

# Evaluation of Test Data in Distributed Research Networks: A Sentinel System Example



Emily C. Welch, MPH, Tiffany S. Woodworth, MPH, Talia J. Menzin, BS, Ting-Ying Jane Huang, BPharm, PhD

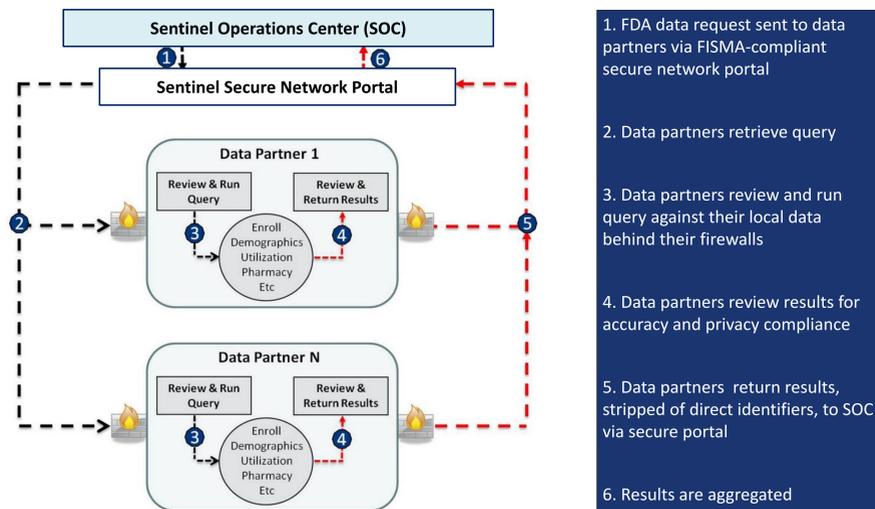
Dept. of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA

33rd INTERNATIONAL CONFERENCE ON PHARMACOEPIDEMOLOGY & THERAPEUTIC RISK MANAGEMENT, Palais de Congres de Montreal, Montreal, Canada, August 26-30, 2017

## BACKGROUND

- Distributed Research Networks (DRNs) have strengths over single data sources in identifying rare exposures and outcomes by executing distributed analytic programs in a standardized data structure using common data model (CDM)
- DRN users do not have direct access to data, requiring preliminary analysis to be performed in a test dataset to specify analytic program details (such as error identification in analytic programs and final model decisions) before program package distribution to data partners (Figure 1)
- Representativeness of the test dataset to the target DRN is important in ensuring the appropriateness of analytic decisions

Figure 1. Structure of Sentinel's Distributed Network



## OBJECTIVE

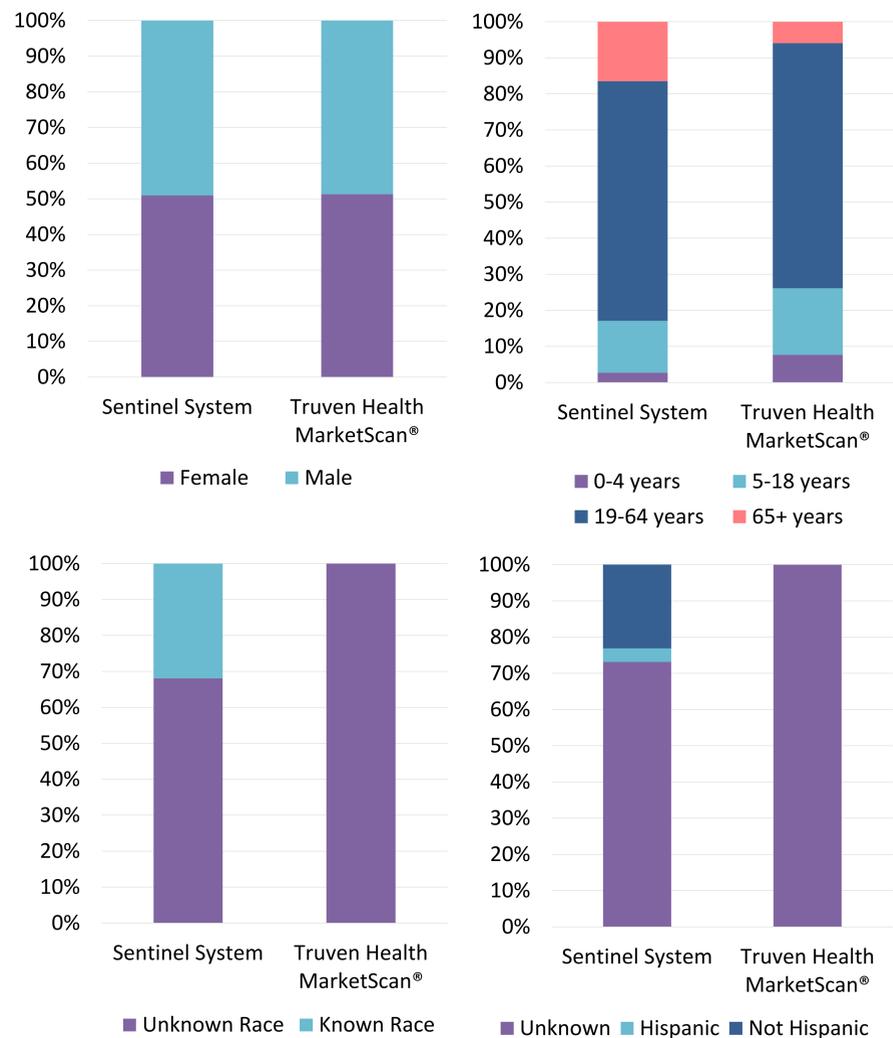
To assess the comparability of Sentinel Distributed Database (SDD) to its test dataset

## RESULTS

### Demographic Characteristics (Figure 2)

- Majority of members were 19-64 years of age in both the SDD (66%) and test dataset (68%)
- SDD had substantially more members aged 65 and older (17% vs 6%)
- Both datasets had even sex distributions and primarily unknown race and ethnicity information

Figure 2. Demographic Characteristics of Enrolled Members in the SDD versus Sentinel's Test Dataset



## METHODS

### Sentinel Distributed Database

- 4 national data partners using claims data
  - ~199.1 million members total
  - 86% of SDD
- 12 regional data partners using claims and integrated delivery systems
  - ~31.2 million members total
  - 14% of SDD
- 1 data partner with electronic medical record-based data was excluded
- Data remain at each data partner site with aggregate data available to Sentinel Operations Center

### Sentinel's Test Dataset

- 5% randomized sample of two data streams:
  - Truven Health MarketScan® Commercial Claims and Encounters Database
  - Truven Health MarketScan® Medicare Supplemental Database
- Transformed into the Sentinel CDM
- ~6.0 million members total

### Data Analysis

- We conducted descriptive analyses on member demographics, health plan enrollment, characteristics of diagnoses and prescription fills, and length of stay for inpatient hospital and non-acute institutional admissions

Table 1. Composition of SDD and Sentinel's Test Dataset

	Sentinel System	Truven Health MarketScan®
Data Start Date	1/1/2000	1/1/2010
Data End Date	12/31/2016	3/31/2016
Number of Members	230,354,010	6,051,872

### Health Plan Enrollment (Figure 3)

- Members in the test dataset had a shorter median (IQR) enrollment of 14 (27) months
  - 18 (30) – 30 (51) months for the 4 national data partners
  - 24 (36) – 58 (111) months for the 12 regional data partners
- Members in the test dataset were more likely to have both drug and medical coverage (73% vs 64%)

### Health Care Utilization (Figure 4)

- Diagnoses and prescription fills in the SDD and test dataset:
  - ~90% of diagnoses came from outpatient settings for both datasets
  - Median days' supply was 30 (2) and 30 (3) days
- Median length of stay in the SDD and test dataset:
  - Similar for inpatient hospital admissions with 4 (3) vs 3 (3) days
  - Different for non-acute institutional admissions with 13 (24) vs 6 (9) days

Figure 3. Health Plan Enrollment Characteristics of Enrolled Members in the SDD versus Sentinel's Test Dataset

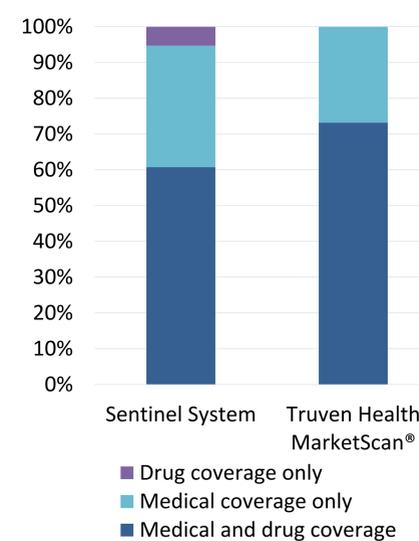
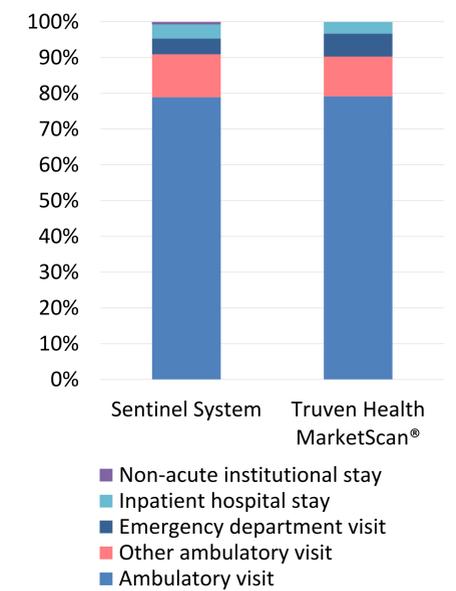


Figure 4. Health Care Utilization Characteristics of Enrolled Members in the SDD versus Sentinel's Test Dataset



## CONCLUSIONS

- Overall, the SDD and test dataset are comparable in most member demographics, as well as dispensing and health care encounter characteristics
- Before finalizing analytic specifications, SDD users should be aware of limitations on test results based on data with lower proportions of older adults and shorter enrollment periods, compared with SDD results
- Future upgrades of Sentinel's test datasets should address these issues to improve generalizability to SDD

## ACKNOWLEDGEMENTS/DISCLOSURES

- This work was supported by the U.S. Food and Drug Administration through the Department of Health and Human Services Contract # HHSF2232014000301
- Many thanks are due to data partners who provided data used in the analysis
- The authors have no conflicts of interest to disclose