RWD26

Tokenization-Linked Social Determinants of Health (SDoH) Data: A Gateway to Enhanced Understanding of Rare Disease Clinical Trial Populations

Leah Broder, MPH[1], Ryan Moog, MBA[2]

[1]Socially Determined, Los Angeles, CA; [2]Datavant, Lake Lotawana, MO

Introduction

Social determinants of health (SDoH)—including socioeconomic status, education, and access to healthcare—significantly influence clinical trials, especially in rare diseases, and affect patient recruitment, retention, and outcomes. Despite the potential value of inclusion of these detailed data – beyond race and ethnicity - they are typically overlooked in trial designs due the complexity of accurate and consistent collection in traditional clinical trial settings.

The emergence of reliable, normalized social determinants of health data sources, collected from federal, state and local data combined with commercial and consumer data, can open opportunities to use these data in clinical trial settings. Additionally, the ability to link these data at the patient level, while still maintaining de-identification, via emerging Privacy-Preserving Record Linkage (PPRL) or "tokenization" technologies opens opportunities to passively add rich SDoH data to other primary-collected clinical trial data.

This poster examines a methodology for creating PPRLs and linking validated SDoH measures on clinical trials for rare diseases, proposing strategies to improve trial effectiveness and inclusivity. By integrating SDoH into clinical development, we aim to enhance data robustness and promote equitable healthcare outcomes for diverse patient populations.

Objectives

- Discuss the ability to generate Privacy-Preserving Record Linkages (PPRL) in a rare disease clinical trial for linkage to Real-World Data (RWD) sources such as SDoH data
- Validate overlapping patient-level data able to be passively included into clinical trial data sets in a rare disease trial
- Demonstrate the types of data available in real-world SDoH data sources able to be linked with clinical trial data from primary data sources like electronic data capture and patient outcome measurement tools
- Discuss considerations for including these components into future trials, including ethical, privacy and adoption considerations

Methodology

- Trial patients in an ongoing Phase IV interventional study of spinal muscular atrophy patients were consented for the creation of PPRLs (unique, encrypted, de-identified patient keys that enable for patient-level linkages of other data sets with PPRLs) to support complementary analyses of clinical trial data with relevant real-world data
- PPRLs/Tokens are generated at clinical sites through an application converting basic PII components (Subject ID, First Name, Last Name, Date of Birth, Gender at Birth) to encrypted, certified de-identified tokens. Tokens generated for this trial are specifically encrypted with a unique key for this trial *(Figure 1)*
- PPRLs/Tokens are also created for a Social Determinants of Health (SDoH) data set in a separate encryption scheme using PII held by the SDoH data vendor *(Figure 1), who* offers normalized social risk data organized by visibility into both communities and individuals through actionable SDoH insights. Specifically, the social determinants dataset encompasses community and individual-level, modeled social risk scores, drivers and related SDoH metrics across seven domains (financial strain, food insecurity, housing instability, transportation barriers, health literacy, digital landscape and social connectedness)
- Prior to creation of a new linked data set of both trial and real-world data, tokens from both an SDoH data set as well as the trial cohort are uploaded to a web-based portal to confirm availability of data on adult clinical trial patients

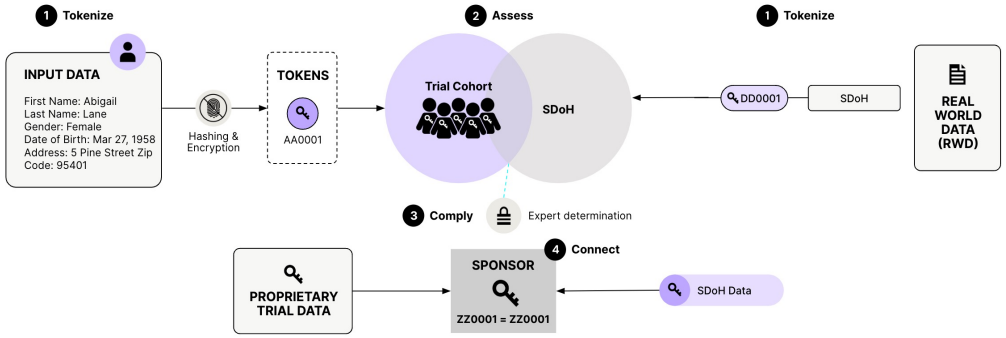

Figure 1. Overview of PPRL/token creation and subsequent linkage of clinical trial and SDoH data

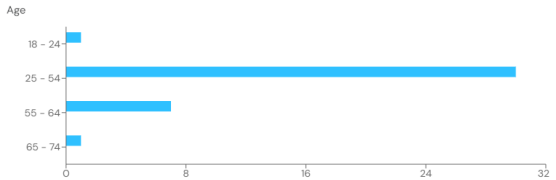

Figure 2. Age distribution of overlapping adult trial patients with linked SDoH available

References

- Bernstam EV, Applegate RJ, Yu A, Chaudhari D, Liu T, Coda A, Leshin J. Real-World Matching Performance of Deidentified Record-Linking Tokens. Appl Clin Inform. 2022 Aug;13(4):865-873. doi: 10.1055/a-1910-4154. Epub 2022 Jul 27. PMID: 35896508; PMCID: PMC9474266.
- Mirel LB, Resnick DM, Aram J, Cox CS. A methodological assessment of privacy preserving record linkage using survey and administrative data. Stat J IAOS. 2022 Jun 7;38(2):413-421. doi: 10.3233/sji-210891. PMID: 35910693; PMCID: PMC9335262.

Results

- PPRLs/Tokens were successfully generated on all US trial patients who consented to the sub-study
- PPRLs/Tokens were also successfully generated on all 317 million patients in a commercially-available SDoH data set
- Overlap of the trial cohort determined that all US adults in the clinical trial cohort have SDoH data available in a normalized SDoH data set. *(Figure 2 demonstrates age distribution of overlapping patients – patients known to exist in both the trial cohort and the SDoH data set)*
- Normalized de-identified data domains now available via additional privacy-preserving linkage on these patients include: financial data, food insecurity, housing status, transportation availability, health literacy, digital and social connectivity
- Precision of linked-data matches on de-identified PPRLs/tokens have been extensively studied and are able to be used with high confidence in consented clinical settings

Conclusions

- This overlap confirms an initial hypothesis that SDoH data are consistently available on US adult populations – even patients enrolled in rare disease clinical trials
- Ultimately, a normalized SDoH data set can yield several benefits for any trial population, but uniquely impactful use cases for rare disease trials include:
- *Using linked employment status data in concert with trial outcomes to support assessment of adherence and affordability of care options, to support tailoring financial support and intervention strategies*
- *Using healthcare access information linked to trial retention and adherence data to identify gaps in care and drive initiatives to improve access to necessary medical expertise*
- *Using health literacy data linked to trial outcome data to understand impacts and opportunities to improve patient education and adherence*
- Despite small data set sizes in rare disease, an expert determination-centered approach will enable investigators to link and remediate integrated trial and SDoH data