



Published in final edited form as:

*Med Care*. 2020 October ; 58(10): 903–911. doi:10.1097/MLR.0000000000001362.

## Beyond *Gender Identity Disorder* Diagnosis Codes: An Examination of Additional Methods to Identify Transgender Individuals in Administrative Databases

Guneet K. Jasuja, PhD<sup>1,2,12</sup>, Alexander de Groot, MPH<sup>3</sup>, Emily K. Quinn, MA<sup>3</sup>, Omid Ameli, MD, MPH<sup>4</sup>, Jaclyn M.W. Hughto, PhD, MPH<sup>5,6,7</sup>, Michael Dunbar, PhD<sup>8</sup>, Madeline Deutsch, MD, MPH<sup>9</sup>, Carl G. Streed Jr, MD, MPH, FACP<sup>10,11</sup>, Michael K. Paasche-Orlow, MD, MA, MPH<sup>10</sup>, Hill L. Wolfe, MPA<sup>1,2</sup>, Adam J. Rose, MD, MSc, FACP<sup>10</sup>

<sup>1</sup>Center for Healthcare Organization and Implementation Research (CHOIR), ENRM VAMC, Bedford, MA.

<sup>2</sup>Department of Health Law, Policy and Management, Boston University School of Public Health, Boston, MA

<sup>3</sup>Biostatistics and Epidemiology Data Analytics Center, Boston University School of Public Health, Boston, MA.

<sup>4</sup>OptumLabs, Cambridge, MA.

<sup>5</sup>Brown School of Public Health, Departments of Behavioral and Social Sciences; and Epidemiology, Providence RI.

<sup>6</sup>Brown University, Center for Health Equity Research, Providence RI.

<sup>7</sup>The Fenway Institute, Fenway Health, Boston, MA

<sup>8</sup>RAND Corporation, Health Care Division, Pittsburgh, PA

<sup>9</sup>UCSF Transgender Care, Department of Family & Community Medicine, University of California – San Francisco, CA.

<sup>10</sup>Section of General Internal Medicine, BU School of Medicine, Boston Medical Center, Boston, MA.

<sup>11</sup>Center for Transgender Medicine & Surgery, Boston Medical Center, Boston, MA.

<sup>12</sup>OptumLabs Visiting Scholar, Cambridge, MA.

### Abstract

**Background:** Large administrative databases often do not capture gender identity data, limiting researchers' ability to identify transgender people and complicating the study of this population.

---

**Corresponding author:** Guneet K. Jasuja, Ph.D., Center for Healthcare Organization and Implementation Research, Bedford VA Medical Center, 200 Springs Road, Bedford, MA 01730, Phone: 781-687-4799, Fax: 718-687-3106, guneet.jasuja@va.gov.

Conflict of interest and financial disclosure  
The authors have no conflicts of interest.

**Objectives:** To develop methods for identifying transgender people in a large, national dataset for insured adults.

**Research Design:** This was a retrospective analysis of administrative claims data. *After* using gender **identity disorder (GID)** diagnoses codes, the current method for identifying transgender people in administrative data, we used the following two strategies to improve the accuracy of identifying transgender people that involves: 1) Endocrine Disorder Not Otherwise Specified (Endo NOS) codes and a transgender-related procedure code or 2) Receipt of sex hormones not associated with the sex recorded in the patient's chart ("sex-discordant hormone therapy") and an Endo NOS code or transgender-related procedure code.

**Subjects:** 74 million adults 18 years-old enrolled at some point in commercial or Medicare Advantage plans from 2006 through 2017.

**Results:** We identified 27,227 unique transgender people overall; 18,785 (69%) were identified using *GID* codes alone. Using Endo NOS with a transgender-related procedure code, and sex-discordant hormone therapy with either Endo NOS or transgender-related procedure code, we added 4,391 (16%) and 4,051 (15%) transgender persons, respectively. Of the 27,227 transgender people in our cohort, 8,694 (32%) were transmasculine (TM), 3,959 (15%) were transfeminine (TF), and 14,574 (54%) could not be **classified**.

**Conclusions:** In the absence of gender identity data, additional data elements beyond **GID** codes improves the identification of transgender persons in large, administrative claims databases.

## Keywords

transgender; gender dysphoria; *gender identity disorder*; identification methods; administrative data

## Introduction

The percentage of the U.S. population identifying as transgender, or having a gender identity that differs from one's assigned birth sex, is thought to be as high as 0.6%, or 1.5 million adults.<sup>1-3</sup> Transgender people experience healthcare needs<sup>4</sup> differently from other patient populations.<sup>5</sup> Further, this population is at higher risk of certain health conditions (e.g., human immunodeficiency virus (HIV), depression, tobacco use) than cisgender (non-transgender) people,<sup>2,3,6-10</sup> **which stems largely in part from being subjected to stigma.**<sup>11-13</sup>

Lack of information about the healthcare that transgender people are receiving represents a critical barrier to improving health and reducing disparities for these individuals. Few studies have described the care received by transgender people, and the studies that have been conducted are overwhelmingly based on small, non-representative convenience samples, or databases with limited clinical detail.<sup>6,13-17</sup> Currently, large administrative databases, such as those based on insurance claims data, often fail to systematically record self-reported gender identity data,<sup>18-20</sup> limiting researchers' ability to identify transgender people in these data.

Gender Dysphoria, as described in the Diagnostic and Statistical Manual of Mental Disorders-5 (defined as “gender identity disorder” (**GID**) **in the International Classification of Diseases (ICD)-Ninth and Tenth Revision**) describes distress resulting from non-alignment between one’s gender identity and the sex assigned at birth.<sup>21</sup> **GID** identified using ICD diagnosis codes in administrative data have been used as the primary method for identifying transgender people in previous studies using commercial insurance and Medicare claims data.<sup>22–27</sup> The validity of using **GID** codes to identify samples of transgender people has been demonstrated, at least in terms of achieving acceptable positive predictive value.<sup>24,28,29</sup>

The utility of **GID** codes to identify transgender people is limited by the fact that not all transgender people experience **GID**<sup>30</sup> and **GID** codes may be underused by clinicians to avoid labeling the patient as transgender, or to avoid non-payment of insurance claims.<sup>31,32</sup> Given that **GID** codes alone may not be sufficient to identify a full spectrum of transgender people in administrative data, we aimed to develop a method for identifying transgender people in an administrative dataset lacking patient-reported gender identity data. We used multiple types of data, including inpatient and outpatient claims with diagnosis and procedure codes, pharmacy claims, and administrative sex data. Application of our methods can help us better use administrative claims databases to understand the health of transgender people.

## Methods

### Data Source

This was a retrospective analysis of administrative data from the OptumLabs® Data Warehouse (OLDW), which includes de-identified claims data for commercially insured and Medicare Advantage (MA) enrollees. OLDW includes data on approximately 200 million unique individuals with greatest representation in the Midwest and South U.S. Census Regions.<sup>31</sup> The patient-level information in the OLDW comprises enrollment, medical claims, and outpatient pharmacy claims across a variety of care settings. The claims data are fully de-identified. Our study sample included 74 million adults enrolled at some point in commercial or MA plans for a 10-year period from 2006 through 2017 who were 18 years and had a claim initiated during this study period. The study was approved by the Institutional Review Boards of Boston University and the RAND Corporation.

### Strategy for Identifying Transgender Individuals

In addition to identifying individuals with at least one claim containing a *GID* code (Appendix Table 1), we used three additional categories of data for identifying transgender people in the OLDW: 1) Endocrine Disorder Not Otherwise Specified (Endo NOS) codes (Appendix Table 2), **suggested by the content experts on the study, Drs. Hughto and Deutsch**. These codes are often used to bill for transgender services to avoid **the stigma of** labeling the person as transgender, because no **GID** is present, and/or to avoid denials of payment<sup>32,33</sup>; 2) Common Procedural Terminology (CPT) or ICD-9/10 transgender-related procedure codes (Appendix Tables 3a–3f); and 3) receipt of sex hormones **discordant with the sex recorded at that time** (“sex-discordant hormone therapy”), such as receipt of

testosterone by a person with an administrative (registration) sex of female (Appendix Tables 4a–4d).

**We specified a hierarchy in our strategy for identifying transgender individuals and hence, each individual could only identified as transgender by one method.** After using the **GID** code strategy, which has been the primary method for identifying transgender people within claims data **and was the first step in our hierarchical strategy, we then utilized the three additional data sources described above (Endo NOS, transgender-related procedure codes, and, sex-discordant hormone therapy).** We selected **individuals with Endo NOS and a trans-related procedure code, followed by sex-discordant hormone therapy in conjunction with either Endo NOS OR a trans-related procedure code.**

We also considered looking for changes across time in a person’s sex as an additional approach. **However, sex data in OLDW are only recorded at intervals corresponding to a change in insurance status, making these data less granular.** Thus, we concluded that that the examination of administrative sex marker over time was not a reliable method to detect transgender status in the context of this dataset (i.e., OLDW), **although in a different dataset it might be useful.**

### Strategies to Enhance Positive Predictive Value

We pursued several inherently conservative strategies aimed at increasing positive predictive value, at the possible expense of reduced sensitivity. We enhanced the specificity of Endo NOS codes by requiring that they not be followed by a specific non-diabetes endocrine code within one year, such as thyroid or adrenal disease (Appendix Table 2). For procedure codes related to surgeries and procedures indicative of transgender status, **we categorized our compiled list of codes as “almost conclusive” (e.g., creation of vagina), “more suggestive” (e.g., resection of penis), and “less suggestive” (e.g., breast augmentation), with the help of a technical panel of experts in the clinical management of trans patients. “Less suggestive” procedures, which were too nonspecific, were excluded from this list altogether. For some of the “more suggestive” procedures, we did not consider surgeries that might have been due to a cancer diagnosis. For example, vaginectomy in a patient with a diagnosis of vaginal cancer was not included (Appendix Tables 3a–3f).** To identify individuals who received hormone therapy, presumably as a means to medically affirm their gender, we defined a minimum dose for both testosterone (Appendix Table 4b & 4c) and spironolactone (Appendix Table 4d), which may both be used in lower doses to treat conditions in non-transgender people. Specifically, to be included on the basis of sex-discordant hormone therapy prescription, persons administratively listed as female needed an average daily dose of at least 2 mg in testosterone gels and transdermal patches or 7 mg in intramuscular form, consistent with published guidelines.<sup>34</sup> Similarly, they could not have a documented diagnosis of hypoactive sexual desire (Appendix Table 4e) so as to not include any cisgender women receiving testosterone for off-label treatment of low sexual desire.<sup>35</sup> As it is difficult to compute the daily dose equivalent of testosterone administered as (uncommonly used) compounded creams and ointments, prescriptions for these formulations were not included. Further, if a person administratively listed as male received a daily dose

of spironolactone less than or equal to 50 mg, or had a documented diagnosis of cirrhosis or congestive heart failure (for which spironolactone doses > 50 mg can be used as treatment),<sup>36</sup> the individual was not considered to have received transgender-dose spironolactone and was thus excluded (Appendix Table 4f).

### Categorizing Transgender Individuals as Transgender Female or Transgender Male

After identifying our transgender cohort, we further categorized this cohort into transfeminine (TF), transmasculine (TM), or transgender, and unable to classify gender identity from available data (“*unclassified* gender identity”). TF individuals were defined as transgender people who received feminizing hormones and/or had sentinel surgeries such as neovaginoplasty or tracheal shave. TM individuals were defined as transgender people who received masculinizing hormones and/or had sentinel surgeries such as metoidioplasty or phalloplasty. It is likely that both the TF and TM groups included non-binary people (those who identify as neither male nor female). The remaining people, most of whom were classified as transgender based on a *GID* code, did not have hormones or surgeries that implied **same gender identity status**, and so were classified as **unclassified** gender identity. **Some of the people who fall into this unclassified group may be transgender women, transgender men, non-binary people, or transgender individuals with another gender identity who have not accessed medical forms of gender affirmation and do not plan to.** Among those identified transgender people, we also excluded the few individuals **from our overall trans cohort** who had a procedure code and sex hormone use code that pointed to divergent **gender identity status** (e.g., a procedure code suggestive of being TF such as tracheal shave and a sex hormone suggestive of being TM such as testosterone) potentially due to an individual being intersex or, alternatively, due to coding errors at the clinician level.

### Data Analysis

Descriptive statistics were generated to compare TM, TF, and **unclassified** gender identity people as well as to compare **GID** identified people with non-*GID* identified inclusion criteria. P-values were calculated using a chi-square test to assess for significant differences ( $p > 0.05$ ) in descriptive characteristics.

### Results

Based on the methods specified above, we identified a total of 27,227 (0.03%) unique transgender people in the OLDW (Figure 1). Specifically, using ICD-9/10 *GID* codes alone (i.e., the method used by previous studies),<sup>22–27</sup> we identified 18,785 (69%) transgender people. Using Endo NOS with a transgender-related procedure code, we added 4,391 (16%) transgender people. We identified an additional 4,051 (15%) people with an Endo NOS code or suggestive transgender-related procedure code and sex-discordant hormone medication. These 27,227 identified transgender people were then classified into TF, TM, and **unclassified** gender identity. The TF group contained 3,959 (14.7%) people, the TM contained 8,694 (32.3%), and 14,242 (52.9%) people were categorized as having **unclassified** gender identity. Further, 332 persons (1.2%) were removed **from our overall transgender cohort** because they had a procedure code and hormone prescription that did

not point to the **same gender identity status** (i.e., a feminizing procedure code with a masculinizing hormone code, or vice versa).

Figure 2a shows the changes in the proportion identified as transgender people by **GID** codes compared with other strategies over time. **We observed that the number of individuals identified by prevalent GID code, and prevalent GID code and other strategies increased from 27% persons identified in 2006 to 96% in 2017, whereas transgender people only identified by other strategies decreased from 73% to 4% for the same years. Similar patterns were observed among TF and TM subsets of the transgender population, although a higher proportion of TF individuals (87%) were identified by GID codes in 2017 relative to TM individuals (77%) (Figures 2b and 2c).**

A comparison of **the populations identified through GID codes and only through other strategies** is shown in Table 1. A higher proportion of transgender people identified using **GID** code were younger (18–29 years: 35.2% vs. 3.8%; 30–39 years: 24.8% vs 12.6%), had diabetes (11.8% vs. 9.4%), depression (49.3% vs. 39.2%), and HIV (2.8% vs. 0.5%) compared to transgender people identified by other strategies. **A higher proportion of older transgender people above 40 years of age were identified by non-GID strategies compared to GID codes alone** (40–49 years: 27.2% vs. 13.2%; 50–64 years: 44.8% vs. 15.8%) ( $p < 0.0001$  for all).

Table 2 reports the descriptive characteristics for the entire transgender sample, TF, TM, and **unclassified** status. A higher proportion of TF were younger (18–29 years: 24.5% vs. 12.7%; 30–39 years: 24.6% vs 14.7%) compared to TM. Both TF and TM people in this dataset were mostly White (66.3% and 70.6%), and most commonly lived in the South (39.5% and 49.8%). A higher proportion of TF (compared to TM) reported diabetes (10.5% vs. 8.6%), depression (44.8% vs. 42.9%), HIV (3.7% vs. 0.4%), and tobacco use (14.8% vs. 13.5%) compared to TM ( $p < 0.0001$  for all).

## Discussion

Using a comprehensive approach based on multiple types of data, including inpatient and outpatient claims with diagnosis and procedure codes, pharmacy claims, and administrative sex marker data, we identified 27,227 (0.03%) unique transgender people, enrolled in commercial insurance and Medicare Advantage from 2006 to 2017. We found an increase in **GID** code identified transgender people over time, along with a simultaneous decrease in non-**GID** code identified people. Our cohort contained more TM than TF persons. A higher proportion of the TF group and individuals identified by **GID** codes were younger and reported more co-morbidities compared to the TM and non-**GID** code identified groups.

Our study extends prior methods to identify transgender people in administrative data<sup>24</sup> by utilizing a combination of codes for transgender-related procedures, medications, and the use of Endocrine NOS diagnosis. Our use of these data elements was inherently conservative, in that we required multiple data elements to define transgender status in the absence of a **GID** code. Despite these conservative decision rules, we identified 18,785 (69%) transgender persons **through GID codes alone**, a number that increased to 27,227

with additional approaches. Thus, the addition of these new approaches accounted for approximately 31% of the transgender people that we identified. While increased sample size is always desirable, the more important contribution that we observed is that the subset of transgender people without a **GID** code were systematically different (e.g., with respect to age and specific types of conditions) from the people that would be found with a **GID** code. Further, identification of a higher proportion of older transgender people using non-**GID** coded strategies highlights the impact of our approach in contributing to identification of a more diverse and representative sample. **The greater prevalence of depression, diabetes, hyperlipidemia, hypertension, and tobacco use in our identified transgender population using multiple data pieces is consistent with findings from other claim-based studies, which used primarily GID codes.<sup>23,24,27</sup> Applying our method of going beyond GID codes in these claim-based datasets will potentially improve generalizability by facilitating identification of additional likely transgender people who would not be picked up with GID codes alone. For example, in the Veterans Affairs (VA), GID documentation is required before initiation of hormone therapy care for transgender Veterans.<sup>37,38</sup> GID codes would thus, not be sufficient to potentially identify those who do not receive hormones in the VA, thereby implying the need for such additional data-driven methods.**

The overall population proportion of transgender people in this sample of 74 million people was only **0.03%**; however, this prevalence is comparable to the **0.026% prevalence proportion reported in another commercial claims database though using GID diagnoses codes.<sup>27</sup> In two studies based on the VA, GID prevalence was reported to be 0.023% in fiscal year 2011, and 0.058% in 2013.<sup>22,23</sup> Using diagnosis codes related to GID, Proctor and colleagues identified 4098 transgender persons in the Medicare program in 2013,<sup>24</sup> while Ewald et al., identified 10,242 transgender Medicare beneficiaries.<sup>25</sup> However, our estimate is 20 times smaller than the estimate from the Williams Institute<sup>39</sup> and other similar recent data,<sup>2</sup> which can be potentially explained by the disparities in commercial insurance coverage in the transgender population as lack of insurance coverage is frequently reported in this population.<sup>14</sup> Further, it could also relate to our conservative approach to opt more for specificity. In addition, transgender people who have not received a *GID* code or trans-related health services, those who do not want to disclose their transgender experience to their healthcare provider, as well as those who do not desire transgender-related medical care, cannot be captured using our method, which may contribute to underascertainment. Further, the difference in the prevalence of transgender people relative to previous estimates, may be due to the fact that previous estimates have been computed from primary survey data collection using self-reported transgender identity.<sup>2,39</sup> While primary survey data collection can enable the identification of individuals not engaged in care, such methods are subject to sampling bias and thus, it is likely we captured individuals who would not otherwise have participated in research, yielding a potentially more representative sample of transgender people engaged in care **and insured by commercial and Medicare Advantage coverage.****

Our finding of an increasing trend in the proportion of transgender people identified by **GID** code in our study between 2006 and 2017 is consistent with what has been reported

previously though for different time frames. Two studies conducted on the Veteran population found a similar increase in transgender population identified using *GID* codes.<sup>22,40</sup> While Blosnich et al. found a two-fold increase in *GID* code identified transgender Veterans from a prevalence estimate of 12.9 in 2002 to 22.9 per 100,000 in 2011,<sup>22</sup> Kauth and colleagues reported this continuing increasing trend with prevalence in 2013 to be 32.9 per 100,000 Veterans.<sup>40</sup> **Even within the Medicare program, Ewald et al reported a 390% increase in the number of beneficiaries identified as transgender from 2088 beneficiaries in 2010 to 10,242 beneficiaries in 2016.**<sup>25</sup>

Interestingly, we found a higher proportion of TM (32%) than TF (15%) in our study cohort. TF generally have been found to outnumber TM in previous research.<sup>41,42</sup> However, the shift in gender balance towards more TM observed in our study has been suggested in a recent study conducted in Australia in which greater proportions of TM (44.1%) were identified as compared to TF (36.3%).<sup>43</sup> While it is possible that our results may reflect changing gender identities within the transgender community at large, it also seems likely that TM individuals face fewer structural barriers to care (e.g., greater access to employment, insurance coverage, and ability to navigate the healthcare system) relative to TF individuals.<sup>44</sup> **However, there is much unknown about what barriers individuals are facing, systematically, between these groups. This potential ability to better access and utilization of care for TM individuals with commercial and Medicare Advantage insurance coverage, relative to TF individuals, could also explain the fact that though TM individuals in our study were much more likely to be older than the TF and unclassified groups, yet seemed to have a lesser burden of many comorbid conditions.** It is also possible that more TF individuals access hormones and other forms of medical gender affirmation outside of insurance, including **unregulated** market hormone use and medical forms of gender affirmation (i.e., silicone injection)<sup>11</sup> and thus, are not captured by database-driven methods based on ICD or CPT codes.

Besides ascertaining a large sample of transgender people, and our use of innovative strategies to expand our **GID** codes-based sample, our study has other unique strengths. Our data source, OLDW, is representative of the commercially insured US population, specifically in the Midwest and South, including Medicare Advantage enrollees, and is large enough to examine minority populations such as transgender people. This is a contrast with much of the previous work on transgender health, which has focused on small convenience samples.

The study also has potential limitations. First, while the methods that we have developed for identifying transgender people in large administrative datasets represent an important step, we are unable to validate these methods using a gold standard such as patient self-identification or chart review, because OLDW does not contain such data. However, we do note that previous studies have at least validated the **GID** code approach **using clinician's text notes in electronic medical record.**<sup>29</sup> In future research, the work we began in this study can be expanded by validating these methods against a gold standard of **patient self-reported gender identity**. Such validations have been conducted previously to assess agreement between electronic medical records and self-reported gender identity.<sup>45</sup> Second, it is much more likely that there are limitations of sensitivity than of positive predictive value



in identifying transgender people using our approach, given the many conservative assumptions built into our approach. This means that while some transgender people may have been missed by our approach (e.g., individuals who have not disclosed their transgender experience to their healthcare provider, those who have not sought out transgender-related healthcare, **some of those with a non-binary identity**), we feel confident that the people we identified as transgender truly fit that description. Third, since administrative data generally only includes binary designations of sex (“male” and “female”), it is likely that some non-binary people were included in our TF, TM, and **unclassified** groups. Fourth, people who receive care through Medicaid, fee-for-service Medicare, TRICARE, or the VA not be well represented in these data; this limitation of the data would need to be addressed in future studies. **Fifth, we selected the CPT codes used in this study largely based on input from our technical expert panel members and not from the exhaustive list of codes issued by Proctor and colleagues.**<sup>24</sup> However, since gender affirming procedures typically involve multiple associated CPT codes, it is unlikely that more than a small handful of cases would have been missed using our list and methodology.

In the absence of available, systematically measured, self-reported gender identity data in administrative databases, findings from this study suggest promising strategies for identifying transgender people in large claims-based datasets, both increasing the sample size and also potentially the representativeness of transgender people so identified. Improving the health of and care for transgender people requires population level data from large datasets for analysis. By using innovative methods to identify transgender individuals, we accomplished a **critical** step in unlocking the potential of large dataset to improve the health of this vulnerable population.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

1. Meerwijk EL, Sevelius JM. Transgender Population Size in the United States: a Meta-Regression of Population-Based Probability Samples. *Am J Public Health*. 2017;107(2):e1–e8.
2. Meyer IH, Brown TN, Herman JL, et al. Demographic Characteristics and Health Status of Transgender Adults in Select US Regions: Behavioral Risk Factor Surveillance System, 2014. *Am J Public Health*. 2017;107(4):582–589. doi:10.2105/AJPH.2016.303648 [PubMed: 28207334]
3. Streed CG Jr, McCarthy EP, Haas JS. Association Between Gender Minority Status and Self-Reported Physical and Mental Health in the United States. *JAMA Intern Med*. 2017;177(8):1210–1212. [PubMed: 28558100]
4. Houssayni S, Nilsen K. Transgender Competent Provider: Identifying Transgender Health Needs, Health Disparities, and Health Coverage. *Kans J Med*. 2018;11(1):1–18. Published 2018 Feb 28.
5. Roberts TK, Fantz CR. Barriers to quality health care for the transgender population. *Clin Biochem*. 2014 7;47(10–11):983–7. [PubMed: 24560655]
6. Streed CG Jr, McCarthy EP, Haas JS. Self-Reported Physical and Mental Health of Gender Nonconforming Transgender Adults in the United States. *LGBT Health*. 2018;5(7):443–448. doi:10.1089/lgbt.2017.0275 [PubMed: 30183498]

7. Mustanski BS, Garofalo R, Emerson EM. Mental health disorders, psychological distress, and suicidality in a diverse sample of lesbian, gay, bisexual, and transgender youths. *Am J Public Health*. 2010;100(12):2426–2432. [PubMed: 20966378]
8. Clements-Nolle K, Marx R, Guzman R, et al. HIV prevalence, risk behaviors, health care use, and mental health status of transgender persons: implications for public health intervention. *Am J Public Health*. 2001;91(6):915–921. [PubMed: 11392934]
9. Lagos D. Looking at Population Health Beyond “Male” and “Female”: Implications of Transgender Identity and Gender Nonconformity for Population Health. *Demography* 2018; 55(6): 2097–2117. [PubMed: 30255426]
10. Buchting FO, Emory KT, Scout, et al. Transgender Use of Cigarettes, Cigars, and E-Cigarettes in a National Study. *Am J Prev Med*. 2017;53(1):e1–e7. doi:10.1016/j.amepre.2016.11.022 [PubMed: 28094133]
11. Hughto JMW, Reisner SL, and Pachankis JE. Transgender stigma and health: A critical review of stigma determinants, mechanisms, and interventions. *Social science & medicine*, 2015: 147, pp.222–231. [PubMed: 26599625]
12. Shires DA, Jaffee KD: Structural discrimination is associated with smoking status among a national sample of transgender individuals. *Nicotine Tob Res* 2016;18:1502–1508 [PubMed: 26438646]
13. Bockting WO, Miner MH, Swinburne Romine RE, et al. Stigma, mental health, and resilience in an online sample of the US transgender population. *Am J Public Health*. 2013;103(5):943–951. [PubMed: 23488522]
14. Sanchez NF, Sanchez JP, Danoff A. Health care utilization, barriers to care, and hormone usage among male-to-female transgender persons in New York City. *Am J Public Health*. 2009;99(4):713–719. [PubMed: 19150911]
15. Conron KJ, Scott G, Stowell GS, et al. Transgender health in Massachusetts: results from a household probability sample of adults. *Am J Public Health*. 2012;102(1):118–122. [PubMed: 22095354]
16. Feldman J, Romine RS, Bockting WO. HIV risk behaviors in the U.S. Transgender population: Prevalence and predictors in a large internet sample. *J Homosex*. 2014;61(11):1558–1588. [PubMed: 25022491]
17. Baker KE. Findings From the Behavioral Risk Factor Surveillance System on Health-Related Quality of Life Among US Transgender Adults, 2014–2017. *JAMA Intern Med*. 2019. doi:10.1001/jamainternmed.2018.
18. Cahill S, and Makadon H. Sexual Orientation and Gender Identity Data Collection in Clinical Settings and in Electronic Health Records: A Key to Ending LGBT Health Disparities. *LGBT Health*. 2014; 1(1): 34–41. [PubMed: 26789508]
19. The GenIUSS Group. (2014). Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys. Herman JL (Ed.). Los Angeles, CA: The Williams Institute.
20. Grasso C, Goldhammer H, Funk D, et al. Required Sexual Orientation and Gender Identity Reporting by US Health Centers: First-Year Data. *American Journal of Public Health*, 2019: 109, 1111–1118. [PubMed: 31219717]
21. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association; 2013.
22. Blosnich JR, Brown GR, Shipherd JC, et al. Prevalence of gender identity disorder and suicide risk among transgender veterans utilizing veterans health administration care. *A J Public Health*. 2013;103(10):e27–32.
23. Brown GR, Jones KT. Mental Health and Medical Health Disparities in 5135 Transgender Veterans Receiving Healthcare in the Veterans Health Administration: A Case-Control Study. *LGBT Health*. 2016;3(2):122–131. [PubMed: 26674598]
24. Proctor K, Haffer SC, Ewald E, et al. Identifying the transgender population in the medicare program. *Transgender Health* 2016; 1 (1): 250–265. [PubMed: 28861539]
25. Ewald ER, Guerino P, Dragon C, et al. Identifying Medicare Beneficiaries Accessing Transgender-Related Care in the Era of ICD-10. *LGBT Health*. 6 2019. Ahead of print.

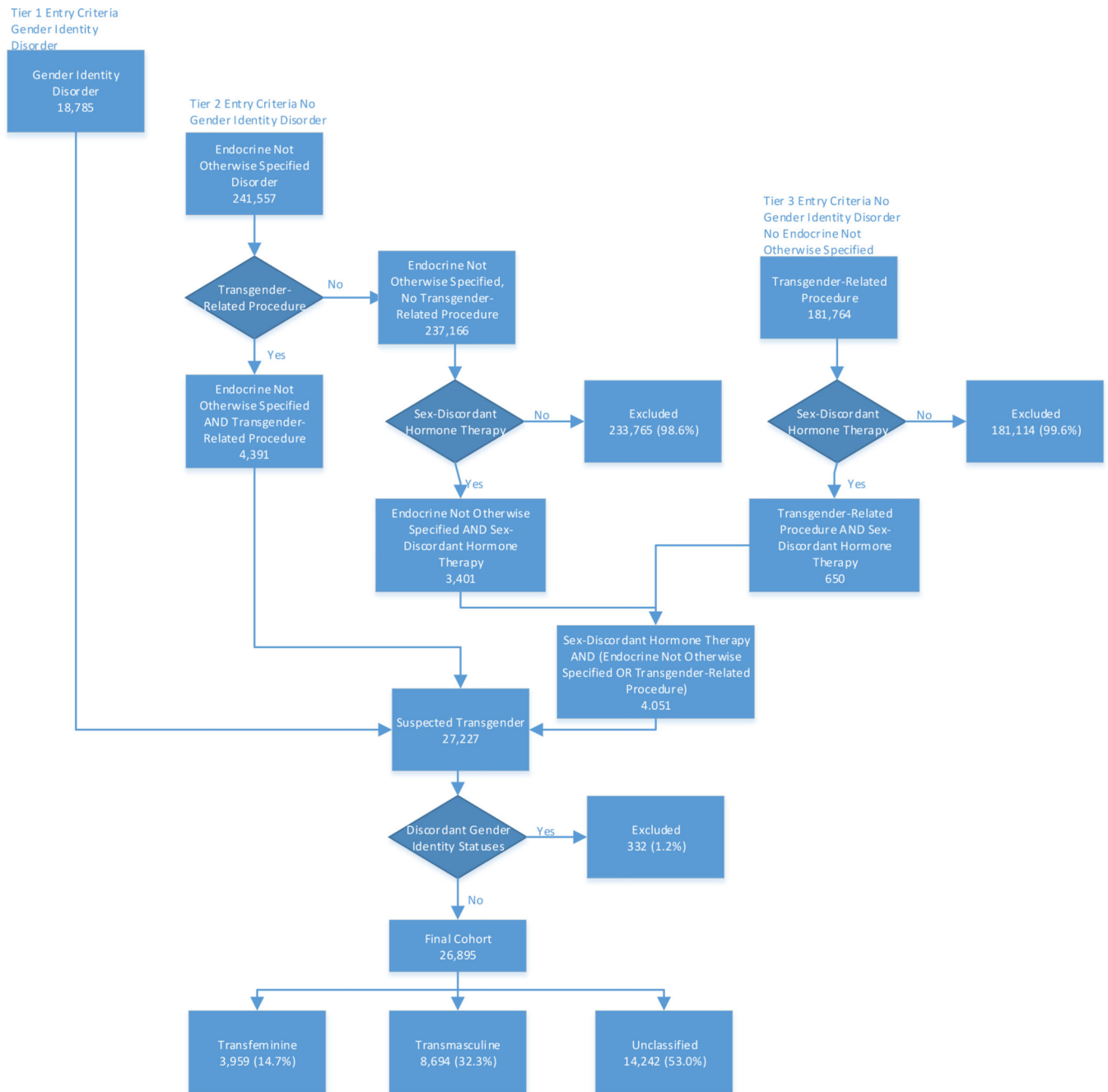
26. Progovac AM, Cook BL, McDowell A, et al. Identifying Gender Minority Patients' Health And Health Care Needs In Administrative Claims Data. *Health Aff (Millwood)* 2018; 37(3): 413–420. [PubMed: 29505378]
27. McDowell A, Progovac AM, Cook BL, & Rose S. Estimating the Health Status of Privately Insured Gender Minority Children and Adults. *LGBT Health*, 2019; 6(6): 289–296. [PubMed: 31314674]
28. Roblin D, Barzilay J, Tolsma D, et al. A novel method for estimating transgender status using electronic medical records. *Ann Epidemiol* 2016; 26 (3): 198–203. [PubMed: 26907539]
29. Blosnich JR, Cashy J, Gordon AJ, 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 7 1; 25(7):905–908. [PubMed: 29635362]
30. Davy Z, Toze M. What Is Gender Dysphoria? A Critical Systematic Narrative Review. *Transgend Health*. 2018;3(1):159–169. Published 2018 Nov 1. doi:10.1089/trgh.2018.0014 [PubMed: 30426079]
31. OptumLabs. OptumLabs and OptumLabs Data Warehouse (OLDW) Descriptions and Citation. Cambridge, MA: n.p., 5 2019. PDF. Reproduced with permission from OptumLabs.
32. Mayer G. Providing Cross-Gender Hormone Therapy for Transgender Patients Boston, MA: The Fenway Institute, 2013. [August 18, 2015]. Available from: [www.lgbthealtheducation.org/wp-content/uploads/Providing-Cross-Gender-Hormone-Therapy-to-Transgender-Patients.pdf](http://www.lgbthealtheducation.org/wp-content/uploads/Providing-Cross-Gender-Hormone-Therapy-to-Transgender-Patients.pdf) (accessed July 20th, 2019)
33. Thompson HM, 2016. Visible Lives: Electronic Health Records and the Biomedical Surveillance of Gender Non-Conformity (Doctoral dissertation). Available from: [https://indigo.uic.edu/bitstream/handle/10027/20205/Thompson\\_Hale.pdf?sequence=1&isAllowed=y](https://indigo.uic.edu/bitstream/handle/10027/20205/Thompson_Hale.pdf?sequence=1&isAllowed=y) (accessed November 29th, 2019)
34. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine Treatment of Gender-Dysphoric/ Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline, *The Journal of Clinical Endocrinology & Metabolism*, Volume 102, Issue 11, 1 11 2017, Pages 3869–3903 [PubMed: 28945902]
35. Basson R. Testosterone therapy for reduced libido in women. *Ther Adv Endocrinol Metab*. 2010;1(4):155–64. [PubMed: 23148160]
36. Schouten J, Michielsens PP. Treatment of cirrhotic ascites. *Acta Gastroenterol Belg*. 2007 4-6;70(2):217–22. [PubMed: 17715638]
37. <http://transveteran.org/wp-content/uploads/2015/03/MtF-HRT-@-VA.pdf>
38. <http://transveteran.org/wp-content/uploads/2015/03/FTM-HRT-@-VA.pdf>
39. Flores AR, Herman JL, Gates GJ, & Brown TNT. How many adults identify as transgender in the United States? Los Angeles: CA: The Williams Institute.
40. Kauth MR, Shipherd JC, Lindsay J, et al. Access to Care for Transgender Veterans in the Veterans Health Administration: 2006–2013. *Am J Public Health*. 2014;104:S532–S4. [PubMed: 25100417]
41. Heylens G, Elaut E, Kreukels BP, et al. Psychiatric characteristics in transsexual individuals: multicentre study in four European countries. *Br J Psychiatry*. 2014;204:151–156. [PubMed: 23869030]
42. Bouman WP, Claes L, Marshall E, et al. Sociodemographic variables, clinical features, and the role of preassessment cross-sex hormones in older trans people. *J Sex Med*. 2016;13:711–719. [PubMed: 26897462]
43. Cheung AS, Ooi O, Leemaqz S, et al. Sociodemographic and Clinical Characteristics of Transgender Adults in Australia. *Transgend Health*. 2018;3(1):229–238. [PubMed: 30596151]
44. Hughto JMW, Rose AJ, Pachankis JE, et al. Barriers to gender transition-related healthcare: Identifying underserved transgender adults in Massachusetts. *Transgender health*, 2017: 2(1), pp.107–118. [PubMed: 29082331]
45. Gerth J, Becerra-Culqui T, Bradlyn A et al. Agreement between medical records and self-reports: Implications of transgender health research. *Rev Endocr Metab Disord*. 2018; 19: 263. [PubMed: 30219985]

Author Manuscript

Author Manuscript

Author Manuscript

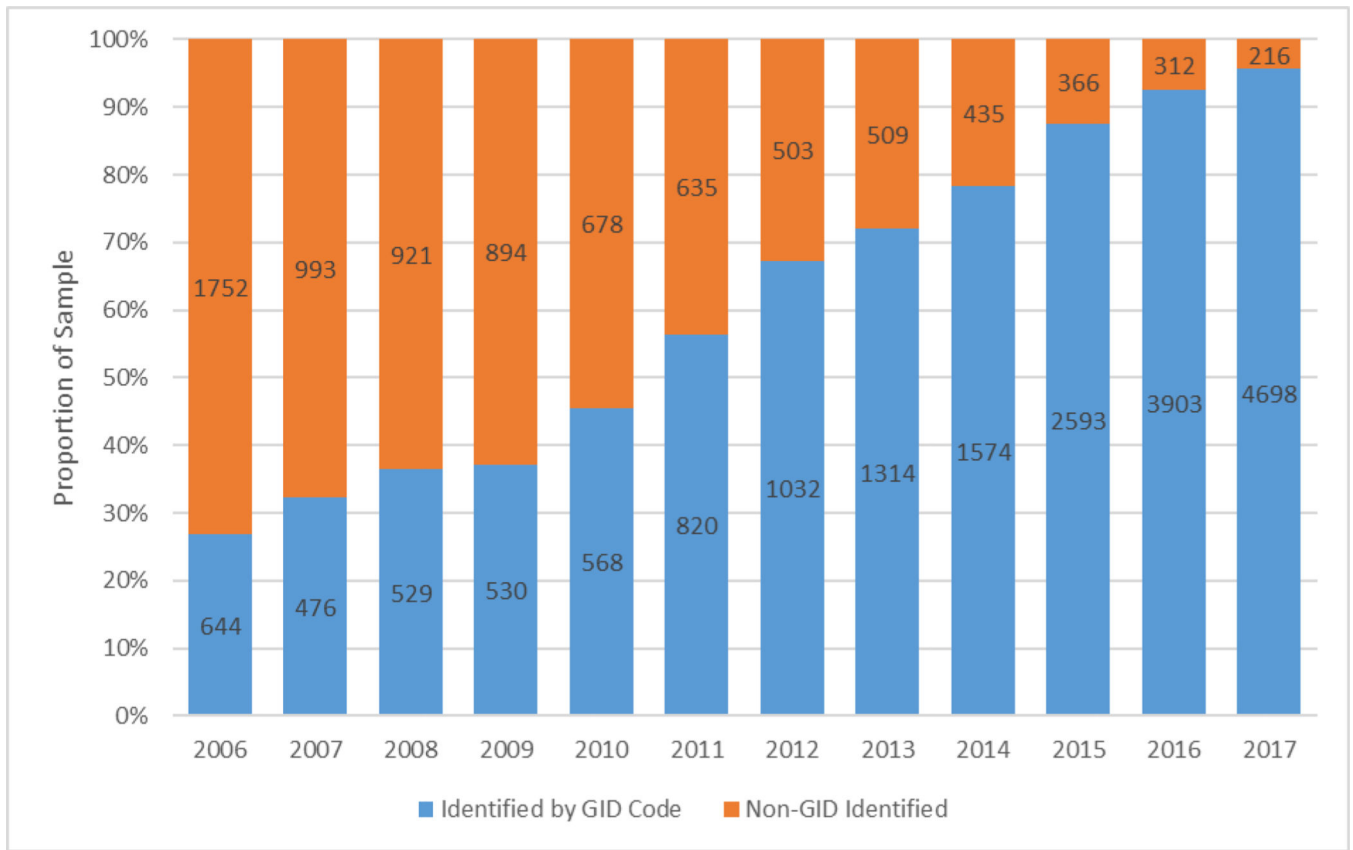
Author Manuscript



**Figure 1. Discordant Gender Identity Statuses:**

332 persons (1.2%) were removed because they had a procedure code and sex-discordant hormone prescription that did not point in the **same gender identity status** (i.e. a procedure code was TF and a sex-discordant hormone prescription was TM, or vice versa).

Cohort Identification

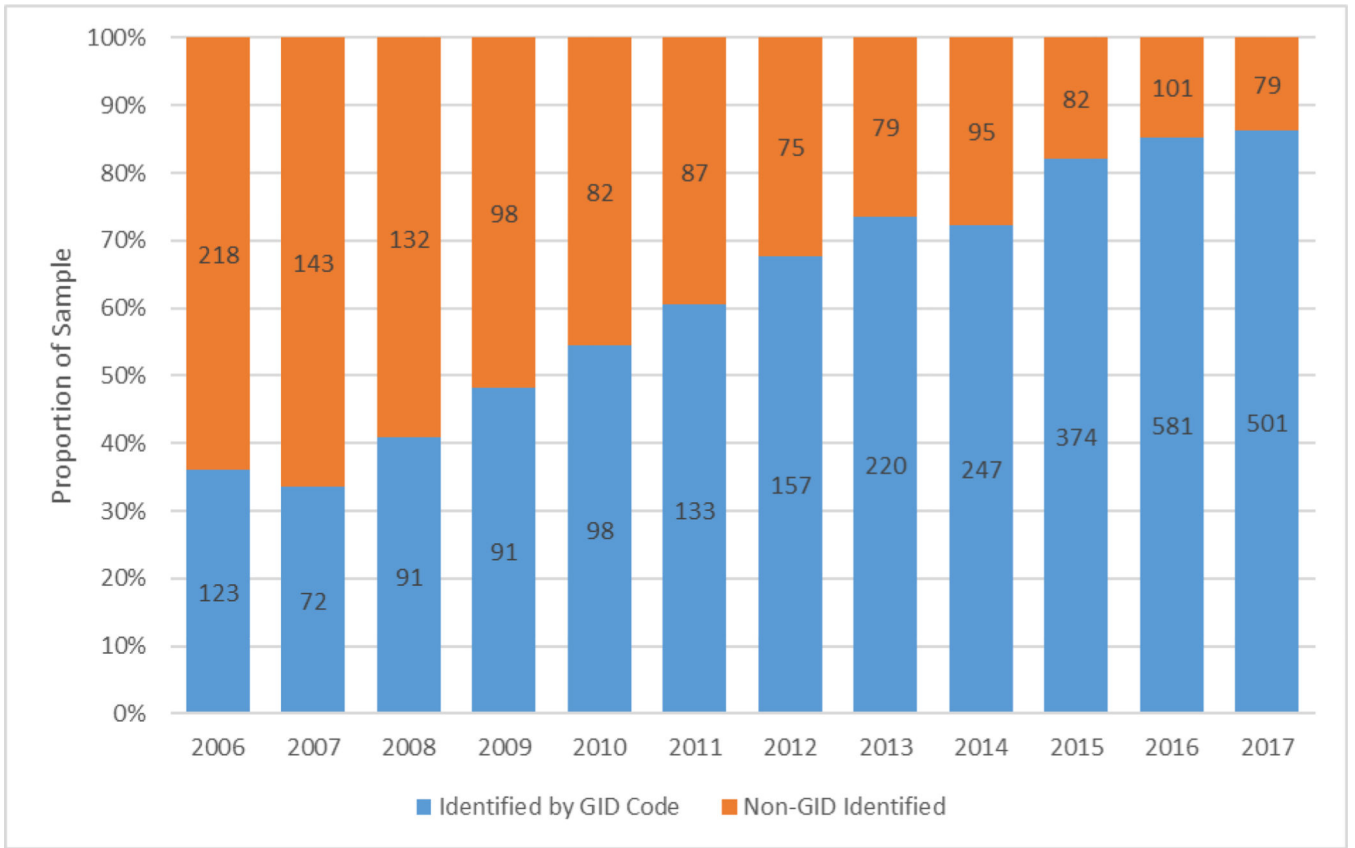


Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

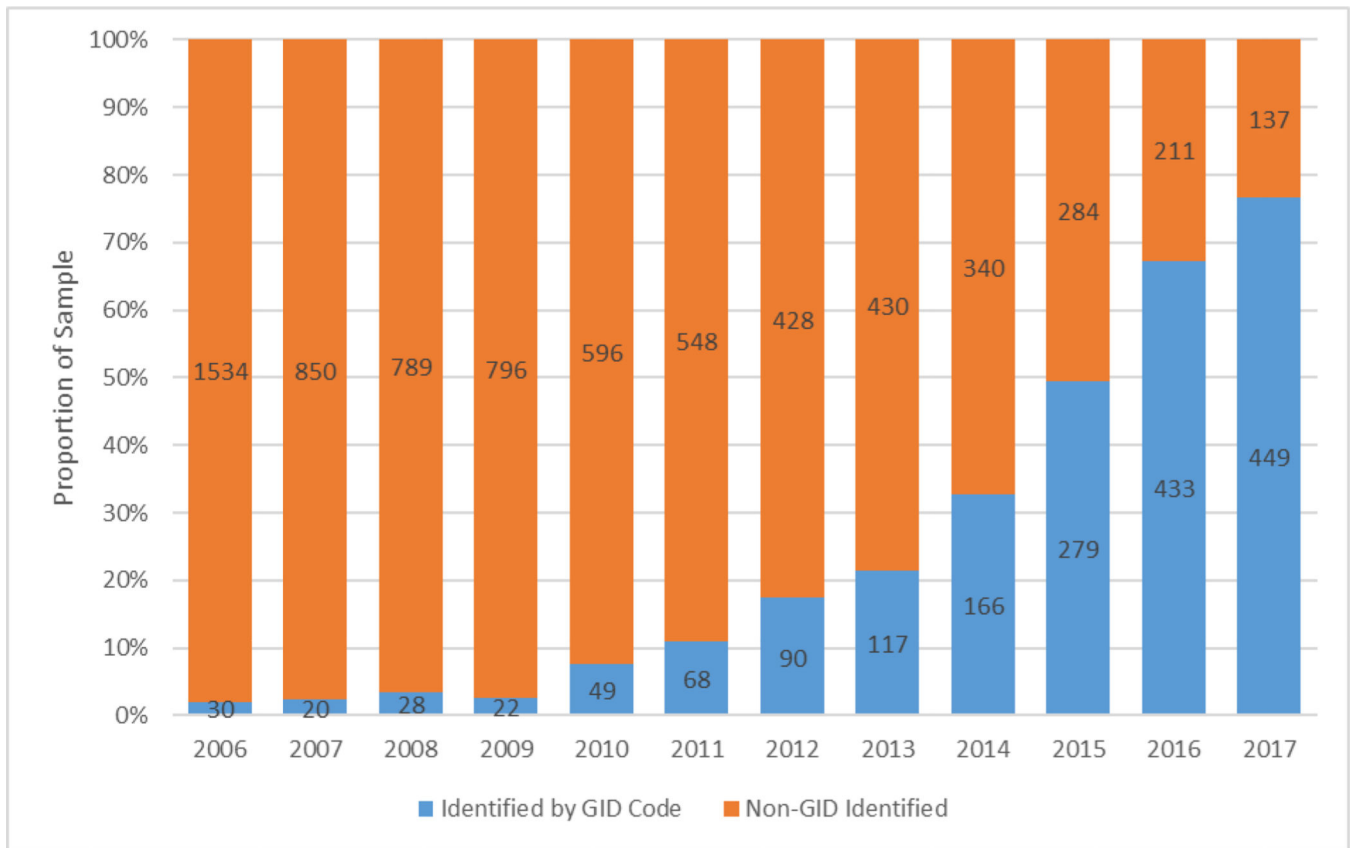


Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 2a, 2b, and 2c.**  
**GID=Gender identity disorder**  
 Identification of Unique Transgender People Over Time  
 Identification of Unique Transfeminine People Over Time  
 Identification of Unique Transmasculine People Over Time

**Table 1.**

Descriptive Characteristics by Transgender Cohort Inclusion Group (N=26,895)

|  | All<br>(N = 26,895) | GID* Code Identified<br>(N = 18,681) | Non-GID Code Identified**<br>(N = 8,214) | p-value |
|--|---------------------|--------------------------------------|--|---------|
| <b>Age</b>                             |                     |                                      |  |         |
| 18–29                                  | 6,884 (25.6%)       | 6,573 (35.2%)                        | 311 (3.8%)                               |         |
| 30–39                                  | 5,661 (21.0%)       | 4,626 (24.8%)                        | 1,035 (12.6%)                            | <0.0001 |
| 40–49                                  | 4,662 (17.3%)       | 2,428 (13.0%)                        | 2,234 (27.2%)                            |         |
| 50–64                                  | 6,636 (24.7%)       | 2,952 (15.8%)                        | 3,684 (44.8%)                            |         |
| 65+                                    | 3,052 (11.3%)       | 2,102 (11.2%)                        | 950 (11.6%)                              |         |
| <b>Race/Ethnicity</b>                  |                     |                                      |  |         |
| White                                  | 16,175 (60.1%)      | 10,488 (56.1%)                       | 5,687 (69.2%)                            |         |
| Black                                  | 2,032 (7.6%)        | 1,535 (8.2%)                         | 497 (6.0%)                               | <0.0001 |
| Hispanic                               | 1,909 (7.1%)        | 1,314 (7.0%)                         | 595 (7.2%)                               |         |
| Asian                                  | 637 (2.4%)          | 447 (2.4%)                           | 190 (2.3%)                               |         |
| Other/Unknown                          | 6,142 (22.8%)       | 4,897 (26.2%)                        | 1,245 (15.2%)                            |         |
| <b>Years of Insurance Coverage</b>     |                     |                                      |  |         |
| 1 to 3 years                           | 3,280 (12.2%)       | 3,091 (16.5%)                        | 189 (2.3%)                               |         |
| 3 to 10 years                          | 11,691 (43.5%)      | 8,999 (48.2%)                        | 2,692 (32.8%)                            | <0.0001 |
| > 10 years                             | 11,924 (44.3%)      | 6,591 (35.1%)                        | 5,333 (64.9%)                            |         |
| <b>Region</b>                          |                     |                                      |  |         |
| Midwest                                | 6,440 (23.9%)       | 4,861 (26.0%)                        | 1,579 (19.2%)                            |         |
| Northeast                              | 3,186 (11.8%)       | 2,576 (13.8%)                        | 610 (7.4%)                               | <0.0001 |
| South                                  | 11,039 (41.0%)      | 6,944 (37.2%)                        | 4,095 (49.8%)                            |         |
| West                                   | >6,208 (>23.1%)     | >4,289 (>23.0%)                      | >1,919 (>23.4%)                          |         |
| Unknown                                | <22 (<0.1%)         | <11 (<0.1%)                          | <11 (<0.1%)                              |         |
| <b>Health Conditions</b>               |                     |                                      |  |         |
| Alcohol Use Disorder                   | 1,514 (5.6%)        | 1,269 (6.8%)                         | 245 (2.9%)                               | <0.0001 |
| Any Diabetes                           | 2,974 (11.1%)       | 2,201 (11.8%)                        | 773 (9.4%)                               | <0.0001 |
| Atherosclerotic Cardiovascular Disease | 1,426 (5.3%)        | 1,096 (5.9%)                         | 330 (4.0%)                               | <0.0001 |
| Cerebrovascular Accident               | 418 (1.5%)          | 314 (1.7%)                           | 104 (1.3%)                               | 0.0336  |



|                                | All<br>(N = 26,895) | GID* Code Identified<br>(N = 18,681) | Non-GID Code Identified**<br>(N = 8,214) | p-value |
|--------------------------------|---------------------|--------------------------------------|--|---------|
| Depression                     | 12,436 (46.2%)      | 9,214 (49.3%)                        | 3,222 (39.2%)                            | <0.0001 |
| Drug Use Disorder              | 2,088 (7.8%)        | 1,729 (9.3%)                         | 359 (4.4%)                               | <0.0001 |
| Human Immunodeficiency Virus   | 560 (2.1%)          | 514 (2.8%)                           | 46 (0.5%)                                | <0.0001 |
| Hyperlipidemia                 | 9,008 (33.5%)       | 6,028 (32.3%)                        | 2,980 (36.3%)                            | <0.0001 |
| Hypertension                   | 7,095 (26.4%)       | 4,799 (25.7%)                        | 2,296 (27.9%)                            | 0.0005  |
| Myocardial Infarction          | 326 (1.2%)          | 274 (1.5%)                           | 52 (0.6%)                                | <0.0001 |
| Post-traumatic Stress Disorder | 1,737 (6.5%)        | 1,467 (7.8%)                         | 270 (3.3%)                               | <0.0001 |
| Tobacco Use                    | 4,064 (15.1%)       | 3,067 (16.4%)                        | 997 (12.1%)                              | <0.0001 |

\* GID=Gender Identity Disorder

\*\* Includes those meeting Endocrine Disorder Not Otherwise Specified Code, Transgender-Related Suggestive Procedure Code, and/or Sex-Discordant Hormone Therapy Criteria

**Table 2.**

Descriptive Characteristics by *Transgender Identity Status* (N=26,895)

|  | TF (N = 3,959) | TM (N = 8,694)  | Unclassified (N = 14,242) | p-value |
|--|----------------|-----------------|---------------------------|---------|
| <b>Age</b>                             |                |                 |                           |         |
| 18-29                                  | 972 (24.5%)    | 1,102 (12.7%)   | 4,810 (33.8%)             |         |
| 30-39                                  | 974 (24.6%)    | 1,276 (14.7%)   | 3,411 (23.9%)             | <0.0001 |
| 40-49                                  | 818 (20.7%)    | 1,996 (22.9%)   | 1,848 (12.9%)             |         |
| 50-64                                  | 834 (21.1%)    | 3,453 (39.7%)   | 2,349 (16.5%)             |         |
| 65+                                    | 361 (9.1%)     | 867 (9.9%)      | 1,824 (12.8%)             |         |
| <b>Race/Ethnicity</b>                  |                |                 |                           |         |
| White                                  | 2,626 (66.3%)  | 6,140 (70.6%)   | 7,409 (52.0%)             |         |
| Black                                  | 364 (9.2%)     | 595 (6.8%)      | 1,073 (7.5%)              | <0.0001 |
| Hispanic                               | 312 (7.9%)     | 639 (7.3%)      | 958 (6.7%)                |         |
| Asian                                  | 135 (3.4%)     | 182 (2.1%)      | 320 (2.2%)                |         |
| Unknown/Other                          | 522 (13.2%)    | 1,138 (13.1%)   | 4,482 (31.5%)             |         |
| <b>Years of Insurance Coverage</b>     |                |                 |                           |         |
| 1 to 3 years                           | 478 (12.1%)    | 443 (5.1%)      | 2,359 (16.6%)             |         |
| 3 to 10 years                          | 1,750 (44.2%)  | 3,088 (35.5%)   | 6,853 (48.1%)             | <0.0001 |
| > 10 years                             | 1,731 (43.7%)  | 5,163 (59.4%)   | 5,030 (35.3%)             |         |
| <b>Region</b>                          |                |                 |                           |         |
| Midwest                                | 1,006 (25.4%)  | 1,764 (20.3%)   | 3,670 (25.8%)             |         |
| Northeast                              | 434 (10.9%)    | 596 (6.9%)      | 2,156 (15.1%)             | <0.0001 |
| South                                  | 1,562 (39.5%)  | 4,328 (49.8%)   | 5,149 (36.1%)             |         |
| West                                   | >946 (>23.9%)  | >1,995 (>22.9%) | >3,256 (>22.9%)           |         |
| Unknown                                | <11 (<0.3%)    | <11 (<0.1%)     | <11 (<0.1%)               |         |
| <b>Health Conditions</b>               |                |                 |                           |         |
| Alcohol Use Disorder                   | 218 (5.5%)     | 309 (3.6%)      | 987 (6.9%)                | <0.0001 |
| Diabetes                               | 415 (10.5%)    | 747 (8.6%)      | 1,812 (12.7%)             | <0.0001 |
| Atherosclerotic Cardiovascular Disease | 170 (4.3%)     | 300 (3.4%)      | 956 (6.7%)                | <0.0001 |
| Cerebrovascular Accident               | 43 (1.1%)      | 92 (1.1%)       | 283 (1.9%)                | <0.0001 |
| Depression                             | 1,772 (44.8%)  | 3,730 (42.9%)   | 6,934 (48.7%)             | <0.0001 |

|                                | TF (N = 3,959) | TM (N = 8,694) | Unclassified (N = 14,242) | p-value |
|--------------------------------|----------------|----------------|---------------------------|---------|
| Drug Use Disorder              | 279 (7.0%)     | 466 (5.4%)     | 1,343 (9.4%)              | <0.0001 |
| Human Immunodeficiency Virus   | 146 (3.7%)     | 37 (0.4%)      | 377 (2.6%)                | <0.0001 |
| Hyperlipidemia                 | 1,252 (31.6%)  | 3,028 (34.8%)  | 4,728 (33.2%)             | 0.0010  |
| Hypertension                   | 1,046 (26.4%)  | 2,241 (25.8%)  | 3,808 (26.7%)             | 0.2762  |
| Myocardial Infarction          | 37 (0.9%)      | 53 (0.6%)      | 236 (1.7%)                | <0.0001 |
| Post-traumatic stress disorder | 227 (5.7%)     | 417 (4.8%)     | 1,093 (7.7%)              | <0.0001 |
| Tobacco Use                    | 585 (14.8%)    | 1,177 (13.5%)  | 2,302 (16.2%)             | <0.0001 |

TF=Transfeminine; TM=Transmasculine