Demographic Representation in Aggregated Diabetes Clinical Trials
Shondelle Wilson-Frederick, PhD; Helena Sviglin, MPH; Lilliam Rosario, PhD; Jonca Bull, MD; ShaAvhrée Buckman-Garnier, MD, PhD
HHS, Office of Minority Health; FDA, Center for Drug Evaluation and Research, Office of Translational Sciences; FDA, Office of the Commissioner, Office of Minority Health

Background: The U.S. Food and Drug Administration (FDA) has a large repository of clinical trial data from regulatory submissions of new drug applications. However, these data exists in disparate formats, with data structure, terminology, and data elements unique to each application. As the FDA increasingly looks for safety signals across applications, overcoming the challenge of assembling data - not originally designed to go together - becomes more important.

The Office of Translational Sciences in the Center for Drug Evaluation and Research (CDER) at FDA recently converted data from 51 pivotal clinical diabetes trials submitted in support of new diabetes drug applications into a standardized format to facilitate patient-centered outcomes research (PCOR). This initiative was funded through the American Recovery and Re-Investment Act (ARRA) to test the usability of converted and stacked data for multiple applications. Data structure was harmonized to a single standard (CDISC, SDTM 3.1.2) to enable an aggregated trial database. We describe the technical challenges and benefits of aggregating data from 6 clinical trials conducted for a diabetes drug that belong to a single drug class.

Methods: We used a subset of 6 of the 51 trials to describe the demographic characteristics of adults patients enrolled in diabetes trials for a single drug class (Thiazolidinediones), stratified by geographic study region (U.S., U.S. and non-U.S., and Non-U.S. studies). Using a scientific workstation with sufficient computing capability, we created SAIL code (v9.3), to join Analysis Data Model (ADaM) subject-level, ADaL, and laboratory, ADLB datasets within an individual trial submission. This resulted in an aggregated dataset with a structure of one record per observation, with the subject (USUBJID) as the unit of observation. We describe the demographic characteristics of patients enrolled in these 6 clinical trials.

Results: We performed a demographic subset analysis and examined baseline laboratory test characteristics of patients enrolled in diabetes trials for a single drug class (Thiazolidinediones), stratified by geographic study region (U.S., U.S. and non-U.S., and Non-U.S. studies). Using a scientific workstation with sufficient computing capability, we created SAIL code (v9.3), to join Analysis Data Model (ADaM) subject-level, ADaL, and laboratory, ADLB datasets within an individual trial submission. This resulted in an aggregated dataset with a structure of one record per observation, with the subject (USUBJID) as the unit of observation. We describe the demographic characteristics of patients enrolled in these 6 clinical trials.

FDA receives a vast amount of data in disparate formats, including paper/pdf. Because these files are not systematically coded or even comparable, merging or concatenating them poses challenges for data users. By applying standardized methods to convert data into a common language, information can be entered and stored into Janus Clinical Trial Repository (CTR). CTR is a data warehouse that transforms data into a usable format recognized by various statistical software programs and tools, such as JReview, MAED, SAS, JMP, etc. Having the data in a standardized format provides a systematic way to address important scientific and regulatory questions.

Table 1. Demographic Characteristics of Hispanic Ethnicity Natives

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age (Mean ± SD)</th>
<th>BMI (Mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>55.8 ± 10.3</td>
<td>32.8 ± 5.2</td>
</tr>
<tr>
<td>Female</td>
<td>59.8 ± 9.3</td>
<td>36.9 ± 6.7</td>
</tr>
</tbody>
</table>

Fig. 1: A Strategy to Handle Massive Amounts of Data

Fig. 2: Approaches for Using Standardized Data

Fig. 3: Racial Inclusion by Region of Pioglitazone Studies (N=9002)

Lessons Learned:
• Able to join multiple SDTM domains however, ADaM datasets facilitate cross study analyses
• Secondary use of these data should be well planned
• Prior to conducting analyses, clearly define research inquiry and strategy

Conclusions:
• Analysis validates successful data integration
• Proof of concept for aggregated clinical trial datasets that lend themselves to analytical purposes
• Populations in these 6 diabetes trials do not reflect the full range of patients with the disease in the U.S population
• -women and racial/ethnic minorities are under-represented but are likely to be treated with these products
• To ensure safety and efficacy of marketed therapies, strategies should be identified to increase participation of these subgroups in clinical research

Acknowledgements: Dr. Shondelle Wilson-Frederick would like to thank the FDA Commissioner’s Fellowship and the FDA Office of Minority Health for supporting this research.