Challenges with Accuracy of Gender Fields in Identifying Transgender Patients in Electronic Health Records

Dinah Foer, MD[1] , David M. Rubins, MD, MS[1], Anthony Almazan, BA[2], Kit Chan, MD[1], David W. Bates, MD[1,2,3], and Ole-Petter R. Hamnvik, MBBChBAO, MMSc, MRCPI[1,2]

[1]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA; [2]Harvard Medical School, Boston, MA, USA; [3]Harvard Chan School of Public Health, Boston, MA, USA.

INTRODUCTION

Healthcare institutions must now demonstrate that their electronic health record (EHR) has the capacity to collect gender identity data [1], and gender-related demographic fields are promoted as best practice for implementation [2], [3]. However, the utility of gender identity fields in retrospectively identifying transgender patients, which is important for clinical care as well as population health management, remains unknown. This represents a critical first step in closing the gender identity gap in healthcare. In this study, we assessed the accuracy of gender identity fields in the EHR by comparing their accuracy with keyword search or problem list diagnoses, which have previously been used to identify transgender patients in the absence of gender identity fields [4], [5].

METHODS

Subjects with a primary care provider (PCP) in either of Partners' two primary academic teaching institutions between years 2015 and 2019 were included in this retrospective chart-based study (n = 1,504,423). At Partners, the Epic EHR (Epic Corporation, Verona, WI) has gender-related fields that include gender identity, legal sex, and sex assigned at birth. Patients were included as possibly transgender if any of the following were found: narrative text from physician notes with the words "transgender," "transsexual," "transvestite," "gender identity," "gender dysphoria," or "gender reassignment" [4] a problem list diagnosis mapped to an F64 ICD10 code ("Gender identity disorders"); a gender identity field entry listed as transgender; any discrepancy between the three gender identity fields when they were present and represented a binary gender classification. We performed chart reviews to validate the classification of patients as the gold standard for transgender classification and calculated the four components' test characteristics.

RESULTS

Of the 1,504,423 included patients, 13,424 (0.89%) were identified as possibly transgender. All patients had a completed legal sex field (100%). The "Sex assigned at birth" field was complete for 6539 patients (48.7%), and 6445 (48.0%) patients had a completed "Gender identity" field; there were 6166 (45.9%) patients with all three fields completed. Of the potentially transgender patients, 2109 (15.7%) were identified as such by at least two components (e.g., diagnosis and keyword). The remaining patients were only identified via a single method: keyword (89%), gender field discrepancy (14%), ICD codes (1.2%), and listed as transgender (5.1%).

A subset of patients (n = 324) were selected for chart review to validate transgender classification. Twenty-six patients (8.0%) were confirmed to be transgender on chart review (Fig. 1). Notably, 24 patients were identified as possibly transgender by gender-related field discrepancy alone and were all found to be cisgender on chart review. Finally, to test the extent to which the lack of the four components could distinguish between trans- and cisgender, we reviewed 100 patient charts that did not meet any component of the algorithm. None was found to be transgender on chart review.]–>

DISCUSSION

Accurate identification of transgender patients directly impacts clinical care, research, and policy. We found significant challenges in the use of gender-related fields alone for identifying transgender patients. These fields, including a "transgender" identifier, were less sensitive and less specific than other methods. In our large health system, narrative text keywords and ICD codes were more sensitive than any gender-related field-based method of identifying transgender patients (Table 1).

Variation in gender-related fields' use and accuracy present a significant barrier to improving clinical care for transgender patients, and further limits the promise of these fields for research endpoints [6]. Overall, the full set of gender-related
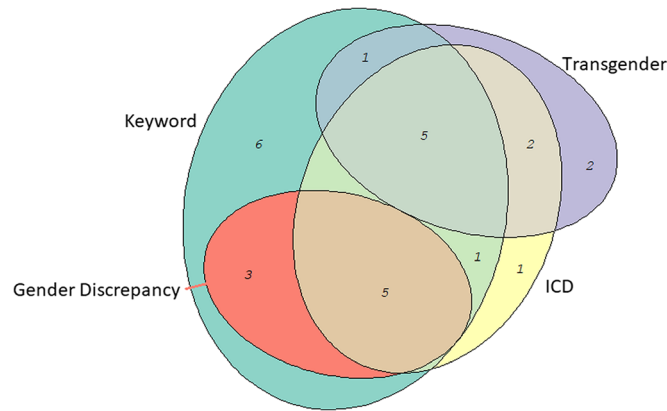
Figure 1. Transgender identification in the EHR. Euler diagram of patients confirmed to be transgender on EHR chart review by method.

Table 1 Test characteristics for transgender patient identification in the electronic health record

Method	Sensitivity	Specificity
Keyword	23%	35%
ICD*	4%	100%
Transgender	8%	99%
Gender discrepancy	0%	94%
ICD + transgender	8%	100%
Keyword + ICD	4%	100%
Keyword + transgender	4%	100%
Keyword + gender discrepancy	12%	99%
Keyword + ICD + transgender	19%	100%
Keyword + gender discrepancy + ICD	19%	100%
Any method	100%	27%

ICD international classification of diseases

fields were used in less than half of patients studied. Most notably, all patients identified by gender-related field discrepancy alone were found to be cisgender on chart review. Further research is needed to understand the reasons for this variability and the impact of EHR structure on the potential to capture patient-centered metrics.

This study was performed in a large integrated hospital system with a common commercial EHR though may not reflect institution-specific fields in other EHRs. In the context of increasing federal and institutional pressures to collect gender data, our findings uncover early challenges that have significant implications for patients as well as clinicians, administrators, and policy makers. Post-implementation review of gender identity fields in the EHR should be a critical first step towards improving care particularly for the transgender population.

Corresponding Author: *Dinah Foer, MD; Department of Medicine-Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115, USA (e-mail: dfoer@bwh.harvard.edu).*

REFERENCES

1. Centers for Medicare & Medicaid Services. Medicare and Medicaid Programs; Electronic Health Record Incentive Program–Stage 3 and Modifications to Meaningful Use in 2015 Through 2017; Final Rule. In: Department of Health and Human Services. Federal Register 2015; vol. 8(200), p. 1-195.
2. **Callahan EJ**, **Hazarian S**, **Yarborough M**, **Sanchez JP**. Eliminating LGBTIQQ Health Disparities: The Associated Roles of Electronic Health Records and Institutional Culture. The Hastings Center report. 2014;44 Suppl 4:S48-52.
3. **Deutsch MB**, **Buchholz D**. Electronic health records and transgender patients–practical recommendations for the collection of gender identity data. J Gen Intern Med. 2015;30(6):843-7.
4. **Roblin D**, **Barzilay J**, **Tolsma D**, **Robinson B**, **Schild L**, **Cromwell L**, et al. A novel method for estimating transgender status using electronic medical records. Annals of epidemiology. 2016;26(3):198-203.
5. **Blosnich JR**, **Cashy J**, **Gordon AJ**, **Shipherd JC**, **Kauth MR**, **Brown GR**, et al. Using clinician text notes in electronic medical record data to validate transgender-related diagnosis codes. Journal of the American Medical Informatics Association : JAMIA. 2018;25(7):905-8.
6. **Grasso C**, **McDowell MJ**, **Goldhammer H**, **Keuroghlian AS**. Planning and implementing sexual orientation and gender identity data collection in electronic health records. Journal of the American Medical Informatics Association. 2018;26(1):66-70.