

Objective Fit-4-Purpose Assessment of Real-World Data for Evidence Generation in Type 2 Diabetes Mellitus: A Trial Tokenization Approach

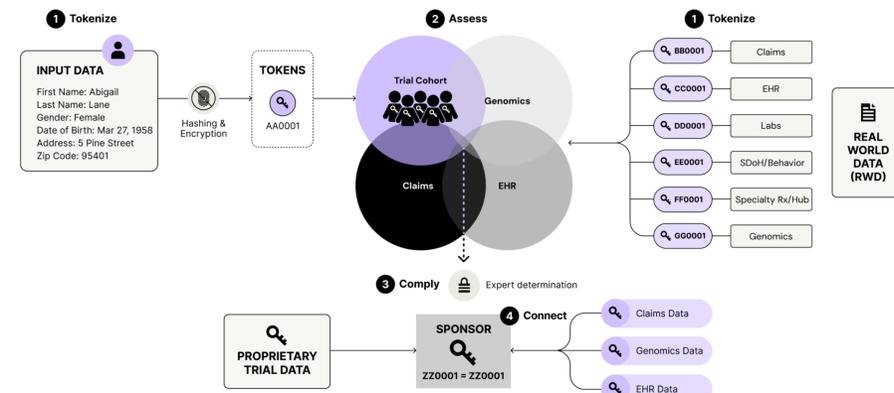
Aaron Kamaau¹, Amanda Shields², Craig G. Parker³, Ryan Moog⁴, Kathleen M. Gavin⁵

¹Navidence LLC, Bountiful, UT, USA; ²Navidence LLC, Chapel Hill, NC, USA; ³Navidence LLC, Salt Lake City, UT, USA; ⁴Datavant, Lake Lotawana, MO, USA; ⁵Datavant, Denver, CO, USA

Introduction

- The integration of real-world data (RWD) strategies together with clinical trials has many advantages to accelerate real-world evidence (RWE) generation.
- Tracking trial participants through RWD to generate RWE after a clinical trial is a growing strategy for understanding the long-term effectiveness and safety of medical intervention.
- One effective approach for this is "trial tokenization," using privacy preserving record linkage (PPRL) to connect clinical trial data with de-identified RWD at the patient level (Figure 1).
- One of the many use cases of this methodology is post-trial health outcomes assessment.
- Ensuring the underlying real-world data (RWD) is relevant and reliable is a critical step in the project planning process.
- Objective:** Develop an objective process to evaluate RWD sources for a Type 2 diabetes mellitus (T2DM) example use case (e.g., tracking weight and cardiovascular outcomes over time), which is critical to ensure selection of RWD qualified as relevant and reliable to the specific research questions.

Figure 1. Trial Tokenization is the use of privacy preserving record linkage to connect clinical trial patients/data with de-identified RWD.



Results

- 15 standards-based CODEfs were developed:
 - 5 determined cohort and subcohort eligibility, including age, T2DM (Cohort 1), overweight or obesity (Cohort 2), and weight-related illnesses (Subcohorts 2 and 3)
 - 10 described key outcomes, including, height, weight, BMI, Hb1A1c, glucose, and MACCE safety events.
- The 15 CODEfs encompassed 18 value sets including 3,269 values (Table 1). CODEfs can include more than 1 value set.

Table 1. Number of value sets and values used for CODEfs.

	Medication (NDC)	Diagnosis (ICD-10-CM)	Medication (name)	Labs (LOINC)	Physiological measure	Demographic measure
# values	2813	385	63	4	3	1
# value sets	1	8	1	4	3	1

- 39 study-relevant data variables were assessed for availability (e.g. demographics, vitals, medical history, treatments, and labs)
- 6 data sources completed partial or full assessment; representing RWD including inpatient and outpatient settings, EHR, claims, medications, procedures, and labs.

Methods

- Computable operational definitions (CODEfs) were modeled for key eligibility criteria and outcomes and a use case-specific assessment plan was developed.
- Candidate RWD sources were identified and assessed on sample size and demographics, as well as availability and reliability of outcomes, specifically weight and cardiovascular outcomes.
- An abbreviated overview of this process is presented in Figure 2.

Figure 2. Overview of CODEf modeling and use case specific assessment plan.

Element	Conceptual Definition	Operational Definition	Justifications, Algorithms, Limitations, & Notes
Inclusion 3: Diagnosis of diabetes mellitus, type 2 (T2DM)	Diagnosis of diabetes mellitus, type 2 (T2DM) AND Treatment with antihyperglycemic therapy	≥ 1 Diagnosis Record from Diabetes Mellitus, Type 2, (CCI) Diagnoses (any time prior to the index date AND any encounter type AND any diagnosis position) AND ≥ 1 Medication Record from Antihyperglycemic Therapies any time prior to the index date	Limitations Does not take into account duration of therapy. This broad definition may have the potential for false positives
Source Wording: Diagnosis of T2DM			
Exclusion 1: Diagnosis of diabetes mellitus, type 1 (T1DM)	Diagnosis of diabetes mellitus, type 1 (T1DM)	≥ 1 Diagnosis Record from Diabetes Mellitus, Type 1, (CCI) Diagnoses (any time prior to the index date AND any encounter type AND any diagnosis position)	Limitations This broad definition may have the potential for false positives
Endpoint: Outcome: BMI	Body mass index (BMI)	Body Mass Index any value kg/m2 within the 36 months (3 years) after the index date	
Endpoint: Outcome: Safety: Ischemic Stroke	Diagnosis of ischemic stroke	≥ 1 Diagnosis Record from Ischemic Stroke Diagnoses (within the 36 months (3 years) after the index date AND any encounter type AND any diagnosis position)	Limitations This broad definition may have the potential for false positives

Adults	Of the total patients from cell D7, # of patients who meet the following operational definition: Age ≥ 18 years old on the query date
Obesity OR Overweight	Of the total patients from cell D10, # of patients who meet the following operational definition: (Body Mass Index ≥ 30 kg/m2 within the 12 months prior to the index date AND most proximal documented value to the index date) OR ≥ 1 Diagnosis Record from Obesity Diagnoses (refer to Obesity_Dx (ICD-10-CM) tab) within the 12 months prior to the index date AND any encounter type AND any diagnosis position AND most proximal documented value to the index date) OR Body Mass Index = 27.0-29.9 kg/m2 within the 12 months prior to the index date AND most proximal documented value to the index date
Diagnosis of diabetes mellitus, type 2 (T2DM)	Of the total patients from cell D11, # of patients who meet the following operational definition: ≥ 1 Diagnosis Record from Diabetes Mellitus, Type 2, (CCI) Diagnoses (refer to T2DM_CCI_Dx (ICD-10-CM) tab) any time prior to the index date AND any encounter type AND any diagnosis position

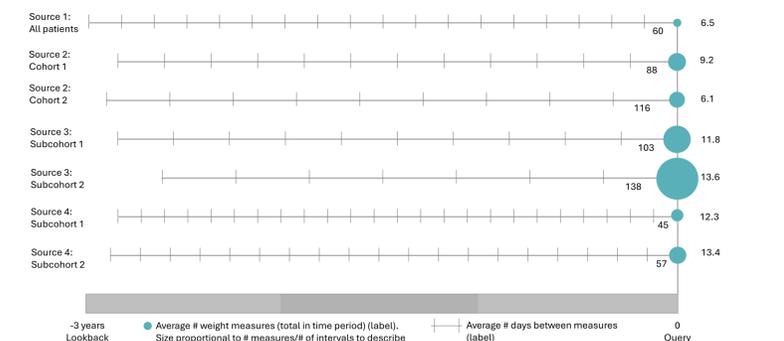
DATA VARIABLES (all require associated dates)	Is this data available?	Structured vs. Unstructured	What Coding Schemes Are Used?
Socioeconomic Status			
Employment/Occupation			
Height *			
Weight *			
Body Mass Index (BMI) *			
Waist Circumference			
Glycemic control (HbA1c, FPG) *			
Liver function (LDL, HDL, triglycerides, cholesterol) *			
Renal function			
Haematology (e.g. eosinophils, neutrophils, basophils, monocytes, lymphocytes)			
Biochemistry			
Other (e.g., hs-CRP, IL-6, Fibrinogen)			

Label	Code
Age	age
Body Mass Index	body_mass_index
Diagnosis Record	diagnosis_record
Height	height
Medication Record	medication_record
Weight	weight

	# of Unique Patients	% of Total
Physiological Data		100%
Height Data		
Weight Data		
BMI Data		
Lab Data		
Hemoglobin A1c/Hemoglobin total in Blood % (HbA1c; LOINC code: 4548-4)		
Glucose [Mass/volume] in Serum or Plasma (LOINC code: 2345-7)		
Safety Outcomes		
Diagnosis Record from Cardiovascular Diagnoses (refer to CVD_Dx (ICD-10-CM) tab)		

- 369,000–2.5 million patients meeting the T2DM cohort criteria were found across data sources
- Weight was measured at frequent intervals, averaging ~2.0 - 4.5 times a year (w/ avg of ~90-140 days between measurements), across data sources and patient populations (Figure 3).
- Data origin varied across sources—patterns of weight measurement reflect this variance, with larger, less frequent clusters of measurements documented in inpatient EHR data. Average number of days between measurements exceeds the average number of documented measures, suggesting unobserved clustering that warrants further evaluation.

Figure 3. Patterns of available weight measurements over time



- MACCE outcomes were documented in all assessed sources, though results varied substantially. Depending on the data source utilized, for a given subcohort, 0.08%-10.2% of patients had evidence of an acute MACCE outcome within 3 years prior to the query date.

Conclusion

Objective assessment using standards-based CODEfs ensures the reliability of data to support post-trial RWE generation. Outpatient and inpatient data are both important to assess for T2DM patient populations. Linking strategies may be useful to provide a more complete view.

Future Directions: This provides stakeholders like regulatory agencies the confidence to consider such data as evidence and supports the use of trial tokenization for future research.

Abbreviations: CODEfs: Computable operational definitions; EHR: electronic health record; MACCE: major adverse cardiac and cerebrovascular event; PPRL: privacy preserving record linkage; RWD: real-world data, RWE: real-world evidence; SDoH: social determinants of health, T2DM: Type 2 diabetes mellitus

Contact Information: Aaron Kamaau - aaron@navidence.com