

Using the OMOP CDM to identify gender discordant administrative codes

Kristine E Lynch PhD^{1,2}, Benjamin Viernes MPH^{1,2}, Jodie Katon PhD^{3,4}, Scott L DuVall PhD^{1,2}, Chad A Dorn⁵, Michael E Matheny MD^{5,6}

¹VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT; ²University of Utah, Department of Internal Medicine; ³Health Services Research and Development, Center of Innovation for Veteran-Centered and Value-Driven Care, US Department of Veterans Affairs Puget Sound Health Care System, Seattle, WA; ⁴Department of Health Services, University of Washington, Seattle, WA; ⁵Vanderbilt University Medical Center, Nashville, TN; ⁶VA Informatics and Computing Infrastructure, Tennessee Valley Healthcare System, Nashville, TN

Abstract

Administrative codes sourced from electronic medical record (EMR) data can potentially create bias in epidemiologic research due to data entry errors. The causes of errors are multifaceted, but incomplete knowledge of the existence of gender-specific coding schemas may be a factor. We sought to quantify the misalignment between records of gender-specific administrative codes and patient gender, and determine differences between a federal healthcare system and a private system. We used EMR data from both the Department of Veterans Affairs (VA) and Vanderbilt University Medical Center (VUMC) that had been transformed into the Observational Medical Outcomes Partnership Common Data Model. Using a combination of string matching and ancestor matching of SNOMED condition concepts, we identified female-male SNOMED dyads. We calculated frequency of gender-code misalignment on an instance level overall and for each clinical concept comparing differences between the two systems. Gender-code misalignment occurred for 7.7% of records in VA and 2.7% at VUMC. Differences in reimbursement/motivations for administrative coding as well as population demographics may explain differences between the two health care systems.

Background

Erroneous data entry errors in research databases are common and the etiologies are multifaceted.¹ It is reasonable to believe that some degree of error in administrative coding is due to providers being unaware that gender specific coding exists, but the extent to which this occurs is unknown. The vast majority of administrative codes are gender agnostic, meaning that they can be used to describe a patient regardless of their gender, but a minority are gender specific. Of the gender specific codes, some have a parallel code belonging to the opposite gender. For instance, there is no male equivalent code to SNOMED code 72892002: “Normal Pregnancy.” In contrast, the equivalent female SNOMED code to “primary malignant neoplasm of the male breast” (93884005) is “primary malignant neoplasm of the female breast” (93796005). Codes may also have semantic differences, while being similar within the hierarchical structure of the vocabulary—that is, having similar ancestry. Figure 1 demonstrates an exact ancestry match of two SNOMED codes.

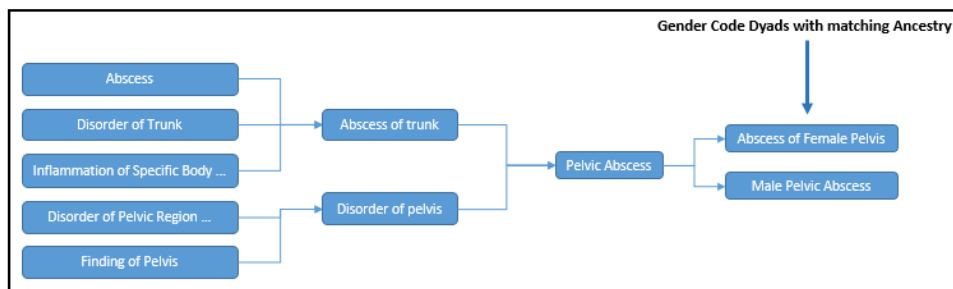


Figure 1. Code dyads that match exactly on SNOMED CT ancestral hierarchy

Using an informatics approach, we sought to identify female and male condition dyads, quantify the degree of misalignment between code and patient gender, and determine if the degree of misalignment differs between the EMR of a federal healthcare system (VA) and a private healthcare system (VUMC).

Methods

We used a two-prong approach to identify two different SNOMED codes that represent similar clinical concepts, but for opposite genders (gender code dyads). First, we considered codes if their text descriptions matched on all characters except for the words ‘female’ or ‘male’. Second, using the Concept Ancestor table we identified codes with equivalent family trees, that is—both codes have the exact same ancestor codes, but one code had the word ‘female’ in its description while the other contained the word ‘male’ (Figure 1).

We identified all instances of the gender codes in the Condition Occurrence of VA from 2001-2017 and of VUMC from 1990-2019. We used the Person table in each OMOP instance to ascertain gender of each patient. We calculated the proportion of male codes recorded for female patient and female codes recorded for male patients (gender-code misalignment) overall and for each SNOMED code and compared differences between VA and VUMC.

Results

We found 481 gender code dyads (962 distinct codes) in the SNOMED vocabulary, of which 66 codes were found in the VA Condition Occurrence table and 83 in VUMC. There were 1,247,505 records of the 66 VA codes and 400,500 records of the 83 VUMC codes. Gender-code misalignment occurred for 96,223 of the VA gender specific records (7.71%) and 10,886 (2.72%) of VUMC. Ordered by the absolute number of discordant records, the top 10 VA and VUMC codes share 5 of the same codes. In VA the top misaligned codes are mostly cancer codes and other illnesses related to aging, while the VUMC codes contain a number of fertility related codes (Table).

Top 10 VA SNOMEDs by # of discordant instances			Top 10 VUMC SNOMEDs by # of discordant instances		
SNOMED Description	VA Discordant count (%)	VUMC Discordant Count (%)	SNOMED Description	VUMC Discordant Count (%)	VA Discordant count (%)
Primary malignant neoplasm of female breast	49,407 (9.1%)	1,359 (0.5%)	Primary malignant neoplasm of male breast	4,470 (83.6%)	736 (2.3%)
Secondary malignant neoplasm of female genital organ	19,505 (97.1%)	0	Female stress incontinence	1,953 (5.1%)	10,866 (7.9%)
Female stress incontinence	10,866 (7.9%)	1,953 (5.1%)	Primary malignant neoplasm of female breast	1,359 (0.5%)	49,407 (9.1%)
Male urinary stress incontinence	5,364 (4.4%)	189 (17.3%)	Female infertility	780 (3.9%)	593 (1.2%)
Male pattern alopecia	2,290 (50.2%)	771 (53.8%)	Male pattern alopecia	771 (53.8%)	2,290 (50.2%)
Injury of female external genital organs	1,681 (95.9%)	16 (2.3%)	Male sterility	493 (47.1%)	0
Overlapping malignant neoplasm of female breast	1,557 (6.4%)	70 (0.3%)	Male infertility	205 (7.9%)	334 (0.7%)
Secondary malignant neoplasm of female breast	1,203 (38.0%)	3 (0.9%)	Male urinary stress incontinence	189 (17.3%)	5,364 (4.5%)
Malignant neoplasm of nipple and areola of female breast	899 (20.0%)	49 (1.2%)	Female sterility	108 (13.7)	0
Primary malignant neoplasm of male breast	736 (2.3%)	4,470 (83.6%)	Congenital anomaly of female genital system	79 (1.8%)	0

Discussion/Conclusions

We used a simple approach facilitated by the OMOP CDM to identify gender code dyads. Gender-code misalignment is relatively rare but occurs in both a federal and private health care system. Although rare overall, for certain clinical concepts the majority of instances misalign with patient gender. The VA has relatively more misalignment for female-to-male misalignment, perhaps explained by its capitation system and the older, more male demographic. Studies using EMR data should be aware of this nature of coding error and consider ways to mitigate misclassification.

References

1. Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases. *AMIA Annual Symposium Proceedings*. 2008;2008:242-246.