Subject + Site Data**; bottom is **Target = Unblinded Events** (note it would be same format for **Blinded Events**), **Input = Subject Data Only**) are provided in Figure 6.188.

Percentile Summary

| Percentile | Sample Size | Study Duration | Sites Opened |
|---|---|---|---|
| 5.00% | 804 | 42.1003 | 127 |
| 25.00% | 804 | 43.0285 | 127 |
| 50.00% | 804 | 43.7222 | 127 |
| 75.00% | 804 | 44.4059 | 127 |
| 95.00% | 804 | 45.4795 | 127 |

Percentile Summary

| Percentile | Events | Sample Size | Dropouts | Accrual Duration | Study Duration |
|---|---|---|---|---|---|
| 5.00% | 374 | 587 | 4 | 36.5928 | 36.6565 |
| 25.00% | 374 | 601 | 5 | 37.2794 | 37.3398 |
| 50.00% | 374 | 610 | 6 | 37.7507 | 37.8121 |
| 75.00% | 374 | 620 | 7 | 38.2445 | 38.3030 |
| 95.00% | 374 | 634 | 9 | 38.9979 | 39.0529 |

**Figure 6.188:** Percentile Summary Section Examples

Note as mentioned in **Inputs Column Section**, if the **Exponential Response Distribution** is used with a **Number of Hazard Pieces > 1** (i.e. piecewise exponential model) for the **Events Model** and/or **Dropout Model** then the piecewise exponential table will be shown below the results sections discussed above (examples given in **Inputs** section above).

**Report Toolbar** In nQuery Predict, each **Report** has a toolbar which allows the user to edit, save, print and export a report. The toolbar options available for nQuery Predict reports is shown in Figure 6.189.

**Figure 6.189:** Report Toolbar

These options are defined as follows:

1. Open: Open another nQuery Predict Report (.prnx)
2. Save: Save current nQuery Predict Report (.prnx)
3. Help: Unused
4. URL: Unused
5. Search: Search for text in current report. Select 14 to remove
6. Print: Open Print menu for current report
7. Quick Print: Print report using system defaults
8. Page Setup: Open Page Setup window to edit page size, orientation and margin sizes
9. Scale: Open Scale window to rescale size of report text by percentage or page size
10. First Page: Return to first page if report has multiple pages
11. Page Up: Go up one page if report has multiple pages
12. Page Down: Go down one page if report has multiple pages
13. Last Page: Go to last page if report has multiple pages
14. Navigation Pane: Open and close Navigation Pane. Contains Pages list and Search Results window
15. Zoom Out: Zoom out report by one increment
16. Zoom: Change the current zoom from list of options
17. Zoom In: Zoom in report by one increment
18. Page Layout: Set out how multiple pages are displayed (Single Page, Two Pages, Wrap Around)
19. Continuous Scolling: Enable or disable continuous scrolling in report
20. Show Cover Page: Show cover page

21. Export: Export report as other file type. Select arrow to see list in-report or select icon to open Export window
22. Send: Send report as email in export file type using system default email client.
23. Watermark: Add watermark to result using Watermark menu

The Export option can be used to save the report in the following file formats: PDF, HTML, MHT, RDF, DOCX, XLS, XLSX, CSV, Plain Text File (TXT), Image File (PNG, JPEG, BMP, GIF, EMF, WMF, TIFF).

#### 6.2.8.3 Tables

nQuery Predict provides a number of tables to better explore and understand the results of a prediction. Note that these tables will be exported (as .csv files) when an nQuery Predict workspace is saved (see subsubsection 6.2.9.1). To open a given table, select the required table from under the **Tables** header in the Workspace Navigation Bar on the left-hand side of the nQuery Predict workspace after a prediction has been completed.

nQuery Predict can provide up to seven tables in the results output. These are:

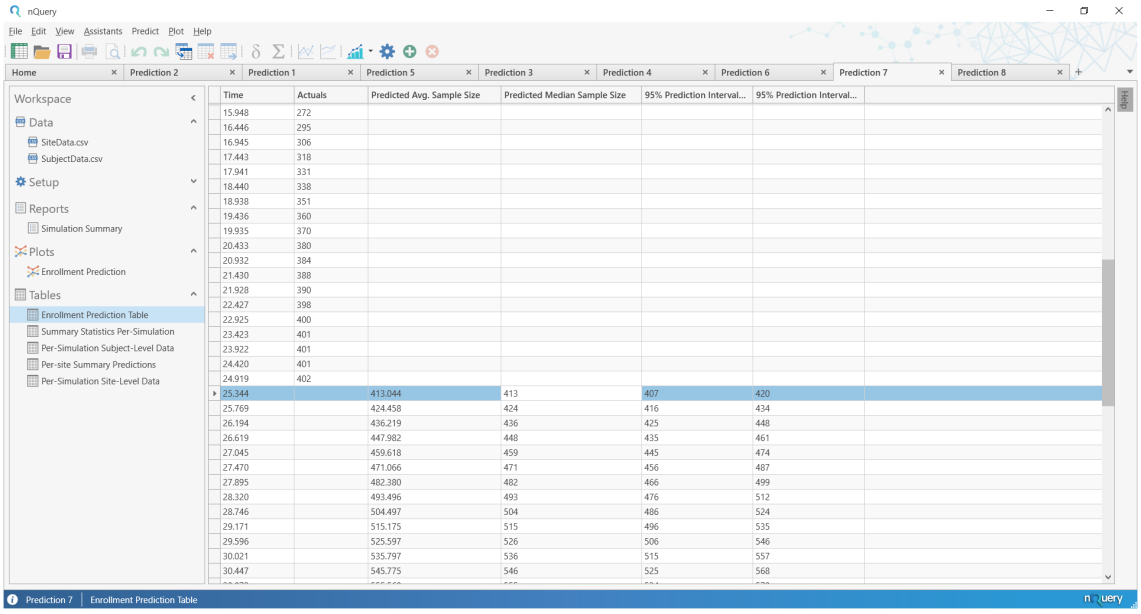
- Enrollment Prediction Table (All Cases)
- Events Prediction Table (Target = Unblinded Events or Blinded Events)
- Dropout Prediction Table (Target = Unblinded Events or Blinded Events + Current Dropout or Dropout Hazard Rate(s) > 0)
- Summary Statistics Per-Simulation (Optional - All Cases)
- Per-Simulation Subject-level Data (Optional - All Cases)
- Per-Site Summary Statistics (Optional - Input = Subject + Site Data)
- Per-Simulation Site-level Data (Options - Input = Subject + Site Data)

The Enrollment Prediction, Events Prediction and Dropout Prediction tables are all self-similar and associated with their respective plots.

The Summary Statistics Per-Simulation and Per-Simulation Subject-level Data provide additional information for specific simulations on the subject-level. The Per-Site Summary Statistics and Per-Simulation Site-level Data provide additional information for specific simulations at the site-level. All four of these tables are optional and can be added or removed to the **Tables** header at the **Simulation Controls** step of **Setup** (subsubsection 6.2.9.4).

Multiple examples of each these tables is provided in the worked example above (Enrollment Demonstrations: subsection 6.2.6, Events Prediction: subsection 6.2.7). This section will provide a description of each table.

**Enrollment Prediction Table** The Enrollment Prediction Table provides a tabular summary of the actual and predicted accrual pattern alongside a 95% prediction interval for the projected accrual pattern. This table will consist of 100 **Time** points of which half are assigned before and after the **Current Time**. An example of an nQuery Prediction workspace with the Enrollment Prediction Table displayed is provided in Figure 6.190 in which the first projected future **Time** value of 25.344 has been highlighted.




| Time | Actuals | Predicted Avg. Sample Size | Predicted Median Sample Size | 95% Prediction Interval... | 95% Prediction Interval... |
|---|---|---|---|---|---|
| 15.948 | 272 | | | | |
| 16.446 | 295 | | | | |
| 16.945 | 306 | | | | |
| 17.443 | 318 | | | | |
| 17.941 | 331 | | | | |
| 18.440 | 338 | | | | |
| 18.938 | 351 | | | | |
| 19.436 | 360 | | | | |
| 19.935 | 370 | | | | |
| 20.433 | 380 | | | | |
| 20.932 | 384 | | | | |
| 21.430 | 388 | | | | |
| 21.928 | 390 | | | | |
| 22.427 | 398 | | | | |
| 22.925 | 400 | | | | |
| 23.423 | 401 | | | | |
| 23.922 | 401 | | | | |
| 24.420 | 401 | | | | |
| 24.919 | 402 | | | | |
| 25.344 | | 413.044 | 413 | 407 | 420 |
| 25.769 | | 424.458 | 424 | 416 | 434 |
| 26.194 | | 436.219 | 436 | 425 | 448 |
| 26.619 | | 447.982 | 448 | 435 | 461 |
| 27.045 | | 459.618 | 459 | 445 | 474 |
| 27.470 | | 471.066 | 471 | 456 | 487 |
| 27.895 | | 482.380 | 482 | 466 | 499 |
| 28.320 | | 493.496 | 493 | 476 | 512 |
| 28.746 | | 504.497 | 504 | 486 | 524 |
| 29.171 | | 515.175 | 515 | 496 | 535 |
| 29.596 | | 525.597 | 526 | 506 | 546 |
| 30.021 | | 535.797 | 536 | 515 | 557 |
| 30.447 | | 545.775 | 546 | 525 | 568 |

**Figure 6.190:** Enrollment Prediction Table Example

The Enrollment Prediction Table has six columns: **Time, Actual, Predicted Avg. Sample Size, Predicted Median Sample Size, 95% Prediction Interval LL, 95% Prediction Interval UL.**

**Time** is the length of time after the study start at which the enrollment process is being evaluated. The maximum **Time** value will be related to the largest **Accrual Duration** found across all simulations.

**Actual** is the actual number of subjects enrolled at a given past **Time**. Actual values are only shown for all values of **Time** less than or equal to the **Current Time**. All subsequent **Actual** cells are left blank.

**Predicted Avg. Sample Size** is the predicted (arithmetic) average (across all simulations) number of subjects enrolled at a given future **Time**. **Predicted Avg. Sample Size** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**Predicted Median Sample Size** is the predicted median (i.e. 50% percentile over all simulations) number of subjects enrolled at a given future **Time**. **Predicted**

**Median Sample Size** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**95% Prediction Interval LL** is the lower limit of the 95% prediction interval (i.e. the 2.5% percentile over all simulations) for the number of subjects enrolled at a given future **Time**. **95% Prediction Interval LL** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**95% Prediction Interval UL** is the upper limit of the 95% prediction interval (i.e. 97.5% percentile over all simulations) for the number of subjects enrolled at a given future **Time**. **95% Prediction Interval LL** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**Events Prediction Table** The Events Prediction Table provides a tabular summary of the actual and predicted event pattern alongside a 95% prediction interval for the projected event pattern. This table will consist of 100 **Time** points of which half are assigned before and after the **Current Time**. An example of an nQuery Prediction workspace with the Events Prediction Table displayed is provided in Figure 6.191 in which the first projected future **Time** value of 24.958 has been highlighted.

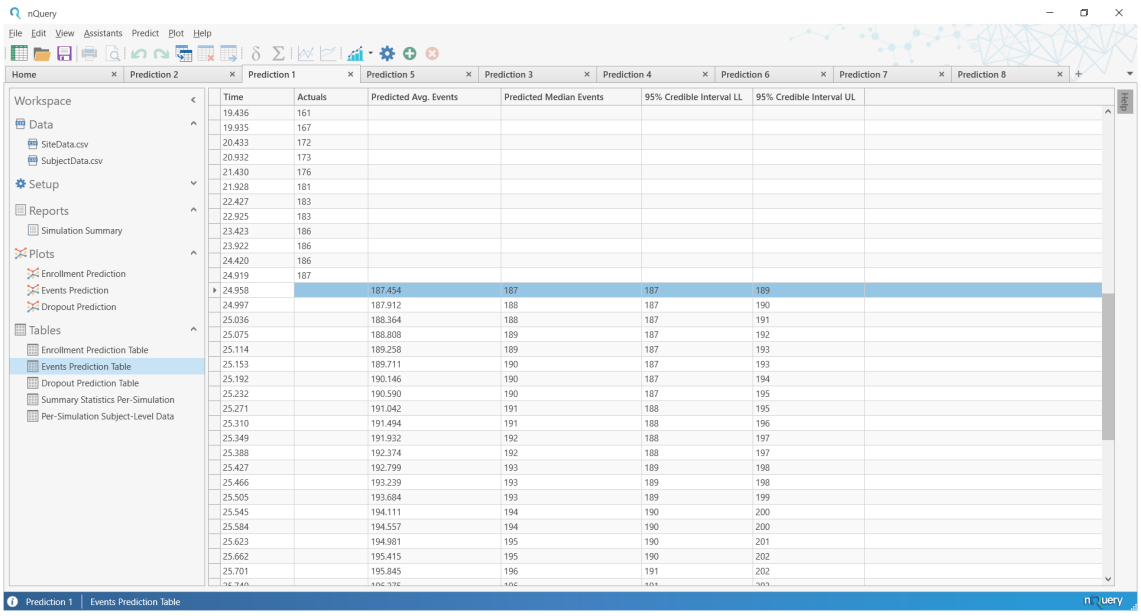



**Figure 6.191:** Event Prediction Table Example

The Events Prediction Table has six columns: **Time, Actual, Predicted Avg. Events, Predicted Median Events, 95% Prediction Interval LL, 95% Prediction Interval UL.**

**Time** is the length of time after the study start at which the event process is being evaluated. The maximum **Time** value will be related to the largest **Study Duration** found across all simulations.

**Actual** is the actual number of subjects who have had the event of interest at a given past **Time**. Actual values are only shown for all values of **Time** less than or equal to the **Current Time**. All subsequent **Actual** cells are left blank.

**Predicted Avg. Events** is the predicted (arithmetic) average (across all simulations) number of subjects who have had the event of interest at a given future **Time**. **Predicted Avg. Events** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**Predicted Median Events** is the predicted median (i.e. 50% percentile over all simulations) number of subjects who have had the event of interest at a given future **Time**. **Predicted Median Event** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**95% Prediction Interval LL** is the lower limit of the 95% prediction interval (i.e. the 2.5% percentile over all simulations) for the number of subjects who have had the event of interest at a given future **Time**. **95% Prediction Interval LL** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**95% Prediction Interval UL** is the upper limit of the 95% prediction interval (i.e. 97.5% percentile over all simulations) for the number of subjects who have had the event of interest at a given future **Time**. **95% Prediction Interval LL** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**Dropout Prediction Table** The Dropout Prediction Table provides a tabular summary of the actual and predicted dropout pattern alongside a 95% prediction interval for the projected accrual pattern. This table will consist of 100 **Time** points of which half are assigned before and after the **Current Time**. Note that if there are no dropouts in the prediction (i.e. no dropouts in **Subject-level Data** or **Summary Data** and **Dropout Hazard Rate(s)** equal to zero) then this table will not be provided.

An example of an nQuery Prediction workspace with the Dropout Prediction Table displayed is provided in Figure 6.192 in which the first projected future **Time** value of 24.958 has been highlighted.

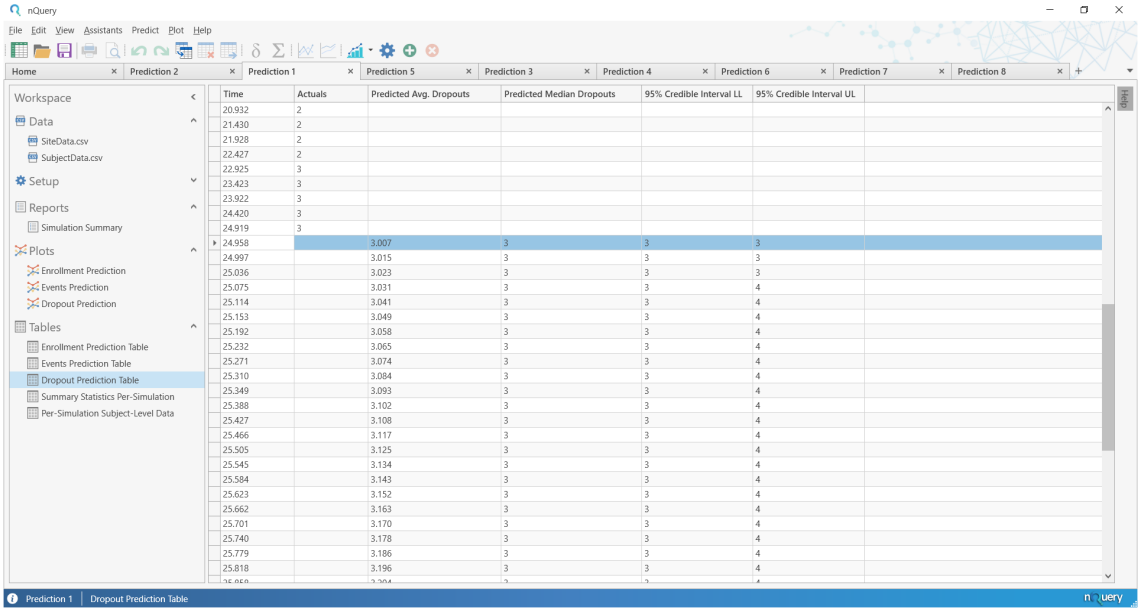



**Figure 6.192:** Dropout Prediction Table Example

The Dropout Prediction Table has six columns: **Time, Actual, Predicted Avg. Dropouts, Predicted Median Dropouts, 95% Prediction Interval LL, 95% Prediction Interval UL.**

**Time** is the length of time after the study start at which the dropout process is being evaluated. The maximum **Time** value will be related to the largest **Study Duration** found across all simulations.

**Actual** is the actual number of subjects who have dropped out at a given past **Time**. Actual values are only shown for all values of **Time** less than or equal to the **Current Time**. All subsequent **Actual** cells are left blank.

**Predicted Avg. Dropouts** is the predicted (arithmetic) average (across all simulations) number of subjects who have dropped out at a given future **Time**. **Predicted Avg. Dropouts** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**Predicted Median Dropouts** is the predicted median (i.e. 50% percentile over all simulations) number of subjects who have dropped out at a given future **Time**. **Predicted Median Dropouts** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**95% Prediction Interval LL** is the lower limit of the 95% prediction interval (i.e. the 2.5% percentile over all simulations) for the number of subjects who have dropped out at a given future **Time**. **95% Prediction Interval LL** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**95% Prediction Interval UL** is the upper limit of the 95% prediction interval (i.e. 97.5% percentile over all simulations) for the number of subjects who have

dropped out at a given future **Time**. **95% Prediction Interval LL** is only shown for **Time** values greater than the **Current Time**. All prior cells are left blank.

**Optional Tables** The next section will cover the optional tables which can be added or removed from a simulation in the **Output Options** section of the **Simulation Controls** step, with these options highlighted in Figure 6.193.

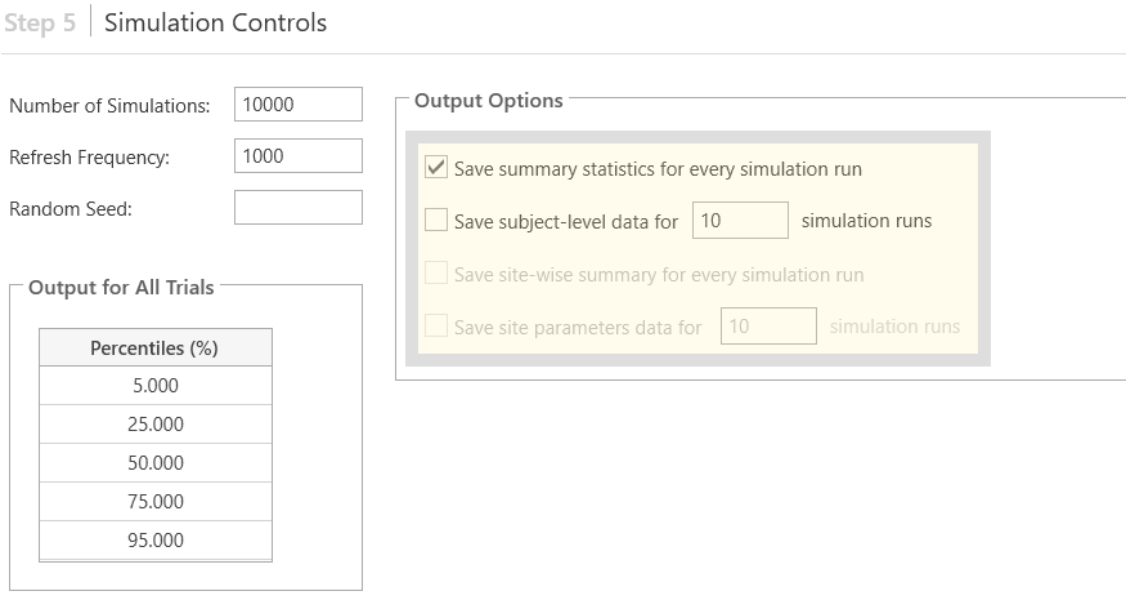



**Figure 6.193:** Optional Table fields in Simulation Controls

The **Output Options** for each table is as follows:

- Save summary statistics for every simulation run = Summary Statistics Per-Simulation table
- Save subject-level data for (X) simulation runs = Per-Simulation Subject-level Data table
- Save site-wise summary for every simulation run = Per-site Summary Predictions table
- Save site parameters data for (X) simulation runs = Per-Simulation Site-level Data table

The **(X)** value for the **Save subject-level data for (X) simulation** runs and **Save site parameters data for (X) simulation runs** allows the user the specify the number of individual simulations to output in their respective tables.

**Save site-wise summary for every simulation run** and **Save site parameters data for (X) simulation runs** will only be available if **Input = Subject + Site-level Data**. Otherwise, these options will be greyed out.

**Summary Statistics Per-Simulation Table** The Summary Statistics Per-Simulation Table provides the important inputs and summary statistic outputs for each individual simulation. The fields displayed will depend on the **Target** selected at the **Select the type of prediction** step of **Setup**. Definitions for each possible field are provided in the Help window on the right-hand side while the table is displayed in the nQuery Predict workspace.

Examples of the Summary Statistics Per-Simulation Table are provided in Figure 6.194 and examples can also be found during the Enrollment Prediction Demonstrations (subsection 6.2.6) and Events Prediction Demonstrations (subsection 6.2.7).

| Simulation ID | Current Time | Current Sample Size | Target Sample Size | Accrual Duration |
|---|---|---|---|---|
| 1 | 24.919 | 402 | 804 | 42.346 |
| 2 | 24.919 | 402 | 804 | 44.396 |
| 3 | 24.919 | 402 | 804 | 42.590 |
| 4 | 24.919 | 402 | 804 | 43.500 |
| 5 | 24.919 | 402 | 804 | 44.239 |
| 6 | 24.919 | 402 | 804 | 42.541 |

Enrollment Only

| l Censored | Control Sample Size | Control Events | Control Dropouts | Treatment Sample Size | Treatment Events | Treatment Dropouts | Accrual Duration | Accrual Duration Uncens... | Target Events Reached |
|---|---|---|---|---|---|---|---|---|---|
| | 200 | 100 | 8 | 200 | 100 | 8 | 19.593 | 19.593 | ✔ |
| | 200 | 104 | 11 | 200 | 96 | 15 | 21.294 | 19.856 | ✔ |
| | 196 | 101 | 8 | 190 | 99 | 6 | 19.559 | 19.374 | ✔ |
| | 200 | 103 | 9 | 200 | 97 | 5 | 18.270 | 18.270 | ✔ |
| | 190 | 99 | 8 | 185 | 101 | 8 | 19.250 | 19.155 | ✔ |
| | 198 | 94 | 7 | 197 | 106 | 8 | 20.484 | 19.495 | ✔ |

Unblinded Events

| Current Events | Current Dropouts | Current Available | Target Sample Size | Target Events | Study End Time | Total Sample Size | Total Events | Total Dropouts | Total Censored | Accrual Duration |
|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 0 | 150 | 400 | 100 | 17.889 | 349 | 100 | 0 | 249 | 17.879 |
| 50 | 0 | 150 | 400 | 100 | 17.281 | 349 | 100 | 0 | 249 | 17.213 |
| 50 | 0 | 150 | 400 | 100 | 17.694 | 344 | 100 | 0 | 244 | 17.657 |
| 50 | 0 | 150 | 400 | 100 | 18.190 | 376 | 100 | 0 | 276 | 18.025 |
| 50 | 0 | 150 | 400 | 100 | 18.709 | 368 | 100 | 0 | 268 | 18.632 |
| 50 | 0 | 150 | 400 | 100 | 17.572 | 366 | 100 | 0 | 266 | 17.524 |
| 50 | 0 | 150 | 400 | 100 | 18.667 | 364 | 100 | 0 | 264 | 18.585 |

Blinded Events

**Figure 6.194:** Summary Statistics Per-Simulation Table Examples

When **Target = Enrollment Only**, the Summary Statistics Per-Simulation Table will contain the following fields: **Simulation ID, Current Time**, **Current Sample Size, Target Sample Size**, **Accrual Duration**.

When **Target = Unblinded Events**, the Summary Statistics Per-Simulation Table will contain the following fields: **Simulation ID, Current Time, Current Sample Size, Current Events, Current Dropouts, Target Sample Size, Target Events, Study End Time, Total Sample Size, Total Events, Total Dropouts, Total Censored, Control Sample Size, Control Events, Control Dropouts, Treatment Sample Size, Treatment Events, Treatment Dropouts, Accrual Duration, Accrual Duration Uncensored, Target Events Reached**

When **Target = Blinded Events**, the Summary Statistics Per-Simulation Table will contain the following fields: **Simulation ID, Current Time, Current Sample Size, Current Events, Current Dropouts, Target Sample Size, Target Events, Study End Time, Total Sample Size, Total Events, Total Dro-**

**pouts, Total Censored, Accrual Duration, Accrual Duration Uncensored, Target Events Reached**

For the common **Sample Size, Event** and **Dropout** columns, the "Current" columns describe the number of subjects with a given status at the **Current Time** (based on the **Subject-level Data** or **Summary Data Fixed Parameters** inputs), the "Target" columns describe the targets specified for the simulation during **Setup**, the "Total" columns describe the total number of subjects with that status at the end of that simulation and the "Control" / "Treatment" columns give the total number subjects with that status at the end of that simulation for **Target = Unblinded Events**.

Other fields definitions are as follows:

**Simulation ID** is the simulation number/identifier, **Total Censored** is subjects who were administratively censored at study end due to not having an event or dropout at that point. **Study End Time** is the length of time after study start (Time = 0) at which the trial ended as **Target Number of Events** was reached, **Accrual Duration** is the length of time after study start at which the accrual process ended as the **Target Sample Size** was reached unless ended early for events prediction. **Accrual Duration Uncensored** is the accrual duration if the enrollment process had not been cut short due to the **Target Number of Events** being reached before the **Target Sample Size** during events prediction. **Target Events Reached** is an indicator for whether the **Target Number of Events** was reached in that simulation.

If **Target Events Reached** indicator is negative (X) then the **Study End Time** becomes the length of time at which there were no longer any remaining subjects left to have the event (e.g. due to high dropout rates, due to **Fixed Followup**). **Accrual Duration** and **Accrual Duration Uncensored** will be equal unless the enrollment process is ended early due to the **Target Number of Events** being reached before the **Target Sample Size** was reached.

**Per-Simulation Subject-level Data Table** The Per-Simulation Subject-level Data Table provides the simulated outcomes for each subject from a pre-specified number of simulations. The number of simulations is specified by editing the (X) value in the **Save subject-level data for (X) simulation runs** field of the **Output Options** section of the **Simulation Controls** step of **Setup.**

The fields and number of subjects (rows) displayed will depend on the **Target** and **Input** selected at the **Select the type of prediction** step of **Setup**. Definitions for each possible field are provided in the Help window on the right-hand side while the table is displayed in the nQuery Predict workspace.

Examples of the Per-Simulation Subject-level Data Table are provided in Figure 6.195 and examples can also be found during the Enrollment Prediction Demonstrations (subsection 6.2.6) and Events Prediction Demonstrations (subsection 6.2.7).

| Simulation ID | Subject ID | Arrival Time | Site ID |
|---|---|---|---|
| 1 | 399 | 22.717 | 366 |
| 1 | 400 | 22.866 | 101 |
| 1 | 401 | 22.942 | 204 |
| 1 | 402 | 24.919 | 322 |
| 1 | 403 | 24.980 | 127 |

Enrollment Only
Subject + Site Data

| Simulation ID | Subject ID | Arrival Time | Event Time | Dropout Time | Final Status | Site ID | Group |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.186 | 5.479 | | 1 | 103 | 0 |
| 1 | 2 | 0.226 | 3.728 | | 1 | 108 | 1 |
| 1 | 3 | 0.341 | 6.546 | | 1 | 121 | 0 |
| 1 | 4 | 0.726 | 25.275 | 662.214 | 1 | 103 | 0 |
| 1 | 5 | 1.495 | 9.656 | | 1 | 103 | 1 |
| 1 | 6 | 2.209 | 5.447 | | 1 | 103 | 0 |

Unblinded Events, Subject + Site Data

| Simulation ID | Subject ID | Arrival Time | Event Time | Dropout Time | Final Status |
|---|---|---|---|---|---|
| 1 | 56 | 10.000 | 0.027 | 105.428 | 1 |
| 1 | 57 | 10.000 | 18.747 | 364.962 | 0 |
| 1 | 58 | 10.000 | 0.655 | 79.015 | 0 |
| 1 | 59 | 10.000 | 2.096 | 309.335 | 0 |
| 1 | 60 | 10.000 | 6.171 | 327.186 | 0 |

Blinded Events, Subject Data Only

**Figure 6.195:** Per-Simulation Subject-level Data Table Examples

When **Target = Enrollment Only** and **Input = Subject Data Only** or **Summary Data**, the Per-Simulation Subject-level Data Table will contain information for each simulation with the following fields: **Simulation ID, Subject ID, Arrival Time**

When **Target = Enrollment Only** and **Input = Subject + Site Data**, the Per-Simulation Subject-level Data Table will contain information for each simulation with the following fields: **Simulation ID, Subject ID, Arrival Time, Site ID**

When **Target = Unblinded Events** and **Input = Subject Data Only** or **Summary Data**, the Per-Simulation Subject-level Data Table will contain information for each simulation with the following fields: **Simulation ID, Subject ID, Arrival Time, Event Time, Dropout Time, Final Status, Group**

When **Target = Unblinded Events** and **Input = Subject + Site Data**, the Per-Simulation Subject-level Data Table will contain information for each simulation with the following fields: **Simulation ID, Subject ID, Arrival Time, Event Time, Dropout Time, Final Status, Site ID, Group**

When **Target = Blinded Events** and **Input = Subject Data Only** or **Summary Data**, the Per-Simulation Subject-level Data Table will contain information for each simulation with the following fields: **Simulation ID, Subject ID, Arrival Time, Event Time, Dropout Time, Final Status, Group**

When **Target = Blinded Events** and **Input = Subject + Site Data**, the Per-Simulation Subject-level Data Table will contain information for each simulation with the following fields: **Simulation ID, Subject ID, Arrival Time, Event Time, Dropout Time, Final Status, Site ID**

**Simulation ID** is the simulation number/identifier, **Subject ID** is the subject identifier based on the order of **Arrival Time** (which the length of time after study

start (Time = 0) at which a subject entered the study), **Event Time** is the actual or projected length of time after **Arrival Time** after which a subject had the event of interest, **Dropout Time** is the actual or projected length of time after **Arrival Time** after which a subject had dropped out, **Final Status** is the final status (**Event, Dropout** or **Censored**) of a subject in the current simulation, **Site ID** is the site number/identifier a subject was recruited from and **Group** is the treatment group (**Control** or **Treatment**) a subject was assigned to.

Note that for **Input = Summary Data**, only the simulated subjects are provided in this table. When subject-level data is used, this will contain the subjects from the subject-level data and the fully simulated subjects.

For **Target = Enrollment Only**, simulated **Arrival Time** values are provided for all subjects alongside the **Arrival Time** column values from the subject-level data if used.

For **Target = Unblinded Events** or **Blinded Events**, the simulated **Arrival Time** values are generated for all fully simulated subjects (i.e. did not exist in subject-level data) and are provided alongside the subject-level data **Arrival Time** column values. The **Event Time** and **Dropout Time** are simulated for all fully simulated subjects AND those subjects who had **Censored Status** in the **Status Indicator** column in the subject-level data.

The **Final Status** is taken from the subject-level data for those subjects who had a **Event Status** or **Dropout Status** in the **Status Indicator** column. The **Event Time** will be empty if a subject was **Dropout Status** in the subject-level data and the **Dropout Time** will be empty if a subject was **Event Status** in the subject-level data.

For fully simulated subjects or **Censored Status** subjects from the subject-level data (i.e. subjects with a simulated **Event Time** and **Dropout Time**), the **Final Status** is based on the relationship between the simulated **Event Time** and **Dropout Time** and the **Study Duration/Study End Time** in that **Simulation ID** (**Study Duration/Study End Time** for given **Simulation ID** can be found in the **Summary Statistics Per-Simulation Table**). If a subject has an **Event Time** less than the **Dropout Time** and **Study Duration**, the **Final Status** will be **Event Status**. If a subject has a **Dropout Time** less than the **Event Time** and **Study Duration**, the **Final Status** will be **Dropout Status**. If a subject has both **Event Time** and **Dropout Time** greater than the **Study Duration**, the **Final Status** will be **Censored Status**. The one exception is if **For a Fixed Period** is selected from the **Subjects are Followed:** field in the **Accrual Options** step of **Setup** where in the rules above replace **Study Duration** with the specified **Followup Time**.

**Site ID** and **Group** will be generated for simulated **Arrival Time** subjects and taken from the subject-level data otherwise.

**Per-site Summary Predictions table** The Per-site Summary Predictions table provides the summary statistic outputs for each site from each individual simulation. This table is only available if **Input = Subject + Site Data**. Definitions for each field are provided in the Help window on the right-hand side while the table is displayed in the nQuery Predict workspace.

An example of the Per-site Summary Predictions Table is provided in Figure 6.196 and examples can also be found during the Enrollment Prediction Demonstrations (subsection 6.2.6) and Events Prediction Demonstrations (subsection 6.2.7).

| Site ID | Avg. Accrual | Avg. Max Time | Avg. Start Time | Avg. Duration | Avg. Rate | Times Open |
|---|---|---|---|---|---|---|
| 441 | 5.353 | 35.598 | 6.414 | 33.194 | 0.168 | 10000 |
| 442 | 11.903 | 40.817 | 5.428 | 38.609 | 0.309 | 10000 |
| 443 | 2.225 | 37.901 | 25.251 | 18.781 | 0.119 | 10000 |
| 444 | 4.455 | 39.426 | 9.013 | 33.264 | 0.138 | 10000 |
| 445 | 7.908 | 40.211 | 13.257 | 30.770 | 0.257 | 10000 |
| 446 | 7.999 | 27.623 | 5.829 | 21.828 | 0.371 | 10000 |
| 447 | 7.752 | 37.771 | 14.737 | 23.407 | 0.344 | 10000 |
| 501 | 4.823 | 40.299 | 6.842 | 37.189 | 0.130 | 10000 |
| 502 | 6.684 | 38.721 | 9.414 | 34.585 | 0.194 | 10000 |
| 503 | 2.343 | 31.142 | 10.395 | 33.643 | 0.070 | 10000 |
| 520 | 7.000 | 19.506 | 3.579 | 15.927 | 0.440 | 10000 |
| 521 | 5.940 | 32.185 | 8.783 | 24.012 | 0.258 | 10000 |
| 522 | 2.206 | 28.597 | 9.145 | 34.863 | 0.063 | 10000 |
| 523 | 2.080 | 26.995 | 7.184 | 36.853 | 0.056 | 10000 |
| 540 | 4.068 | 34.415 | 6.579 | 37.458 | 0.109 | 10000 |
| 541 | 2.049 | 26.518 | 6.579 | 37.368 | 0.055 | 10000 |
| 542 | 1.957 | 37.426 | 15.822 | 28.215 | 0.069 | 10000 |
| 543 | 4.786 | 42.280 | 34.577 | 9.451 | 0.512 | 7739 |
| 601 | 4.201 | 34.981 | 7.664 | 36.074 | 0.118 | 10000 |
| 602 | 3.928 | 34.546 | 4.868 | 39.169 | 0.100 | 10000 |
| 603 | 8.654 | 39.969 | 8.421 | 35.616 | 0.243 | 10000 |
| 604 | 5.739 | 37.025 | 4.178 | 39.631 | 0.146 | 10000 |
| 605 | 1.875 | 23.095 | 3.263 | 40.773 | 0.046 | 10000 |
| 606 | 10.191 | 40.445 | 6.579 | 37.459 | 0.272 | 10000 |
| 607 | 9.858 | 35.358 | 7.859 | 28.043 | 0.363 | 10000 |
| 608 | 5.289 | 40.491 | 10.066 | 33.672 | 0.159 | 10000 |
| 609 | 12.000 | 30.302 | 19.375 | 10.927 | 1.120 | 10000 |
| 610 | 6.514 | 38.392 | 8.684 | 35.353 | 0.184 | 10000 |
| 611 | 6.264 | 38.245 | 7.664 | 35.880 | 0.176 | 10000 |
| 620 | 2.507 | 38.188 | 9.243 | 34.793 | 0.072 | 10000 |
| 621 | 2.061 | 26.832 | 7.270 | 36.768 | 0.056 | 10000 |
| 622 | 1.992 | 27.124 | 5.691 | 38.347 | 0.052 | 10000 |
| 623 | 1.985 | 24.839 | 5.559 | 38.478 | 0.052 | 10000 |

**Figure 6.196:** Per-site Summary Predictions Table Example

The Per-site Summary Predictions Table has the following fields:

- Site ID: Site Identifier/Number
- Avg. Accrual: Average Number of accruals from this site across all simulations
- Avg. Max Time: Average time of the final enrollment from this site across all simulations

- Avg. Start Time: Average time a site opened at across all simulations
- Avg. Duration: Average length of time a site was open for across all simulations
- Avg. Rate: Average accrual rate (subjects enrolled per unit time) in a site across all simulations
- Times Open: The number of simulations in which a site opened

Note that the **Avg. Start Time** is the same for all simulations for **Opened Sites** in the **Site-level Data** but is randomly simulated for **Unopened Sites**. **Times Opened** will be equal to the number of simulations from the **Simulation Controls** step for all **Opened Sites** but will depend on if the simulated opening time was less than or greater than the **Study Duration** for **Unopened Sites**.

**Per-Simulation Site-level Data Table** The Per-Simulation Site-level Data Table provides the simulated summary outcomes for each site from a pre-specified number of individual simulations. This table is only available if **Input = Subject + Site Data**. The number of simulations is specified by editing the (X) value in the **Save site parameters data for (X) simulation runs** field of the **Output Options** section of the **Simulation Controls** step of **Setup.**

An example of the Per-Simulation Site-level Data Table is provided in Figure 6.197 and examples can also be found during the Enrollment Prediction Demonstrations (subsection 6.2.6) and Events Prediction Demonstrations (subsection 6.2.7).

| Simulation ID | Site ID | Site Activation Time | True Enrollment Rate | Accruals | Observed Enrollment Rate | Final Enrollment Time | Enrollment Duration | Site Open |
|---|---|---|---|---|---|---|---|---|
| 6 | 522 | 9.145 | 0.063 | 4 | 0.118 | 39.160 | 33.976 | ✔ |
| 6 | 523 | 7.184 | 0.056 | 1 | 0.028 | 8.902 | 35.936 | ✔ |
| 6 | 540 | 6.579 | 0.109 | 2 | 0.055 | 13.202 | 36.541 | ✔ |
| 6 | 541 | 6.579 | 0.055 | 2 | 0.055 | 32.148 | 36.541 | ✔ |
| 6 | 542 | 15.822 | 0.102 | 2 | 0.073 | 40.753 | 27.298 | ✔ |
| 6 | 543 | 0.000 | 0.500 | 0 | 0.000 | 0.000 | 0.000 | ✘ |
| 6 | 601 | 7.664 | 0.116 | 3 | 0.085 | 42.027 | 35.456 | ✔ |
| 6 | 602 | 4.868 | 0.100 | 3 | 0.078 | 38.043 | 38.252 | ✔ |
| 6 | 603 | 8.421 | 0.242 | 8 | 0.231 | 41.917 | 34.699 | ✔ |
| 6 | 604 | 4.178 | 0.145 | 7 | 0.180 | 38.998 | 38.943 | ✔ |
| 6 | 605 | 3.263 | 0.046 | 3 | 0.075 | 32.311 | 39.857 | ✔ |
| 6 | 606 | 6.579 | 0.273 | 8 | 0.219 | 35.518 | 36.541 | ✔ |
| 6 | 607 | 7.859 | 0.352 | 10 | 0.318 | 39.347 | 31.488 | ✔ |
| 6 | 608 | 10.066 | 0.281 | 3 | 0.091 | 42.778 | 33.055 | ✔ |
| 6 | 609 | 19.375 | 2.224 | 12 | 1.101 | 30.274 | 10.899 | ✔ |
| 6 | 610 | 8.684 | 0.185 | 7 | 0.203 | 41.625 | 34.436 | ✔ |
| 6 | 611 | 7.664 | 0.174 | 5 | 0.141 | 42.215 | 35.456 | ✔ |
| 6 | 620 | 9.243 | 0.131 | 1 | 0.030 | 30.714 | 33.877 | ✔ |
| 6 | 621 | 7.270 | 0.057 | 1 | 0.028 | 9.123 | 35.851 | ✔ |
| 6 | 622 | 5.691 | 0.052 | 3 | 0.080 | 41.439 | 37.430 | ✔ |
| 6 | 623 | 5.559 | 0.052 | 1 | 0.027 | 6.304 | 37.561 | ✔ |
| 7 | 101 | 3.184 | 0.506 | 15 | 0.510 | 32.567 | 29.383 | ✔ |
| 7 | 102 | 1.283 | 0.127 | 5 | 0.116 | 43.568 | 43.047 | ✔ |
| 7 | 103 | 0.000 | 1.605 | 45 | 1.710 | 26.308 | 26.308 | ✔ |
| 7 | 104 | 2.533 | 0.938 | 21 | 1.175 | 20.412 | 17.879 | ✔ |
| 7 | 105 | 3.257 | 0.185 | 7 | 0.170 | 35.205 | 41.073 | ✔ |
| 7 | 106 | 1.645 | 0.773 | 27 | 0.937 | 30.467 | 28.822 | ✔ |
| 7 | 108 | 0.213 | 0.526 | 13 | 0.609 | 21.542 | 21.329 | ✔ |
| 7 | 112 | 5.822 | 0.105 | 2 | 0.052 | 7.697 | 38.507 | ✔ |
| 7 | 113 | 6.941 | 0.222 | 5 | 0.134 | 28.421 | 37.389 | ✔ |
| 7 | 114 | 2.993 | 0.182 | 7 | 0.266 | 29.273 | 26.280 | ✔ |

**Figure 6.197:** Per-Simulation Site-level Data Table Example

The Per-Simulation Site-level Data has the following fields:

- Simulation ID: Simulation Identifier/Number
- Site ID: Site Identifier/Number
- Site Activation Time: Time at which site opened for enrollment in the current simulation
- True Enrollment Rate: User specified enrollment rate for site
- Accruals: Number of subjects enrolled in that site for this simulation
- Observed Enrollment Rate: Enrollment rate achieved in this site for this simulation based on number of Accruals
- Enrollment Duration: Length of time site was open for enrollment in that site for this simulation
- Site Open: Did site open during this simulation

**Site Activation Time** is taken from the **Subject-level Data** for **Opened Sites** (simulated for **Unopened Sites**). **True Enrollment Rate** is taken from the **Accrual Rates/Site** column found in the **Accrual Infos/Site** tab at the **Accrual Options** step of **Setup,** with **Observed Enrollment Rate** being the actual rate achieved in a site in a given simulation. The **Enrollment Duration** will equal the **Accrual Duration** minus the **Site Activation Time** unless the **Enrollment Cap** is reached in a site. **Site Open** will always be a tick for **Opened Sites**.

## 6.2.9 Other Features

### 6.2.9.1 Saving and Exporting Workspace

nQuery provides the file format of **.nqp** to save and export an nQuery Predict Workspace. To save an nQuery Predict Workspace, select **Save** from the File Menu or select the **Save** icon from the toolbar or use the **CTRL+S** keyboard shortcut. This will open the **Save Workspace** window which is shown in Figure 6.198.

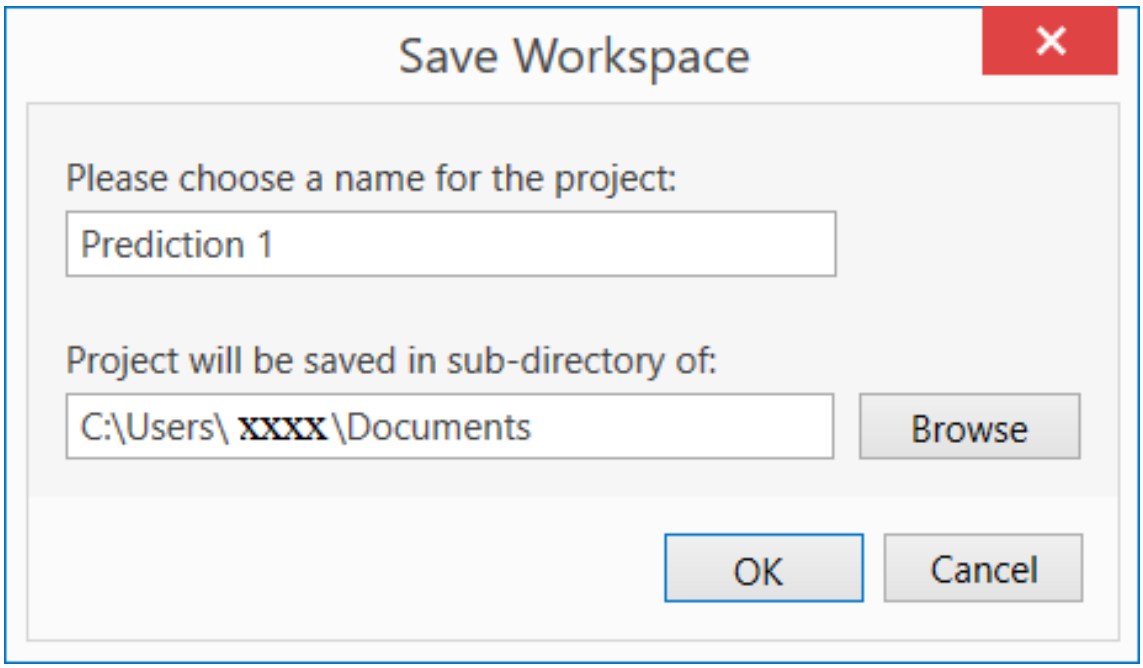



**Figure 6.198:** Save Workspace Window

The **Save Workspace** window allows the user to specify the name of the .nqp file in the **Please choose a name for the project:** field and the location to save the .nqp file in the **Project will be saved in sub-directory of:** field. The user can also edit the **Project will be saved in sub-directory of:** using the Browse button, which will open the **Browse for Folder** window. To save the project, select the OK button.

When a project is saved, a folder with the **Please choose a name for the project:** name will be present in the folder specified in the **Project will be saved in sub-directory of:** field. This project folder will contain the .nqp file alongside any datasets in the current workspace and all tables under the **Tables** header generated in the current prediction (all as .csv files).

After an nQuery Workspace is saved if a change is made to the workspace the user can save over the prior save using the Save options (see above) or save a separate version using Save As from the File menu.

To open a saved .nqp file, select the file in Window Explorer or find the file from the Open Menu which can be selected from the File Menu, the toolbar icon or **CTRL+O** keyboard shortcut. Opening a .nqp file works as for other nQuery file types (see subsection 1.7.5).

#### 6.2.9.2 Editing and Saving Plots

nQuery Predict provides numerous tools to edit the plots and save those plots as separate image files. To edit, save or print a plot, open the context menu by right-clicking directly on plot while it is displayed in the workspace. An example of the context menu being open is shown in Figure 6.199.

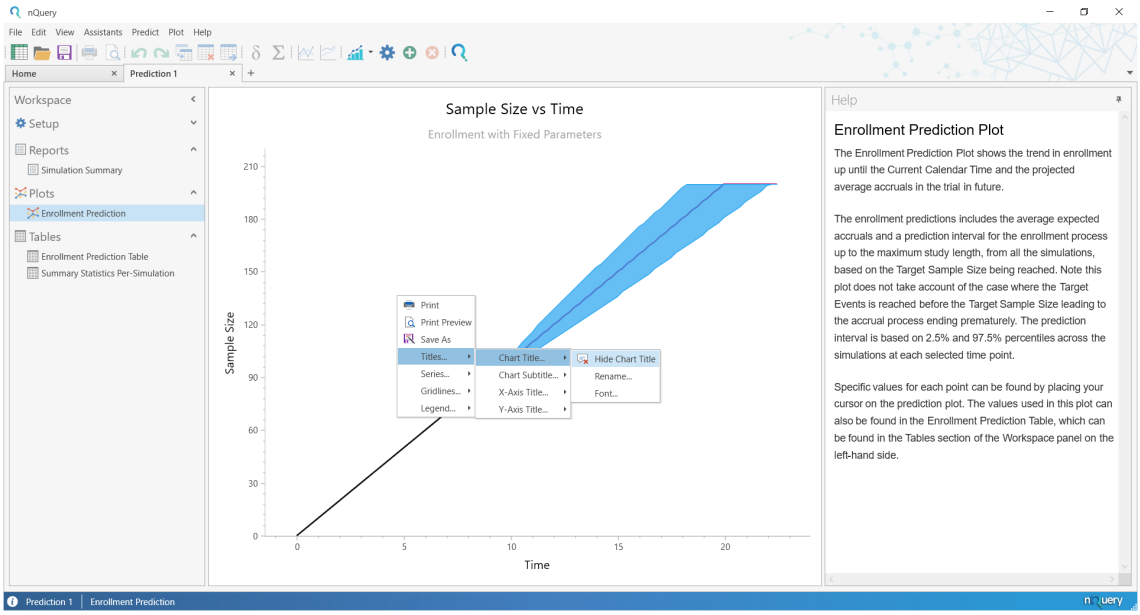



**Figure 6.199:** Plot Context Menu

The context menu contains the following options:

- Print: Open Print screen for current plot
- Print Preview: Create Print Preview of current plot
- Save as: Save current image as .pdf or .jpeg image file
- Titles: Hide, Rename or change font of all plot titles
- Series: Change color, add marker (including size and style) or rename any series line
- Gridlines: Add or remove X-axis or Y-axis gridlines (major only)
- Legend: Add/Remove legend and change location and outline

The majority of options for editing the plot under **Titles, Series, Gridlines** and **Legend** are selected directly from the context menu with ‣ indicating that a sub-menu is available if this option is selected. **Rename** and **Font** options will open additional pop-out windows to change these fields.

**Save as** will open the nQuery Save As screen where file type can changed from the **Save as Type** field at the bottom of the window.

**Print** and **Print Preview** will open up new windows for printing.

#### 6.2.9.3 Error Checking

nQuery Predict provides error checking to ensure that all inputs during the **Setup** stage are appropriate. There are two primary error checks: data field checks, input field checks.

**Data Field Checks** relate to which columns can be selected from the subject or site-level data for a particular field at a **Data Field Selection** selection step when **Input = Subject Data Only** or **Subject + Site Data**. The **Subject-level Data** (Step 2 where available) and **Site-level Data** (Step 3 where available) steps are both **Data Field Selection** steps where the user chooses the required dataset and the columns which correspond to the fields required for the current prediction.

**Data Field Checks** ensure that only columns which have the appropriate characteristics are available for selection for a given field. For example, the **Arrival Time** field must all be numeric values greater than zero with at least 4 distinct values or the **Status Indicator** field must have no more than 3 unique alphanumeric values. For example, in Figure 6.200 the only two columns in **SubjectData.csv** which consist of numeric values only are **Arrival** (arrival times) and **Followup** (time on study).

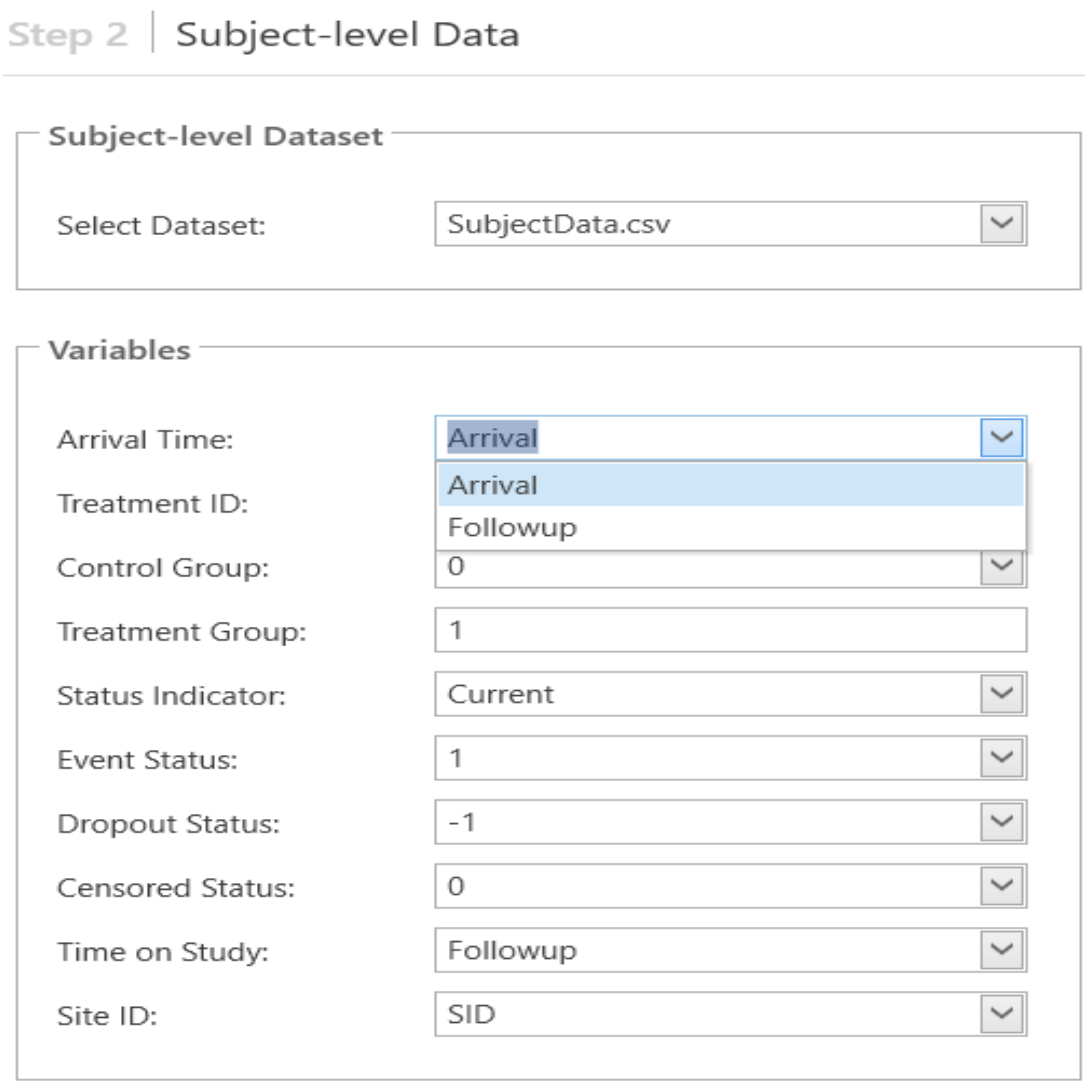



**Figure 6.200:** Data Field Check Example

If a required input is not being shown, the user may need to check if the criteria for that field (see subsection 6.2.4, subsection 6.2.6, subsection 6.2.7 for field requirements) have been fulfilled by, for example, checking and editing the data from the **Data** header in the Workspace Navigation Bar on the left side of the workspace window.

Note that an **Out of Range** error (see subsection 2.2.1) will be shown if the same column is selected in multiple fields at a **Data Field Selection** step.

**Input Field Checks** show an **Out of Range** error in any non-data field selection when a value that is not allowed is entered into a field. For example, **Target Number of Events** must be greater than the **Current Number of Events** for

event prediction. These errors work the same as for standard nQuery errors (see subsection 2.2.1). An example is provided in Figure 6.201.

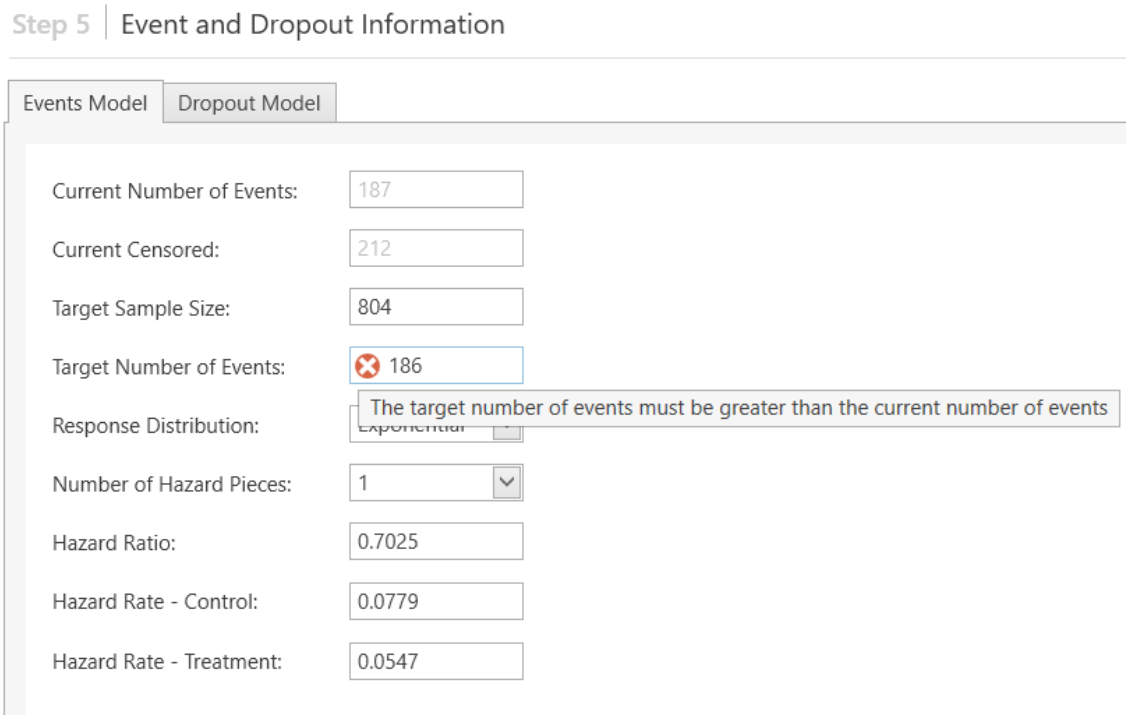



**Figure 6.201:** Input Field Check Example

#### 6.2.9.4 Simulation Controls Overview

The **Simulation Controls** step is the common final step across all simulations during **Setup**. This step controls the overall simulation parameters, the percentiles of interest and the optional tables required by the user.

The **Simulation Controls** step consists of three main sections:

1. Simulation Controls
2. Output for all Trials (Percentile Controls)
3. Output Options

These sections are highlighted in Figure 6.202.

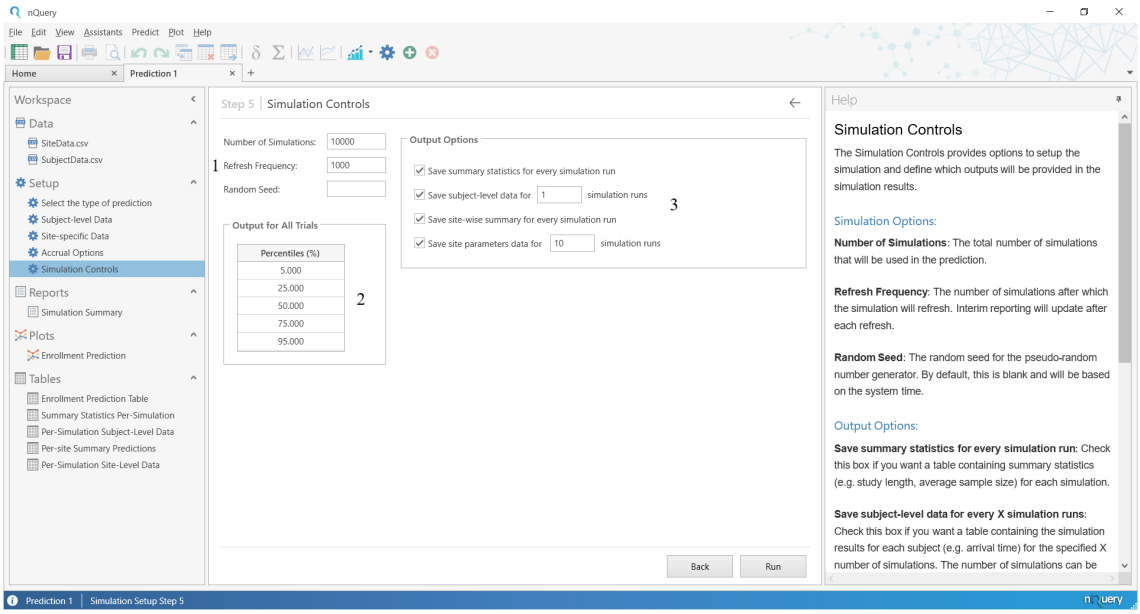



**Figure 6.202:** Simulation Controls

The **Simulation Controls** (1) section consists of the following fields (range in brackets): Number of Simulations (Integer > 100), Refresh Frequency (Integer > Number of Simulations) , Random Seed (Integer).

**Number of Simulations** is the required total number of simulations which will be run in the current simulation. **Refresh Frequency** is the number of simulations after which the summary statistics will be updated in the **Simulation in Progress** screen when Run is selected. **Random Seed** is used to specify the random seed for the pseudo-random number generator. If **Random Seed** is unspecified a random seed is selected for the user and will be shown in the **Simulation Summary** section of the **Simulation Summary Report**.

The **Output for all Trials / Percentile Table** (2) section consist of a 5 cell vertical table which specifies the percentiles of interest to display in the **Percentile Summary** section of the **Simulation Summary Report** results. By default this is set to 5th, 25th, 50th, 75th, 95th percentiles of the results across all simulations. This corresponds to the 90% Interval (5 - 95), Interquartile Range (25 - 75) and median (50) for the primary outcomes of interest. These values can be edited with any value between 0 and 100.

**Output Options** specifies the optional tables required in the current simulation. Details on these tables is provided in subsubsection 6.2.8.3. The **(X)** value for the **Save subject-level data for (X) simulation** runs and **Save site parameters data for (X) simulation runs** allows the user the specify the number of individual simulations to output in their respective tables.

#### 6.2.9.5 nQuery Predict Options

There are additional options available to edit how nQuery Predict works from the Options menu. The Options menu can be opened by selecting Options from the File Menu. nQuery Predict specific options can be found under the Predict header and are shown in Figure 6.203.

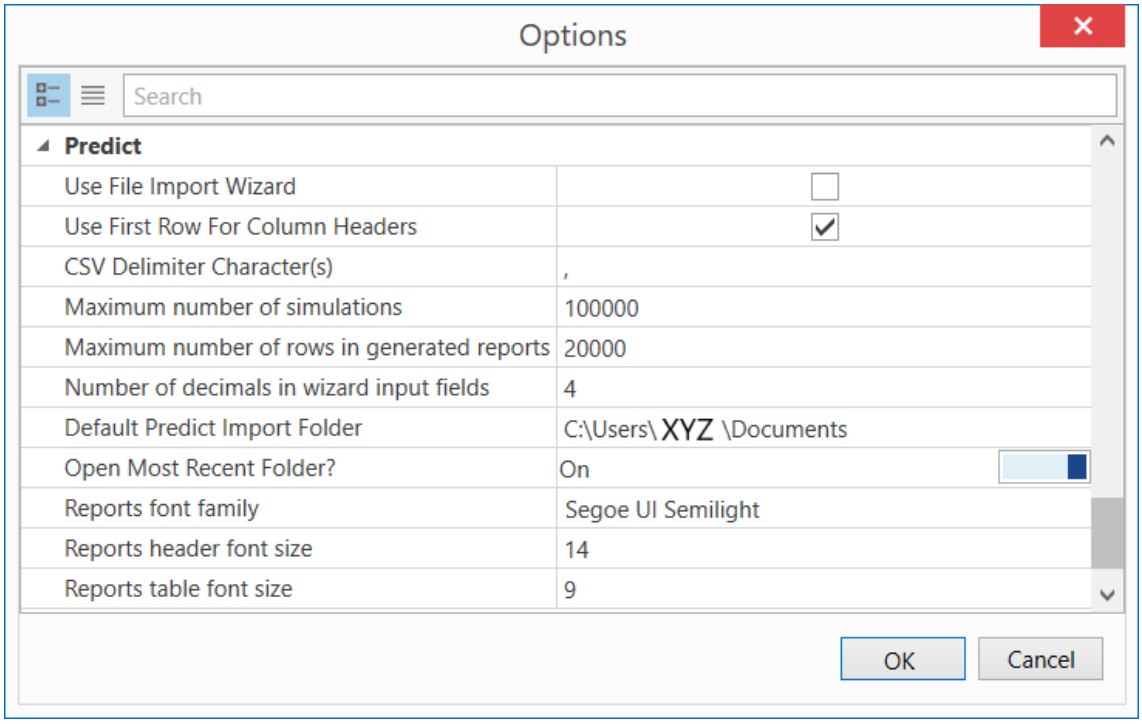



**Figure 6.203:** nQuery Predict Options

The options for nQuery Predict are as follows:

1. Use File Import Wizard: Toggle if Advanced File Import Wizard is needed
2. Use First Row for Column Headers: Toggle if first row of imported datasets should be assigned as column headers
3. CSV Delimiter Character(s): Enter delimiter between entries in imported CSV files
4. Maximum Number of Simulations: Maximum number of simulations allowed in nQuery Predict workspace
5. Maximum Number of Rows in Generated Reports: Maximum number of rows displayed in generated spreadsheet reports
6. Number of decimals in wizard input fields: Number of decimal places displayed in nQuery Predict numeric input fields and Report output fields

7. Default Predict Import Folder: Default folder displayed when importing datasets
8. Open Most Recent Folder?: Open most recently used folder in nQuery Predict workspace when importing datasets
9. Reports Font Family: Font used in nQuery Predict Reports
10. Reports Header Font Size: Font size of headers in nQuery Predict Reports
11. Reports Table Font Size: Font size of values in tables contained in nQuery Predict Reports

# 7 nQuery Qualification Tools

nQuery provides automated qualification scripts to quickly allow a user to verify that their nQuery application is installed and operating to the manufacturers specifications. nQuery specifically has two separate tools for installation qualification (IQ) and operation qualification (OQ) to assist you in verifying nQuery in regulatory industries.

Note that operational and performance qualification (PQ) are considered interchangeable based on our correspondence with the relevant stakeholders. However, you may wish to perform within-application testing for PQ purposes.

## 7.1 Installation Qualification (IQ) Tool

The nQuery IQ tool assists you in demonstrating that nQuery has been installed and maintained to the manufacturer's specifications. nQuery IQ verifies the integrity of each file in the nQuery 8 system and provides the user a set of reports detailing the results.

The nQuery Installation Qualification (IQ) Tool validates an nQuery installation by verifying that each installed file is correct with a report generated detailing all file results. This determination is made using the SHA-1 algorithm to create a hash value for each file. This thus checks that each file is present and that its integrity has been maintained.

Important: IQ validation represents the expected state of the system as it is at the time an installation is completed. If changes are made to the original files subsequently then these may show as failed when IQ is run on the system at a later date.

### 7.1.1 Running nQuery IQ

To run nQuery IQ, select "Installation Qualification" from the Help file menu. This will automatically run the nQuery.Tools.InstallationQualification.exe application (found in your installation folder, default of C:/Program Files (x86)/Statistical Solutions Ltd/nQuery) and output a HTML report which will display the results automatically in the machine's default HTML viewer (e.g. default internet browser).

### 7.1.2 nQuery IQ Results

The IQ report file will automatically be saved on the machine when the IQ tool runs. It will be saved to the C:\Users\<User Account>\AppData\Local\nQuery\Reports folder where <User Account> is the named User which currently logged into on the machine. An example of an IQ report is shown in Figure 7.1.

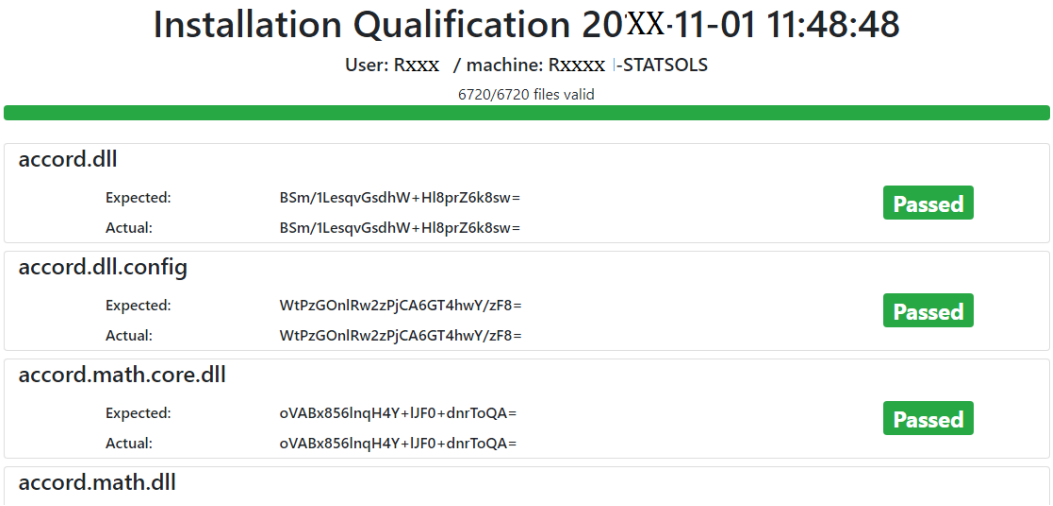



**Figure 7.1:** Installation Qualification Report Example

## 7.2 Operational Qualification (OQ) Tool

The nQuery OQ tool assists you in demonstrating that nQuery 8 is operational. nQuery 8 uses the nQuery solver functions over each design table for a set of validated design parameter inputs and will execute, process, and report the solver results.

Note that the files used for OQ are available in the TestData folder in your installation folder (default is C:/Program Files (x86)/Statistical Solutions Ltd/nQuery)

*Important: All nQuery results have been validated exactly against the results from nQuery + nTerim 4.0 and the original validation documentation and references. All results should be identical to nQuery + nTerim 4.0 except for simulation tables. There are slight difference with simulation tables due to an upgraded random number generator being used in nQuery. For those tables a 5% precision rule was used for validation. Results were also compared to EAST 6.5, PASS 2021 and SAS 9.4 Proc Power where the same or similar methods were implemented in those software for a design table.*

### 7.2.1 Running nQuery OQ

To run nQuery OQ, select "Operational Qualification" from the Help file menu. This will open the OQ Validation Tool menu which is shown in Figure 7.2.

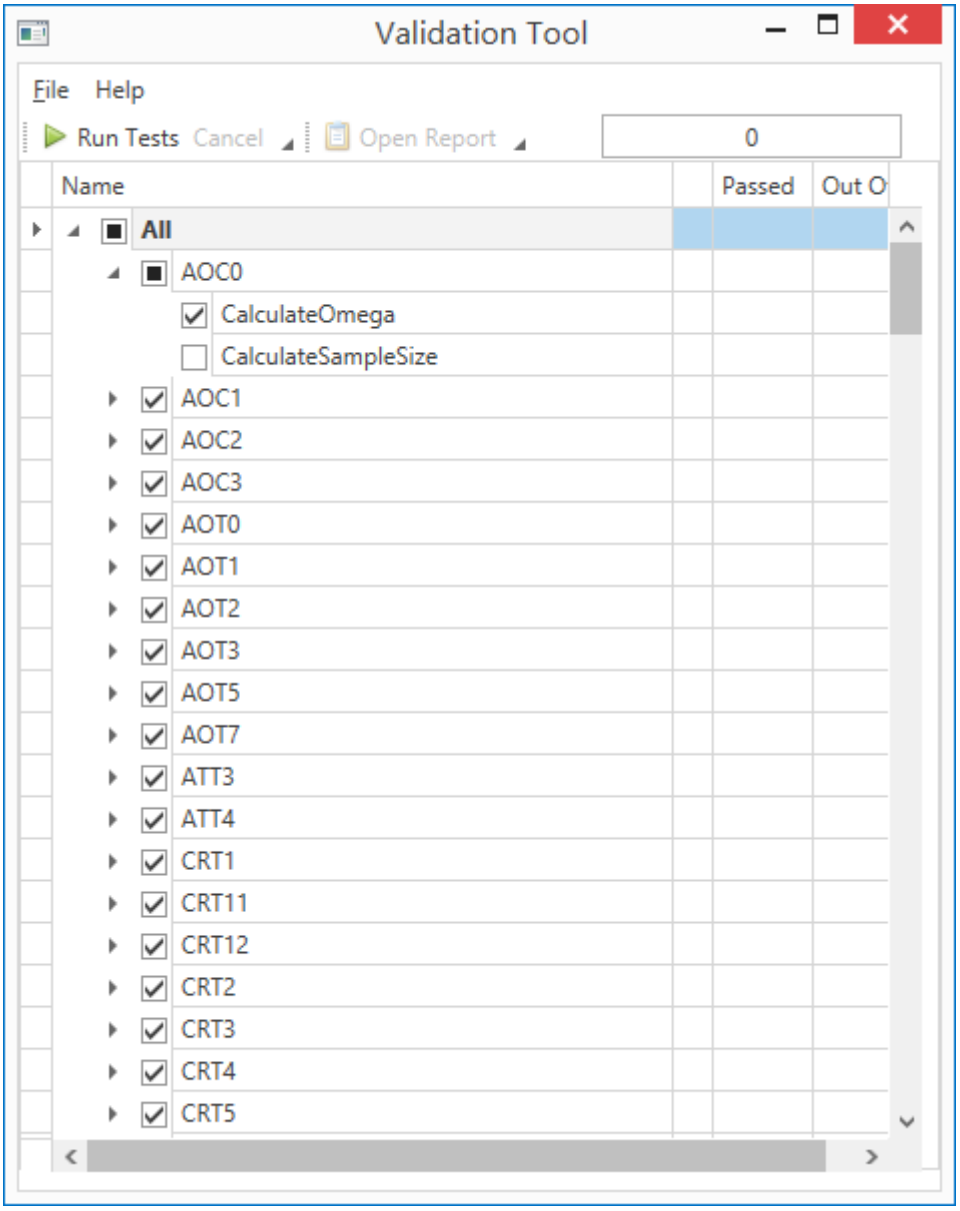



**Figure 7.2:** Operational Qualification Tool

The nQuery OQ window provides a list of all of the nQuery design tables by code (AOT1, MTE1 etc.) and has a three level menu hierarchy: All Tables, Per Table, Per Solver. Selecting the arrow ▶ to the left of the "All" option at the top of the menu will open and collapse the per-table options. Selecting the arrow ▶ to the left of a table code option will open and collapse the solver options for that table.

For all options, the user can include or exclude a table or solver by selecting or de-selecting the check-box to the left of that option. Selection is indicated with a tick icon.

Note that the tables selected by default will depend on when the OQ tool is opened. If the OQ tool is opened on the Getting Started tab or when no tab is open then all tables will be selected by default. If the OQ tool is opened on a specific design table then only that table will be selected by default.

Note that a small number of shortcuts are available for the OQ tool. Select the Help file menu item to see these.

Once the user has selected the tables which will be qualified, select the "Run Tests" button in the window menu bar. Note that since the nQuery testing data is specifically designed to include extreme values to test robustness that some tests may take some time to run. If the user wishes to stop qualification before it is finished, select the "Cancel" button in the menu bar.

As the OQ tool progresses, each item in the menu will show the Qualification results per-feature as it is completed in the columns to the right of the code-name.

The first column will show a tick symbol if the feature has passed and an "x" symbol if it has failed. The next "Passed" column will show the number of passed individual table/solver tests. The final "Out of" column will display the total number of tests ran for that individual table/solver tests. If a table/solver passes then the tests passed should equal the total number of tests.

A progress bar in the menu bar will display the total number of tables tests and indicate overall progress by the amount of the bar filled.

An example of the OQ tool in progress is shown in Figure 7.3.

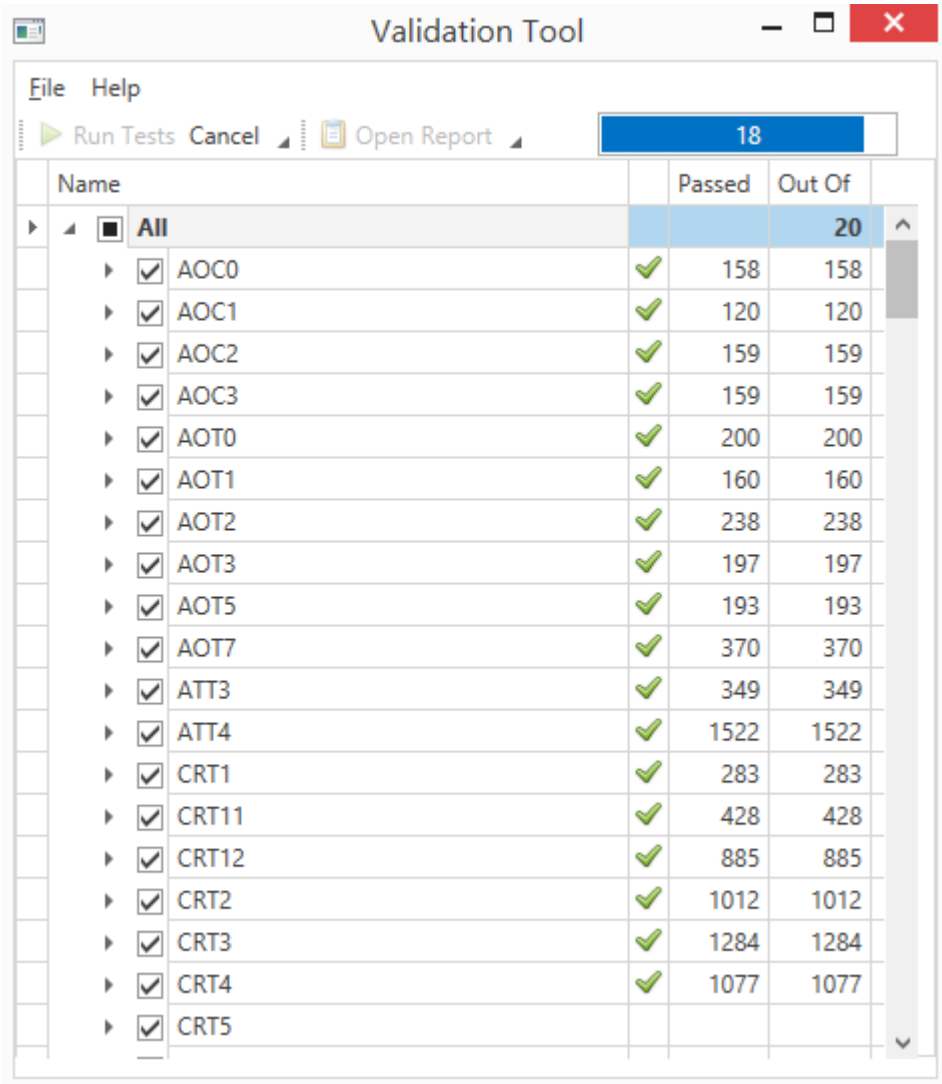



**Figure 7.3:** Operational Qualification Tool - In Progress

### 7.2.2 nQuery OQ Results

To view the full report of the OQ results, select the "Open Report" option in the menu bar. This will automatically output a HTML report which will display the results automatically in the machine's default HTML viewer. This will often be the users default internet browser.

To OQ report file will automatically be saved on the machine when a "Open Report" is used. It will be saved to the C:\Users\<User Account>\AppData\Local\nQuery\Reports folder where <User Account> is the named User which currently logged into on the machine. Alternatively, you can save a report to an alternative folder using the "Save as" option in the File menu of the OQ window.

If OQ report is opened in a supported application, the structure of the report will be as follows:

At the start, the report will contain the time and date the tests were run, the number of passed design table tests and the time taken for testing.

Within the report, there will be an individual report for each design table test run. These will show the table code, table name, an indication whether the table passed overall, the time taken to test that table and the number of passed tests for that table.

Within each design table report, there will be a report for each solver in that design table. This will show the solver name, an indication whether the table solver passed overall, the time taken to test that table and the number of passed table solver tests. When shown within a browser, it allows a user to see the individual design parameters used for each test applied to a solver. To open these, select the "Details" option to the right of the solver name. An example an OQ report with a Details menu open is shown in Figure 7.4.

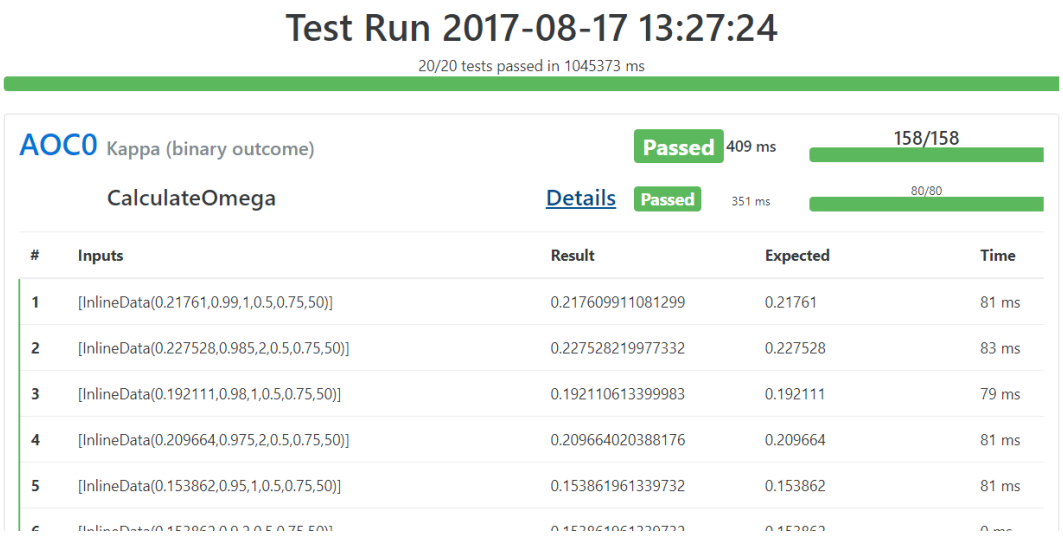



**Figure 7.4:** Operational Qualification Tool Report

# 8 nQuery Updates and Licensing

## 8.1 Renewing a License

*Important: Internet Access is required to renew nQuery using the Activation key. If you need to renew offline, please use the URL from "Web Registration" option (`http://statistical-solutions-software.net/eleckey/registration_activation.aspx`) in the bottom-right of the Product Activation dialog to obtain a License key on an internet enabled device.*

*If still unable to renew your software, contact support@statsols.com or login to your online account*

To renew an nQuery License, select the Activate/Renew License option from the Help menu. This will open the Product Activation dialog. Enter your provided nQuery Activation Key in the Activation Key field. This is shown in Figure 8.1.

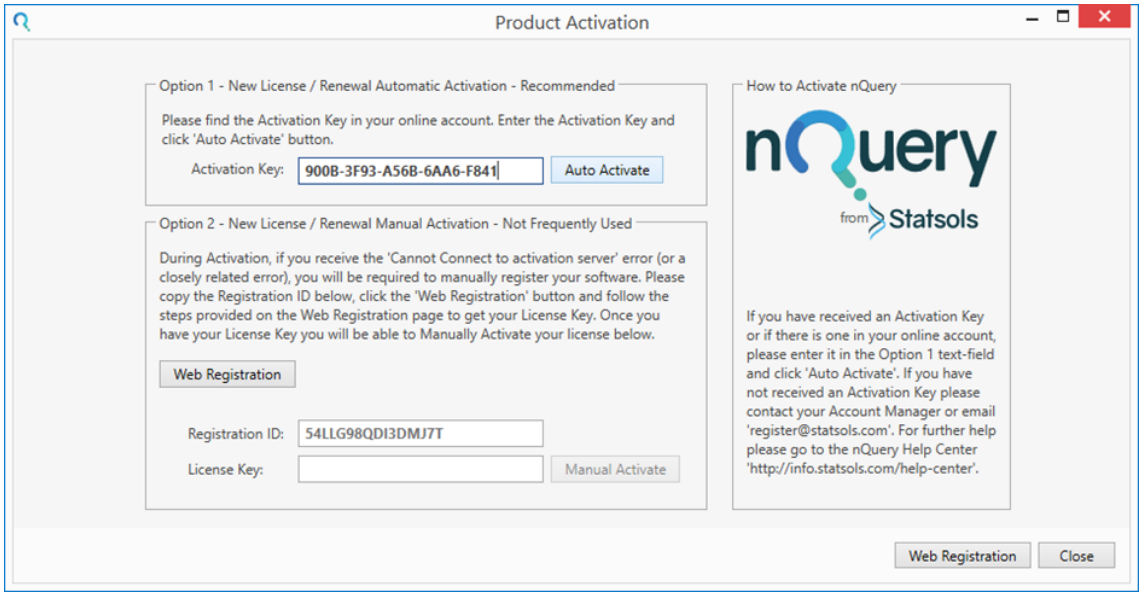



**Figure 8.1:** Renewal Example

If you have issues you can contact support@statsols.com or login to your online account and create a support ticket for fastest response.

## 8.2 Activating an Add-on Module

*Important: Internet Access is required to add an Add-on module nQuery. If you need to add a module offline, please contact support@statsols.com or login to your online account*

To purchase an add-on module, either use the link provided in the Getting Started screen or visit the Statsols website at www.statsols.com

To activate an add-on module, select the Enable Modules option in the Help file menu. The Enable Modules dialog box should open. This is shown in Figure 8.2.

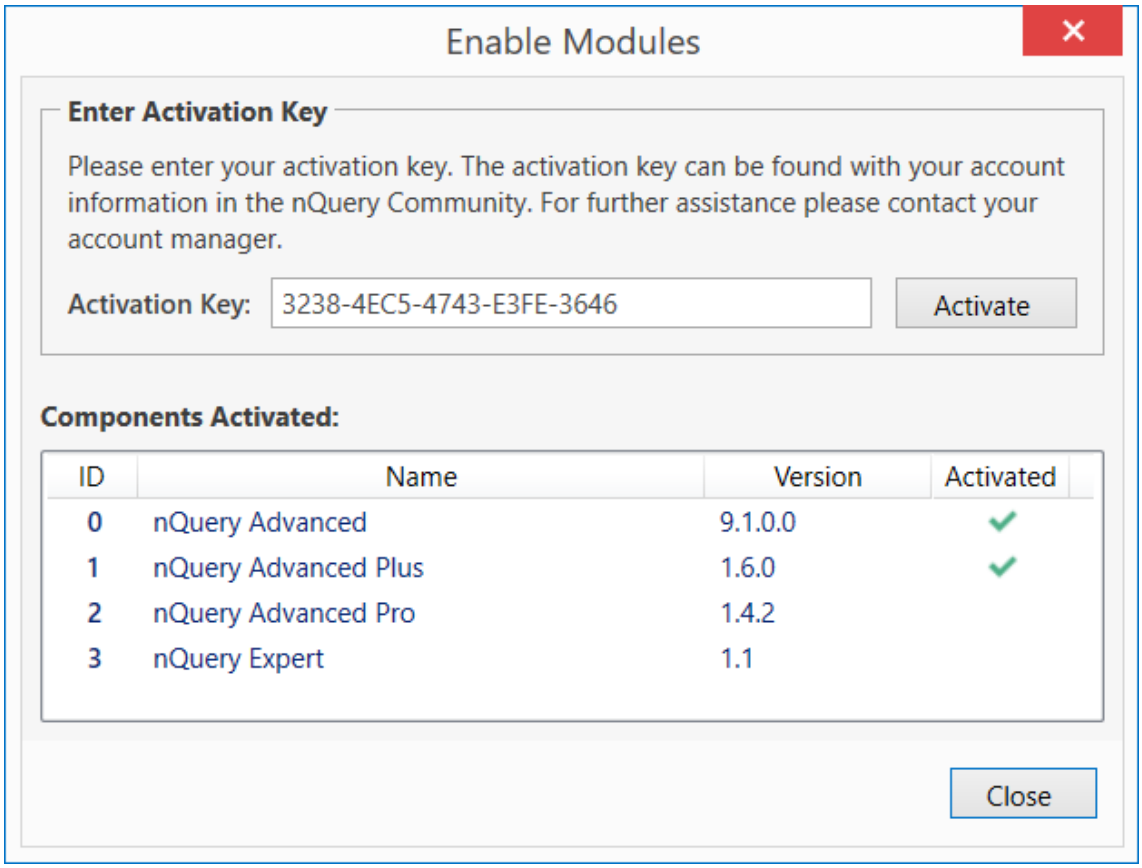



**Figure 8.2:** Module Activation Screen

If you have used a valid nQuery Activation Key in nQuery before, this key will appear in the Activation Key area. Otherwise, you will need to enter the key provided by Statsols.

To activate the module, click "Activate". The Components Activated area will refresh after a couple of seconds and the nQuery Module(s) should have a tick in the Activated column for the selected module(s) .

All tables in the module(s) will be enabled in nQuery Advanced

If you have issues you can contact support@statsols.com or login to your online account and create a support ticket for fastest response.

## 8.3 Checking for Software Updates

*Important: Internet Access is required to update nQuery. If you want to manually update, please contact support@statsols.com or login to your online account*

### 8.3.1 Checking for Updates

nQuery provides automated tools to allow the user to upgrade their nQuery application to the latest version. There are two methods to update your nQuery application: using the Help menu option or using the system tray option.

To select the Help menu option, open the Help file menu and select "Check for Updates". To use the system tray option, find the Update Statsols - nQuery system tray item, right-click and select "Open Updater".

If there are no updates, these will open the Update Statsols - nQuery dialog which will say that no upgrades are available.

If updates are available, a prompt will open asking the user if they wish to upgrade. An example of this prompt is shown in Figure 8.3.

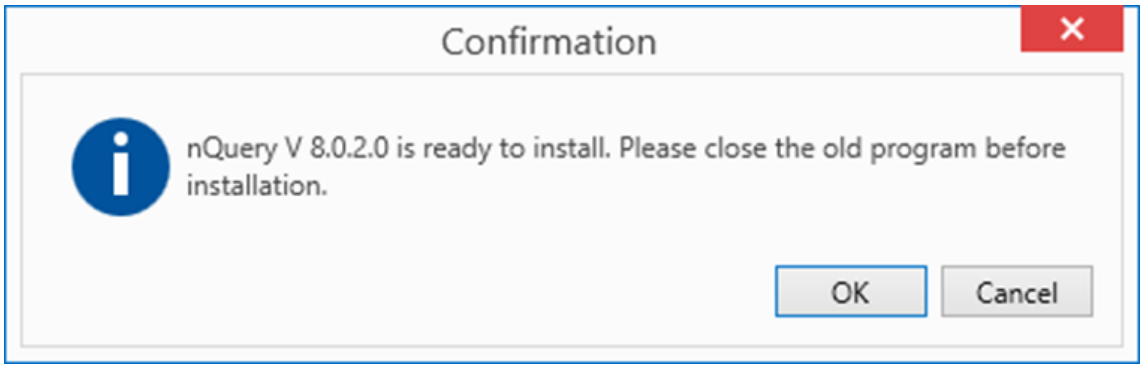



**Figure 8.3:** Update Prompt Example

Select OK to install the update or Cancel to close this dialog. If you select OK, the Update Statsols - nQuery dialog will appear and show the download progress. When downloading is completed, the Figure 8.3 dialog will appear again. Select OK and this will close any currently open versions of nQuery. An install screen will automatically appear and update your nQuery application. On completion, your nQuery application will be updated and can be used again.

# Bibliography


[Ahrens and Dieter, 1972] Ahrens, J. H. and Dieter, U. (1972). Computer methods for sampling from the exponential and normal distributions. *Communications of the ACM*, 15(10):873–882.

[Bretz et al., 2005] Bretz, F., Pinheiro, J. C., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, 61(3):738–748.

[Chen et al., 2004] Chen, Y. J., DeMets, D. L., and Gordon Lan, K. (2004). Increasing the sample size when the unblinded interim result is promising. *Statistics in medicine*, 23(7):1023–1038.

[Cohen, 1988] Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *NJ: Lawrence Earlbaum Associates*, 2.

[Cui et al., 1999] Cui, L., Hung, H. J., and Wang, S.-J. (1999). Modification of sample size in group sequential clinical trials. *Biometrics*, 55(3):853–857.

[Demets and Lan, 1984] Demets, D. L. and Lan, G. K. (1984). An overview of sequential methods and their application in clinical trials. *Communications in Statistics-Theory and Methods*, 13(19):2315–2338.

[Demets and Lan, 1994] Demets, D. L. and Lan, K. G. (1994). Interim analysis: the alpha spending function approach. *Statistics in medicine*, 13(13-14):1341–1352.

[Fang and Su, 2011] Fang, L. and Su, Z. (2011). A hybrid approach to predicting events in clinical trials with time-to-event outcomes. *Contemporary clinical trials*, 32(5):755–759.

[Gao et al., 2008] Gao, P., Ware, J. H., and Mehta, C. (2008). Sample size re-estimation for adaptive sequential design in clinical trials. *Journal of Biopharmaceutical Statistics*, 18(6):1184–1196.

[Hwang et al., 1990] Hwang, I. K., Shih, W. J., and De Cani, J. S. (1990). Group sequential designs using a family of type i error probability spending functions. *Statistics in medicine*, 9(12):1439–1445.

[Jennison and Turnbull, 1999] Jennison, C. and Turnbull, B. W. (1999). *Group sequential methods with applications to clinical trials*. Chapman and Hall/CRC.

[Lee and Zelen, 2000] Lee, S. J. and Zelen, M. (2000). Clinical trials and sample size considerations: another perspective. *Statistical science*, 15(2):95–110.

[Mehta and Pocock, 2011] Mehta, C. R. and Pocock, S. J. (2011). Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Statistics in medicine*, 30(28):3267–3284.

[Nielsen, 2011] Nielsen, M. A. (2011). *Parameter estimation for the two-parameter Weibull distribution*. Brigham Young University.

[O'Brien and Fleming, 1979] O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, pages 549–556.

[Pampallona et al., 1995] Pampallona, S., Tsiatis, A., and Kim, K. (1995). Spending functions for the type i and type ii error probabilities of group sequential tests. *J Statist Plan Inference*, 42(19):1994–35.

[Pampallona et al., 2001] Pampallona, S., Tsiatis, A. A., and Kim, K. (2001). Interim monitoring of group sequential trials using spending functions for the type i and type ii error probabilities. *Drug Information Journal*, 35(4):1113–1121.

[Pinheiro et al., 2006] Pinheiro, J., Bornkamp, B., and Bretz, F. (2006). Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. *Journal of biopharmaceutical statistics*, 16(5):639–656.

[Pocock, 1977] Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199.

[Tukey, 1985] Tukey, J. (1985). The problem of multiple comparisons. *The Collected Works of John W. Tukey*, 2:1965–1984.

[Walke, 2010] Walke, R. (2010). Example for a piecewise constant hazard data simulation in r. *Max Planck Institute for Demographic Research*.