TARGETED MAPPING TO SDTM ENABLES LEGACY DATA REUSE WITHOUT FULL CONVERSION

Bereket Tesfaldet, Helena Sviglin, Tejas Patel, Eileen Navarro Almario, for MATIG

CDER/OTS/OCS and DCVS/NHLBI/NIH

Introduction

The Meta-AnalyTical Interagency Group (MATIG) is composed of multidisciplinary scientists from the National Heart, Lung, and Blood Institute (NHLBI), Food and Drug Administration (FDA), and other academic institutions. Our research interest is to advance public health outcomes in cardio-metabolic disorders using existing data from publicly available sources such as NHLBI data repository (BioLINCC). Disparate formats, inadequate metadata, and limited traceability from data-as-collected to analysis data are practical challenges to this reuse. We outline the approach developed to overcome these challenges by building a common data model (CDM) based on the essential CDISC standard domains to answer a specific research question, enabling data reuse with limited legacy transformation.

1. State Research Hypothesis

In subjects with T2DM, serum HDL-cholesterol (HDL-C) raising and/or triglycerides (TG) lowering is associated with a reduced risk of heart failure (HF).

2. Develop a Statistical Analysis Plan

- Time-to-first-event analyses to compare the occurrence of HF among 4 groups based on % change in HDL-C and TG from baseline to 6 months
  - Achieves both HDL-C and TG targets
  - Achieves neither target
  - Achieves only HDL-C target
  - Achieves only TG target.
- Perform multivariate analysis to adjust for age, gender, body mass index (BMI), waist circumference, established cardiovascular disease (CVD), history of hypertension, diabetes, or metabolic syndrome, and estimated glomerular filtration rate [eGFR].

3. Identify Relevant Data Elements

- Identify the data elements required to answer the research question and build a CDM based on the SDTM standard.
- Identifying the elements needed to perform the analyses facilitates the identification, extraction and harmonization of the core data from individual studies that are in their legacy data model (LDM).

4. Build a Compendium to Identify Relevant Studies

- A compendium is a framework that describes details needed to identify and stack data elements across databases for the specific research purpose.
- It facilitates a priori determination whether all needed data elements that support the meta-analytic research concepts are present in each new data source and whether they are sufficiently uniform to enable extraction into the pooled database.

5. Build a Data Mapping Scheme

- For each data source, create a data mapping spreadsheet that maps legacy data elements to CDM data elements.

<table>
<thead>
<tr>
<th>CDM Variables</th>
<th>LDM Variables</th>
<th>Stackability*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint</td>
<td>Research Concept</td>
<td>Data Element (Definition)</td>
</tr>
<tr>
<td>HF</td>
<td>RCT</td>
<td>CRF, binary variable</td>
</tr>
<tr>
<td>HF</td>
<td>Trial 1</td>
<td>CRF, binary variable</td>
</tr>
<tr>
<td>HF</td>
<td>Data not available</td>
<td>Study is not able to support the research question.</td>
</tr>
<tr>
<td>Population</td>
<td>T2D status (based on 1997 ADA criteria)</td>
<td>CRF, binary variable</td>
</tr>
<tr>
<td>Population</td>
<td>T2D status (patient self-report, medical record, current treatment w/ insulin or OHAs)</td>
<td>CRF, binary variable</td>
</tr>
<tr>
<td>Population</td>
<td>T2D status (patient self-report, medical record, current treatment w/ insulin or OHAs)</td>
<td>CRF, binary variable</td>
</tr>
</tbody>
</table>

6. Implement the Data Mapping and Run the Analyses

Conclusion

- A targeted mapping of data elements from clinical trials identified in a compendium to the SDTM format enables patient-level meta-analysis without full conversion of trial data.
- This approach can save the cost and time needed to do full conversion of disparate clinical trials data into common data format.

Disclaimer: The views expressed in this presentation are those of the authors and do not represent FDA policy.