Data Quality – How Much is Enough?

Detlef Nehrdich
Director Statistics, Data Management &
EDC Project Office Europe
Abbott GmbH & Co KG
Ludwigshafen, Germany

Disclaimer

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to Drug Information Association, Inc. (“DIA”), its directors, officers, employees, volunteers, members, chapters, councils, Special Interest Area Communities or affiliates, or any organization with which the presenter is employed or affiliated.

These PowerPoint slides are the intellectual property of the individual presenter and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. Drug Information Association, DIA and DIA logo are registered trademarks or trademarks of Drug Information Association Inc. All other trademarks are the property of their respective owners.
Data Quality Definitions (1)

- A degree of excellence?
- Conformance with requirements (e.g. less than 50, 20, 10, or 5 errors in 10000 data points?)
- Fit for analysis?
Data Quality Definitions (2)

- The main purpose of quality assurance methods applied to CT should be to protect the right and safety of trial participants and to reduce the likelihood that the trial results are affected by bias and thus affecting the safety of future patients

Some quotes (Mats Lörstad)

- The traditional interpretation of data quality does not consider the initial, preparatory work necessary:
  - To ensure standardized, valid, accurate and reliable measurement
  - Same rigor used by all investigators
- Even the full application of GxP methodology does not guarantee perfection
Some quotes (continued Mats Lörstad)

• GCP is incomplete as its start with recording of measurement instead of starting with the training of procedures, creation of accurate measurements
• The notion is that the education of investigators is sufficient to prevent them from making errors and that they will retain this ability for ever
• Wasted time on symbolic checking routines
• Scientifically meaningless formalities are polished to perfection but the quality management activities which matter are mistreated

=> Quality Declaration needed

Assumptions

• Every transcription process has an error rate > 0
• There is NO clinical database with an error rate = 0 (BP: 124/84 vs. 142/84)
• All error rate detection is based on redundant information
Assumptions (1)

• There are different sources of errors:
  – Design errors
  – Procedural errors
  – Recording errors
  – Fraud
  – Analytical errors
• What is the impact of all our activities like data cleaning, SDV and others?
• Analysis of 17 locked databases
• Four therapeutic areas:
  – Immunology
  – Neuroscience
  – Dyslipidemia
  – Antivirals
• 2.183 Mill. datapoints
Impact on Analysis (1)

• Danish Breast Cancer group investigated two studies by a complete check of hospital data of patients who went off drug due to recurrence

• The group found 16.2% of this selection of data being incorrect

=> But: Statistical results were “not significantly influenced”

Impact on Analysis (2)

• Vermont Oxford Trials Network (neonatal intensive care units)

• The group found 19.3% of data being incorrect

=> "Despite the disagreements between database and medical records… for 4341 infants… the overall proportions (calculated) … were close to the values estimated. This suggests that database reports of overall event frequencies are reliable"
Impact on Analysis (3)

• Simulation of key efficacy and safety variables different error rates (15%, 10%, 5%, 1%) for a large (>333000 data points) Abbott study

• Comparison of original results with simulated data

Abbott Case Study Results (1)
Lesson learned

• We look at our data quality measures on a regular basis
• Collecting only data that are required for the clinical trial is simple, sound logic that is sometimes ignored
• Re-think our efforts on SDV: Targeted SDV
• Re-think our efforts on data cleaning: Remote Monitoring
Monitoring Differently

WHY?
Utilize $ more efficiently w/o negatively impacting quality
Increase number of studies we can fund

HOW?
Focus more on process than data points
Reduce time spent on SDV

WHAT?
Both internally and externally monitored studies

Achieve Cost Savings By Managing Monitoring Interval

Increase onsite monitoring interval to average of 10 weeks for Phase 2-4

- Less frequent monitoring visits for sites with fewer patients
- More frequent monitoring visits for sites with more patients or where required for quality concerns
Increased Monitoring Interval Made Possible by EDC

<table>
<thead>
<tr>
<th>Example Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross check con meds and AEs</td>
</tr>
<tr>
<td>Medical history for incl/excl criteria</td>
</tr>
<tr>
<td>Review gaps in dosing</td>
</tr>
<tr>
<td>Enrollment activity</td>
</tr>
<tr>
<td>Reason for termination versus rest of supporting data</td>
</tr>
<tr>
<td>Review reasons for screen failure</td>
</tr>
<tr>
<td>Resolution of data issues</td>
</tr>
</tbody>
</table>

Remote Monitoring

Expectations for Data Managers, Safety Reviewer and Field Monitors:

- Enter the system(s) at least once a week
  - EDC, IVRS, Central Labs, etc.
- Review each site’s data per IDRP
- Perform all remote monitoring tasks on the remote monitoring checklist
At Onsite Visits, Monitors will Perform Targeted SDV

**Targeted SDV**
- All Adverse Events
- Endpoints – predefined by study team based on the protocol
- Reason for termination
- Reason for screen failure Stratification variables (if applicable)
- Verify that CRF and source updated for any discrepancies found during drug accountability

**Obstacles**

If this was good it wouldn't be possible here –
if it is indeed that good others would have already done it!