Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Hannah K. Moriarty Date

Utilization of SOGI in the Confirmation of Transgender and Gender Diverse Patient Populations by Inter-observer Agreement in the Kaiser Permanente System

By

Hannah Moriarty
Master of Public Health

Department of Epidemiology

_____

Michael Goodman, MD, MPH
Committee Chair

Utilization of SOGI in the Confirmation of Transgender and Gender Diverse Patient Populations by Inter-observer Agreement in the Kaiser Permanente System

By

Hannah Moriarty

Bachelor of Science
Emory University
2015

Thesis Committee Chair: Michael Goodman, MD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in the Department of Epidemiology
2021

Abstract

Utilization of SOGI in the Confirmation of Transgender and Gender Diverse Patient Populations by Inter-observer Agreement in the Kaiser Permanente System

By Hannah Moriarty

Background: Despite ongoing efforts to systematically capture gender identity information, accurate identification of transgender and gender diverse (TGD) people based on data from electronic health records (EHR) is still a challenge. One way of improving identification of TGD individuals in the EHR is to use a combination of free text keywords (KWD) and diagnostic codes (DX). In recent years, health systems have also begun to capture Sexual Orientation and Gender Identity (SOGI) data. The aim for the current analysis is to assess the utility of SOGI data in identifying TGD people enrolled in two large health systems.

Methods: The Study of Transitions, Outcomes and Gender (STRONG) data in conjunction with newly available SOGI information was used to assess the identification of TGD patients in Kaiser Permanente health systems in Northern and Southern California. For each patient, two reviewers examined free text to determine TGD status. In the event of disagreement between reviewers, a third reviewer served as adjudicator. Extent of agreement between reviewers was assessed by calculating a kappa statistic and the corresponding 95% confidence interval (CI), both overall and across subgoups of study participants. Factors associated with final confirmation of TGD status were analyzed by multiple logistic regression with results expressed as odds ratios (OR) and 95% CI.

Results: Inter-observer agreement for TGD status was higher among persons with SOGI data [kappa=0.29; 95% CI: 0.24, 0.34] compared to patients without SOGI data [kappa=0.18; 95% CI: 0.14, 0.22]. The odds of inter-reviewer disagreement were significantly lower [OR=0.32; 95% CI 0.27, 0.37] amongst persons whose records contained all three data elements (DX, KWD, and SOGI) compared to health plan members with only two of the three data elements available. Patient's age also predicted final TGD status confirmation, with those in age groups 19-44 years having reduced odds of disagreement compared to patients ≤18 years of age.

Conclusions: Availability of SOGI data in addition to other data elements improved inter-reviewer agreement and increased the likelihood of TGD status confirmation. Complete and accurate SOGI data collection is necessary for informing evidence-based equitable health care delivery among sexual and gender minority populations.

Utilization of SOGI in the Confirmation of Transgender and Gender Diverse Patient Populations by Inter-observer Agreement in the Kaiser Permanente System

By

Hannah Moriarty

Bachelor of Science
Emory University
2015

Thesis Committee Chair: Michael Goodman, MD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Department of Epidemiology
2021

Introduction

In clinical research, identification of eligible study subjects and collection of information on participant characteristics are often subject to variability and error. Concerns about variability and error are especially warranted in areas that require observer's interpretation of the evidence. One approach towards reducing variability and decreasing the amount of resulting error is to employ more than one observer. The level of inter-observer agreement can then be measured, and the reasons for discrepancy can be examined through additional investigation (1, 2). In the event of disagreement, an additional (usually most experienced) observer acts as adjudicator thereby reducing the likelihood of error (3). Examination of inter-observer agreement plays an important role in both clinical research and clinical practice, and is especially well-documented within the fields of pathology and radiology (4, 5).

In modern clinical research, a large proportion of patient information is stored in electronic health records (EHR) containing copious clinical notes. Thus, researchers are often asked to make observer judgements based upon their reading and interpretation of these clinical notes. This approach is particularly valuable in situations where the goal is to identify patient population groups that may not be accurately identified through the use of diagnostic codes or other standardized data elements.

An example of such a hard-to-ascertain population group are transgender and gender diverse (TGD) individuals. Despite ongoing efforts to systematically capture gender identity

information, accurate identification of TGD people based on EHR data is still a challenge (6). Studies have shown that many TGD patients feel they need to infer whether routinely used intake forms are asking for their gender identity or sex assigned at birth (7). Similarly, the traditional demographic gender variable in administrative records is of limited use. For example, one study found that adult TGD people have about a 50% likelihood of being identified as “female” or “male” in the EHR, which sometimes reflects gender identity and sometimes sex assigned at birth with no way of telling which is which (6). Thus, a critical methodological issue facing clinical epidemiology studies on transgender health is the need to use reliable and accurate means for determining TGD identity in the context of EHR data extraction.

One way of improving identification of TGD individuals in the EHR is to use a combination of free text keywords and diagnostic or treatment codes (6). In recent years, several health systems have begun capturing Sexual Orientation and Gender Identity (SOGI) data. SOGI has been highlighted as a key data element for charting longitudinal TGD health outcomes, de-stigmatizing sexual and gender diversity, and optimizing research within the TGD community (8, 9). The United States Office of Disease Prevention and Health Promotion has identified the collection of SOGI data in health records as an important priority in reducing LGBT health disparities (10).

Although the importance of accurately identifying TGD people is gaining recognition, empirical data pertaining to the application of various available methods is lacking. To begin closing this

knowledge gap, the present study sought to address two research aims. The first aim is to examine the extent and predictors of agreement between two reviewers charged with identification of TGD people in the EHR of two large integrated health systems. The second aim is to assess factors associated with confirmed TGD status determined by final adjudication by expert review. Of particular interest in both of these aims is the utility of the newly available SOGI data.

Methods

The Study of Transitions, Outcomes and Gender (STRONG) was initiated in 2013 to improve identification of TGD patients within the EHR and to better assess health outcomes in this population. The STRONG cohort includes patients who enrolled in Kaiser Permanente Northern California (KPNC), Kaiser Permanente Southern California (KPSC) and Kaiser Permanente Georgia between January 1, 2006, and December 31, 2014. Two computer programs were applied to the EHR data for this timeframe to ascertain a cohort of TGD individuals. The first computer program assessed for International Classification of Disease, Ninth edition (ICD-9) codes suggestive of TGD status. The second computer program assessed for specific keywords from free-text clinical notes that indicated TGD status. The resulting cohort was then validated by two reviewers who examined the free text clinical notes to make a judgement on patients' TGD status. If these two reviewers disagreed, a third and more-experienced adjudicator reviewed the free text and made a final determination of TGD status (6).

The current analysis represents an extension of the STRONG study and takes advantage of the newly available SOGI data for participants enrolled in two of the study sites – KPNC and KPSC – from January 2019 through July 2020. This relatively narrow time interval was selected because the goal of the cohort expansion was to examine the impact of the coronavirus disease 2019 (COVID-19) pandemic on the TGD community. IRB approval for this study was received from Emory University.

The text strings for all newly identified cohort members underwent review and adjudication using the same protocol as the original STRONG cohort (6). For the purpose of the current analysis, only individuals with keywords and at least one other data source (diagnostic codes or SOGI information) were considered eligible.

The main independent variable for this project was the source of data used for identification of the cohort candidates. This variable was categorized into three groups: diagnostic codes (DX) + keywords (KWD), KWD + SOGI, and DX + KWD + SOGI. Covariates included current age (≤18, 19-24, 25-34, 35-44, and ≥45 years) recorded gender (male vs. female), study site (KPNC vs. KPSC), and index interval. Index interval was divided into three categories (2006-2015, 2016-2018, 2019-2020) representing the year of the first evidence of TGD status in the EHR.

The data analysis addressed two research aims. The first aim assessed the extent and determinants of disagreement between the reviewers, and the second aim investigated the factors associated with the final assignment of TGD status among cohort candidates.

The extent of agreement and disagreement between observers was analyzed using kappa statistic. Kappa values can range from -1.00 to 1.00, where a value of 0.00 indicates an agreement that is no different than expected by chance alone. By convention, kappa statistic values of <0.20, 0.21-0.40, 0.41-0.60, 0.61-0.80, and 0.81-1.00 indicate poor, fair, moderate, good, and excellent levels of agreement, respectively (11). The kappa statistics and the

corresponding 95% confidence intervals (CI) were calculated both overall and separately by study site and presence of SOGI data. The determinants of disagreement (dependent variable) were assessed with the use of multiple logistic regression model, which included study site, data source, age group, recorded gender, and index interval as the independent variables of interest.

Factors associated with the final assignment of TGD status among cohort candidates were first examined by comparing the distribution of variables among persons with and without TGD status confirmation. This was followed by a second multiple logistic regression analysis, where the dependent variable was final TGD status assignment (confirmed vs. not confirmed) and the independent variables were the same as in the first model.

The results for each model were reported as adjusted odds ratios (OR) and the corresponding 95% CI, and both models were assessed for collinearity and interactions. All analyses were carried out using IBM® SPSS ® Statistics v. 26 (IBM Corporation Armonk, NY; 2019).

Results

Table 1 assesses the inter-observer agreement between the two reviewers as pertains to the overall population, as well as by study site, and according to SOGI data availability. The kappa statistic for the overall population was 0.23 [95% CI: 0.20, 0.26] indicating a fair degree of agreement between reviewers. The site-specific kappa [95% CI] estimates were similar: 0.23 [0.19, 0.27] for KPNC and 0.23 [0.17, 0.28] for KPSC. When agreement was examined according to the availability of SOGI data, the resulting kappa was higher among persons with SOGI data [0.29; 95% CI: 0.24, 0.34] compared to the cohort members for whom SOGI information was not available [0.18; 95% CI: 0.14, 0.22].

As shown in Table 2, the odds of inter-reviewer disagreement were significantly lower [OR=0.32; 95% CI: 0.27, 0.37] among persons who had all three sources of information (DX+KWD+SOGI) and significantly higher in the KWD+SOGI group, relative to cohort candidates whose source of relevant data was limited to TGD-specific diagnostic codes and keywords (KWD+DX group). The odds of having disagreement amongst those with recorded male gender was 15% lower [95% CI: 0.73, 0.97] than the odds of having disagreement if recorded gender was female. Other factors associated with lower odds of disagreement included age over 18, but under 45 years (reference group ≤18 years), and enrollment in KPSC (vs. KPNC) health system (Table 2).

Table 3 compares characteristics of cohort candidates whose TGD status was confirmed to those who were deemed not TGD and thus not eligible for inclusion in the study. Participants with SOGI data had a higher proportion of confirmed TGD cases (96%) compared to the KWD+DX group (92%). Those in the youngest (≤18 years old) and oldest (≥45 years old) age groups had lower proportions of persons with confirmed TGD status than the other age groups. The differences with respect to study site, recorded gender, and index interval were less evident (Table 3).

Table 4 presents the results of the logistic regression analyses which investigated the factors independently associated with final confirmation of TGD status. Factors associated with higher odds of confirmed TGD status included availability of all three lines of evidence (DX+KWD+SOGI) relative to the DX+KWD reference group [OR=2.72; 95% CI: 2.24, 3.31], and age groups of 19-24, 25-34, and 35-44 years, relative to the reference age group of ≤18 years (OR range 1.57-1.98; all 95% CI intervals excluded 1.0). Cohort members in the KWD+SOGI group had significantly lower odds of eligibility compared to the KWD+DX group [OR=0.30; 95% CI: 0.22, 0.40]. By contrast, there was no evidence that TGD status confirmation was independently associated with study site, recorded gender, or index interval (Table 4).

Discussion

The results of this analysis produced several important findings in the consideration of how to best identify TGD patients from current EHR data. We observed only fair agreement between reviewers overall; however, the agreement improved when the data was limited to participants whose EHR contained SOGI information. Notable predictors of lower inter-reviewer disagreement and higher likelihood of TGD status confirmation included availability of all three data elements (KW, DX ,and SOGI), and an age range of 19 to 44 years old.

Important to note, is that although the overall kappa statistic for reviewer agreement was only fair, the percent agreement was very high. If the expected agreement is >90% (as was the case in our study), then kappa values are typically lower (12, 13).

SOGI information represents a relatively recent element of data collection (14). Previous estimates of the proportion of TGD people in a population differed by orders of magnitude depending on what definition of TGD was used (15). Therefore, SOGI collection has emerged in recent years as a tool for improving sensitive discussions between patient and provider, identifying transgender patients within the larger EHR systems, and ultimately improving transgender patient health outcomes through better planning and delivery of care (16). To our knowledge, this paper is the first attempt to evaluate reviewer agreement of TGD status based on SOGI characteristics within a large health system that mimics a population size study.

Additionally, our study also adds value to the medical literature as its study design uses reviewer comprehension of free text clinical notes to arrive at inter-observer agreement. Few articles in the literature have conducted inter-observer agreement studies based on reviewers' reading comprehension (17, 18). These previous studies have used inter-observer agreement to judge the quality of scientific abstracts for inclusion in scientific meetings (17, 18). These studies found that agreement can be improved when quality assessment items are applied to improve the reading comprehension of reviewers (18). However, further development of these tools for improving inter-observer agreement based on reading comprehension of clinical notes is needed.

A notable limitation of this study is the inability to take into account the characteristics of individual reviewers. The only information available about reviewers is that the initial two reviewers were graduate students and that the adjudicator was a faculty member specializing in transgender health. Nevertheless, the reviewers performed their tasks using the same protocol and identical explicit criteria, all of them underwent the same training, and their performance was monitored on a weekly basis to ensure maximum standardization of TGD status assignment. Another limitation of this study is that relatively little is known about participants' characteristics. For instance, we do not have data on patients' race/ethnicity, prior treatment status, or current medications. It is unclear how this missing information may have influenced inter-reviewer agreement and final TGD status confirmation. It is also worth keeping in mind that Kaiser Permanente health systems were amongst the first to begin collecting SOGI data in the United States (19). Although the importance of collecting SOGI data is gaining

recognition, its actual implementation outside of Kaiser Permanente remains relatively sporadic (16). For this reason, the findings from our study may not be generalizable to other healthcare organizations.

Despite the aforementioned limitations, the results of this study indicate that SOGI information, in addition to other data elements, substantially improves inter-reviewer agreement and increases the likelihood of TGD status confirmation. These findings may serve as additional motivation for instituting collection of SOGI data across health systems and for perhaps incorporating these data in future rounds of the US Census (20). The availability of complete and accurate SOGI data is expected to improve evidence-based planning and funding of equitable health care among gender minority populations.

References

1. Goodman M, Steenland NK, Almon ML, *et al*. Prostate cancer treatment ascertained from several sources: analysis of disagreement and error. Ann Oncol 2012; 23(1): 256-263.

2. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012; 22(3): 276-282.

3. Goodman M, Ward KC, Osunkoya AO, *et al*. Frequency and determinants of disagreement and error in Gleason scores: a population-based study of prostate cancer. Prostate 2012; 72(13): 1389-1398.

4. Benchoufi M, Matzner-Lober E, Molinari N, *et al*. Interobserver agreement issues in radiology. Diagn Interv Imaging 2020; 101(10): 639-641.

5. Gültekin A, Yaylalı O, Şengöz T, *et al*. Intraobserver and interobserver agreement for the interpretation of 68Ga-prostate-specific membrane antigen-I&T positron emission tomography/computed tomography imaging. Nucl Med Commun 2019; 40(12): 1250-1255.

6. Quinn VP, Nash R, Hunkeler E, *et al*. Cohort profile: Study of Transition, Outcomes and Gender (STRONG) to assess health status of transgender people. BMJ Open 2017; 7(12): e018121.

7. Cicero EC, Reisner SL, Silva SG, *et al*. Health Care Experiences of Transgender Adults: An Integrated Mixed Research Literature Review. ANS Adv Nurs Sci 2019; 42(2): 123-138.

8. Grasso C, Goldhammer H, Brown RJ, *et al.* Using sexual orientation and gender identity data in electronic health records to assess for disparities in preventive health screening services. Int J Med Inform 2020; 142: 104245.

9. Suen LW, Lunn MR, Katuzny K, *et al*. What Sexual and Gender Minority People Want Researchers to Know About Sexual Orientation and Gender Identity Questions: A Qualitative Study. Arch Sex Behav 2020; 49(7): 2301-2318.

10. Office of Disease Prevention and Health Promotion. Lesbian, Gay, Bisexual, and Transgender Health. 2020. Retrieved April 2021 from https://www.healthypeople.gov/2020/topics-objectives/topic/lesbian-gay-bisexual-and-transgender-health.

11. Fleiss, J. Statistical methods for rates and proportions. New York, NY: John Wiley and Sims; 1981.

12. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990; 43(6): 543-549.

13. Vetter TR, Schober P. Agreement Analysis: What He Said, She Said Versus You Said. Anesth Analg 2018; 126(6): 2123-2128.

14. NASTAD. Modernizing programs to collect sexual orientation and gender identity data. Policy and Legislative Affairs (Issue Brief). June 2017. Retrieved April 2021 from https://www.nastad.org/sites/default/files/sogi.issue_.brief_.final_.pdf.

15. Collin L, Reisner SL, Tangpricha V, *et al*. Prevalence of Transgender Depends on the "Case" Definition: A Systematic Review. J Sex Med 2016; 13(4): 613-626.

16. Streed Jr. CG, Grasso C, Reisner S, *et al*. Sexual Orientation and Gender Identity Data Collection: Clinical and Public Health Importance. American Journal of Public Health 2020; 110(7): 991-993.

17. Kemper KJ, McCarthy PL, and Cicchetti DV. Improving participation and interrater agreement in scoring Ambulatory Pediatric Association abstracts. How well have we succeeded? Arch Pediatr Adolesc Med 1996; 150(4): 380-383.

18. Timmer A, Sutherland LR, and Hilsden RJ. Development and evaluation of a quality score for abstracts. BMC Med Res Methodol 2003; 3: 2.

19. The Permanente Medical Group. Transgender Health: Delivering a Full Spectrum of Care. Permanente Medicine—The Permanente Medical Group. 2017. Retrieved April 2021 from https://doctorsatkaisertpmg.com/2018/11/09/transgender-health-delivering-full-spectrum-care/

20. Goodman M, Adams N, Corneil T, *et al*. Size and Distribution of Transgender and Gender Nonconforming Populations: A Narrative Review. Endocrinol Metab Clin North Am 2019; 48(2): 303-321.

Tables

Table 1: Agreement between reviewers by subject category

Overall					
			Reviewer 2		
			Not eligible	Eligible	Total
	Reviewer 1	Not eligible	197	818	1015
		Eligible	230	9340	9570
		Total	427	10158	10585
	Percent agreement = 90% Kappa = 0.23, 95% CI [0.20, 0.26]				
By study site					
			Reviewer 2		
			Not eligible	Eligible	Total
KPNC	**Reviewer 1**	Not eligible	133	514	647
		Eligible	165	5251	5416
		Total	298	5765	6063
	Percent agreement = 89% Kappa = 0.23, 95% CI [0.19, 0.27]				
			Reviewer 2		
			Not eligible	Eligible	Total
KPSC	**Reviewer 1**	Not eligible	64	304	368
		Eligible	65	4089	4154
		Total	129	4393	4522
	Percent agreement = 92% Kappa = 0.23, 95% CI [0.17, 0.28]				
By presence of SOGI data					
			Reviewer 2		
			Not eligible	Eligible	Total
SOGI present	**Reviewer 1**	Not eligible	105	290	395
		Eligible	149	5880	6029
		Total	254	6170	6424
	Percent agreement = 93% Kappa = 0.29, 95% CI [0.24, 0.34]				
			Reviewer 2		
			Not eligible	Eligible	Total
SOGI absent	**Reviewer 1**	Not eligible	92	528	620
		Eligible	81	3460	3541
		Total	173	3988	4161
	Percent agreement = 85% Kappa = 0.18, 95% CI [0.14, 0.22]				

Abbreviations: CI=confidence interval, KPNC=Kaiser Permanente Northern California, KPSC=Kaiser Permanente Southern California, SOGI=sexual orientation gender identity

Table 2: Factors associated with reviewer disagreement

Participant Characteristics	OR	95% CI
Study Site		
KPNC	1.0 (reference)	
KPSC	0.54	[0.47, 0.62]
Text Evidence		
DX+KWD	1.0 (reference)	
KWD+SOGI	1.73	[1.32, 2.26]
DX+KWD+SOGI	0.32	[0.27, 0.37]
Age Group (years)		
≤18	1.0 (reference)	
19-24	0.78	[0.66, 0.93]
25-34	0.75	[0.62, 0.90]
35-44	0.67	[0.52, 0.86]
≥45	0.91	[0.71, 1.16]
Recorded Gender		
Female	1.0 (reference)	
Male	0.85	[0.73, 0.97]
Index Interval		
2006-2015	1.0 (reference)	
2016-2018	1.21	[0.97, 1.52]
2019-2020	1.20	[0.95, 1.51]

Abbreviations: CI=confidence interval, DX= diagnostic codes, KPNC=Kaiser Permanente Northern California, KPSC=Kaiser Permanente Southern California, KWD=keywords, OR=odds ratio, SOGI=sexual orientation gender identity

Table 3: Final adjudication of TGD status by subject category

Participant Characteristics	TGD confirmed		TGD not confirmed	
	N	%	N	%
Study Site				
KPNC	5736	94.61	327	5.39
KPSC	4270	94.43	252	5.57
SOGI Data Availability				
SOGI present	6169	96.03	255	3.97
SOGI absent	3837	92.21	324	7.79
Age Group (years)				
≤18	2480	92.02	215	7.98
19-24	2799	95.40	135	4.60
25-34	2715	95.80	119	4.20
35-44	1102	96.24	43	3.76
≥45	910	93.14	67	6.86
Recorded Gender				
Female	3938	95.42	189	4.58
Male	6010	93.96	386	6.04
Index Interval				
2006-2015	1229	94.54	71	5.46
2016-2018	4793	94.72	267	5.28
2019-2020	3984	94.30	241	5.70

Abbreviations: KPNC=Kaiser Permanente Northern California, KPSC=Kaiser Permanente Southern California, N=number of individuals, SOGI=sexual orientation gender identity, TGD= transgender or gender diverse

Table 4: Factors associated with final confirmation of TGD status

Participant Characteristics	OR	95% CI
Study Site		
KPNC	1.0 (reference)	
KPSC	1.16	[0.97, 1.38]
Text Evidence		
DX+KWD	1.0 (reference)	
KWD+SOGI	0.30	[0.22, 0.40]
DX+KWD+SOGI	2.72	[2.24, 3.31]
Age Group (years)		
≤18	1.0 (reference)	
19-24	1.57	[1.25, 1.97]
25-34	1.64	[1.29, 2.09]
35-44	1.98	[1.39, 2.80]
≥45	1.04	[0.95, 1.37]
Recorded Gender		
Female	1.0 (reference)	
Male	1.14	[0.95, 1.37]
Index Interval		
2006-2015	1.0 (reference)	
2016-2018	1.13	[0.86, 1.50]
2019-2020	1.06	[0.80, 1.41]

Abbreviations: CI=confidence interval, DX= diagnostic codes, KPNC=Kaiser Permanente Northern California, KPSC=Kaiser Permanente Southern California KWD=keywords, OR=odds ratio, SOGI=sexual orientation gender identity, TGD= transgender or gender diverse