Patterns

CellPress
OPEN ACCESS

Preview

Bringing FAIRness to FDA data through visualization

Juan Espinoza Salomon[1,2,*]
[1]Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
[2]Northwestern University Feinberg School of Medicine, Chicago, IL, USA
*Correspondence: jespinozasalomon@luriechildrens.org

https://doi.org/10.1016/j.patter.2023.100755

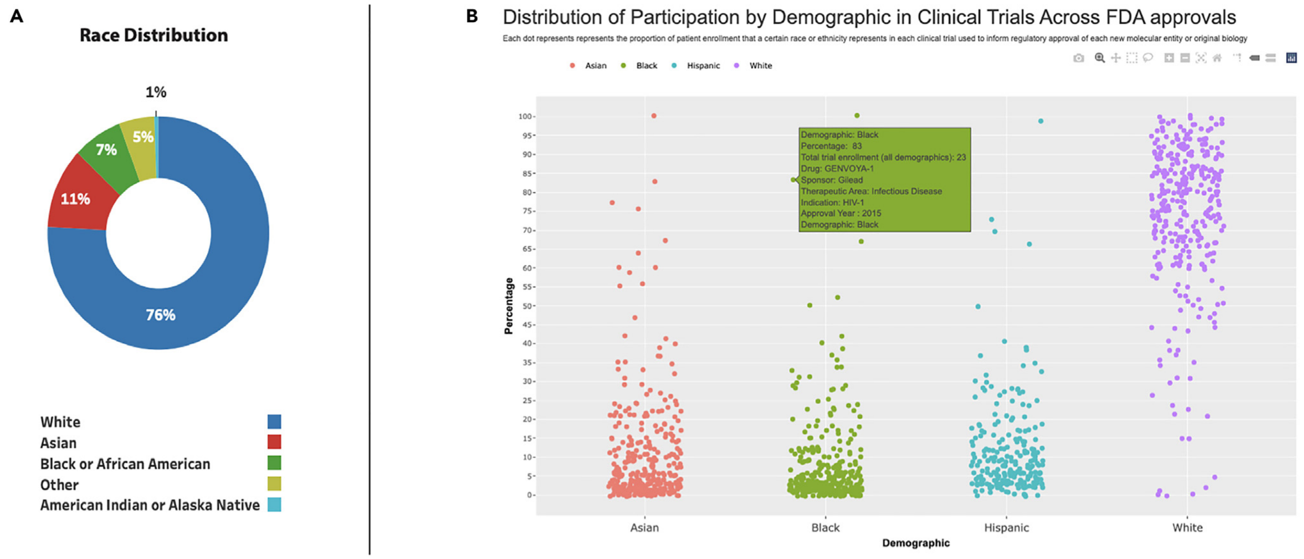
Improving representation and inclusion in clinical trials for new medicinal products has been a high priority for federal agencies for over 2 decades, but data to evaluate progress have been difficult to access. In this issue of *Patterns*, Carmeli et al. provide a novel approach to aggregating and visualizing existing data to improve transparency and research.

Women, underrepresented minorities (URMs), older adults, and children were *de jure* or *de facto* excluded from clinical research for decades. Even when explicit barriers were addressed, a legacy of harm and misconduct has created significant mistrust between URMs and the research community, which combined with systemic racism, has kept their participation in clinical trials low.[1] Since at least the NIH Revitalization Act of 1993, which mandated the appropriate inclusion of minorities in all National Institutes of Health-funded research, making clinical trials more inclusive and representative of the US population has been a priority across several Health and Human Services agencies, though progress in the 30 years since has been slow at best.[2,3]

Today, NIH, FDA, and others have initiatives intended to increase the number of individuals from URMs who participate in and benefit from clinical trials. One such initiative is the Drug Trial Snapshots (DTS), launched in 2015 by the FDA.[4] DTS is focused on public reporting and accountability and provides information on the diversity of participants in clinical trials for drugs and biologics (collectively referred to as medicinal products). The data are presented on the FDA website and are available for individual medicinal products approved since 2015 and in aggregate for each calendar year. In general, DTS are an important step forward in transparency and are written to be accessible to a general audience. However, this resource suffers from a number of important limitations. In the pursuit of delivering data that are accessible to a lay audience, they have failed in adhering to FAIR (findable, accessible, interoperable, reusable) data principles.[5] There are essentially no metadata nor machine-readable data, and the decision to release it as web text and PDFs means that the data are almost impossible to reuse without significant effort.

Enter Carmeli et al., who discuss in this issue of *Patterns* the interactive data visualization tool they have developed to explore DTS data.[3] The research team manually scraped data from all 339 medicinal products included in the DTS from 2015 to 2021, to create a structured, machine-readable database, and built the Data Visualization Explorer on top of it. Where data were missing, they were gathered from other FDA sources, and the dashboard is enriched with incidence data from the CDC and SEER for certain therapeutic areas (infectious diseases and oncology). In doing so, they made a number of valuable contributions: (1) created a structured DTS dataset that is available to download from the visualization tool website; (2) gave the end user significant control to explore the data across a number of dimensions, including race, ethnicity, gender, age group, sponsor, and therapeutic area; and (3) provided distributions and medians for much of the data, which are often more helpful than averages when group distributions are skewed, as is often the case with the demographics of clinical trials participants. Once built, as a demonstration Carmeli et al. pose and answer three questions that would be difficult to answer from DTS but are relatively straightforward with the data visualizer (Figure 1). They are able to show that almost no progress has been made in representation and inclusion over time, and across some therapeutic areas, the issue of underrepresentation is particularly marked given the incidence of diseases in those same communities. Because they have both aggregate and discrete trial data, they are also able to find key insights, like 25% of all trials enrolled ≤1% Black participants or that the least diverse trials are often also the smallest trials. Overall, Carmeli et al.'s work is interesting and genuinely useful, providing an opportunity to generate new questions about representation in clinical trials for medicinal products. The mark of good design in this project is that, for an end user, the tool feels immediately obvious, and you are left wondering why the FDA didn't build the DTS like this from the beginning.

The authors point out a number of limitations of the DTS data (and therefore, their findings) that are worth mentioning, besides the technical ones discussed above. There is a high degree of data missingness and issues with data quality, which has been noted about FDA data sources by other authors.[6,7] Collecting and aggregating race and ethnicity data can be inherently challenging, particularly when data are collected in other countries with different social constructs of these concepts, or when data are sourced from other agencies with different working nomenclature.[1] There is no DTS equivalent for medical devices, the third critical class of medical products regulated by the FDA, and so we have little insight into representation and inclusion in MedTech trials. Carmeli et al. have shown us what is possible in terms of meaningful transparency with public data, and their main limitations are the source. To achieve the goals of the US Department



Distribution of Participation by Demographic in Clinical Trials Across FDA approvals

Each dot represents represents the proportion of patient enrollment that a certain race or ethnicity represents in each clinical trial used to inform regulatory approval of each new molecular entity or original biology

Demographic: Black
Percentage: 83
Total trial enrollment (all demographics): 23
Drug: GENVOYA-1
Sponsor: Gilead
Therapeutic Area: Infectious Disease
Indication: HIV-1
Approval Year : 2015
Demographic: Black

Figure 1. Distribution of participants by race and ethnicity across clinical trials
Shown in the (A) *2015–2019 Drug Trials Snapshots Summary Report* compared to a screenshot from the (B) Data Visualization Explorer by Carmeli et al. The Data Visualization Explorer provides significantly more information, has interactive filters, and includes mouse-over text to provide contextual trial data about each individual data point.

of Health and Human Services' (HHS) 2023–2026 Evidence-Building Plan,[8] the FDA and other agencies will need to further modernize their data infrastructure and policies, both for how they collect and how they share data. This will need to include mandating common data elements, implementing data-quality frameworks, and conforming to common data models.[9] A great place to start would be for government agencies to follow the same data management and sharing policy they require of the researchers they fund.[10] According to NIH, "Data sharing enables researchers to rigorously test the validity of research findings, strengthen analyses through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery." I would be inclined to agree.

ACKNOWLEDGMENTS

Dr. Espinoza's effort is supported by the Food and Drug Administration under award number P50FD006425 for The West Coast Consortium for Technology & Innovation in Pediatrics (PI: Espinoza). The funding sources had no involvement in the development of this manuscript or in the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the FDA.

DECLARATION OF INTERESTS

Dr. Espinoza receives grant funding from FDA, NIDDK, NCATS, and NICHD, none of whom participated in the development of this manuscript or in the decision to submit the paper for publication. He is also a paid consultant for Sanofi. Sanofi played no role in the design, execution, analysis, or write up of this work. Sanofi did not play a role in the decision to publish this manuscript and had no editorial input.

REFERENCES

1. Cook, L., Espinoza, J., Weiskopf, N.G., Mathews, N., Dorr, D.A., Gonzales, K.L., Wilcox, A., and Madlock-Brown, C.; N3C Consortium (2022). Issues With Variability in Electronic Health Record Data About Race and Ethnicity: Descriptive Analysis of the National COVID Cohort Collaborative Data Enclave. JMIR Med. Inform. *10*, e39235. https://doi.org/10.2196/39235.
2. Chen, M.S., Jr., Lara, P.N., Dang, J.H.T., Paterniti, D.A., and Kelly, K. (2014). Twenty years post-NIH Revitalization Act: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual: renewing the case for enhancing minority participation in cancer clinical trials. Cancer *120 Suppl 7*, 1091–1096. https://doi.org/10.1002/cncr.28575.
3. Carmeli, A.B., Meloney, L., and Bierer, B.E. (2023). Data visualization explorer: A tool for participant representation in pivotal trials of FDA-approved medicinal products. Patterns *4*, 100713.
4. U.S. Food and Drug Administration. Drug Trials Snapshots. https://www.fda.gov/drugs/drug-approvals-and-databases/drug-trials-snapshots.
5. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data *3*, 160018. https://doi.org/10.1038/sdata.2016.18.
6. Lee, S.J., Cho, L., Klang, E., Wall, J., Rensi, S., and Glicksberg, B.S. (2021). Quantification of US Food and Drug Administration Premarket Approval Statements for High-Risk Medical Devices With Pediatric Age Indications. JAMA Netw. Open *4*, e2112562. https://doi.org/10.1001/jamanetworkopen.2021.12562.
7. Duggirala, H.J., Tonning, J.M., Smith, E., Bright, R.A., Baker, J.D., Ball, R., Bell, C., Bright-Ponte, S.J., Botsis, T., Bouri, K., et al. (2016). Use of data mining at the Food and Drug Administration. J. Am. Med. Inform. Assoc. *23*, 428–434. https://doi.org/10.1093/jamia/ocv063.
8. U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation. FY 2023-2026 HHS Evidence-Building Plan. https://aspe.hhs.gov/reports/fy-2023-2026-hhs-evidence-building-plan.
9. Espinoza, J.C. (2021). The Scarcity of Approved Pediatric High-Risk Medical Devices. JAMA Netw. Open *4*, e2112760. https://doi.org/10.1001/jamanetworkopen.2021.12760.
10. National Institutes of Health. "Data Management & Sharing Policy." https://sharing.nih.gov/data-management-and-sharing-policy.