

**Title:** Improving the Efficiency and Ensuring the Quality of Data Assembly for Pharmacometric Analysis

**Authors:** Thaddeus Grasel\* (1), Jill Fiedler-Kelly (1), Darcy Hitchcock (1)

**Institutions:** (1) Cognigen Corporation; Buffalo, NY, USA

**Introduction:** As modeling and simulation activities become more critical to the decision-making process, and model-based development becomes a more typical drug development paradigm, the pharmacometrics process must be able to withstand increasing regulatory and quality management scrutiny. Considerations such as analysis planning, data pooling, analysis dataset creation, quality of the analysis-ready data, and documentation that the data meet the analysis requirements are only a few critical components of such a discussion.

The data acquisition and management tasks required for the preparation of analysis-ready datasets for pharmacometric analysis have always been a challenge. Unlike traditional statistical analyses, in which the focus is on the data from a single study, pharmacometric analyses typically use data arising from multiple sources (trials) across various phases of drug development. It is not uncommon for data from multiple Phase 1, 2 and/or 3 studies to be combined for an exposure-response analysis.

The creation of an analysis-ready dataset consists of preparing a time-ordered sequence of events for each subject, based on a statement of clear and concise specifications for the analyses. Once data programming begins, the programmer is likely to face a host of issues that arise from deficiencies in the specifications or inconsistencies between the specifications and raw data. These issues typically spawn a series of e-mails and discussions between the project team members. In the process of answering these questions, more specific questions are formulated as the team members clarify issues and resolve uncertainties. The cycle of questioning and discussion is a valuable source of information on how to improve specifications and reduce the time and effort required for data assembly.

**Objectives:** The objectives of this presentation are to describe a process for:

- 1) Analyzing email communications between team members to identify common sources of miscommunication and errors
- 2) Incorporating findings into formalized programming specifications for data assembly
- 3) Continually refining and expanding the scope of these programming specifications

**Methods:** A systematic analysis of unstructured e-mail communications between programmers and pharmacometricians generated during the course of data assembly for numerous independent projects was performed. Information extraction and discovery techniques were utilized to uncover the type of information being communicated and the most frequently discussed topics, issues and problems. These issues and the attendant queries and answers were categorized into knowledge domains for further analysis. This categorization was then used to develop a series of formal programming specification forms that captured the necessary informatic elements suggested by the systematic analysis.

**Results:** Inadequate or incomplete information regarding analysis population selection and dosing were common issues necessitating team discussion. Of these, issues pertaining to dose were most important. For example, proper dosing specification requires instructions on handling first dose versus multiple dose, criteria for ascertaining steady-state, and instructions for managing multiple doses prior to achieving steady-state.

The specification forms were deployed in conjunction with a secure, collaborative website (wiki) to capture communications between scientists and programmers in a consistent semi-structured format. System-generated and user-provided metadata serve to facilitate the discussion and resolution of residual issues. Subsequent review of these team communications will enable future refinement and expansion of the specification forms.

**Conclusions:** The development of these specification forms is anticipated to improve the performance characteristics of data assembly in terms of consistency, reliability, timeliness and quality of the work product. Future work will involve quantifying the benefits of these improved programming specification forms.