Distribution Agreement

Signature:

_______________________________ 05/03/2021
Kristine MaWhinney Date

Availability of sexual orientation and gender identity (SOGI) information in a cohort of transgender and gender diverse people: An analysis of electronic health records

By

Kristine MaWhinney
MSPH

Department of Epidemiology

Dr. Michael Goodman
Committee Chair

Availability of sexual orientation and gender identity (SOGI) information in a cohort of transgender and gender diverse people: An analysis of electronic health records

By

Kristine MaWhinney

B.S.
Brigham Young University
2018

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 Science in Public Health
in Epidemiology
2021

Abstract

Availability of sexual orientation and gender identity (SOGI) information in a cohort of transgender and gender diverse people: An analysis of electronic health records
By Kristine MaWhinney

ABSTRACT:

Background: In 2015, the U.S. Department of Health and Human Services issued a directive that electronic health record (EHR) systems should enable providers to systematically record patients' sexual orientation and gender identity (SOGI). Kaiser Permanente (KP) health systems began implementing SOGI data capture in 2019. The objective of this study was to investigate what proportion of transgender and gender diverse (TGD) people enrolled in KP plans have SOGI information in their health records, and to identify factors associated with SOGI status availability.

Methods: This analysis used EHR data from KP health plans in Northern and Southern California. TGD persons with and without SOGI data were compared with respect to their age, race, recorded sex, and study site. A separate analysis examined a subset of participants with SOGI data to compare demographic characteristics among TGD persons reporting binary (e.g., transman or transwoman) and non-binary gender identities. Each comparison was carried out using multiple logistic regression models.

Results: Among 19,057 study participants, 9,190 (48%) had SOGI information indicative of TGD status. Participants with SOGI data were younger, and more likely to be Non-Hispanic White, reside in Northern California and have female sex recorded in the EHR files. Among participants with SOGI data, those identified as non-binary were more likely to be 18-35 years of age, and have female gender assigned at birth, but were less likely to be African American or Hispanic.

Conclusion: Although KP plans in California adopted SOGI data collection relatively early, approximately half of their TGD members still do not have SOGI information recorded in the EHR. SOGI data availability appears to differ by age, race ethnicity.

Availability of sexual orientation and gender identity (SOGI) information in a cohort of transgender and gender diverse people: An analysis of electronic health records

By

Kristine MaWhinney

B.S.
Brigham Young University
2018

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 Science in Public Health
in Epidemiology
2021

INTRODUCTION

The term ’transgender and gender diverse (TGD)’ describes a diverse group of people who experience incongruence between their sex assigned at birth and their gender identity (Roblin et al., 2016; Safer, Coleman, & Hembree, 2016). The suffix “trans” means “across”, which can imply identifying with the gender that is exactly opposite of the natal sex. For example, the term transwoman is often used to describe and individual who was assigned male sex at birth but identifies as a female. By analogy, the suffix “cis”, which means “on this side of", is applied to individuals (i.e., cismen and ciswomen) who feel no mismatch between their sex and gender.

It is important to point out that many individuals whose gender identity does not match their birth sex reject the male-female dichotomy and instead self-identify as ‘non-binary’ or ‘genderqueer’ (Bosse, Leblanc, Jackman, & Bjarnadottir, 2018; Fiani & Han, 2019). For this reason, the term ‘TGD’ is preferred to the term ‘transgender’, and the terms ‘transfeminine’ (TF) and ‘transmasculine’ (TM) are preferred to the terms ‘transman’ or ‘transwoman’ because they do not rely on binary categories, but rather consider that gender identity may have a wide range of manifestations. (Gerth et al., 2018; Roblin et al., 2016).

TGD people are a vulnerable population facing several health issues. For example, many TGD persons seek to affirm their gender through hormone therapy or surgery to achieve desired femininity or masculinity (Coleman et al., 2012). Receipt of gender affirming treatments is associated with both risks and benefits that are unique to TGD people and require careful consideration (Getahun et al., 2018). Other health issues in TGD health are more general, and are usually attributable to minority stress or inadequate access to health care (Gonzales & Henning-Smith, 2017; James, 2016; Safer, Coleman, Feldman, et al., 2016). Specific health concerns among TGD people include high rates of mental health problems, including suicide attempts, lack of adequate screening and preventive care, as well as elevated risk of thrombotic events associated with feminizing hormone therapy (Deutsch, 2016; Feldman et al., 2016).

An important barrier to establishing adequate care for TGD people is the lack of systematically collected population-level information on gender identity. Accurate estimates of the number and the proportion of TGD people in the general population are necessary for allocating health care resources and for ensuring its adequate coverage (Goodman et al., 2019). One of the best ways to identify TGD people is by administering the two-step gender identity measure, which is increasingly used in both research and clinical practice. The two step-measure first asks individuals to indicate their sex assigned at birth, and then inquires about their current gender identity (Zhang, Goodman, et al., 2020).

The two-step identification of TGD people is one component of the Sexual Orientation and Gender Identity (SOGI) data collection (Bosse et al., 2018; Cahill & Makadon, 2014). In 2015, the U.S. Department of Health and Human Services issued a directive that electronic health record (EHR) systems should enable providers to systematically record SOGI information. (Grasso, McDowell, Goldhammer, & Keuroghlian, 2019; HHS, 2015).

Kaiser Permanente (KP) has been leading the way in providing care to TGD members, opening one of the first comprehensive, transgender care clinics in the US in 2013 (Seto, 2017). In 2019, KP began implementing SOGI data capture, which is expected to further improve ascertainment of the TGD population with more accurate determination of TM and TF status. KP is also one of the leading organizations in TGD health research. Three KP plans in Georgia, Northern California and Southern California serve as sites for the “Study of Transition, Outcomes and Gender” (STRONG). STRONG is a longitudinal cohort study of TGD KP health plan members identified using an EHR-based algorithm that includes both standard diagnostic codes and analysis of digitized provider notes (Quinn et al., 2017). With the introduction of SOGI data collection, STRONG cohort offers an opportunity to examine the extent to which SOGI information can be used for identifying TGD individuals. With these considerations in mind, the goal of the present study is to investigate what proportion of TGD people enrolled in KP plans have the corresponding SOGI information captured in their health records and to identify factors that are associated with documented SOGI status and with binary vs. non-binary gender identity.

METHODS

This analysis is based on the EHR data obtained from two KP health systems in Northern and Southern California. The enrollee populations of the two participating KP sites are generally representative of the communities in the surrounding areas and together include approximately 9 million individuals (Gordon, 2006; Koebnick et al., 2012).

The data collected for the current analysis were obtained as part of the planned STRONG cohort expansion with the goal of evaluating feasibility of using SOGI data to facilitate identification and more complete characterization of new study participants. As described elsewhere, eligibility for inclusion in the current STRONG cohort was based on the presence of diagnostic codes and free-text keywords indicative of TGD status (Quinn et al., 2017). Based on results of validation analyses, over 95% of individuals with at least one diagnostic code and at least one keyword were confirmed as eligible (Roblin et al., 2016; Xie et al., 2021). For this reason, the data for the current analyses were limited to KP enrollees who had both a TGD-specific diagnostic code and a relevant keyword to minimize inclusion of non-eligible persons.

Each person included in the present study was characterized with respect to the availability of SOGI data indicative of TGD status. The relevant components of the SOGI data used for this purpose included two variables “gender identity” and “sex assigned at birth”. Any disagreement between the two variables was deemed consistent with TGD status. Persons whose SOGI data confirmed TGD status were further sub-categorized based on their gender identity into one of the following groups: man (coded as ‘MM’), woman (‘FF’), trans-man (‘FM’), transwoman (‘MF’), non-binary/genderqueer (‘GQ’), and other/unknown. Additional data obtained on each study participant included age at index date (defined as the date of the first evidence of TGD status in the EHR), race/ethnicity, study site, and the value of ‘sex’ variable (male or female) recorded in the administrative records.

The initial data analyses compared distributions of demographic variables among TGD persons with and without supporting SOGI data. These demographic variables were then included in a logistic regression model to assess their independent associations with SOGI data (the outcome variable of interest). A separate analysis was used to examine data on a subset of participants with SOGI data to compare the distributions of demographic factors among TGD persons assigned male at birth (AMAB) and persons assigned female at birth (AFAB). The same subset of study participants with available SOGI information provided data for the logistic regression model evaluating the associations of various participant characteristics and binary vs. non-binary gender identity.

The descriptive comparisons of variable distributions across groups were performed using chi-square tests. All logistic regression analysis results were expressed as crude and adjusted odds ratios (OR) with the corresponding 95% confidence intervals (CI). All models were evaluated for interaction and collinearity. The data analyses were performed using SAS® Software version 9.4 (SAS Institute Inc., Cary, NC).

RESULTS

Table 1 summarizes demographic characteristics of all 19,057 study participants of whom 9,190 individuals (48%) had SOGI information indicative of TGD status. Compared to TGD persons with no relevant SOGI information, participants with SOGI data included greater proportion of persons from Northern California (72% vs. 45%) and slightly higher percentage of Non-Hispanic Whites (56% vs. 50%). By contrast, cohort members with SOGI data were less likely to be over 35 years of age (19% vs. 25%) and have male sex recorded in the administrative EHR files (42% vs. 45%).

Table 2 shows distribution of participant characteristics among 9,156 subjects with SOGI data. Of those, 34 participants were excluded from further analyses because their SOGI data were inconsistent with TGD status – AFAB with MF identity or AMAB with FM identity – likely due to coding errors. Of 3,797 AMAB cohort members, 60% identified as women (FF), 29% identified as transwomen (MF), 8% identified as non-binary or genderqueer (GQ), and 3% identified as other or unknown. Among 5,359 AFAB participants, 48% identified as men (MM), 25% identified as transmen (FM), 23% identified as non-binary or genderqueer (GQ), and 4% identified as other or unknown. A comparison of AMAB and AFAB participants demonstrated that the former group included a much lower proportion of persons under the age of 18 years (14% vs. 33%), and a greater percentage of individuals over 55 years of age (7% vs. 1%). With respect to sex variable in the administrative EHR files, about half of AMAB cohort members were recorded as males and half as females. The corresponding proportions in the AFAB group were 36% and 64%, respectively.

The logistic regression models evaluating the crude and adjusted associations between various demographic factors and SOGI data availability are presented in **Table 3**. These analyses showed that SOGI data were less likely to be available among participants from Southern California compared to Northern California adjusted OR (aOR) of 0.32 (95% CI: 0.30-0.34). Relative to the youngest age group (<18 years), the odds of having SOGI data were higher in persons 18-25 and 26-35 years of age with aOR (95% CI) estimates of 1.15 (1.06-1.25) and 1.18 (1.08-1.28), respectively. By contrast, the odds were significantly lower than 1.0 for persons 46-55 years of age (aOR=0.76; 95% CI: 0.67-0.87) and especially for those over the age of 55 years (aOR=0.51; 95% CI: 0.44-0.59). Using Non-Hispanic Whites as the reference group, the odds of having SOGI data were significantly lower among African Americans (aOR=0.76; 95% CI: 0.67-0.86). Persons whose administrative data indicated ‘female’ sex were more likely to have SOGI data than person whose sex was recorded ‘male’ (aOR=1.12; 95% CI: 1.05-1.19).

Table 4 presents crude and adjusted OR for various factors associated with non-binary identity among persons with SOGI data. Participants from the Southern California site were less likely to self-identify as non-binary (aOR = 0.63; 95% CI: 0.55-0.73). Using the youngest age group (<18 years) as reference, the odds of identifying as non-binary were higher in all age categories. This difference was especially pronounced in persons 18-25 years of age (aOR= 2.09; 95% CI: 1.78-2.45) and 26-35 years of age (aOR=2.49; 95% CI: 2.11-2.93), but gradually became less evident in older age groups. Compared to Non-Hispanic Whites, all race/ethnicity categories were less likely to identify as non-binary, with statistically significant results observed for African Americans (aOR = 0.67; 95% CI: 0.52-0.86), Asian/Pacific Islanders (aOR = 0.73; 95% CI: 0.59-0.90), and Hispanics (aOR = 0.65; 95% CI: 0.55-0.76).

DISCUSSION

In this study, we observed that only about half of TGD participants had SOGI data available in their health records. Among those with SOGI data, less than one-fifth of participants self-identified as non-binary. Younger age groups as well as those who indicated ‘female’ sex were more likely to have SOGI data available. Persons of young age and especially AFAB individuals were more likely to identify as non-binary.

The literature on population estimates of non-binary persons is sparse with most studies combining non-binary and transgender people into a single group. Recent research shows that the proportion of non-binary identifying persons is increasing over time (Claahsen-van der Grinten et al., 2021; National Academies of Sciences et al., 2020; Whyte, Brooks, & Torgler, 2018; Zhang, Rechler, et al., 2020).

Routine collection of SOGI data allows accurate estimates of the size and composition of the TGD population (Baker, Streed, & Durso, 2021; Callahan, Hazarian, Yarborough, & Sánchez, 2014; Callahan et al., 2015). Early reports about implementation of SOGI data collection indicate initial resistance from health care providers. However, experience also showed that proper education on how to initiate SOGI conversations with patients and how to cultivate an LGBTQ friendly environment helps overcome these barriers (Callahan et al., 2015).

Another study conducted in the emergency department reported that health professionals tended to overestimate the level of patients’ discomfort in disclosing their sexual orientation and estimated that 80% of patients would refuse to share this information (Haider et al., 2017). It turned out that 90% of patients were willing to respond to questions about their sexual orientation(Haider et al., 2017). Other surveys also found that most patients fully understood the importance of SOGI data collection and were willing to disclose their SOGI status to health care providers (Cahill et al., 2014; Pinto et al., 2019).

Several limitations in the present study should be considered when interpreting its findings. As the data were obtained from two KP sites in California, this restricted the geographic representation of the results. Further, the study population included TGD persons only, and for this reason, it is possible that SOGI data availability will be quite different among cisgender KP members. In addition, the current analyses did not consider a number of demographic and clinical factors that may influence SOGI data availability; these factors likely include socioeconomic status, education, and health care utilization.

These limitations notwithstanding, our study allowed several useful conclusions. First, although KP plans in California adopted SOGI data collection relatively early, approximately half of their TGD members still do not have SOGI information recorded in the EHR; this leaves considerable room for improvement. Second, the availability of SOGI information appears to differ considerably by age and race/ethnicity; these and other factors need to be considered in planning improvements in SOGI data collection. Third, there are marked differences in the distribution of non-binary gender identities across age categories and AMAB/AFAB groups; this finding requires confirmation and if confirmed deserves further study.

In summary, further efforts are required to ensure complete and accurate SOGI data collection. These data will be critical for planning and funding of equitable health care delivery and research, and for developing data-based policy aimed at improving lives of TGD people.

BIBLIOGRAPHY

Baker, K. E., Streed, C. G., & Durso, L. E. (2021). Ensuring That LGBTQI+ People Count — Collecting Data on Sexual Orientation, Gender Identity, and Intersex Status. *New England Journal of Medicine, 384*(13), 1184-1186. doi:10.1056/NEJMp2032447

Bosse, J. D., Leblanc, R. G., Jackman, K., & Bjarnadottir, R. I. (2018). Benefits of Implementing and Improving Collection of Sexual Orientation and Gender Identity Data in Electronic Health Records. *Comput Inform Nurs, 36*(6), 267-274. doi:10.1097/cin.0000000000000417

Cahill, S., & Makadon, H. (2014). Sexual Orientation and Gender Identity Data Collection in Clinical Settings and in Electronic Health Records: A Key to Ending LGBT Health Disparities. *LGBT Health, 1*(1), 34-41. doi:10.1089/lgbt.2013.0001

Cahill, S., Singal, R., Grasso, C., King, D., Mayer, K., Baker, K., & Makadon, H. (2014). Do ask, do tell: high levels of acceptability by patients of routine collection of sexual orientation and gender identity data in four diverse American community health centers. *PLoS One, 9*(9), e107104. doi:10.1371/journal.pone.0107104

Callahan, E. J., Hazarian, S., Yarborough, M., & Sánchez, J. P. (2014). Eliminating LGBTIQQ Health Disparities: The Associated Roles of Electronic Health Records and Institutional Culture. *Hastings Cent Rep, 44 Suppl 4,* S48-52. doi:10.1002/hast.371

Callahan, E. J., Sitkin, N., Ton, H., Eidson-Ton, W. S., Weckstein, J., & Latimore, D. (2015). Introducing sexual orientation and gender identity into the electronic health record: one academic health center's experience. *Acad Med, 90*(2), 154-160. doi:10.1097/acm.0000000000000467

Claahsen-van der Grinten, H., Verhaak, C., Steensma, T., Middelberg, T., Roeffen, J., & Klink, D. (2021). Gender incongruence and gender dysphoria in childhood and adolescence-current insights in diagnostics, management, and follow-up. *Eur J Pediatr, 180*(5), 1349-1357. doi:10.1007/s00431-020-03906-y

Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., . . . Zucker, K. (2012). Standards of Care for the Health of Transsexual, Transgender, and Gender-Nonconforming People, Version 7. *International Journal of Transgenderism, 13*(4), 165-232. doi:10.1080/15532739.2011.700873

Deutsch, M. B., MD, MPH. (2016). Guidelines for the Primary and Gender-Affirming Care of Transgender and Gender Nonbinary People. Retrieved from https://transcare.ucsf.edu/guidelines

Feldman, J., Brown, G. R., Deutsch, M. B., Hembree, W., Meyer, W., Meyer-Bahlburg, H. F., . . . Safer, J. D. (2016). Priorities for transgender medical and healthcare research. *Curr Opin Endocrinol Diabetes Obes, 23*(2), 180-187. doi:10.1097/med.0000000000000231

Fiani, C. N., & Han, H. J. (2019). Navigating identity: Experiences of binary and non-binary transgender and gender non-conforming (TGNC) adults. *Int J Transgend, 20*(2-3), 181-194. doi:10.1080/15532739.2018.1426074

Gerth, J., Becerra-Culqui, T., Bradlyn, A., Getahun, D., Hunkeler, E. M., Lash, T. L., . . . Goodman, M. (2018). Agreement between medical records and self-reports: Implications for transgender health research. *Rev Endocr Metab Disord, 19*(3), 263-269. doi:10.1007/s11154-018-9461-4

Getahun, D., Nash, R., Flanders, W. D., Baird, T. C., Becerra-Culqui, T. A., Cromwell, L., . . . Goodman, M. (2018). Cross-sex Hormones and Acute Cardiovascular Events in

Transgender Persons: A Cohort Study. *Ann Intern Med, 169*(4), 205-213. doi:10.7326/m17-2785

Gonzales, G., & Henning-Smith, C. (2017). Barriers to Care Among Transgender and Gender Nonconforming Adults. *Milbank Q, 95*(4), 726-748. doi:10.1111/1468-0009.12297

Goodman, M., Adams, N., Corneil, T., Kreukels, B., Motmans, J., & Coleman, E. (2019). Size and Distribution of Transgender and Gender Nonconforming Populations: A Narrative Review. *Endocrinol Metab Clin North Am, 48*(2), 303-321. doi:10.1016/j.ecl.2019.01.001

Gordon, N. (2006). How Does the Adult Kaiser Permanente Membership in Northern California Compare with the Larger Community? Oakland, CA, Kaiser Permanente Northern California Division of Research.

Grasso, C., McDowell, M. J., Goldhammer, H., & Keuroghlian, A. S. (2019). Planning and implementing sexual orientation and gender identity data collection in electronic health records. *J Am Med Inform Assoc, 26*(1), 66-70. doi:10.1093/jamia/ocy137

Haider, A. H., Schneider, E. B., Kodadek, L. M., Adler, R. R., Ranjit, A., Torain, M., . . . Lau, B. D. (2017). Emergency Department Query for Patient-Centered Approaches to Sexual Orientation and Gender Identity : The EQUALITY Study. *JAMA Internal Medicine, 177*(6), 819-828. doi:10.1001/jamainternmed.2017.0906

HHS. (2015). Medicare and Medicaid Programs; Electronic Health Record Incentive Program--Stage 3 and Modifications to Meaningful Use in 2015 Through 2017. Final rules with comment period. *Fed Regist, 80*(200), 62761-62955.

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. . (2016). *The Report of the 2015 U.S. Transgender Survey*. Retrieved from Washington, DC:

Koebnick, C., Langer-Gould, A. M., Gould, M. K., Chao, C. R., Iyer, R. L., Smith, N., . . . Jacobsen, S. J. (2012). Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. *Perm J, 16*(3), 37-41. doi:10.7812/tpp/12-031

National Academies of Sciences, E., Medicine, Division of, B., Social, S., Education, Committee on, P., . . . Gender Diverse, P. (2020). The National Academies Collection: Reports funded by National Institutes of Health. In J. White, M. J. Sepúlveda, & C. J. Patterson (Eds.), *Understanding the Well-Being of LGBTQI+ Populations*. Washington (DC): National Academies Press (US) Copyright 2020 by the National Academy of Sciences. All rights reserved.

Pinto, A. D., Aratangy, T., Abramovich, A., Devotta, K., Nisenbaum, R., Wang, R., & Kiran, T. (2019). Routine collection of sexual orientation and gender identity data: a mixed-methods study. *Cmaj, 191*(3), E63-e68. doi:10.1503/cmaj.180839

Quinn, V. P., Nash, R., Hunkeler, E., Contreras, R., Cromwell, L., Becerra-Culqui, T. A., . . . Goodman, M. (2017). Cohort profile: Study of Transition, Outcomes and Gender (STRONG) to assess health status of transgender people. *BMJ Open, 7*(12), e018121. doi:10.1136/bmjopen-2017-018121

Roblin, D., Barzilay, J., Tolsma, D., Robinson, B., Schild, L., Cromwell, L., . . . Goodman, M. (2016). A novel method for estimating transgender status using electronic medical records. *Ann Epidemiol, 26*(3), 198-203. doi:10.1016/j.annepidem.2016.01.004

Safer, J. D., Coleman, E., Feldman, J., Garofalo, R., Hembree, W., Radix, A., & Sevelius, J. (2016). Barriers to healthcare for transgender individuals. *Curr Opin Endocrinol Diabetes Obes*, *23*(2), 168-171. doi:10.1097/med.0000000000000227

Safer, J. D., Coleman, E., & Hembree, W. (2016). There is reason for optimism: an introduction to the special issue on research needs in transgender health and medicine. *Curr Opin Endocrinol Diabetes Obes, 23*(2), 165-167. doi:10.1097/med.0000000000000228

Seto, B. (2017). A Coordinated Transgender Person Care Pathway. Retrieved from https://permanente.org/coordinated-transgender-person-care-pathway/

Whyte, S., Brooks, R. C., & Torgler, B. (2018). Man, Woman, "Other": Factors Associated with Nonbinary Gender Identification. *Arch Sex Behav, 47*(8), 2397-2406. doi:10.1007/s10508-018-1307-3

Xie, F., Getahun, D., Quinn, V. P., Im, T. M., Contreras, R., Silverberg, M. J., . . . Goodman, M. (2021). An automated algorithm using free-text clinical notes to improve identification of transgender people. *Inform Health Soc Care, 46*(1), 18-28. doi:10.1080/17538157.2020.1828890

Zhang, Q., Goodman, M., Adams, N., Corneil, T., Hashemi, L., Kreukels, B., . . . Coleman, E. (2020). Epidemiological considerations in transgender health: A systematic review with focus on higher quality data. *Int J Transgend Health*, *21*(2), 125-137. doi:10.1080/26895269.2020.1753136

Zhang, Q., Rechler, W., Bradlyn, A., Flanders, W. D., Getahun, D., Lash, T. L., . . . Goodman, M. (2020). Changes in Size and Demographic Composition of Transgender and Gender Non-Binary Population Receiving Care at Integrated Health Systems. *Endocr Pract*. doi:10.1016/j.eprac.2020.11.016

TABLES

Table 1: Distribution of participant characteristics by SOGI data status

Participant characteristics	**All participants** N	%	**SOGI data available** N	%	**SOGI data not available** N	%	**P-value**
Site							<.001
KPNC	11133	58.4%	6658	72.4%	4475	45.4%	
KPSC	7924	41.6%	2532	27.6%	5392	54.6%	
Age group (years)							<.001
<18	4877	25.6%	2342	25.5%	2535	25.7%	
18-25	5575	29.3%	2782	30.3%	2793	28.3%	
26-35	4424	23.2%	2312	25.2%	2112	21.4%	
36-45	1943	10.2%	906	9.9%	1037	10.5%	
46-55	1257	6.6%	522	5.7%	735	7.4%	
56+	981	5.1%	326	3.5%	655	6.6%	
Race/ethnicity							<.001
Non-Hispanic White	10128	53.1%	5152	56.1%	4976	50.4%	
African American	1223	6.4%	525	5.7%	698	7.1%	
Asian/Pacific islander	1588	8.3%	837	9.1%	751	7.6%	
Hispanic	4365	22.9%	1938	21.1%	2427	24.6%	
Other	571	3.0%	297	3.2%	274	2.8%	
Unknown	1182	6.2%	441	4.8%	741	7.5%	
Sex in the EHR							<.001
Male	8377	44.0%	3897	42.4%	4480	45.4%	
Female	10680	56.0%	5293	57.6%	5387	54.6%	
TOTAL	**19057**		**9190**		**9867**		

Abbreviations: SOGI= sexual orientation and gender identity, EHR= electronic health records, KPNC=Kaiser Permanente Northern California; KPSC= Kaiser Permanente Southern California

Table 2: Distribution of participant characteristics by sex assigned at birth with SOGI data

Participant characteristics	All subjects with SOGI data		AMAB		AFAB		P-value
	N	%	N	%	N	%	
SOGI category (gender identity)							<.001
Woman	2271	24.8%	2271	59.8%	0	0.0%	
Man	2586	28.2%	0	0.0%	2586	48.3%	
Trans-man	1317	14.4%	0	0.0%	1317	24.6%	
Trans-woman	1087	11.9%	1087	28.6%	0	0.0%	
Non-binary	1558	17.0%	313	8.2%	1245	23.2%	
Other/unknown	337	3.7%	126	3.3%	211	3.9%	
Site							0.500
KPNC	6651	72.6%	2744	72.3%	3907	72.9%	
KPSC	2505	27.4%	1053	27.7%	1452	27.1%	
Age group (years)							<.001
<18	2334	25.5%	547	14.4%	1787	33.3%	
18-25	2770	30.3%	1114	29.3%	1656	30.9%	
26-35	2302	25.1%	1056	27.8%	1246	23.3%	
36-45	904	9.9%	495	13.0%	409	7.6%	
46-55	520	5.7%	326	8.6%	194	3.6%	
56+	326	3.6%	259	6.8%	67	1.3%	
Race/ethnicity							<.001
Non-Hispanic White	5134	56.1%	2147	56.5%	2987	55.7%	
African American	525	5.7%	187	4.9%	338	6.3%	
Asian/Pacific islander	835	9.1%	422	11.1%	413	7.7%	
Hispanic	1927	21.0%	728	19.2%	1199	22.4%	
Other	296	3.2%	108	2.8%	188	3.5%	
Unknown	439	4.8%	205	5.4%	234	4.4%	
Sex in the EHR							<.001
Male	3881	42.4%	1944	51.2%	1937	36.1%	
Female	5275	57.6%	1853	48.8%	3422	63.9%	
TOTAL	**9156**		**3797**		**5359**		

Abbreviations: AMAB=Assigned Male at Birth; AFAB=Assigned Female at Birth; SOGI= sexual orientation and gender identity, EHR= electronic health records, KPNC=Kaiser Permanente Northern California; KPSC= Kaiser Permanente Southern California

Table 3: Crude and multivariable models investigating factors associated with SOGI data availability

Participant characteristics		Crude results		Adjusted result	
		OR	95% CI	OR	95% CI
Site					
	KPNC	1	ref	1	ref
	KPSC	0.32	(0.30, 0.34)	0.32	(0.30, 0.34)
Age group (years)					
	<18	1	ref	1	ref
	18-25	1.08	(1.0, 1.16)	1.15	(1.06, 1.25)
	26-35	1.19	(1.09, 1.29)	1.18	(1.08, 1.28)
	36-45	0.95	(0.85, 1.05)	0.92	(0.82, 1.03)
	46-55	0.77	(0.68, 0.87)	0.76	(0.67, 0.87)
	56+	0.54	(0.47, 0.62)	0.51	(0.44, 0.59)
Race/ethnicity					
	Non-Hispanic White	1	ref	1	ref
	African American	0.73	(0.65, 0.82)	0.76	(0.67, 0.86)
	Asian/Pacific islander	1.08	(0.97, 1.20)	1.03	(0.92, 1.15)
	Hispanic	0.77	(0.72, 0.83)	0.98	(0.91, 1.06)
	Other	1.05	(0.88, 1.24)	0.98	(0.82, 1.17)
	Unknown	0.58	(0.51, 0.65)	0.59	(0.52, 0.67)
Sex in the EHR					
	Male	1	ref	1	ref
	Female	1.13	(1.07, 1.20)	1.12	(1.05, 1.19)

Abbreviations: SOGI= sexual orientation and gender identity, EHR= electronic health records, KPNC=Kaiser Permanente Northern California; KPSC= Kaiser Permanente Southern California; OR=odds ratio; CI= confidence interval

Table 4: Crude and multivariable models investigating factors associated with non-binary identity among persons with SOGI data

Participant characteristics	**Crude results** OR	95% CI	**Adjusted result** OR	95% CI
Site				
KPNC	**1**	ref	**1**	ref
KPSC	**0.60**	(0.52, 0.68)	**0.63**	(0.55, 0.73)
Age group (years)				
<18	**1**	ref	**1**	ref
18-25	**1.65**	(1.41, 1.92)	**2.09**	(1.78, 2.45)
26-35	**1.89**	(1.61, 2.21)	**2.49**	(2.11, 2.93)
36-45	**1.32**	(1.07, 1.64)	**1.88**	(1.50, 2.35)
46-55	**0.91**	(0.68, 1.22)	**1.38**	(1.02, 1.87)
56+	**0.64**	(0.43, 0.96)	**1.19**	(0.78, 1.82)
Race/ethnicity				
Non-Hispanic White	**1**	ref	**1**	ref
African American	**0.77**	(0.60, 0.98)	**0.67**	(0.52, 0.86)
Asian/Pacific islander	**0.71**	(0.58, 0.87)	**0.73**	(0.59, 0.90)
Hispanic	**0.62**	(0.54, 0.73)	**0.65**	(0.55, 0.76)
Other	**0.87**	(0.64, 1.20)	**0.81**	(0.59, 1.12)
Unknown	**0.90**	(0.70, 1.17)	**0.95**	(0.73, 1.24)
Sex assigned at birth				
AMAB	**1**	ref	**1**	ref
AFAB	**3.42**	(3.00, 3.91)	**3.81**	(3.32, 4.37)

Abbreviations: AMAB=Assigned Male at Birth; AFAB=Assigned Female at Birth; SOGI= sexual orientation and gender identity; KPNC=Kaiser Permanente Northern California; KPSC= Kaiser Permanente Southern California; OR=odds ratio; CI= confidence interval