Data Management considerations in Observational studies

Valerie Alward, Manager RWLPR Data Management
Zia Haque, Senior Director RWLPR Data Management

Improve your probability of success™

Better outcomes
Connecting insights
Superior delivery
Your Presenters

**Valerie Alward**  
*Manager, Clinical Data Management, Real-World Late Phase at Quintiles*

Valerie Alward has been with Quintiles Real-World Late Phase Research DM group for a little over 2 years and has been a Data Team Lead for many studies. Prior to Quintiles, Valerie worked at a senior level in Data Management for several companies, including CROs and Pharmaceutical companies. Valerie started her career in clinical research in 1996 as a site coordinator and then moved on to the data management field in 1998. She has both a Masters in Management from Rensselaer University and a BA in Psychology from North Carolina State University.

**Zia Haque**  
*Senior Director of Data Management, Real-World Late Phase Research*

Zia has been with Quintiles Real World and Late Phase (RWLPR) DM Management team since the past three years. He has worked in clinical data management for 18 years in roles of increasing responsibility and has led global DM teams in early and late phase arenas. Zia holds a BS in Chemistry & Zoology from Bangalore University, a MA in English from Karnatak University and is completing a MS in clinical research from Campbell University.
Today’s Webinar Audience

- Academia: 26.0%
- Biostatistician: 12.0%
- Clinical Operations: 13.0%
- Epidemiology: 11.0%
- Health Economics/Health Outcomes: 13.0%
- Market Access: 23.0%
- Medical Affairs: 2.0%
- Risk Management: 2.0%
- Other: 5.0%
Polling Questions

A small number of polling questions have been added to today’s webinar to make the session more interactive.
Agenda

- Research Study Designs
- Types of Late Phase studies
- Data Management strategies on Observational studies
- Q&A
Typical Hierarchy of Research Designs

- **RCT**
- Prospective Observational Cohort Studies & Pragmatic Trials
- Case-control studies
- Case-reports
- Expert opinion

* Randomized Controlled Trials
Commercial Product Continuum
Interventional Studies, Observational Studies & Registries

Provide real-world data on:

- natural history of disease
- burden of illness
- treatment patterns
- disease management

to inform development, launch strategy, and market access (product not included).

Provide real-world data on:

- pharmaceutical use (on/off-label)
  - safety
  - effectiveness
  - compliance, adherence, persistence
  - treatment satisfaction
- competitor brands
  - comparative effectiveness
- disease management
to inform treatment use.
RCTs cannot answer all research questions

• **Hypothesis-driven** nature of experimental design requires substantial knowledge at the study outset and limits the potential for discovering new information

• Atypical behavior, patients, and settings
  – Protocol-driven behavior in **highly selective patients**
  – May not be usual physician or usual practice
  – Optimal patients should have best outcomes

• Do not give insights into why patients and/or clinicians use products as they do or about off-label or risky situations

• Also
  – Can be hard to recruit patients
  – May be small, with imprecise results
  – Intermediate endpoints may not be clinically meaningful
# Types of Late Phase Studies

<table>
<thead>
<tr>
<th>Interventional late phase studies</th>
<th>Allow for combining generalizability of Observational studies with the validity of RCTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-interventional (Observational) studies</td>
<td>Assess safety of approved products in real world settings, under current standard of care</td>
</tr>
<tr>
<td>Patient Registries</td>
<td>Evaluate specified outcomes for a population defined by a particular disease, condition, or exposure</td>
</tr>
<tr>
<td>Post-Marketing Surveillance</td>
<td>Mandated by regulatory agencies to verify safety, tolerability and effectiveness of approved products</td>
</tr>
<tr>
<td>Pharmacovigilance Studies</td>
<td>Aimed at detecting, assessing and preventing prevention of short and long term Adverse effects or side effects of an approved product</td>
</tr>
</tbody>
</table>
# RCTs, Open Label and Observational studies

<table>
<thead>
<tr>
<th>RCTs</th>
<th>Randomized, Open Label Studies</th>
<th>Observational Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well defined, tightly controlled visit structure</td>
<td>Open label studies involving randomization</td>
<td>Broadly defined visit schedules that reflect real world settings</td>
</tr>
<tr>
<td>Strict Inclusion / Exclusion criteria</td>
<td>Expanded Inclusion / Exclusion criteria</td>
<td>Reduced barriers for patients to enter studies</td>
</tr>
<tr>
<td>Clinical research savvy PIs and site staff</td>
<td>Clinical research savvy PIs and site staff</td>
<td>PIs and site staff are typically therapeutic focused medical practitioners</td>
</tr>
<tr>
<td>Focus on safety and efficacy in controlled study environments</td>
<td>Comparator and dosage compliance studies</td>
<td>Studies designed to collect data from real world settings, with none to minimal interventional procedures</td>
</tr>
<tr>
<td>Data collection and cleaning occurs in a controlled, experimental environment</td>
<td>Data collection and cleaning along the lines of RCTs</td>
<td>Data collection approach should reflect real world settings, and not force fit RCT expectations on Observational studies</td>
</tr>
</tbody>
</table>

*The nature of data on observational studies should be a true representation of the area of research from where the data has been collected*
Efficacy vs. Effectiveness

Late-Phase … Focus Shift

Does it work?  Does it work in the real world?
### ‘Real-World’ Studies

<table>
<thead>
<tr>
<th>Reality</th>
<th>Real-world practice and outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Applicability</strong></td>
<td>Physician practice and resulting outcomes of that behavior</td>
</tr>
<tr>
<td><strong>Generalizability</strong></td>
<td>Broader inclusion/exclusion criteria resulting in diverse study populations, often including many subgroups not traditionally studied in RCTs. May have longer follow-up</td>
</tr>
<tr>
<td><strong>Availability</strong></td>
<td>Add knowledge. Relatively few RCTs are conducted relative to the number of decisions that need to be made</td>
</tr>
</tbody>
</table>
Data Management Approach on Observational studies
Observational Studies – Data Management Differentiators

• Compared to a clinical trial, observational studies tend to have fewer required data elements since we cannot mandate assessments or visits

• In the ideal world we would have a smart CRF with only the must-have data elements

  • eCRFs with optimized conditional branching of data fields
  • Maximizing data collection via drop down menus and radio button options
  • Cleaning majority of the data utilizing real time, front end edit checks

• Epidemiology and Biostatistics teams need to be involved upfront in the eCRF and validation processes
Data cleaning Considerations on Observational Studies

• PIs and site staff are likely to log into the EDC database infrequently

• Site staff may decide to batch enter data after certain number of patients are enrolled in the study at their sites

• Level of familiarity with clinical research translates to limited knowledge with eCRF completion, query resolution turnaround times, and the need for precise responses to queries

• Heavy dependence on listing review will lead to expanded query turnaround times, as sites may not log into EDC for extended periods of time to resolve manual queries

• CRFs should be designed to ensure that maximum data cleaning activity can be achieved at time of data entry into the EDC platform

• Easy to understand query text and avoiding complex edit checks

These factors should be considered during the CRF design process to create a simple, straightforward data collection tool
‘Light’ onsite monitoring and data quality

Enabling data quality

- Observational studies typically have infrequent on-site monitoring since there is no IP involved and data is collected as part of routine practice.

- The model presents risks to the timeliness and quality of data collection at sites.

- Data Management needs to take a lead role (in collaboration with remote site monitoring) in defining metric reports to monitor performance and progress at individual site level. Examples of site reports include: Complete & incomplete eCRF reports, Query ageing reports and site activity reports to gauge site level EDC activity.

- Data Management + Monitoring teams need to function in unison to avoid data collection lag and quality issues.
Patient Reported Outcomes (PRO)

- The need to maintain visit schedules that replicate real world settings can be complimented by appropriate PRO tools.

- If the study patient population can respond to electronic patient reported outcomes (ePRO), this should be the modality for PRO data collection.

- ePRO enables timely collection of patient reported data as per protocol driven time points, and avoids the ‘parking lot’ effect.

- Preferred mode of ePRO collection should be via a web based module that is integrated within the EDC platform, rather than from a third party ePRO vendor.

- Dynamic data validation and reporting enables real time compliance monitoring.
Safety Reporting – Risks and Solutions

Risks associated with safety reporting

Lack of familiarity with safety reporting requirements

Under or over reporting of Serious Adverse Events (SAE)

Inadvertent entry of SAEs in free text fields may lead to under reporting of safety events

Solutions

Educate PIs on safety reporting requirements

Program EDC to send automated alert e-mails when a site enters an event and classifies this as a SAE

Program EDC to launch SAE forms when an Adverse Event is entered, and classified as a SAE

Periodic medical review of relevant listings to ensure appropriate reporting of safety related events
Optimizing Data Quality in Observational studies
Data Quality Challenges

Missing Source Data

- Sites do not always record all data from routine practice into medical records
- Sites may not understand importance of maintaining source data

Missing Data in CRF

- Sites do not all follow same routine practice and therefore may not have same data available
- Sites may need to enter data at multiple locations, and by multiple users if patients treated across different specialties
- Sites may not have adequate staffing needed for data entry
- Site omission to collect data if small target population with low study activity between visits
- Participation is a burden to sites with too much data to collect and/or CRF not intuitive

Poor Quality Data

- Erroneous data
- Data not plausible
- Unanswered queries
## Data Quality

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Risk Mitigation Plan - Keys to Success</th>
</tr>
</thead>
</table>
| Quality data from research naive sites | - Designing and implementing a precise data collection and cleaning model  
  - Identifying precise data elements and appropriate data cleaning strategies  
  - Real time data validation and less dependence on manual listings  
  - Involve Epidemiology and Biostats teams in the CRF and validation plan efforts  
  - Developing targeted eCRF Completion Guidelines (CCG)  
  - CCGs should explain how each field in the eCRF should be completed and which fields need to be completed during each visit  
  - Sites should be trained on the eCCGs during the Site Initiation Visit (could be remote visit) |
## Data Quality

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Risk Mitigation Plan - Keys to Success</th>
</tr>
</thead>
</table>
| **Establishing appropriate level of data quality** | - Data Management Plan (DMP) Outlines plan for concurrent data review and cleaning to include:  
  - Real time data cleaning efforts need to be maximized!  
  - Critical variables need to be clearly identified with input from appropriate teams  
  - Data and queries reviewed on an ongoing basis to identify error trends  
  - Data consistency checks: Comparison across sites and over time often uncovers systemic issues that would otherwise go unnoticed  
  - SAE/AESI occurrences are monitored on an ongoing basis  
  - Ongoing reporting/metrics analysis to identify error trends |
# Data Quality

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Risk Mitigation Plan - Keys to Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site training</td>
<td>▪ <strong>Training</strong></td>
</tr>
<tr>
<td></td>
<td>• Comprehensive training on data collection provided to sites during the Site Initiation Visit including;</td>
</tr>
<tr>
<td></td>
<td>− What data is to be recorded in CRF and at what time points</td>
</tr>
<tr>
<td></td>
<td>− How to complete CRF (CCG and if eCRF-an EDC User Manuel)</td>
</tr>
<tr>
<td></td>
<td>− What is considered as source data, and identify source data location(s)</td>
</tr>
<tr>
<td></td>
<td>− Data quality and managing queries</td>
</tr>
<tr>
<td></td>
<td>− How to request EDC support</td>
</tr>
<tr>
<td></td>
<td>• During Site Monitoring Visits;</td>
</tr>
<tr>
<td></td>
<td>− Any additional training will be provided during SMV</td>
</tr>
<tr>
<td></td>
<td>− Resolutions to any data quality issues issues that have been identified previously will be discussed, with clear follow up plans</td>
</tr>
<tr>
<td></td>
<td>• Escalation process for data quality issues within project team</td>
</tr>
<tr>
<td></td>
<td>• Lead Data Manager alerts CPM on Data Quality issue/s</td>
</tr>
<tr>
<td></td>
<td>• Study team creates appropriate plan to address Data Quality issue/s at site</td>
</tr>
<tr>
<td></td>
<td>• Follow up meetings scheduled as needed to determine results of quality improvement plans</td>
</tr>
<tr>
<td></td>
<td>• Ongoing support and training to sites throughout the study</td>
</tr>
</tbody>
</table>
## Data Quality

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Risk Mitigation Plan - Keys to Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimizing level of onsite Monitoring</td>
<td>▪ Develop Monitoring Plan using a Risk Based Monitoring Approach</td>
</tr>
<tr>
<td></td>
<td>– Outline what, when and how data will be monitored through-out the study</td>
</tr>
<tr>
<td></td>
<td>– Define how monitoring will be reported and develop templates for monitoring</td>
</tr>
<tr>
<td></td>
<td>– Define escalation process if data quality issues</td>
</tr>
<tr>
<td></td>
<td>– Define what standard procedures/guidelines will be followed</td>
</tr>
</tbody>
</table>

Whether collecting data as part of a regulatory requirement or not, adherence to the core tenants of the ICH GCP and/or GPP Guidelines is a best practice that should be followed.

To ensure the appropriate level of compliance and data integrity, evaluate on a per-study basis individual practices, national and local regulations and guidelines.
Handling Missing Data
Challenges in Observational Studies

• **What:** For Observational Studies, we can encounter missing *exposure* variables, *outcome* variables, *confounders* and *effect* modifiers!
  › In Randomized Clinical Trials (RCTs) the *exposure* variable (treatment) should never be missing

• **Why:** Data may be missing due to an assessment not being conducted, unknown or not applicable

• **Magnitude:** Because we have less control over the subject’s treatment and often rely on data collected in medical charts, we encounter missing data more often

• **Impact:** All truly Observational studies will have some elements of missing data – this is a reality and true reflection of the research space. Analysis driving critical data elements must be identified precisely to ensure that targeted data is collected, thereby reducing site burden in completing eCRFs
How Do We Define Missing Data?

Missing is not the same as unknown.

- Missing = no measurement and no known reason why
- Unknown = no measurement but a known reason why

Missing data can result from:

- Refusal of a test or questionnaire
- Loss to follow-up
- Revocation of consent
- Removal from study
- Study design problems
- Field work errors
- Data handling errors
- Test or procedure not done (e.g., lab test not ordered by the physician)?
- Death (e.g., “missing” quality of life measures if we know the patient died)?
"(Missing data are) values that are not available and that would be meaningful for analysis if they were observed".

- Ware, NEJM, 2012; 367;14
The Big Picture

When faced with missing data we need to answer four questions:

- **How many** data points are missing?
- How do patients with missing data **differ** from those without?
- **Why** are the data missing?
- What can we **do** about it?
What to do about Missing Data!

- **Prevent** Missing Data from the start of the study with careful field work and data collection, along with data management and monitoring
- **Quantify** the amount of missing data
- **Assess** differences between patients with versus without missing data if needed
- **Identify** the possible causes of missing data
- **Evaluate** the missing data mechanisms
  - **Missing Completely at Random (MCAR)**: may be ignorable - no bias; only reduces statistical power
  - **Missing at Random (MAR)**: consider likelihood-based approaches, multiple imputation
  - **Missing Not at Random (MNAR)**: consider pattern mixture models and selection models
- **Choose** the appropriate analytic methods
- **Determine** the impact of missing data on results using sensitivity analyses, imputation, and/or multiple imputation

It is paramount to involve the Epidemiology and Biostatistics teams for their expertise in addressing this topic!
# Addressing Missing Data

<table>
<thead>
<tr>
<th>REGISTRY PLANNING</th>
<th>OPERATIONS</th>
<th>ANALYSIS AND REPORTING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important data elements should be required during the data collection process – <strong>Balance</strong> between collecting must have vs nice to have data elements.</td>
<td>Plan for follow-up contact for participants with missing PROs and ongoing monitoring to prevent patient loss to follow-up.</td>
<td>Understand nuances of the data source and impact of registry transitions over time, such as changes in data collection forms, data elements, and disease management and treatment.</td>
</tr>
<tr>
<td>Choose the most appropriate mode(s) of administration for PROs – consider the target population.</td>
<td>Ensure data management efforts include ongoing data quality review for data completeness and quality.</td>
<td>Evaluate missing data to ensure that the Missing At Random (MAR) assumption is reasonable.</td>
</tr>
<tr>
<td>Choose data elements that are routinely collected and reflect usual care, including PROs.</td>
<td>Define clear abstraction guidelines for medical record data sources to allow sufficient interpretation of missing fields.</td>
<td>Report the amount of missing fields and how missing data were handled.</td>
</tr>
</tbody>
</table>

Plan for Missing Data

When choosing a study design, consider:
- Stakeholder needs and expectations
- Potential risks and harm of making a wrong decision

- **When choosing a study design, consider**
  - Stakeholder needs and expectations
  - Potential risks and harm of making a wrong decision
Thank you
Previous & Upcoming Events

Quintiles experts run regular webinars on Real-World & Late Phase services.

Topics include:

- PATIENT RETENTION & REPORTING
- VACCINES STUDIES
- HTA & MARKET ACCESS
- MAXIMIZING VALUE AND QUALITY IN PHASE IV
- HYBRID DESIGNS
- DATA MANAGEMENT
- PATIENT ASSOCIATIONS

Visit Quintiles to learn more at one of the following upcoming meetings:

- INTERNATIONAL CONFERENCE ON PHARMA COEPIDEMIOLOGY AND THERAPEUTIC RISK MANAGEMENT
- ISPOR 5TH LATIN AMERICA CONFERENCE
- ASA BIOPHARMACEUTICAL SECTION FDA-INDUSTRY STATISTICS WORKSHOP
- BIG DATA IN CLINICAL DEVELOPMENT/EHEALTH FORUM

To register or view previous webinars please go to http://www.quintiles.com/landing-pages/real-world-and-late-phase-research-webinars