

Building a Cohort of Transgender and Non-Binary Patients from the Electronic Medical Record

Paige N Hackenberger MD¹ & Lauren B Beach JD PhD^{2,3}, Mona Ascha MD¹, Natalie Luehmann MD⁴, Dylan Felt BA^{2,3}, Kareem Termanini MD¹, Christopher Benning MBA⁵, Danny Sama MBA⁵, Cynthia Barnard PhD MBA MSJS⁶, Sumanas W Jordan MD PhD¹

¹Division of Plastic and Reconstructive Surgery, ²Department of Medical Social Sciences, ³Institute for Sexual and Gender Minority Health and Wellbeing, ⁴Division of Breast Surgery, ⁵Northwestern Memorial Healthcare, Information Services, ⁶Northwestern Memorial Healthcare, Department of Quality

Background

Understanding the epidemiology of health outcomes, disease processes, and health disparities among transgender and non-binary (TGNB) patients is important. Unfortunately, questions to identify assigned sex at birth, sexual orientation, and gender identity (ASAB/SOGI) have been routinely excluded from demographic and health data collection efforts. Specifically, many healthcare organizations do not house structured data capture elements for ASAB/SOGI within the electronic medical record (EMR) system. This lack of documentation capacity results in the invisibility and misclassification of sexual and gender minority (SGM) patients.

Both the World Professional Association for Transgender Health (WPATH) and Fenway Institute have published recommendations for EMR developers, vendors, and users to standardize and collect ASAB/SOGI data within the EMR. Several studies have used EMR data to identify and study TGNB populations but these have mostly used only International Classification of Disease (ICD) codes and natural language processing (NLP).

Research Objectives

The present study evaluates the effectiveness of identifying a cohort of TGNB patients at our institution using several methods, including ICD code, text mining, and structured ASAB/SOGI data and their combinations. Both algorithm performance and ethical implications for each method are discussed

Methods

- Medical records flagged for review included:
- 1) An ICD-10 code indicating diagnosis of gender dysphoria
 - 2) Any record with completed ASAB/SOGI questions
 - 3) Any clinical note containing a TGNB-related keyword through structured query language filtering

The final TGNB population included patients who were flagged by any of these three methods (right). Positive predictive value (PPV) of each identification method was also calculated (Figure 1)

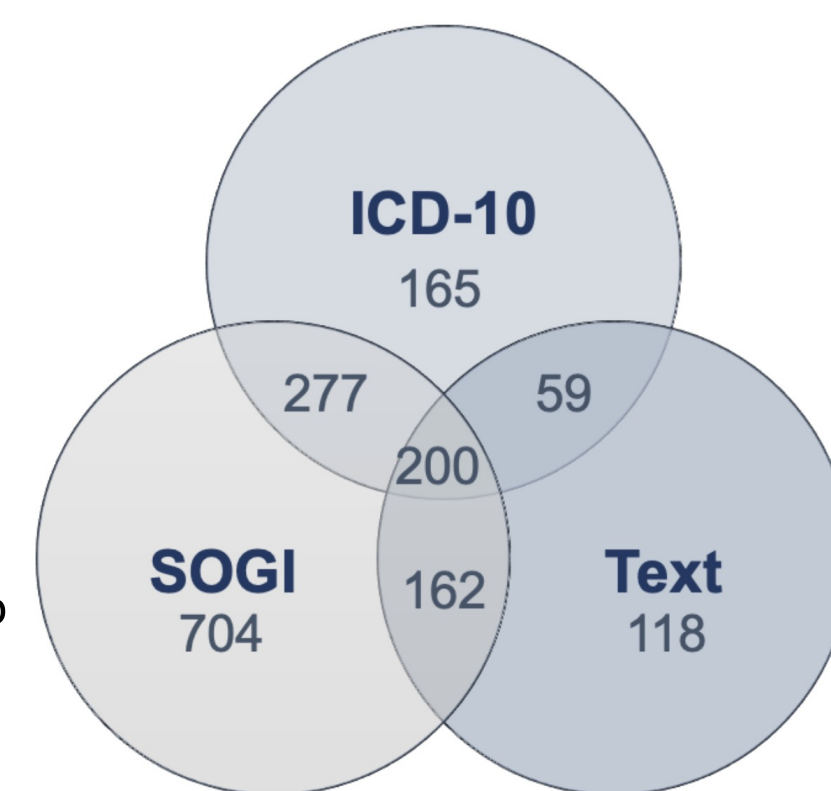


Figure 1. Positive Predictive Value of Each Method

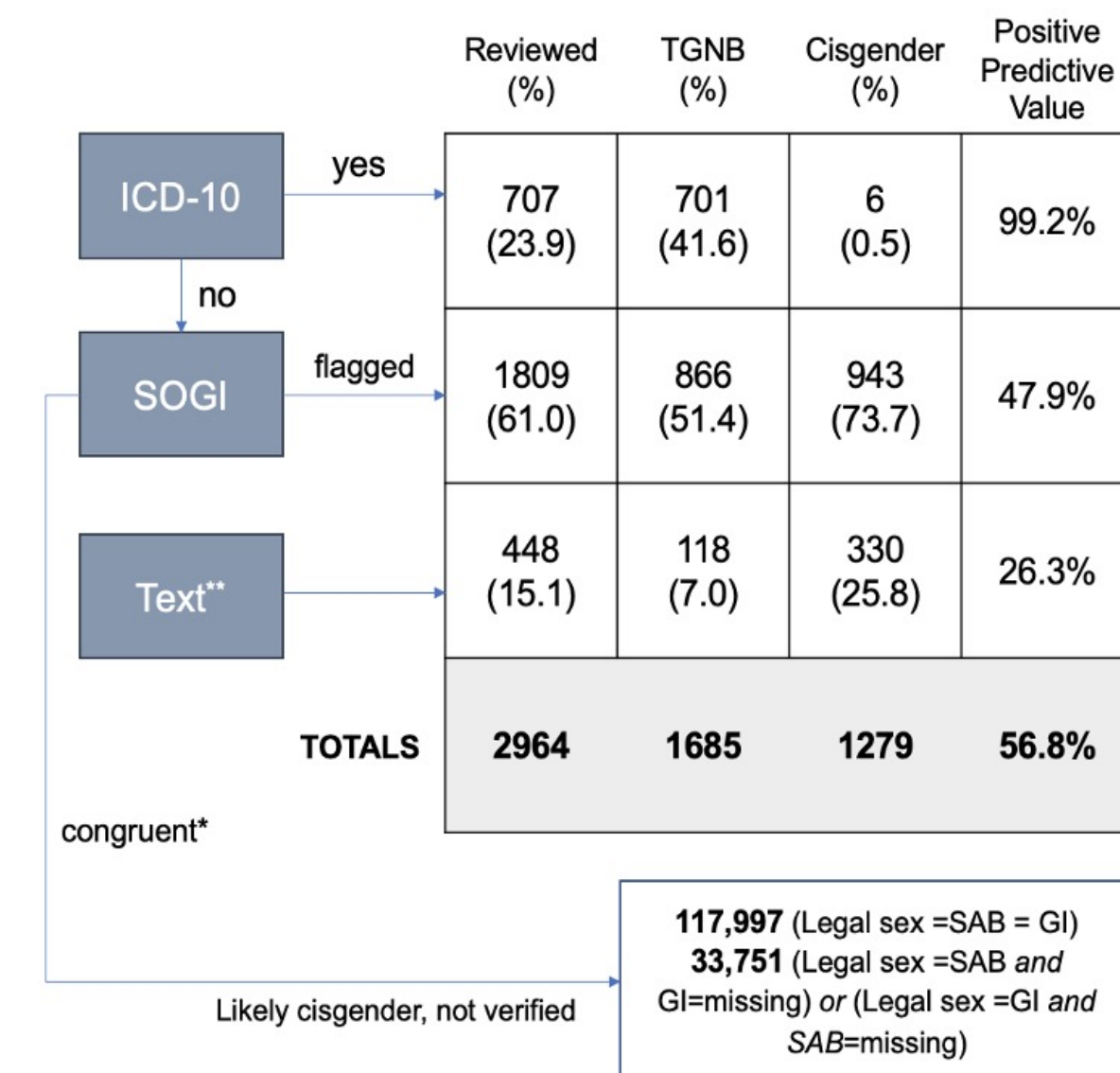


Table 1. Text Mining Results

Search term	All lines	All unique records	Transgender and/or non-binary	Cisgender	Positive predictive value
transgender or trans gender	1454	397	296	101	74.6%
gender dysphoria	787	196	181	15	92.3%
non-binary or nonbinary	163	42	28	14	66.7%
trans woman or transwoman	32	8	6	2	75.0%
trans man or transman	15	2	1	1	50.0%
MTF	512	187	5	182	2.7%
FTM	40	16	9	7	56.3%
gender fluid	23	9	6	3	66.7%
gender queer or genderqueer	4	1	3	0	100.0%
transsexual	17	6	4	2	66.7%
Totals	3052	866	539	327	62.2%

FTM = female-to-male; MTF = male-to-female

Results

1,530,154 distinct medical records of patients treated during the two-year period were queried. 154,712 records had relevant data. Of those, 2,964 met criteria for manual review and 1,685 patients were determined to be TGNB (714 transgender men, 662 transgender women, 307 non-binary individuals, and 2 people with unknown gender identities). Figure 1 illustrates the manual verification process with associated PPV of each identification method. Table 1 outlines the PPV of each specific text term searched.

A total of 1,279 false positive records were tabulated. The majority of false positives were identified when the gender identity field was indeterminate (n=554, 43.3%); that is, when the gender identity field stated “choose not to disclose” or “unknown.” After indeterminate patients were removed, the ASAB/SOGI and overall PPVs improved from 47.9% to 68.4% and from 56.8% to 69.6%, respectively.

Limitations

- Structured ASAB/SOGI data was only available for 10% of all patients in our EMR system
- We were unable to determine the source of ASAB/SOGI data entry, which could be via patients, registrars, providers, or external sources
- ASAB/SOGI documentation may not capture the full spectrum of gender identities, particularly among non-binary and gender diverse patients
- The percent of TGNB patients in our health system was 0.1%, beneath the expected TGNB population in Chicago of 0.5%, and the national estimate of 0.6%

Conclusions

- TGNB patients can be identified in the EMR through the combination of three methods to build a comprehensive cohort in the context of a health system
- Patient-driven ASAB/SOGI data capture enables effective identification of TGNB patient populations to facilitate quality improvement efforts, while respecting patient privacy and autonomy
- Providers should be encouraged to review answers to ASAB/SOGI questions with patients and conduct organ inventories only when such information is directly relevant for patient care. The medical relevance of this information should be respectfully communicated to patients