FY2018 GDUFA Science and Research Report: Data Analytics

This section contains only new information from FY2018. For background scientific information and outcomes from previous years on this research topic, please refer to:

- FY2015 GDUFA Science and Research Report: Database and Knowledge Management (https://www.fda.gov/ForIndustry/UserFees/GenericDrugUserFees/ucm508097.htm)
- FY2016 GDUFA Science and Research Report: Database and Knowledge Management (https://www.fda.gov/ForIndustry/UserFees/GenericDrugUserFees/ucm549179.htm)
- FYs 2013-2017 GDUFA Science and Research Report: Database and Knowledge Management (https://www.fda.gov/ForIndustry/UserFees/GenericDrugUserFees/ucm597035.htm)

Introduction

FDA's research in data analytics is focused on leveraging FDA data sources to make better and more efficient decisions throughout the generic drug program. We are using data analytics to plan science and research, meet our Generic Drug User Fee Amendments (GDUFA) II commitments, review new types of analytical data in abbreviated new drug application (ANDA) submissions, and monitor generic substitution of approved products.

As a part of the GDUFA II commitments, FDA facilitates generic drug development by publishing product-specific guidances (PSGs) for non-complex new chemical entities (NCEs) within 2 years of approval, and conducting pre-ANDA meetings with generic drug developers for complex drug products that do not have PSGs. Data analytics research supports the implementation of the commitments through the creation of a knowledge base regarding complexity of approved drug products that integrates knowledge from three categories: (1) drug product information (including complexity classification); (2) regulatory information (e.g., new drug application (NDA) approval date and PSG publishing date); and (3) pharmacoeconomic information (e.g., drug sales). This knowledge base helps identify future science and research needs and plan the development of PSGs.

Research

FDA continues to develop the capacities of data analytics (e.g., machine learning, nature language processing and process mining, etc.) and data management application tools (e.g., in **Figure 1**) to support OGD's mission. One example is application of machine learning-based time-to-event analysis to predict the time to the first submission of ANDA referencing NCEs. This predictive application will facilitate estimation of OGD's workload. In terms of analytical methods, the machine learning methodologies are designed as data-adaptive approaches, bearing the least assumption for data. Our analysis also shows that the machine learning methodologies for the time-to-event analysis offer overall better performances than the conventional regression-based methods (e.g., Cox regression model) which often work under over-simplified assumptions.

Another domain of FDA's efforts in data analytics is developing equivalence testing methods for in-vitro bioequivalence (BE) studies. One example is developing equivalence methods for comparing the complex particle size distributions (PSDs) profiles (e.g., multiple peaks) between test and the reference products. The conventional PSD analysis methods (e.g., D50/SPAN) is sufficient to evaluate the monomodal PSD profile (i.e., one peak), but not suitable for analyzing complex PSD profiles. The ORS-developed approach employs a statistical distribution comparing algorithm, namely Earth Mover's Distance (EMD) to assess the difference between the whole PSD profiles, and then applies the population bioequivalence method to draw statistical

conclusions. This approach has been included in the PSGs of cyclosporine emulsion and barium sulfate suspension, recommended for assessing equivalence of PSD profiles between test and the reference products. Finally, the data analytics efforts in the ORS also support application review tools to enhance efficiency. A Shiny-R-based application was developed to facilitate the usability of received PK data for reviewers. This application can save 4-8 hours per study for reviewers for managing PK data. Its user-friendly web interface enables the application as an easy-to-use tool for most reviewers. More functions (e.g., statistical bioequivalence test) are being added to further enhance the review efficiency.

Several external projects were focused on evaluating brand name to generic substitution products by developing appropriate analytical tools. Due to the observational nature of post-marketing studies using real-world data, one of most important steps in generating real-world evidence from real-world data is to control the bias from the confounding factors. For instance, if you investigate the association between physical inactivity (exposure) and heart disease (outcome) using observational data (i.e., data not from a well-controlled trial), the variable of 'Age' can be a confounding factor because it is associated with both the exposure and the outcome, where the exposure can be severely biased in term of 'Age'. However, this bias can be mitigated by balancing the exposure across all the age groups. Thus, without properly dealing with the confounding factors, the conclusions in the studies using the observational data can be misleading.

Grants #U01FD005555 and U01FD005556 used different methods for reducing bias, especially selection and temporal bias (e.g., high dimension propensity score, disease risk score, and regression discontinuity), whereas instrumental variable analysis will be implemented in Grant #1U01FD005875 to assess brand name and generic drugs used to treat hypothyroidism (**Figure 2**). One finding from grant #U01FD005555 is about the comparative analysis for the outcomes (i.e., the composite cardiovascular and all-cause ER visits) between brand-name and generic atorvastatin initiators. In the analyses without controlling confounding factors, the brand-name initiators had a similar rate of a composite cardiovascular events, but a lower rate of all-case ER visits, compared to generic initiators. After adjusting the confounding factors by the propensity score fine stratification and weighting, for both of outcomes there is no statistically significant difference found between the brand-name and generic products.

Figure 1. Two Application Tools of Data Management Developed by ORS in FY2018.

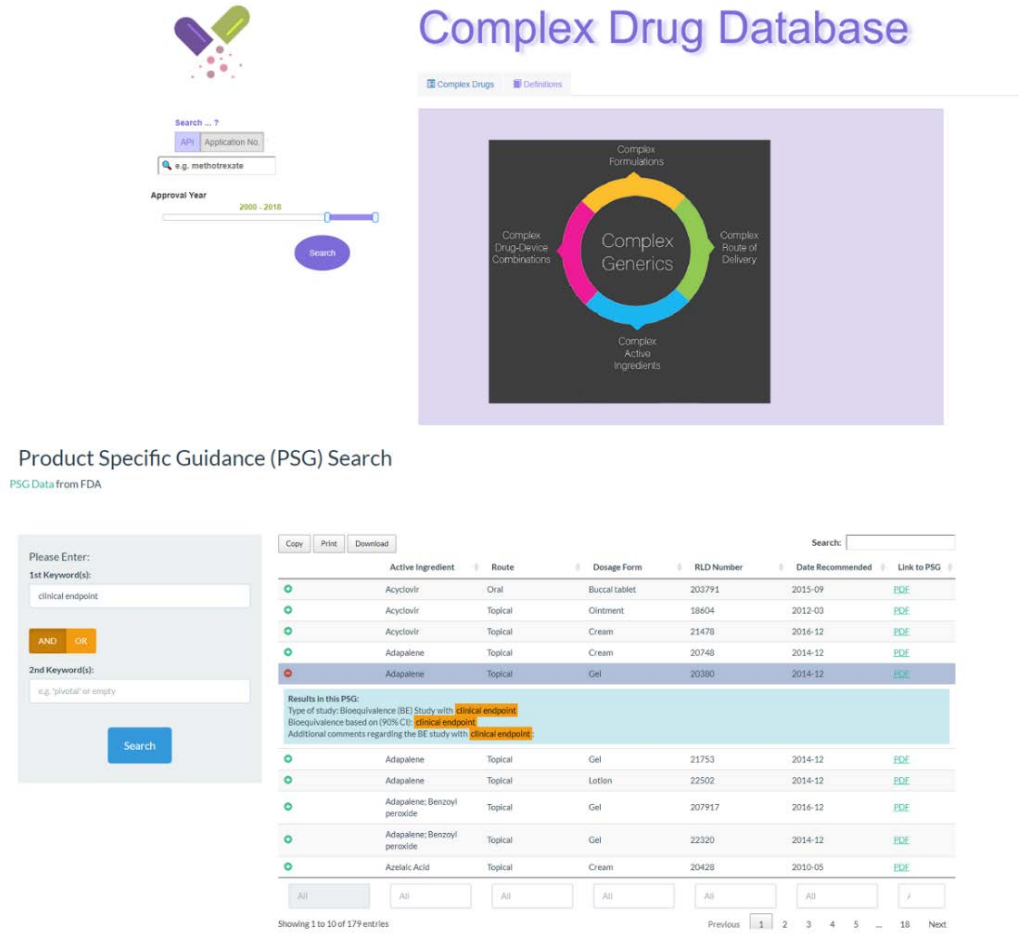

Figure 1: Complex Drug Database (top), and search tool for published PSGs (bottom).

Figure 2. Post Marketing Research Methods, Evaluated Drugs and Associated Outcomes.

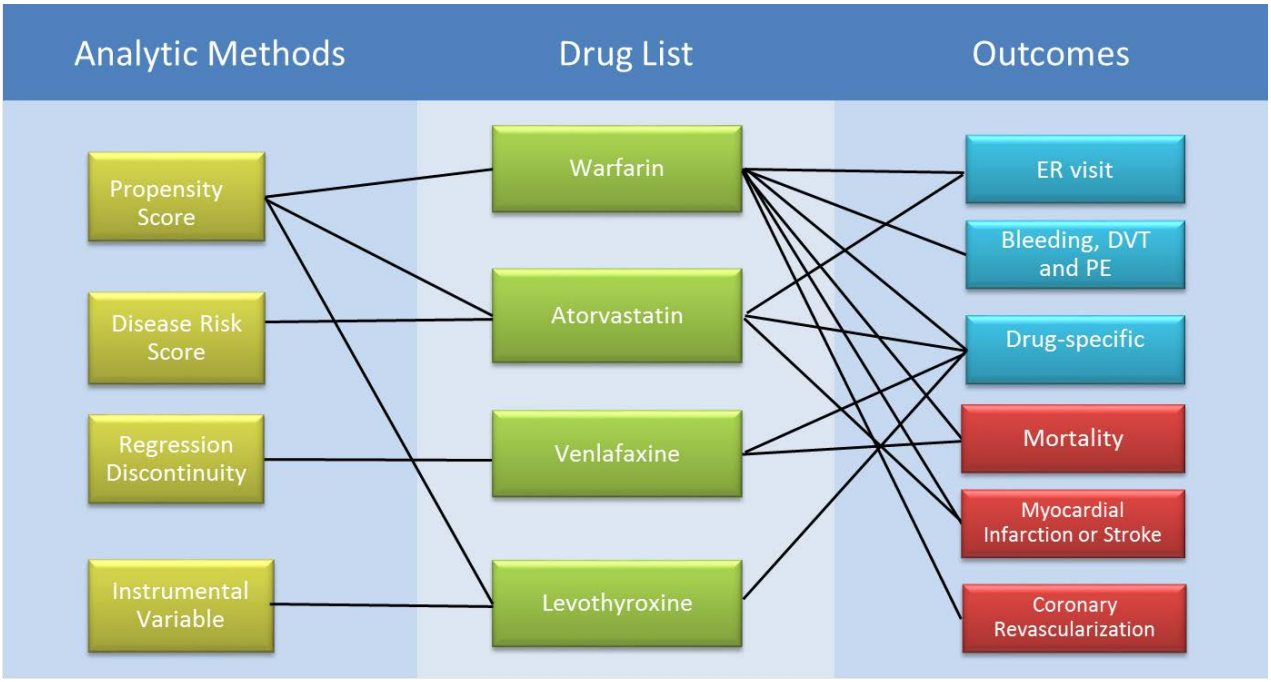

Figure 2: Grants #U01FD005555 and #U01FD005556 used propensity score, disease risk score, and regression discontinuity to study warfarin, atorvastatin and venlafaxine, whereas instrumental variable analysis will be implemented in Grant #1U01FD005875 to assess brand-name and generic levothyroxine used to treat hypothyroidism. Results from the current grants showed that there are no significant differences between the brand and generic products following statistical adjustments with the adopted methods.

Research Projects and Collaborations

Continuing Grants and Contracts

- Active Grant (3U01FD004979-02S3-P1) *Molecular Properties of Excipients* with Brian Shoichet at University of California San Francisco
- Active Grant (1U01FD005556) *Structural Nested Models for Assessing the Safety and Effectiveness of Generic Drugs* with Ravi Varadhan at Johns Hopkins University
- Active Grant (1U01FD005555) *Novel Approaches for Confounding Control in Observational Studies of Generic Drugs* with Rishi J Desai and Joshua J Gagne at Brigham & Women's Hospital
- Active Contract (HHSF223201510112C) *Comparative Surveillance of Generic Drugs by Machine Learning* with Peggy Peissig at Marshfield Clinic, Inc.

Active FDA Research

- *Development of the Earth Movers Distance for Particle Size Distribution Comparisons*
- *Data Analysis for Product Specific Guidances*
- *Development and Analysis of a Complex Product Database*
- *Machine Learning for Generic Drug Analysis*
- *Development of PK Data Warehouse for BE Analysis*

Outcomes

Publications

- Alatawi, Y., Rahman, M., Cheng, N., Qian, J., Peissig, P., Berg, R., Page, C., and Hansen, R. *Brand Vs Generic Adverse Event Reporting Patterns: An Authorized Generic-Controlled Evaluation of Cardiovascular Medications*. J Clin Pharm Ther. (2017) **43**(3):327–335. doi: 10.1111/jcpt.12646. PMID: 29092097.
- Cheng, N., Rahman, M., Alatawi, Y., Qian, J., Peissig, P., Berg, R., Page, D., and Hansen, R. *Mixed Approach Retrospective Analyses of Suicide and Suicidal Ideation for Brand Compared with Generic Central Nervous System Drugs*. Drug Safety. (2018) **41**(4):363. doi: 10.1007/ s40264-017-0624-0. PMID: 29196989.
- Desai, R. J., Sarpatwari, A., Dejene, S., Khan, N. F., Lii, J., Rogers, J. R., Dutcher, S. K., Raofi, S., Bohn, J., Connolly, J., Fischer, M. A., Kesselheim, A. S., and Gagne, J. J. *Differences in Rates of Switchbacks After Switching From Branded to Authorized Generic and Branded to Generic Drug Products: Cohort Study*. BMJ. (2018) **361**:k1180. doi:10.1136/bmj.k1180. PMID: 29615391.
- Fang, L., Kim, M. J., Li, Z., Wang, Y., DiLiberti, C. E., Au, J., Hooker, A., Ducharme, M. P., Lionberger, R., and Zhao, L. *Model-Informed Drug Development and Review for Generic Products: Summary of FDA Public Workshop*. Clin Pharmacol Ther. (2018) **104**(1):27–30. doi: 10.1002/ cpt.1065. PMID: 29603191.
- Gagne, J., Popovic, J., Nguyen, M., Sandhu, S., Greene, P., Izem, R., Jiang, W., Wang, Z., Y, Z., Petrone, A., Wagner, A., and SK, D. *Evaluation of Switching Patterns in Fda's Sentinel System: A New Tool to Assess Generic Drugs*. Drug Safety. (2018) **41**(12):1313–1323. doi: 10.1007/s40264-018-0709-4. PMID: 30120741.
- Gagne, J., Polinski, J., Jiang, W., Dutcher, S., Xie, J., Lii, J., Fulchino, L., and Kesselheim, A. *Correction To: Outcomes Associated with Generic Drugs Approved Using Product-Specific Determinations of Therapeutic Equivalence*. Drugs. (2018) **78**(4):523–4. doi: 10.1007/s40265-0180890-x. PMID: 29520639.
- Gong, X., Hu, M., and Zhao, L. *Big Data Toolsets to Pharmacometrics: Application of Machine Learning for Time-to-Event Analysis*. Clin Transl Sci. (2018) **11**(3):305–311. doi: 10.1111/cts.12541. PMID:29536640.
- Hansen, R., Qian, J., Berg, R., Linneman, J., Seoane-Vazquez, E., Dutcher, S., Raofi, S., Page, C., and Peissig, P. *Comparison of Clinical Outcomes Following A Switch from A Brand to an Authorized Vs Independent Generic Drug*. Clin Pharmacol Ther. (2018) **103**(2):310–317. doi: 10. 1002/cpt.591. PMID: 27981563.
- Hu, M., Jiang, X., Absar, M., Choi, S., Kozak, D., Shen, M., Weng, Y., Zhao, L., and Lionberger, R. *Equivalence Testing of Complex Particle Size Distribution Profiles Based on Earth Mover's Distance*. AAPS J. (2018) **20**(3):62. doi: 10.1208/s12248-018-0212-y. PMID: 29651627.

Presentations

- Dutcher, S. *Use of Regulatory Science Research to Support Post-Marketing Surveillance of Generic Drug Products*. Presentation at Leveraging Quantitative Methods and Modeling to Modernize Generic Drug Development and Review. White Oak, MD, Oct. 3, 2017.
- Hu, M. *Prediction of the First ANDA Submission for NCE Utilizing Machine Learning Methodology*. Presentation at Leveraging Quantitative Methods and Modeling to Modernize Generic Drug Development and Review. White Oak, MD, Oct. 3, 2017.

- Hu, M. *Predictive Analysis of First ANDA Submission for NCEs Based on Machine Learning Methodology*. Presentation at Drug Information Association (DIA) Annual Meeting. Boston, MA, June 24, 2018.
- Hu, M. *Equivalence Testing of Complex Particle Size Distribution Profiles Based on Earth Mover's Distance*. Presentation at Complex Generic Drug Product Development Workshop. Silver Spring, MD, Sept. 12, 2018.
- Zhao, L. *Big Data Application in Life Sciences*. Presentation at Chinese Biopharmaceutical Association. Rockville, MD, Feb. 3, 2018.

Posters

- Desai, R., Gopalakrishnan, C., Dejene, S., Sarpatwari, A., Levin, R., Dutcher, S., Wang, Z., Wittayanukorn, S., Franklin, J., and Gagne, J. *Comparative Outcomes of Treatment Initiation with Brand Versus Generic Warfarin: A Medicare Cohort Study*. Poster Presentation at 34th International Conference on Pharmacoepidemiology & Therapeutic Risk Management. Prague, Cech, Aug. 22, 2018.
- Gong, X., Hu, M., and Zhao, L. *Big Data Toolsets to Pharmacometrics: Application of Machine Learning for Time-to-Event Analysis*. Poster Presentation at ASCPT Annual Meeting. Orlando, FL, Mar. 22, 2018.
- Gong, X., Hu, M., Liu, J., and Zhao, L. *Revealing Association Between Kinases and Adverse Events for Small-Molecule Tyrosine Kinase Inhibitors Using Machine Learning Method*. Poster Presentation at ASCPT Annual Meeting. Orlando, FL, Mar. 22, 2018.
- Hu, M., Jiang, X., Absar, M., Choi, S., Shen, M., Weng, Y., Zhao, L., and Lionberger, R. *A Statistical Bioequivalence Approach Based On Earth Mover's Distance for Equivalence Testing of Particle Size Distribution*. Poster Presentation at AAPS Annual Meeting. San Diego, CA, Nov. 12, 2017.
- Hunt, L., Murimi, I., Scharfstein, D., Mojtabai, R., Segal, J. *Assessing Therapeutic Equivalence of Brand and Generic Drugs Using Observational Data*. Poster Presentation at Atlanta Causal Inference Conference. Pittsburgh, PA, May 22, 2018.
- Hunt, L. Murimi, I., Scharfstein, D., Mohtabai, R., Varadhan, R., Segal, J. *Overcoming Temporal Confounding in the Assessment of Therapeutic Equivalence of Brand and Generic Drugs*. Poster Presentation at International Society for Pharmacoepidemiology. Lyon, France. August 22, 2018.