

CARRA Working Paper Series

Working Paper #2015-03

**Likely Transgender Individuals in U.S. Federal Administrative
Records and the 2010 Census**

Benjamin Cerf Harris
U. S. Census Bureau

Center for Administrative Records Research and Applications
U.S. Census Bureau
Washington, D.C. 20233

Paper Issued: May 4, 2015

Disclaimer: This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

Likely Transgender Individuals in U.S. Federal Administrative Records and the 2010 Census

Benjamin Cerf Harris*[†]

U.S. Census Bureau

May 2015

Abstract

This paper utilizes changes to individuals' first names and sex-coding in files from the Social Security Administration (SSA) to identify people likely to be transgender. I first document trends in these transgender-consistent changes and compare them to trends in other types of changes to personal information. I find that transgender-consistent changes are present as early as 1936 and have grown with non-transgender consistent changes. Of the likely transgender individuals alive during 2010, the majority change their names but not their sex-coding. Of those who changed both their names and their sex-coding, most change both pieces of information concurrently, although over a quarter change their name first and their sex-coding 5-6 years later. Linking individuals to their 2010 Census responses shows my approach identifies more transgender members of racial and ethnic minority groups than other studies using, for example, anonymous online surveys. Finally, states with the highest proportion of likely transgender residents have state-wide laws prohibiting discrimination on the basis of gender identity or expression. States with the lowest proportion do not.

Keywords: Administrative Records, Gender Dysphoria, Gender Identity, Record Linkage, Transgender, Transsexual

1 Introduction

Transgender issues and images have been increasingly present in popular media, literature, journalism, and the legal landscape of the United States. The popular television show, *Glee*, featured a transgender high school student in 2012; the 2013 series, *Orange is the New Black*, alerted viewers to unique issues of race and gender identity that inmates and their families face by featuring a

*E-mail: benjamin.c.harris@census.gov. Center for Administrative Records Research and Applications, 4600 Silver Hill Road, Washington, DC 20233. This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

[†]*Acknowledgments:* I am grateful to Lee Badgett, Kitt Carpenter, Richard Gabryszewski, Amy O'Hara, Sonya Rastogi, and Justin Wolfers for helpful comments on earlier drafts; remaining mistakes are mine alone. In addition, I wish to thank the transgender men and women whose data I am lucky to study and whose lived experiences are more complex than I can hope to illustrate herein.

black transgender prisoner played by transgender actor Laverne Cox; and in the same year, retired Navy SEAL Kristin Beck published her memoir about coming out as a transgender woman after 20 years of decorated service (Votta, 2012; Ryan, 2013; Stanglin, 2013). At the same time, California passed a 2013 law protecting the rights of transgender elementary and high school students to utilize restrooms and participate in sports teams that reflect their asserted gender identities; Cook County Jail—and jails around the country—have begun to update their policies around transgender detainees; and a \$1.35 million grant was recently opened to assess how transgender soldiers could serve openly in the military (Lovett, 2013; Lu, 2011; Sosin, 2013).

Despite increasing awareness of transgender issues and numerous qualitative studies on transgender experiences, there is very little quantitative evidence about the social and economic lives of transgender people in the United States. The main challenge—but certainly not the only challenge—is one of data. Random sample surveys are not feasible, since sample sizes would need to be impractically large to ensure adequate coverage of the transgender population.¹ Another important challenge, however, is definitional. Even if there were a large enough survey to ensure a large random sample of transgender people, it is not clear what question (or questions) would address the various understandings of the word ‘transgender’ or the many incarnations that gender identity takes once we abandon the framework of binary sex categories.

This paper identifies likely transgender individuals in a large, confidential, national data source and then links those individuals to their 2010 Census records. Specifically, I analyze the Social Security Administration (SSA) Numerical Identification System (the “Numident”), which contains information on first and middle names, sex-coding, and date of birth for every holder of a Social Security Number (SSN). I identify all adult SSN holders who permanently change their first names—or their middle names, in cases of gender-neutral first names—from traditionally male to traditionally female names (or vice versa). Because the Numident includes information on sex-coding, I also identify those who permanently change their sex-coding in the same direction as their name change. From October 2002 through the summer of 2013, the SSA required evidence of genital sexual reassignment surgery (SRS) for a person to change the sex-coding on his or her records. Therefore, while looking at name changes alone corresponds to a person’s presentation of gender in society, looking at coincident changes in names and sex-coding during certain periods in the SSA’s history corresponds to a definition of transgenderism that depends on surgical intervention.²

Using these two main identification strategies, I explore the following research questions: First, how many adult SSN holders have changed the personal information on their accounts in ways that are consistent with a gender transition, and to what extent have the rates of transgender-consistent changes grown or decreased since the creation of the SSA in 1936? Second, among those who were alive on April 1 2010—the 2010 Census day—how many change their names only, and how many change both their names and their sex-coding? Of those who change both their name and their sex-coding, do they typically change their name first and their sex-coding later, or do they typically change both their name and their sex-coding simultaneously? Finally, by linking to the 2010 Census, I am able to answer questions about basic demographic characteristics and residential patterns.

¹A notable exception is the Massachusetts Behavioral Risk Factor Surveillance System Survey, which is a telephone survey that began to include a question about transgender identity in 2007. Conron *et al.* (2012) pool several years of the survey to obtain estimates of the transgender population in the state; 0.5 percent of their sample identified as transgender, although the survey does not permit researchers to distinguish between transgender men and transgender women.

²Transgender people who undergo SRS are sometimes referred to as ‘transsexuals’, although not all transsexual people undergo surgery.

I find that since the SSA’s inception in 1936, as many as 135,367 individuals changed their name or sex-coding in ways that are consistent with a gender transition. While transgender-consistent claims registered with the SSA are evident in the agency’s earliest years, the frequency of such claims increased dramatically during the SSA’s history. This growth primarily mirrored growth in all other types of claims, reflecting both a growing population as well as an increased use of the SSN as a universal identifier. Of those 135.4 thousand likely transgender individuals, 89,667 were alive during the 2010 Census. Most had changed their name from one gender to another, but 21,833 had also changed their sex-coding.

People who change both their names and their sex-coding most commonly change both pieces of information concurrently, although just over a quarter change their name first and their sex-coding 5–6 years later. Most are in their mid-thirties when they begin to register these changes with the SSA, although (male to female) transgender women frequently begin the process somewhat later in life.

Linking individuals to their responses in the 2010 Census provides even more detail. While the likely transgender individuals in the SSA data are more likely to report their race as White alone, the proportion who identify as Black alone is greater than in other studies that tend to underrepresent non-white populations. Similarly, I am able to identify a greater proportion of likely transgender individuals who report Hispanic origin than in previous work, suggesting my approach may be beneficial in identifying transgender individuals who are members of racial and ethnic minority groups.

Another important finding from the linked data is that individuals in the SSA files who are likely to be transgender are much less likely than non-transgender individuals to respond to the Census question on sex. Those who do respond to the question are much more likely than non-transgender individuals to check both options (that is, both “M” and “F”) on the Census questionnaire. Nearly all surveys, censuses, and forms impose binary responses to questions about gender. These findings point both to the limitations of framing questions about gender in this way as well as to the need for more research on how transgender individuals interpret such questions.

Finally, the linked data allow me to investigate patterns of residential sorting across states. I find that likely transgender individuals in the SSA files are overrepresented in states with laws banning discrimination on the basis of gender identity or expression. In addition, the states where likely transgender individuals are most underrepresented have no anti-discrimination laws.

I am unaware of any other study that looks at the history of transgender individuals’ interactions with the SSA, the ways likely transgender individuals respond to the question about sex on the Census, or residential sorting patterns of transgender individuals in the U.S. While previous research has documented in-sample demographic characteristics and transition pathways, none is based on so many observations. To my knowledge, each of the paper’s findings—particularly in the U.S. context—is an empirical contribution.

This paper also represents a major methodological contribution. The use of administrative records in transgender-related research is rare. I know of only two other studies that use administrative data: Veale (2008) uses New Zealand passport data, and Weitze and Osburg (1996) uses German court records. This is the first North American study to use administrative files to learn about transgender individuals. It is also the first study, to the best of my knowledge, to link the administrative records to other data that has more information about likely transgender individuals’ demographic and residential characteristics. This approach is a major innovation in its own right, but it also paves the way for future research using linked data to learn about earnings and employment, marriage and divorce rates, household composition, and incarceration rates of likely transgender individuals.

The remainder of the paper is as follows. Section 2 provides a brief overview of how we think

about gender identity, the way gender is encoded (and changed) with the SSA, and a review of the literatures on incidence, prevalence, and demographics of transgender populations. Section 3 describes the data and methodology, Section 4 describes the empirical results, and Section 5 provides concluding remarks.

2 Literature Review and Background

2.1 Background

A fundamental distinction in transgender studies is between the concepts of sex and gender. Generally speaking, the term *sex* is used to refer to the biological characteristics that are used to categorize individuals as *male*, *female*, or *intersex*. The concept of *gender*, on the other hand, refers to “the socially constructed roles, behaviors, activities, and attributes that a given society considers appropriate for men and women” (WHO, 2014). Typically, while words like *male* and *female* refer to a person’s sex, words like *masculine*, *feminine*, *man*, and *woman* all refer to expressions of gender. Usually, a person’s gender is assigned by society at birth, based—correctly or incorrectly—on the assessment of that person’s biological sex. The term *transgender* describes anyone who lives their life identifying as and expressing a different gender than the one assigned to them at birth (Spade, 2008). Transgender men are men who were assigned a feminine gender at birth. Transgender women are women who were assigned a masculine gender at birth.³ Transgender individuals may use hormones, elect for gender-affirming surgery (also referred to as *sex change* or *sex reassignment surgery*), some may change their names, some might not. As noted above, my approach will only identify transgender men and women who have legally changed their first names and, in some cases, their sex-coding with the SSA.

Such updates to legal information are likely to be common for transgender individuals in the U.S. It is very important for a person’s documented identity to match the basic characteristics of that person’s social and physical identity. This was recognized as early as 1967, when one of the first U.S. surgeons to perform sex reassignment surgeries (SRS), Harry Benjamin, wrote “[Transsexuals] want a change of their legal status...constant fear of discovery, arrest, and prosecution makes life miserable for them before the operation...even afterwards they have to fight for the necessary legal changes.” (Benjamin, 1967). Although the quotation gives greater weight to SRS than is typically given today, the point is clear. Barriers to updating the sex-coding and name on a driver’s license, birth certificate, passport, or social security card can lead to involuntary disclosure of transgender status to police, sales clerks, employers, and any of the many other people who ask for identification throughout daily life. This can in turn expose transgender people to discriminatory treatment and violence (Grant *et al.*, 2011; Spade, 2008). Thus, the incentives for transgender people to update their identity documentation are strong; however changing gender documentation with several states and federal institutions can be extremely difficult, and this can result in a single person’s gender being inconsistently documented. In a comprehensive inventory of gender documentation policies across various record-keeping agencies at both the state and federal levels, Spade (2008) illustrates this point. For example, until 2013 a transgender man who underwent hormone therapy and chest surgery, but not genital surgery, could change the sex-coding on his passport but not in his SSA records.

To change information associated with an SSN, a person must complete and submit a form SS-5 to the SSA. The SS-5 is used for surname changes, corrections to information on date of birth, corrections to erroneously entered sex-coding, and of course changes to first names and

³People who identify as *genderqueer* do not identify with either of the dominant gender roles.

changes to sex-coding following gender transitions. Attaining original documentation on historic policies about the requirements for sex-coding changes for transgender individuals was very difficult, however the SSA’s History Museum and Archives proved to be helpful in this regard. Originally, individuals could apply for, obtain, and update the information associated with a Social Security card by allegation alone. Indeed, Puckett (2009) points out that at the beginning of the SSA, the information on a person’s record was based exclusively on what that person asserted to be true. During the 1970s, policies were developed to tighten security and limit fraud, however, the first written policy for changing sex information associated with a person’s account appeared in November, 1980. This policy required “...clinical or medical records, or other combination of documents showing the sex change, or any medical record showing the sex-change surgery has been *started*” (Richard Gabryszewski, personal communication, 4 December 2014). This policy did not change until October 2002, when the SSA adopted the stricter requirement that the surgery needed to be completed prior to the application (Spade, 2008). Less than a year later, in August 2003, the SSA further required that a surgeon or attending physician provide a letter verifying the surgery had been completed (Richard Gabryszewski, personal communication, 4 December 2014). This was the policy in place during the 2010 Census, although the SSA recently updated the policy and no longer requires surgery (SSA, 2013).

2.2 Literature

2.2.1 Prevalence studies

The earliest scientific studies of transgender issues began by asking the most fundamental questions: how many people are transgender, and what is the ratio of transgender women (male to female) to transgender men (female to male)? These studies are almost exclusively concerned with quantifying the number of people who request or undergo SRS. Benjamin (1967), drawing on records from his own practice, indicated that for every 8 male to female patients, he had one female to male patient, and this basic relationship persists in subsequent studies. Using information on applications for SRS in the late 1960s, Pauly (1968) estimated that 2,000 transgender women and 500 transgender men were living in the United States (for a ratio of 4:1). The vast majority of these kinds of studies, however, are based on European data. Wålinder (1968) conducted a census of all psychiatric practices in Sweden and found 1 in 54,000 individuals asked for or received SRS, and a much lower ratio of 2.5:1 transgender women to men. Other studies based on national medical registries or censuses of practitioners have been conducted in Australia, Belgium, Japan, the Netherlands, Scotland, Singapore, and the United Kingdom (Bakker *et al.*, 1993; De Cuypere *et al.*, 2007; Landén *et al.*, 1996; Okabe *et al.*, 2008; Reed *et al.*, 2009; Ross *et al.*, 1981; Wilson *et al.*, 1999; Tsoi, 1988). The estimates range somewhat broadly, and it may be helpful to put them in the context of the U.S. during the 2010 Census. Using the estimates above and the 2010 Census figures for the population over 15, this earlier research suggests there would be as few as 370 up to 56.8 thousand transgender individuals in the U.S. during the 2010 Census (US Census Bureau, 2010). The variation in estimates is due not only to the variety of contexts, methodologies, and time periods under which the studies were conducted, but also to the range of definitions used for considering a person as transgender. Notably, each of these studies requires a person to at least request SRS.

The two studies that are methodologically most similar to this one are by Witte and Osburg (1996) and Veale (2008). Both utilize administrative or court records, and in both cases their counts are not limited to individuals who have undergone SRS. Using court records from (then West) Germany from 1981 to 1991, Witte and Osburg (1996) count the number of petitions for legal name

and sex-coding changes under the “Act on the Changing of First Names and the Determination of Sex Membership in Special Cases”, also known as *Das Transsexuellen-Gesetz* or *Transsexuals’ Act*. The authors estimate that up to 2.4 per 100,000 of the adult German population submitted an application for either a name change or a sex-coding change (or both); of these, for every 2.3 transgender women there was one transgender man. Extrapolating these figures to the U.S. data, as above, yields estimates of approximately 2,000 transgender women and 900 transgender men, a relatively small number. Despite their low estimate, a unique feature of the German data is that it also allows Weitze and Osburg (1996) to calculate the percentage of those who requested a name change but not a sex-coding change. The authors find that up to 30 percent asked only for a name change.

Veale (2008) evaluates administrative records from the New Zealand Department of Internal Affairs Passport Office. Beginning in 1995, New Zealand passport holders have been allowed to change the sex-coding on their passports from “M” or “F” to “X” if they declared they lived as a gender other than the one reflected on their passport and if they had legally changed their first name. Passport holders may change the sex-coding from “M” to “F” (or vice versa) if they undergo SRS. Veale (2008) estimates prevalence rates at 1 out of 3,639 for transgender women and 1 out of 22,714 for transgender men, which would indicate approximately 38.7 thousand transgender adults reside in the U.S. Importantly, only 28.6 percent of the people in the New Zealand passport data change their sex-coding to an “M” or “F”, suggesting very few transgender individuals actually undergo SRS.

Indeed, focusing solely on those who have requested or undergone SRS may severely underrepresent the size of transgender population. In a recent study, Gates (2011) models the size of the U.S. transgender population using results from two state-level population surveys that allow transgender respondents to self-identify as such. Gates (2011) estimates that just over 697 thousand transgender individuals reside in the U.S. In both cases, the surveys used broad definitions of what it means to be transgender and placed no restriction that respondents must have undergone SRS.

2.2.2 Demographics

Most U.S. studies that are based on small-scale samples from focus groups, personal interviews, or similar research designs include descriptions of the sample demographics (e.g., Bender-Baird, 2013; Factor and Rothblum, 2007; Schilt, 2006; Schilt and Wiswall, 2008). Typically, as authors are quick to point out, the samples are disproportionately white and tend to have higher educational attainment than the general population. Such research designs give extremely detailed information about respondents’ experiences; and despite the hazards associated with extrapolating from such samples, most of what we know comes from this body of research. Very little research using large data sets or representative survey data explores the demographic characteristics of transgender individuals living in the U.S. Of those that do, such studies tend to use two broad categories of data.

The first and more common research design uses targeted sampling (e.g., convenience, snowball, or internet sampling), but the scale is substantially larger than in the more qualitative studies described above. Among these, the best known is probably the National Transgender Discrimination Survey (NTDS), a national online and paper-based survey of 6,456 transgender and gender non-conforming individuals during 2008 (Grant *et al.*, 2011). Survey respondents were recruited from several community centers, social organizations, list serves, and some in-person paper surveys. The NTDS has extremely detailed information on current and past experiences of the transgender individuals who participated in the survey, and it is of great value. Still the NTDS sample, like those described above, underrepresents non-white populations and over-represents highly educated

populations. Additional differences between the in-person and online respondents demonstrate the importance of data collection methods for the demographic composition of the sample (Reisner *et al.*, 2014). Other examples of large-scale convenience samples include an internet survey by Rosser *et al.* (2007), the California Transgender Economic Health Survey, and the Massachusetts Department of Public Health MassEquality Survey (Hartzell *et al.*, 2009; Landers and Gilsanz, 2009). Each of these studies provides important information, but the samples appear to be quite skewed in terms of respondents’ demographic characteristics.

The second set of data come from surveys that use random sampling. While these data are typically regarded as providing the most representative sample, it is extremely expensive to carry out a survey with a large enough sample of transgender individuals to make inferences about the population, and one must always consider the role of response bias. I am aware of only two such surveys, both of which are at the state level. The California LGBT Tobacco Use Survey employed stratified random sampling in a telephone survey of the LGBT population in California during 2003 and 2004. While the survey report did not present demographic characteristics for transgender individuals and instead pooled all lesbian, gay, bisexual, and transgender respondents together, it is important to note that the sample was less likely to report non-white race or Hispanic origin than the rest of the population (Bye *et al.*, 2005). The Massachusetts Behavioral Risk Factor Surveillance System (MA-BRFSS) is a representative sample of households in Massachusetts, and, to my knowledge, the only survey that is representative of the entire population while also permitting identification of transgender respondents. These data are unique in that the transgender respondents are *less* likely to report their race as white and *much more* likely to report Hispanic origin than the non-transgender population (Conron *et al.*, 2012).⁴ Unfortunately, the MA-BRFSS does not allow researchers to distinguish between transgender men and transgender women.

3 Analytic Framework

3.1 SSA administrative records

The data used in this paper are administrative records from the SSA, provided to the Census Bureau under Titles 5, 13, and 42 of the U.S. Code.⁵ The SSA Numident is an administrative database containing the name, date of birth, and a sex-coding for every SSN holder. Furthermore, the Numident contains a record for every claim that changes the information associated with a given SSN. The Census Bureau acquires the Numident through a data sharing agreement with SSA, and uses the data to facilitate record linkage and statistical operations. In addition, the SSA sends quarterly updates to inform the Census of the creation of new SSNs as well as changes and corrections of information associated with existing SSNs. Thus, the Numident contains a record of every claim for the *population* of SSN holders as of the most recent update. This study evaluates all the Numident records from 1936 to the end of 2010. These administrative records, alone, are sufficient to answer the first three research questions about the historic trends in likely-gender transitions, the number of likely transgender SSN holders as of the 2010 Census, and the paths people take with regard to timing of name and sex-coding changes. Later, I will discuss how linking

⁴Conron *et al.* (2012) find that 32.4 percent of the transgender sample in the MA-BRFSS reported Hispanic origin, while only 9.0 percent of the non-transgender sample reported Hispanic origin. A relatively small sample of transgender respondents surely plays a role in this surprising result, but the lower bound of the 95 percent confidence interval for the transgender sample is still 5.2 percentage points higher than the upper bound for the non-transgender sample, indicating a need for further investigation into these estimates.

⁵Specifically, 5 U.S.C. §552a (b) (4), 13 U.S.C. §6, and 42 U.S.C §902 and §1306. Ensuring confidentiality of the data are the primary concern at the Census Bureau; data stewardship training is required annually for all staff, and severe penalties exist for any misuse of data.

individuals to their 2010 Census responses will help answer the remaining research questions on the demographics and geographic characteristics of the likely transgender individuals in the SSA data.⁶

Table 1 shows the layout of the SSA files, using fictitious data and variable names for illustrative purposes. The first column shows each record’s Protected Identification Key (PIK), or unique person identifier used by the Census Bureau to link individuals’ records across data sets while simultaneously protecting their confidential information. Section 3.2 discusses PIK assignment and record linkage in greater depth. Within a given PIK, each row represents a claim. Any time an SSN is created or information associated with an existing SSN is changed, that event is registered as a claim. Table 1 provides examples of claims associated with changes to given names, surnames, date of birth, and sex-coding. For instance, John Doe (PIK=01) corrected a transposed month and day in his date of birth record; Jane Smith (PIK=02) changed her surname to Doe, consistent with the popular convention of married women adopting their husbands’ surnames; and John Miller’s (PIK=08) apparently mis-entered sex-coding was corrected when he was under 2 years old.

In order to identify likely transgender people, the first step is to distinguish people whose first name or sex-coding changed from those whose first name and sex-coding were stable. In the example in Table 1, PIKs 01–06 are “stable”, while PIKs 07–10 feature a change in the first name, sex-coding, or both.

Table 2 shows the total number of claims, the total number of unique PIKs, and the average number of claims per PIK within the data. These figures are also given for the set of PIKs with stable first name and sex-coding information, by sex, and the set of PIKs featuring a change in first name, sex-coding, or both. After dropping those who were not at least 16 years old as of April 1 2010 (Census day), my data include 828.0 million claims associated with 374.2 million unique PIKs. The average person had 2.2 claims associated with their Numident record. For the vast majority of PIKs—347.8 million, or 93.0 percent—we observe no first name or sex-coding changes. Nevertheless, these “stable” records have 2.2 claims associated with their Numident record on average. Changes to these records might include corrections to erroneous dates of birth (as in John Doe’s (PIK=01)), changes of surnames (as in Jane Smith’s record (PIK=02)), or records of death (as in Jane Johnson’s record (PIK=09)). There is, however, substantial heterogeneity across males and females with regard to the average number of claims per person. While the average male with a stable record makes 1.8 claims, the average female makes 2.4 claims. This is likely due to the common custom of adopting the surname of one’s husband. The remaining 26.3 million “non-stable” records exhibit at some point a change in their first names, sex-coding, or both. Of the non-stable records, there were 24.7 million changes to first names but not sex-coding (93.6 percent), 1.4 million changes to sex-coding but not name (5.2 percent), and only 328.4 thousand changes to both (1.2 percent).

3.1.1 Modeling gender transitions

Only some of the people who have changes to their first names or sex-coding (or both) are likely to be transgender. To identify those who are most likely to have undergone a gender transition, I look first for those who changed their name from a traditionally male name to a traditionally female

⁶The Census Bureau is analyzing the use of administrative records to improve data quality and reduce respondent burden in its surveys and censuses. Record linkage in Census Bureau processes or products requires administrative records with consistent and high-quality information on name, date of birth, and sex. Researching the transgender population provides valuable insight to researchers and practitioners by exploring how name and sex-coding changes legitimately appear in the SSA file and other administrative data sources. This is an important step in understanding that changes in first name and sex data are not necessarily anomalies or errors.

name, or vice versa. Among those whose name changes are consistent with gender transition, I then identify people whose sex-coding changed in the same direction as their first name. To simplify my analysis and to reduce the likelihood of false-positive assignment of transgender status, I ignore those who changed their first name or sex-coding before turning 16 years old as well as those who later changed their first name or sex-coding back to the original version. Note that these rules need not apply in principle. People can and do transition before turning 16. Further, gender can be a fluid concept; people can (and do) transition from one gender to another and back again.⁷ These rules are instead meant to limit the possibility of falsely identifying a person as transgender.

3.1.2 Determining the gender of a name

Using the 347.8 million stable records, I construct name-sex crosswalk tables showing, for every name, the proportion whose sex-coding is “M” and the proportion that is “F”. Since the gender of names changes over time, I generate these crosswalks by birth-decade (Barry III and Harper, 1993; Lieberman *et al.*, 2000; Rossi, 1965).⁸ I then link the crosswalk, by name and birth decade, onto the 25.0 million records with *non-stable* first names, first by the original first name and then by the new first name. I can identify name changes that are consistent with gender transitions by comparing the likelihood that the original first name is male to the likelihood the new first name is male.

Within the context of Table 1, PIKs 1–6 are used to construct the name-sex crosswalks. The name John is associated with an “M” sex-coding 100 percent of the time, and the name Jane is never associated with an “M.” However, the name Val is associated with an “M” 50 percent of the time.

Figure 1 shows the distribution of likelihoods that a name is male. The sharply bi-modal shape of the distribution makes it immediately clear that most names are either strongly male or strongly female. Approximately 40 percent of all names in the 347.8 million stable records *always* belonged to people whose sex-coding was “M”, while about 52 percent of all names belonged to people whose sex-coding was “F”. Only about 3 percent of all names were perfectly gender-neutral. While Figure 1 characterizes the distribution of names, Table 3 presents the number of individuals with stable records whose names were male or female 90 percent, 95 percent, and 99 percent of the time. Panel A shows that most males had strongly male names. Only 6.3 percent of males had names that were associated with males in fewer than 90 percent of the cases, and 81.5 percent of males had names that were associated with males 99 percent of the time or more. Panel B shows similar results for females, although females are slightly more likely to have names that are female in fewer than 90 percent of the cases. The take-away message is that there is very little ambiguity in determining the gender of a person’s name: very few given names are gender-neutral, and very few people have gender-neutral names.

3.1.3 Determining which name changes are likely transgender

I have no prior notion of what likelihood threshold is sufficient to confidently classify a name as likely male or likely female, so I present results for three progressively demanding sets of thresholds: 10 and 90 percent, 5 and 95 percent, and 1 and 99 percent. That is, for the first set, a name is categorized as likely male (female) if 90 percent or greater (10 percent or lower) of all stable records

⁷For example, Weitze and Osburg (1996) identified a small number of people who transition from one gender to another and back again in Germany.

⁸Some records list the first name as a single letter (e.g., J DOE). Other names are so rare that they appear fewer than 5 times in a decade. These names are not included in the crosswalks.

with that name had a sex-coding that was “M” (“F”). I call this the *90 percent confidence threshold*. I similarly define the *95 percent confidence threshold* and the *99 percent confidence threshold*. Note that name combinations that meet the 99 percent threshold also meet the 90 and 95 percent thresholds, so these are not mutually exclusive categories. If a person’s first name is gender-neutral (i.e., if it does not meet the 90 percent confidence threshold or higher), then I use the gender of that person’s middle name.

This framework considers a name change to be consistent with a gender transition if (a) both the original and the new first name fall within the specified confidence threshold, (b) the gender of the original first name matches the original sex-coding on the account, (c) the gender of the new first name is different than the gender of the original first name, and (d) the name change is not reversed later.⁹ To illustrate the application of these rules, we return to Table 1. Jane Johnson’s (PIK=09) name change from John is considered consistent with a gender transition, since both John and Jane surpass the 90 percent confidence threshold. The same logic applies to John Brown’s (PIK=10) name change from Jane Brown. However, Jane Thompson’s (PIK=07) name change from Val is not classified as consistent with a gender transition, since the name Val does not meet the minimum confidence threshold. If this person’s original middle name was identifiably male, then we would classify Jane Thompson as transgender.

Of course, the actual cases in the Numident are not as straightforward as those described in Table 1. In particular, both accidental perturbations in how an individual’s first name is spelled over the lifecycle (e.g., transposing letters), as well as intentional differences (e.g., using nicknames), will lead that person to be identified as a name-changer and potentially transgender. Falsely identifying someone as a name-changer is problematic in its own right, but it can lead to serious measurement error when we also use the gender of a person’s middle name. For example, it is common for married women to list their maiden surname in the middle name field. Since many surnames are also used as male’s names (e.g., Harrison, Jefferson, etc.), false-positive identification of a woman as a likely transgender man can be an issue. To illustrate using fake names, suppose “MARY JANE RICHARD” gets married to “WILLIAM WILLIAMS.” Mary files an SS-5 to adopt her married name and change her maiden name to her middle name. Either when filling out the form or when transcribing it, two letters in her first name are accidentally transposed, so her new name is encoded as “MAYR RICHARD WILLIAMS”. Because her first name changes from “MARY” to “MAYR”, she falls into the set of potential transgender individuals. “MARY” is a female name, but “MAYR” is neither male nor female. If we use the gender of her new middle name, “RICHARD”, she will be falsely classified as likely transgender, since her first name changed from the traditionally female “MARY” to the traditionally male “RICHARD”. To avoid these scenarios, I use the SAS SPEDIS function to eliminate people with small spelling distances between the new first name and the original first name. I also exclude using the gender of the (new) middle name if that middle name ever appeared as that person’s last name. The benefit of these rules is that they reduce the chances of false-positive classification of individuals as likely transgender. The cost, however, is many transgender people in the data will be identified as non-transgender. Using information on sex-coding, however, can reduce even further the chances of false-positive measurement error.

⁹The requirement that the gender of the original name matches the sex-coding on the account ensures we do not falsely identify people as likely transgender if they change their gender non-conforming names to gender-conforming names later in life. That is, if a boy named Sue changes his name to Stu, requirement (b) in the list above prevents us from accidentally counting this person as likely transgender.

3.1.4 Sex-coding changes

Using information on name changes alone is a more inclusive strategy for identifying likely transgender individuals, since it does not require changes in sex-coding, which typically require costly and not always preferred genital sexual reassignment surgery. Nevertheless, identifying those who changed the sex-coding associated with their SSNs is interesting in its own right *and* can provide additional reductions of false-positive measurement error. Returning to the example in Table 1, we see that four years after Jane Johnson (PIK=09) changed her first name from John, she changed her sex-coding from “M” to “F”. At the same time that John Brown (PIK=10) changed his name from Jane, he also changed his sex-coding from “F” to “M”. John Miller’s (PIK=08) sex-coding change, however, is not consistent with a gender transition for two reasons. First, his first name did not change. Second, his sex-coding changed before he turned 16. Such scenarios are relatively common among sex-coding changes, and appear to reflect corrections to mis-entered sex-coding during the process of applying for an SSN, which is why the age-rule is important for our estimates.

Once likely transgender individuals are identified in the SSA data (allowing the identification to vary by confidence threshold of names and whether or not their sex-coding changes on the account) it is not difficult to calculate the number transgender-consistent claims per year and the number of SSN holders identified as likely transgender who were alive during the 2010 Census. Similarly, it is straightforward to explore how many people change their names and sex-coding at the same time, how many change their names first, and (for the latter group) what the average number of years is between the name and the sex-coding change. In order to learn about race, ethnicity, and geographic characteristics, however, we must link the SSA records to the most recent decennial census.

3.2 Record linkage

To ensure non-disclosure of SSNs other Personally Identifiable Information (PII), the Census Bureau assigns each individual a PIK, which is a unique identifier used internally by the Census to link individuals across data sets for statistical and research purposes. The Census Bureau’s Center for Administrative Records Research and Applications (CARRA) uses probability record linkage techniques and personal information such as name, date of birth, and residential location to assign PIKs to individuals’ census records, where possible, through the Person Identification Validation System (PVS) (see Wagner and Layne, 2014, for more details). The Census Bureau developed the PVS as part of its ongoing research to improve data quality and reduce costs of data collection.

The 2010 Census collected information on 308,745,538 individuals residing in the U.S. during April 2010.¹⁰ Of these, 280,989,153 records (91.0 percent) were assigned PIKs. These 281.0 million records included 10,486,988 duplicate PIKs. This occurs when two census records have the same probability of being a match to the person represented by a given PIK. I am able to resolve 53 percent of the duplicates by dropping records of 5,552,757 individuals who were identical to their duplicates in terms of state of residence, reported sex, race, Hispanic origin, and age (the main census variables used in my analysis). Because the information is identical for these records, dropping all but one can be done without introducing new measurement error.

I next reduce the number of duplicates that need to be resolved by linking the remaining census records (including the remaining 4.9 million duplicates) to the 89,667 individuals in the SSA Numident who were determined to be likely transgender. The match yielded 90,686 potential linkages, of which 88,663 are unique matches and 1,019 are duplicates. I begin by keeping only duplicates with the closest agreement in age between the SSA records and the census records, which

¹⁰This includes residents of group quarters, Puerto Rico, and those under 16 years of age.

eliminates 346 duplicate records. I then randomly select one of the within-state duplicates, which eliminates 397 additional duplicates. This will not introduce bias to my estimates of the geographic distribution of likely transgender individuals, but it will introduce attenuation bias to estimates of race, Hispanic origin, sex-reporting, and age. Finally, I randomly drop the remaining 276 duplicate PIKs. I am left with 89,667 valid links, a 100 percent match rate.

4 Results

4.1 Results from administrative data

4.1.1 Transgender-consistent claims, 1936–2010

Table 4 shows the number of individuals appearing in the SSA Numident between 1936 and 2010 whose records meet the criteria, described above, for being likely transgender. The top Panel A gives results for the entire universe of SSN holders, while Panels B and C give estimates for Male to Female (MTF) and Female to Male (FTM) likely transgender individuals, respectively. The first row within each panel gives the number of people who changed their first name in a way consistent with gender transition, and the second row gives the number who changed their sex-coding in the same direction as their name change. Thus, the second row within a panel imposes more restrictive requirements for being identified as likely transgender than the first row within a panel. Moving from left to right across columns gives the results using the 90 percent, 95 percent, and 99 percent confidence thresholds for the gender of the original and new first names. Since these confidence thresholds are increasingly demanding, the most conservative results appear in the bottom-right cell of each panel, and the most inclusive estimate appears in the top-left cell.

According to the most inclusive criteria, which require a name change consistent with gender transition using the 90 percent confidence threshold, 135,367 individuals in the data are likely to have transitioned gender between 1936 and 2010. As the name confidence thresholds become more demanding, moving rightward across the columns, the figures attenuate; the 95 percent confidence threshold yields a count of 106,550 while 64,738 individuals have name changes that meet the 99 percent confidence threshold. These figures include any individuals who changed their first name from a traditionally female name to a traditionally male name (or vice versa), which is consistent with that person taking steps to be socially recognized as a gender distinct from the sex assigned at birth, regardless of whether that person has taken steps to alter his or her physical appearance through hormone treatment or surgery. The second row within each panel shows results requiring individuals to alter their sex-coding in the same direction as their name change. Between 1936 and 2010, the number of adult SSN holders who changed their first name and sex-coding in the same direction ranges from 21,981 (using the 99 percent confidence threshold) to 30,006 (using the 90 percent confidence threshold). It is difficult to know how the SSA responded to requests to change the sex-coding on a person’s account during the first several decades of its existence, however from 1980 through 2013, evidence of scheduled or completed SRS was necessary to change the sex-coding on a person’s account. The bottom row of each panel, therefore, may include individuals whose gender transition included surgical intervention; uncertainty about how individual cases were handled prior to 1980 as well as how closely later policies were followed should be taken into account when interpreting these figures.

We now turn to the bottom two panels, which break Panel A out by gender. Looking at the bottom rows of Panels B and C, which require corresponding name and sex-coding changes, likely transgender women (MTF) make up a larger share of the total number of likely transgender individuals than do likely transgender men (FTM). This is consistent with much of the previous literature,

which finds that the rate of MTF transitions involving SRS is higher than FTM transitions involving SRS (see, for example, APA, 2000; Bakker *et al.*, 1993; Tsoi, 1988; Pauly, 1968; Wålinder, 1968). However, this relationship is reversed when using name changes but not sex-coding changes. Two possible mechanisms may underlie this reversal. On the one hand, since for over 30 years the SSA required evidence of scheduled or completed genital SRS to change the sex-coding on one’s account, differences in the cost, complexity, and effectiveness of surgical interventions for FTM transitions versus MTF transitions could curtail the number of people who have changed their name *and* sex-coding from female to male, but not the number who have changed their names only. On the other hand, the results using name-changes alone in Panel C may suffer from more measurement error than the corresponding results in Panel B. Table 2 showed that females file more claims than males. Since females names are longer and more complex, variant spellings and errors in writing the first name (such as transposing two letters) can shift that person into the group being checked for likely transgender status (Liebersohn and Bell, 1992). If the “new” first name is unique, then the algorithm will look to their middle name. Since females are more likely to be given male names or to use maiden surnames (which are often male first names) as middle names, they could be falsely identified as likely transgender (Goldin and Shim, 2004). I have taken several steps to minimize this type of measurement error, but I cannot eliminate it entirely. The results from the linked Census data in Section 4.3.3, support the second story, although I cannot rule out the possibility that both mechanisms are at play.

The results in Table 4 do not include those who have not legally changed their names, who have changed their names through common-law, whose name(s) do not meet my minimum confidence threshold, or whose gender transition did not involve a name change at all. Furthermore, my results do not include people who do not possess SSNs, who transitioned before attaining an SSN, or who transitioned before turning 16. For these reasons, these results are likely to undercount the number of transgender people in the data and should not be construed as an estimate of the total population of people who have transitioned gender over the time frame. Nevertheless, the results can inform us of general features of the data and trends in certain types of claims over time.

4.2 Historical rates of transgender-consistent claims

We now turn to the question of how the annual flows of claims that are consistent with gender transitions have changed during the years since the SSA was established. Figures 2–4 investigate changes in the frequency of claims that are consistent with gender transitions relative to all other claims. Transgender-consistent claims are defined as any name changes that meet the 90 percent confidence threshold, and any sex-coding changes that move in the same direction as the name changes. If the name change and sex-coding change occurred separately, then each claim is counted; if the name and sex-coding changes occurred simultaneously, then that is counted as a single claim. All other claims are considered non-transgender-consistent claims and are used as a benchmark of growth in claims due to changes in the population or the legislation and uses of the SSN over time.¹¹

¹¹In 1998, an anomalous spike appeared in the trends of both the non-transgender-consistent claims as well as the transgender-consistent claims. I investigate this and conclude that the spike appeared to be an artifact of the administrative record keeping process, rather than the result of an historical event. Specifically, 1998 was the year that the SSA supplied the Census Bureau with the 100% Numident file, containing every transaction for every SSN, from 1936 onward (USCB-SSA, 2000). In most other years, the share of claims is evenly distributed across months, however this is not the case in 1998. An unusually large number of claims were filed in January. I suspect that a relatively small number of transactions from 1936 onward were missing claim entry dates, and that these missing dates were replaced with the date 1 January of 1998. In order to better display the trends over the rest of the years, I impute values for 1998 by averaging the values in 1997 and 1999. Results using the data prior to the imputation,

We begin by looking at annual counts of claims that are consistent with gender-transitions and counts of all other claims. Figure 2 plots the counts of transgender-consistent claims and all other claims by year, from 1936 to 2010. Transgender-consistent claims are shown with the solid line and measured on the left-hand axis, which ranges from 0 to 4,500 claims. Non-transgender-consistent claims are shown with the dashed line and measured on the right-hand axis, which ranges from 0 to 30 million claims. Immediately evident is that transgender-consistent claims appear as early as the 1930s. In addition, there is marked growth in both the number of transgender-consistent claims as well as the number of all other claims processed each year. For non-transgender-consistent claims, this growth appears to begin in the late 1940s, while transgender-consistent claims grow steadily from the SSA’s inception in 1936. Growth in both types of claims accelerates in the early 1970s. This decade was a time of rapid expansion of the use of the SSN for purposes outside the Social Security Program. For example, new legislation required financial institutions to obtain SSNs of all their customers in 1970; in 1975 all recipients of federal benefits were required to provide an SSN; and 1976 brought legislation that allowed states to require SSNs for taxes, eligibility for state programs, and motor vehicle registrations (Puckett, 2009). In addition, the SSA began to store information on each SSN in a digital Numident in 1972, and the next several years were spent digitizing previous paper records, all of which could contribute to the sharp rise in claims during the 1970s. These factors seemed to have a similar impact on the number of transgender-consistent and non-transgender-consistent claims during that time. By 1979, the number of claims had dropped back to its 1960s levels. The sharp spike in non-transgender-consistent claims in 1987 is likely attributable to the Immigration Reform and Control Act, which was signed into law toward the end of the previous year and resulted in approximately 3 million immigrants gaining legal status (Bean *et al.*, 1989). In 1987, the SSA also introduced Enumeration at Birth (EAB), which allows parents of newborns to register their children for an SSN as part of the birth registration process (Puckett, 2009; SSA, 2010). Interestingly, transgender-consistent claims do not exhibit the same spike in 1987, although a sharp increase in transgender-consistent claims occurred in 1988. During the remainder of the 1990s and early 2000s, the total number of non-transgender-consistent claims per year stayed relatively constant, with small spikes in 1993 and 2001.¹² Transgender-consistent claims peak in 1999 drop in the early 2000s, and then begin to increase again in 2004. While the difference in scale makes it difficult to identify changes that are unique to the transgender population versus the broader population, these annual counts show clearly the growth in both types of claims.

To help identify years in which growth in transgender-consistent deviated from the rate of growth in other types of claims, we next consider Figure 3, which plots the annual percent change in transgender-consistent claims, de-trended by the annual percent change in non-transgender consistent claims. That is, Figure 3 shows the difference between the year-over-year growth rate of transgender-consistent claims and non-transgender-consistent claims. The first thing to note is that from the SSA’s inception until 1948, transgender-consistent claims grew at a faster rate than non-transgender-consistent claims. From 1948 to 1975, transgender-consistent claims grew at a pace that was typically lower than the growth in all other claims. The expansions in the use of the SSN in the 1970s appear to have had larger impact on the growth of transgender-consistent claims than non-transgender-consistent claims. The mid-1970s also saw some of the first political gains for transgender constituents, including the Connecticut District Court’s decision that a transgender person had the right to have the sex-coding changed on her birth certificate; the passing of the

as well as the results of my inspection of 1998, are available upon request.

¹²The former was possibly due to individuals filing for Social Security Benefits in anticipation of the effects of the Omnibus Budget Reconciliation Act, signed into law on 10 August 1993, which increased the potential tax liability of high-income Social Security beneficiaries (DeWitt, 2001).

first city ordinance protecting the civil rights of transgender people in Minneapolis; and the New York State Supreme Court decision that a transgender woman, Renée Richards, was eligible to play in the U.S. Open tennis championships (Darnell v. Lloyd, 1975; City of Minneapolis, 1975; Amdur, 1977). It is possible that those events encouraged the especially high growth in transgender-consistent claims during those years and, to a lesser extent, the decade that followed. Growth in transgender-consistent claims outpace growth in non-transgender-consistent claims in 1982 and 1988, although I was unable to identify major events in U.S. transgender history or in the history of the SSA that could explain this feature of the data. Finally, a jump in the de-trended growth rate of transgender-consistent claims is observed in 1998, followed by a relative drop in 2000 and 2001, which corresponds to the introduction of stricter rules regulating sex-coding changes (Spade, 2008). Soon thereafter, increased concern over security during the onset of the War on Terror led the Department of Homeland Security (DHS) and the SSA to begin sending letters to employers whose employees' stated information on name and sex did not match the records associated with their SSNs. As a result, several individuals' transgender status was involuntarily disclosed at work, which may have increased the incentive for transgender people to change the sex-coding on their files.¹³ To help disentangle the influence of these events, I next present annual counts of each *type* of transgender-consistent claim: first name changes, sex-coding changes, and simultaneous changes.

Figure 4 shows the annual counts of first name changes (shown with the dashed line), sex-coding changes (shown with the solid line), and simultaneous changes (shown with the dotted line). First name changes are measured on the left-hand axis, which ranges from 0 to 4,000. Sex-coding and simultaneous changes are measured on the right-hand axis, ranging from 0 to 1,400. Figure 4 demonstrates that, while the different types of transgender-consistent claims often follow the same patterns of growth and decline, there are also instances in which the patterns vary across the different types of claims. For example, all types of transgender-consistent claims occur during the early decades of the SSA, but are rare. This is especially true for claims that involve changing the sex-coding on one's account. Simultaneous changes become more common in the late 1960s, and all types begin to increase during the 1970s. However, 1982 shows a large jump in simultaneous (sex-coding and first name change) claims, while separate claims for first name changes and sex-coding changes do not have nearly as dramatic an increase in that year. By 1988, though, all three types of claims increased substantially. Finally, between 1999 and 2002, we see a dramatic drop in name changes accompanied by a growth in sex-coding changes and simultaneous changes. Because these trends begin before the tightening of identity documentation oversight during the War on Terror, I do not interpret these results as evidence of individuals responding to the incentives of the War on Terror policies.

4.3 Characteristics of likely transgender individuals during the 2010 Census

The remainder of the paper focuses on the subset of likely transgender individuals in the SSA data who were alive during the 2010 Census. The discussion begins by presenting the number of individuals in the SSA data who meet this additional criterion and then examines the order and timing in which individuals change their names and sex-coding. Finally, I present demographic characteristics and residential distributions that are derived from linking the SSA data to the 2010 Census.

¹³The SSA ended this policy in September, 2011 (NCTE, 2011).

4.3.1 Counts

Table 5 shows counts of individuals in the SSA Numident who were alive during the 2010 Census and who changed their first names and sex-coding in ways that are consistent with transitioning gender. Table 5 mimics the layout of Table 4 by showing results for each of the name-change confidence thresholds broken down by those who change only their names and those who change both their names and their sex-coding. As in Table 4, Panel A shows aggregate results, Panel B shows results for likely transgender women (MTF), and Panel C shows results for likely transgender men (FTM).

Panel A in Table 5 shows that, of all likely transgender individuals in the SSA Numident, about 66 percent of all who changed their names only and 73 percent of all who changed their name as well as their sex-coding were alive during the 2010 Census. Overall, 43,547–89,667 individuals changed their names from male to female or female to male, depending on the confidence threshold used. Of these, 16,155–21,833 also changed their sex-coding in the same direction as their name change. Depending on the confidence threshold, 24.3–37.1 percent of name changers also changed their sex-coding; this is close to but slightly higher than the results in Table 4 (22.2–34.0 percent).

Panels B and C show results for likely transgender women (MTF) and likely transgender men (FTM), respectively. The number of likely transgender women in the SSA Numident range from 19,019–32,027, based on first name changes only, and from 11,028–14,338 based on both the name change and sex-coding change. For likely transgender men, these ranges are 24,528–57,640 and 5,127–7,495. The ratio of those who change both their name and their sex-coding to those who change their name only is approximately the same for transgender men in Table 4 and Table 5, at approximately 0.13 to 0.21. For transgender women, however, the ratio is higher in Table 5, ranging from 0.448 to 0.580, than in Table 4 (0.398–0.530). This suggests that younger likely transgender women in the SSA Numident are more likely to change both their name and their sex-coding than the likely transgender women who were no longer alive at the time of the 2010 Census.

4.3.2 Transition paths of those alive during the 2010 Census

The structure of the SSA Numident allows us to investigate the paths people take in documenting their gender transition. Table 5 demonstrates, among other things, that there are many more people in the data who change their first names in ways that are consistent with gender transitions than there are people who change their first names as well as their sex-coding. But of this second group, what paths do people take to ensure their legal records reflect their asserted gender? At what age do people typically begin to document gender transition? Do name changes and sex-coding changes occur simultaneously, or do people change their names first and their sex-coding later? If they stagger the changes, how much time lapses between changes?

Table 6 helps answer these questions. I present counts of those who changed their name first and counts of those who changed their names and sex-coding simultaneously.¹⁴ For each group, I also show the mean age at the time of the first change (i.e., name or sex-coding), and the mean number of years between that change and the second change. I present this information for the mutually exclusive categories of those whose names met the 90 percent confidence threshold but not the 95 percent threshold, those whose names met the 95 percent threshold but not the 99 percent threshold, and those whose names met the 99 percent confidence threshold or above. As in Table 5, I show total results in Panel A, results for likely transgender women in Panel B, and results for likely transgender men in Panel C.

¹⁴None of the likely transgender individuals in the data changed their sex-coding before their names.

Among those who changed their first names and sex-coding in ways that were consistent with gender-transition, the majority made both changes at the same time. For example, 73.0 percent of those whose original and new names met the 90 percent but not the 95 percent confidence threshold changed both their name and their sex-coding concurrently. As the requirements on names becomes stronger, we see an increase in the percentage of people who concurrently change their name and sex-coding to 79.0 percent (among those whose names meet the 95 percent but not the 99 percent threshold) and 80.0 percent (among those whose names clear the 99 percent threshold). These patterns are similar for the likely transgender women and likely transgender men. The average age at which people make simultaneous name and sex-coding changes ranges from about 34 to 36 years old, depending upon the confidence threshold.

Overall, the remaining 20.0 to 27.0 percent of change their names first, although it is more common for likely transgender men to change their names before their sex-coding than it is for likely transgender women.¹⁵ Among those who change their name before their sex-coding, the mean age at which the name change occurs ranges from 32 to 36, and about 5 to 6 years pass before they change their sex-coding.

4.3.3 Results from linked data

This final section presents results from linking the records of likely transgender individuals in the SSA Numident to their responses in the 2010 Census. For simplicity, this section, like the previous one, will focus on the likely transgender individuals whose name changes met the 99 percent confidence threshold, although results using the other thresholds are similar. After examining basic demographic characteristics, I briefly discuss the distribution of likely transgender individuals across states.

Table 7 compares the distributions of likely transgender individuals' responses to questions about race, Hispanic origin, and age to the distribution of responses given by the rest of the population. Table 8 compares the likely transgender individuals' response patterns to the question about sex to those of the non-transgender population. In both tables, the top row shows characteristics for individuals in the 2010 Census who were at least 16 years old, were assigned a PIK, and were not identified as likely transgender. Panel A shows the same characteristics for all likely transgender individuals whose name change met the 99 percent threshold, Panel B shows results for likely transgender women, and Panel C shows results for likely transgender men. As before, the top row of each panel requires a name change only, while the bottom row requires both the name change as well as a corresponding sex-coding change. Finally, Table 7 drops individuals whose responses were given by proxy.

The top row of Panel A shows that individuals who change the gender of their first names were older and less likely to identify as belonging to racial or ethnic minority groups than the general population. For example, 80.4 percent of the likely transgender individuals in the SSA data reported their race as White Alone, which is 3.9 percentage points greater than the non-transgender population.¹⁶ Likely transgender individuals, on average, were less likely than the rest of the population to report Black or African American Alone, Asian Alone, or Some Other Race Alone. They were, however, more likely to report their race as American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, and Two or More Races. Turning to Hispanic origin,

¹⁵The exception to this is among the group whose names clear the 99 percent confidence threshold.

¹⁶Whether or not a person in the Census receives a PIK is correlated with several factors such as race, Hispanic origin, and age. In particular, PIK assignment rates are higher for respondents who report their race as White, who describe themselves as non-Hispanic, and who are older (Bond *et al.*, 2014; Rastogi and O'Hara, 2012). Therefore, the figures in Table 7 for the 2010 Census will differ somewhat from official statistics.

likely transgender individuals were less likely to report Hispanic than the rest of the population. The mean reported age of 55.6 years is just under 10 years older than the mean age of the general population. The age reported on the Census was, on average, 0.8 years younger than the age calculated from the SSA Numident.

The bottom row of Panel A shows differences between those likely transgender individuals who change their sex-coding as well as their names and those who change their names only. Those who change both their names and their sex-coding were more likely to report their race as White Alone and less likely to report their race as Black or African American Alone. They were more likely to report Hispanic and were approximately 2.5 years younger on average than those who changed their names but not their sex-coding.

Both rows of Panel A in Table 8, finally, show that the likely transgender individuals in the linked data are much more likely than the rest of the population to leave the question about sex blank or to check both “M” and “F” on the questionnaire. While 1.1 percent of the non-transgender population leave the question blank, 1.8 to 2.0 percent of the likely transgender individuals in the data leave the question blank. And while 0.02 percent of non-transgender respondents check both options, 0.13–0.14 of likely transgender respondents check both options. The bottom rows of Panels B and C show that 83.4 percent of transgender women who changed both their name and their sex coding reported “F” on the Census, and 88 percent of transgender men who changed both their name and sex-coding reported “M”, giving strong support for the approach of using administrative files for identifying likely transgender individuals, especially when sex-coding changes are included. The top row of Panel B shows that when we only use name changes, 71.2 percent of likely transgender women report “F” on the Census. The top row of Panel C, on the other hand, shows that only 27.9 percent of those identified as likely transgender men reported “M” on the Census, confirming the suspicion that measurement error is worse for those whose original first names are feminine. Nevertheless, we have no way of knowing how transgender individuals, particularly those whose legal sex may not have changed, interpret the Census question on sex. Nor can we be sure that the way transgender men interpret the question does not differ from the way transgender women interpret it. Ultimately, I am most confident of the measure that requires changes in sex-coding for transgender men. Still, I hand-checked 100 randomly-selected people identified as likely transgender men using their names alone (and the 90 percent confidence interval), and I found an accuracy rate of 82.2 percent, so differences in how the Census question is interpreted and answered may in fact play an important role in the finding that only 27.9 percent of likely transgender men (identified by their name changes only) report “M” on the questionnaire. Certainly further research is needed.

Very little evidence exists on the demographic characteristics of transgender individuals living in the U.S., however it is important to discuss how my results complement the earlier work that has been done. Rosser *et al.* (2007) and Grant *et al.* (2011), both internet-based surveys, are the only studies I am aware of that report demographic characteristics of national samples of transgender individuals. Neither are able to report demographic characteristics separately for transgender men and transgender women, so I compare their results to the results in Table 7, Panel A. Like this paper, both Rosser *et al.* (2007) and Grant *et al.* (2011) report that individuals in their data are disproportionately white and less likely to report their race as black, relative to the general population. For example, 2.8 percent of the sample in Rosser *et al.* (2007) and 5.0 percent of the sample in Grant *et al.* (2011) report their race as black, while just over 11 percent of the general population identifies as Black or African American Alone.¹⁷ While the transgender individuals in my data are also less likely to report their race as Black or African American alone, the race

¹⁷This figure differs slightly from the official report of 12.6 percent Black or African American Alone (Humes *et al.*, 2011). The difference may derive from my age restriction as well as the exclusion of those who are not assigned a PIK.

distribution in my data is closer to the general population distribution than the internet-based samples. This is also the case for those who report Two or More Races as well as Hispanic. In addition, Rosser *et al.* (2007) and Grant *et al.* (2011) both have disproportionately high response rates from those in their 20s and 30s, and low response rates from those over the age of 40. This is not an issue in my data, since it is built off all individuals with an SSN who also appear in the 2010 Census. All told, Table 7 not only provides unique information on the demographic characteristics of those in the SSA files who are likely to be transgender, but it also illustrates the value of a big data approach (i.e., using large administrative files and record linkage techniques) for including members of minority groups and the limitations associated with self-selected sampling strategies.

Finally, Figure 5 maps the distribution of likely transgender individuals in the SSA data across states of residence, showing the number of likely transgender individuals per 100 thousand. As before, I show results for the 99 percent name threshold and corresponding sex-coding change.¹⁸ Cut-offs on the map are based on the distribution of the prevalence of likely transgender individuals across states and correspond to the minimum to the 10th percentile, the 10th to the 25th, the 25th to 50th, the 50th to 75th, the 75th to the 90th, and the 90th percentile and above. Colors darken as the prevalence of likely transgender individuals increases, thus, North Dakota, South Dakota, Louisiana, Alabama, Kentucky, and West Virginia are in the lowest decile, at 1.4–2.7 likely transgender individuals per 100,000. On the other hand, Washington, Oregon, Vermont, and New Hampshire are in the top decile, at 7.3–10.6 likely transgender individuals per 100,000.

Both Rosser *et al.* (2007) and Grant *et al.* (2011) map the geographic distribution of their survey respondents, and in both cases the distribution of respondents closely mirrored the population distribution. This is especially clear in Grant *et al.* (2011). An advantage of showing the prevalence, rather than the raw counts, is that it reveals the states where likely transgender individuals more commonly reside, conditional on how the entire population is distributed across states. For example, Montana is among the least populated states, yet its prevalence of likely transgender individuals falls within the 50th to 75th percentile, at 4.1–5.8 likely transgender individuals per 100,000. We also see relatively high prevalence rates in Nevada, New Mexico, Maine, New Hampshire, and Vermont. On the other hand, Texas is one of the most populated states, but its prevalence falls within the 25th to 50th percentiles.

It is beyond the scope of this paper to make claims about why the likely transgender individuals in the SSA data are more likely to live in certain states than in others. However, it is important to note that each of the states with prevalence rates in the top decile had state-level policies that protected people from discrimination on the basis of gender identity at the time of the 2010 Census. In fact, every state in the 75th percentile and above, except for Arizona and Massachusetts, had a state-level nondiscrimination law in place during the 2010 Census; and Massachusetts enacted a law protecting individuals from discrimination based on gender identity in 2011. On the other hand, no state in the bottom quartile had such a law, although there were some instances of city ordinances and nondiscrimination laws covering public-sector employment. This is not to suggest that the laws (or lack thereof) are the cause of how the likely transgender individuals in the data are distributed across states—the presence of a “critical mass” of transgender individuals may rather be the cause of the laws—but it is a fact that is worth noting and certainly warrants further research.

¹⁸The main findings do not change dramatically under the alternative rules for identifying likely transgender persons. Results are available upon request.

5 Conclusion

Learning about the transgender population is difficult, because very few large-scale data sets exist that allow researchers to directly observe whether or not a respondent has transitioned. Of the large-scale data sets that do permit direct identification, issues associated with convenience sample designs and self-selection into the sample limit researchers’ ability to make inferences about transgender individuals in general. This paper presents a novel method for identifying likely transgender individuals in federal administrative files and learning about some of their basic characteristics. When transgender individuals file a claim to legally change their first names, their sex-coding, or both with the SSA, a record of that transaction is created. By sorting through hundreds of millions of claim records from 1936 to 2010, I am able to identify individuals who changed their names from strongly feminine to strongly masculine names (or vice versa) and also changed their sex-coding in the same direction. I am able to identify 135,367 individuals who are likely to be transgender. Of these, 89,667 were alive during the 2010 Census. Probability record linkage techniques allow the persons in the SSA Numident to be linked to their responses to questions on race, Hispanic origin, age, sex, and residential location in the Census.

This approach surely misses certain transgender individuals—those whose name changes are not as sharply gendered, those who change their names via common law (and not with the SSA), those who do not change their names at all, and those who do not have an SSN—so my data are open to some of the same critiques as in previous studies. Nevertheless, there are several features that suggest these data are a major contribution to the existing literature. First, while the sample I use is not immune to selection bias, the selection occurs (1) when a transgender individual chooses whether or not to register name and sex-coding changes with the SSA and (2) when my algorithm labels transgender individuals as non-transgender because their changes do not fit the criteria I set. Both of these, I argue, are less problematic than when an individual chooses to participate in a survey (especially a survey targeting a specific population) or is directly sought out by the researcher via snowball or other convenience samples. First, because it is extremely important for all individuals to have their legal names and sex-coding match the identity they assume in day-to-day life, people who transition genders have a strong incentive to update their legal documentation; so the first type of selection bias will likely be relatively small. Second, while my method for sorting through name changes will inevitably introduce measurement error, particularly for people with originally feminine, non-western, and otherwise unusual names, my demographic analysis shows that I identify a higher proportion of non-white and Hispanic likely transgender individuals than earlier studies. Furthermore, my most conservative sample of 16,155 individuals who changed both their names (according to the 99 percent threshold) and their sex-coding is still nearly twice as large as the samples of Bye *et al.* (2005) (N=55), Conron *et al.* (2012) (N=131), Grant *et al.* (2011) (N=6,456), Hartzell *et al.* (2009) (N=646), and Rosser *et al.* (2007) (N=1,229) *combined*. My most inclusive sample of 89,667 is over 10 times as large.

I find that transgender-consistent claims registered with the SSA are evident as early as 1936 and increased over time, along with all other types of claims. Most likely transgender individuals alive during the 2010 Census had only changed their name from one gender to another, but 21,833 had also changed their sex-coding. Of those who change both their names and their sex-coding, most update both concurrently, although just over a quarter change their name first and their sex-coding 5–6 years later. Most are in their mid-thirties when they begin to register these changes with the SSA, although transgender women begin the process somewhat later in life.

Analysis of the linked SSA-Census data provides many insights into the characteristics of likely transgender individuals in the SSA data. First, I find the proportion on non-white and Hispanic transgender individuals in the U.S. is greater than what has been found in previous studies. In

addition, individuals in the SSA files who are likely to be transgender are much less likely than the non-transgender population to respond to the Census question on sex, and those who do respond are more likely to check both options. Finally, I find that likely transgender individuals in the SSA files are overrepresented in states with laws banning discrimination on the basis of gender identity or expression. Furthermore, states where likely transgender individuals are most underrepresented have no anti-discrimination laws.

This study showcases the promise of big data techniques like administrative record use and record linkage for investigating questions that conventional data sources would not permit. In particular, the innovative use of these data can permit researchers to identify members of populations that are exceedingly difficult to identify in conventional surveys but are increasingly the subject of popular and political discourse. As data availability increases and as attention to the social and economic experiences of transgender individuals intensifies, the techniques developed in this paper can provide vital information on likely transgender individuals' earnings and employment, marriage and divorce, household composition, access to health care, and even incarceration rates. This information, in turn, can promote informed policy and ultimately cast light on a part of our society traditionally kept in the shadows.

References

- Amdur, N. (1977) Renée richards ruled eligible for u.s. open, *New York Times*, August 17 1977, p. 40.
- American Psychiatric Association (APA) (2000) *Diagnostic and statistical manual of mental disorders: DSM-IV-TR®*, American Psychiatric Pub.
- Bakker, A., "van Kesteren", P., Gooren, L. and Bezemer, P. (1993) The prevalence of transsexualism in the Netherlands, *Acta Psychiatrica Scandinavica*, **87**, 237–238.
- Barry III, H. and Harper, A. S. (1993) Feminization of unisex names from 1960 to 1990, *Names*, **41**, 228–238.
- Bean, F. D., Vernez, G. and Keely, C. B. (1989) *Opening and closing the doors: Evaluating immigration reform and control*, vol. 1, The Urban Insitute.
- Bender-Baird, K. (2013) *Transgender Employment Experiences: Gendered Perceptions and the Law*, State University of New York Press, Albany, NY.
- Benjamin, H. (1967) Transvestism and transsexualism in the male and female, *Journal of Sex Research*, **3**, 107–127.
- Bond, B., Brown, J. D., Luque, A. and O'Hara, A. (2014) The nature of the bias when studying only linkable person records: Evidence from the american community survey, *US Census Bureau Center for Administrative Records Research and Applications Working Paper Series*, no. 2014-08.
- Bye, L., Gruskin, E., Greenwood, G., Albright, V. and Krotke, K. (2005) *California Lesbians, Gays, Bisexual, and Transgender (LGBT) Tobacco Use Survey — 2004*, California Department of Health Services, Sacramento, CA.
- City of Minneapolis (1975) Minneapolis, Minnesota, Code of Ordinances, Title 7 Civil Rights, Chapter 139, <http://library.municode.com/HTML/11490/level2/C00R.TIT7CIRI.html>.
- Conron, K. J., Scott, G., Stowell, G. S. and Landers, S. J. (2012) Transgender health in Massachusetts: results from a household probability sample of adults, *American Journal of Public Health*, **102**.
- Darnell v. Lloyd (1975) 395 F. Supp. 1210, *Civ. No. H-75-33*.
- De Cuypere, G., Van Hemelrijck, M., Michel, A., Carael, B., Heylens, G., Rubens, R., Hoebeke, P. and Monstrey, S. (2007) Prevalence and demography of transsexualism in Belgium, *European Psychiatry*, **22**, 137–141.
- DeWitt, L. (2001) Research Note # 12: Taxation of Social Security Benefits, *Research Notes and Special Studies by the Historian's Office*.
- Factor, R. J. and Rothblum, E. D. (2007) A study of transgender adults and their non-transgender siblings on demographic characteristics, social support, and experiences of violence, *Journal of LGBT Health Research*, **3**.
- Gates, G. J. (2011) How many people are lesbian, gay, bisexual and transgender?, *The Williams Institute*.

- Goldin, C. and Shim, M. (2004) Making a name: Women's surnames at marriage and beyond, *Journal of Economic Perspectives*, pp. 143–160.
- Grant, J. M., Mottet, L. A., Tanis, J., Harrison, J., Herman, J. and Keisling, M. (2011) Injustice at Every Turn: A Report of the National Transgender Discrimination Survey, *Washington: National Center for Transgender Equality and National Gay and Lesbian Task Force*.
- Hartzell, E., Frazer, M. S., Wertz, K. and Davis, M. (2009) *The State of Transgender California: Results from the 2008 California Transgender Economic Health Survey*, Transgender Law Center: San Francisco, CA.
- Humes, K., Jones, N. A. and Ramirez, R. R. (2011) *Overview of race and Hispanic origin, 2010*, US Department of Commerce, Economics and Statistics Administration, US Census Bureau.
- Landén, M., Wålinder, J. and Lundström, B. (1996) Incidence and sex ratio of transsexualism in Sweden, *Acta Psychiatrica Scandinavica*, **93**, 261–263.
- Landers, S. and Gilsanz, P. (2009) *The health of lesbian, gay, bisexual and transgender (LGBT) persons in Massachusetts: A survey of health issues comparing LGBT persons with their heterosexual and non-transgender counterparts*, Commonwealth of Massachusetts, Department of Public Health.
- Lieberson, S. and Bell, E. O. (1992) Children's first names: An empirical study of social taste, *American Journal of Sociology*, pp. 511–554.
- Lieberson, S., Dumais, S. and Baumann, S. (2000) The instability of androgynous names: The symbolic maintenance of gender boundaries, *American Journal of Sociology*, pp. 1249–1287.
- Lovett, I. (2013) California: Rights Guaranteed for Transgender Students, *The New York Times*, August 12 2013, <http://www.nytimes.com/2013/08/13/us/california-rights-guaranteed-for-transgender-students.html?smid=pl-share>.
- Lu, A. (2011) For Transgender Detainees, a Jail Policy Offers Some Security, *The New York Times*, December 23 2011, <http://www.nytimes.com/2011/12/23/us/for-transgender-detainees-a-jail-policy-offers-some-security.html>.
- National Center for Transgender Equality (NCTE) (2011) Social Security Administration No Longer Notifying Employees About Gender, *Press Release*, September 15 2011, <http://transequality.org/news11.html>.
- Okabe, N., Sato, T., Matsumoto, Y., Ido, Y., Terada, S. and Kuroda, S. (2008) Clinical characteristics of patients with gender identity disorder at a Japanese gender identity disorder clinic, *Psychiatry Research*, **157**, 315–318.
- Pauly, I. B. (1968) The current status of the change of sex operation, *The Journal of Nervous and Mental Disease*, **147**, 460–471.
- Puckett, C. (2009) The story of the Social Security Number, *Social Security Bulletin*, **69**, 55–74.
- Rastogi, S. and O'Hara, A. (2012) Census Match Study. 2010 Census Program for Evaluations and Experiments, *Center for Administrative Records Research and Applications*.
- Reed, B., Rhodes, S., Schofield, P. and Wylie, K. (2009) Gender variance in the UK: Prevalence, incidence, growth and geographic distribution, *Gender Identity Research and Education Society*.

- Reisner, S., Conron, K., Scout, N., Mimiaga, M., Haneuse, S. and Austin, B. (2014) Comparing in-person and online survey respondents in the U.S. National Transgender Discrimination Survey: Implications for transgender health research, *LGBT Health*, **1**, 98–106.
- Ross, M. W., Wålinder, J., Lundström, B. and Thuwe, I. (1981) Cross-cultural approaches to transsexualism, *Acta Psychiatrica Scandinavica*, **63**, 75–82.
- Rosser, B. S., Oakes, J. M., Bockting, W. O. and Miner, M. (2007) Capturing the social demographics of hidden sexual minorities: An internet study of the transgender population in the United States, *Sexuality Research & Social Policy*, **4**, 50–64.
- Rossi, A. S. (1965) Naming children in middle-class families, *American Sociological Review*, pp. 499–513.
- Ryan, H. (2013) TV’s Transformative Moment: Netflix’s New Show Takes a Casting Cue from Real Life, *Newsweek*, *July 17 2013*, **161**.
- Schilt, K. (2006) Just One of the Guys? How Transmen Make Gender Visible at Work, *Gender & Society*, **20**.
- Schilt, K. and Wiswall, M. (2008) Before and After: Gender Transitions, Human Capital, and Workplace Experiences, *B.E. Journal of Economic Analysis & Policy*, **8**.
- Social Security Administration (SSA) (2010) Program Operations Manual System, RM 00202.308 Enumeration at Birth Process, <https://secure.ssa.gov/poms.nsf/0/4b2e34a39e55f770852576b30009b02a!OpenDocument&Click=/>.
- Social Security Administration (SSA) (2013) Program Operations Manual System, RM 10212.200 Changing Numident Data for Reasons other than Name Change, <https://secure.ssa.gov/poms.nsf/lnx/0110212200>.
- Sosin, K. (2013) New study to weigh transgender military inclusion, *Windy City Times*, *August 4 2013*, <http://www.windycitymediagroup.com/lgbt/New-study-to-weigh-transgender-military-inclusion/43930.html>.
- Spade, D. (2008) Documenting gender, *Hastings Law Journal*, **59**, 731–842.
- Stanglin, D. (2013) Navy SEAL’s biggest secret: A transgender life, *USA Today*, *June 5 2013*, p. 3A.
- Tsoi, W. F. (1988) The prevalence of transsexualism in Singapore, *Acta Psychiatrica Scandinavica*, **78**, 501–504.
- US Census Bureau (2010) 2010 Census, Profile of General Population and Housing Characteristics: 2010 Demographic Profile Data, Table DP-1.
- U.S. Census Bureau and Social Security Administration (USCB-SSA) (2000) 2000 Memorandum of Agreement Between the Bureau of the Census and the Social Security Administration Numident File, *Internal Document*.
- Veale, J. F. (2008) Prevalence of transsexualism among New Zealand passport holders, *Australasian Psychiatry*, **42**, 887–889.

- Votta, R. (2012) ‘Glee’ Recap: Disco Nightmare? Try Fun Diversion, *Billboard.com*, <http://www.billboard.com/articles/news/493448/glee-recap-disco-nightmare-try-fun-diversion>.
- Wagner, D. and Layne, M. (2014) The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications’ (CARRA) Record Linkage Software , *US Census Bureau Center for Administrative Records Research and Applications Working Paper Series*, no. 2014-01.
- Wålinder, J. (1968) Transsexualism: definition, prevalence, sex distribution, *Acta Psychiatrica Scandinavica*, **43**, 255–258.
- Weitze, C. and Osburg, S. (1996) Transsexualism in Germany: empirical data on epidemiology and application of the German Transsexuals’ Act during its first ten years, *Archives of Sexual Behavior*, **25**, 409–425.
- Wilson, P., Sharp, C. and Carr, S. (1999) The prevalence of gender dysphoria in Scotland: a primary care study., *The British Journal of General Practice*, **49**, 991.
- World Health Organization (WHO) (2014) What do we mean by “sex” and “gender”?, accessed online on 23 January 2014, <http://www.who.int/gender/whatisgender/en/>.

Appendix 1 Tables

Table 1: Sample Record Layout of Social Security Updates

PIK	FIRST NAME (FN)	LAST NAME (LN)	SEX	BIRTH DATE	DEATH DATE	ENTRY DATE	CLAIM TYPE
1	JOHN	DOE	M	19640501		19650101	Creation
1	JONN	DOE	M	19640105		19800501	DOB correction
2	JANE	SMITH	F	19660214		19660224	Creation
2	JANE	DOE	F	19660214		19870715	LN change
3	VAL	JONES	F	19670601		19670616	Creation
4	VAL	WILLIAMS	M	19600909		19601030	Creation
5	JOHN	GREEN	M	19620909		19621015	Creation
6	JANE	WHITE	F	19631102		19640215	Creation
7	VAL	THOMPSON	M	19680704		19680720	Creation
7	JANE	THOMPSON	F	19680704		20071215	FN change
8	JOHN	MILLER	F	19650220		19650320	Creation
8	JOHN	MILLER	M	19650220		19660501	Sex-coding change
9	JOHN	JOHNSON	M	19640207		19640307	Creation
9	JANE	JOHNSON	M	19640207		19980101	FN change
9	JANE	JOHNSON	F	19640207		20020101	Sex-coding change
9	JANE	JOHNSON	F	19640207	20090601	20090801	DOD record
10	JANE	BROWN	F	19690815		19700315	Creation
10	JOHN	BROWN	M	19690815		20050815	FN & Sex-coding change

Notes: (1) Variable names and data are fictitious; (2) PIKs 1–6 are used to generate gender probabilities of names. $P[John = male] = 1$, $P[Jane = male] = 0$, $P[Val = male] = 0.5$; (3) PIK 7 is not identified as likely transgender because $P[Val = male] = 0.5$, and PIK 8 is not identified as likely transgender because the first name did not change and the sex-coding change occurred before the individual turned 18; (4) PIKs 9 and 10 are identified as likely transgender.

Table 2: Number of Claims, Number or Unique Records, and Mean Claim per Record

	Number of Claims	Number of Unique PIKs	Mean Number of Claims per PIK
Total	828,011,724	374,177,477	2.2
Stable Records	735,923,348	347,838,151	2.2
Stable Male Records	325,541,478	178,896,996	1.8
Stable Female Records	410,381,870	168,941,155	2.4
Records with name or sex-coding change	92,088,376	26,339,326	3.5
Name Change only	85,312,019	24,654,048	3.5
Sex-coding change only	5,292,854	1,356,854	3.9
Name and sex-coding change	1,483,503	328,424	4.5

Source: SSA Numident, 1936–2010

Notes: Individuals must be at least 16 years old on 1 April 2010

Table 3: How many people’s names meet the thresholds?

	Count	Percent of Total
Panel A. Males with Stable Records		
Name falls below 90% threshold	12,302,085	6.3
Name falls within 90%–95% threshold	4,421,545	2.3
Name falls within 95%–99% threshold.	19,130,621	9.8
Name falls within 99%–100% threshold	158,437,986	81.5
Panel B. Females with Stable Records		
Name falls below 90% threshold	12,554,593	8.0
Name falls within 90%–95% threshold	3,257,344	2.1
Name falls within 95%–99% threshold.	14,812,561	9.4
Name falls within 99%–100% threshold	127,017,114	80.6

Source: SSA Numident, 1936–2010

Notes: Includes all observations whose name and sex-coding did not change. Total count is less than the total number of stable records in Table 2 because the gender of a name is undefined for records with names that are a single letter or appear fewer than 3 times in a decade.

Table 4: Likely Transgender Individuals in the SSA Numident, 1936–2010

	Confidence Threshold:		
	90 percent	95 percent	99 percent
Panel A. Total			
Name	135,367	106,550	64,738
Name and sex-coding	30,006	28,234	21,981
Panel B. Male to Female			
Name	46,889	39,742	26,682
Name and sex-coding	18,642	17,663	14,139
Panel C. Female to Male			
Name	88,478	66,808	38,056
Name and sex-coding	11,364	10,571	7,842

Source: SSA Numident, 1936–2010

Notes: Individuals must be at least 16 years old on 1 April 2010, at least 16 years old at the point of the name change or sex-coding change, and both changes must be permanent (within the SSA files) to be included in the figures above.

Table 5: Likely Transgender Individuals in the SSA Numident Alive During 2010 Census

	Confidence Threshold:		
	90 percent	95 percent	99 percent
Panel A. Total			
Name	89,667	70,748	43,547
Name and sex-coding	21,833	20,551	16,155
Panel B. Male to Female			
Name	32,027	27,563	19,019
Name and sex-coding	14,338	13,613	11,028
Panel C. Female to Male			
Name	57,640	43,185	24,528
Name and sex-coding	7,495	6,938	5,127

Source: SSA Numident, 1936–2010

Notes: Individuals must be alive on 1 April 2010, at least 16 years old at the point of the name change or sex-coding change, and both changes must be permanent (within the SSA files) to be included in the figures above.

Table 6: Sequences of Transition Among Name and Sex-Coding Changers

Confidence Threshold										
90 percent up to 95 percent				95 percent up to 99 percent				99 percent and above		
Count	Percent of sample	Age at first change	Years between changes	Count	Percent of sample	Age at first change	Years between changes	Count	Percent of sample	Age at first change
Panel A. Total										
Name then sex-coding	343	0.27	33.72	6.27	945	0.22	32.54	5.47	3,258	36.66
Name and sex-coding	939	0.73	34.11	0.00	3,451	0.79	36.22	0.00	12,897	36.06
Panel B. Male to Female										
Name then sex-coding	171	0.24	36.01	5.90	500	0.19	35.12	5.28	2,399	38.50
Name and sex-coding	554	0.76	34.18	0.00	2,085	0.81	36.65	0.00	8,629	35.99
Panel C. Female to Male										
Name then sex-coding	172	0.31	31.44	6.64	445	0.25	29.64	5.69	859	31.51
Name and sex-coding	385	0.69	34.02	0.00	1,366	0.75	35.57	0.00	4,268	36.19

Source: SSA Numident, 1936–2010

Notes: Individuals must be alive on 1 April 2010, at least 16 years old at the point of the name change or sex-coding change, and both changes must be permanent (within the SSA files) to be included.

Table 7: Demographic Characteristics of Likely Transgender Individuals, 99% Threshold

		Race					Two or More Races		Hispanic Origin		Age	
		White Alone	Black Alone	AIAN Alone	Asian Alone	NHPI Alone	SOR Alone		Hispanic	Non-Hispanic	Age (SSA)	Age (Census)
		Count										
Census Population over 16		207,241,145	76.47	11.73	0.81	4.82	0.17	3.99	2.00	13.33	86.67	45.76
Panel A. Total												
Name		26,754	80.44	9.42	1.21	2.61	0.29	2.84	3.20	9.10	90.90	55.56
Name and sex-coding		10,279	83.14	6.66	1.28	1.91	0.26	3.17	3.59	10.54	89.46	52.72
Panel B. Male to Female												
Name		12,193	77.92	9.77	1.41	2.89	0.53	3.67	3.82	11.82	88.18	51.03
Name and sex-coding		7,106	82.46	6.71	1.33	2.13	0.34	3.29	3.74	10.46	89.54	53.67
Panel C. Female to Male												
Name		14,561	82.52	9.13	1.04	2.38	0.09	2.15	2.69	6.79	93.21	59.35
Name and sex-coding		3,173	84.65	6.54	1.17	1.40	0.10	2.90	3.25	10.72	89.28	50.60

Source: SSA Numident, 1936–2010 linked to 2010 Census.

Notes: Proxy responses are not included. Individuals must be alive on 1 April 2010, at least 16 years old at the point of the name change or sex-coding change, and both changes must be permanent (within the SSA files) to be included. Name changes must meet the 99 percent confidence threshold. Corresponding tables for the 90 and 95 percent thresholds are available upon request.

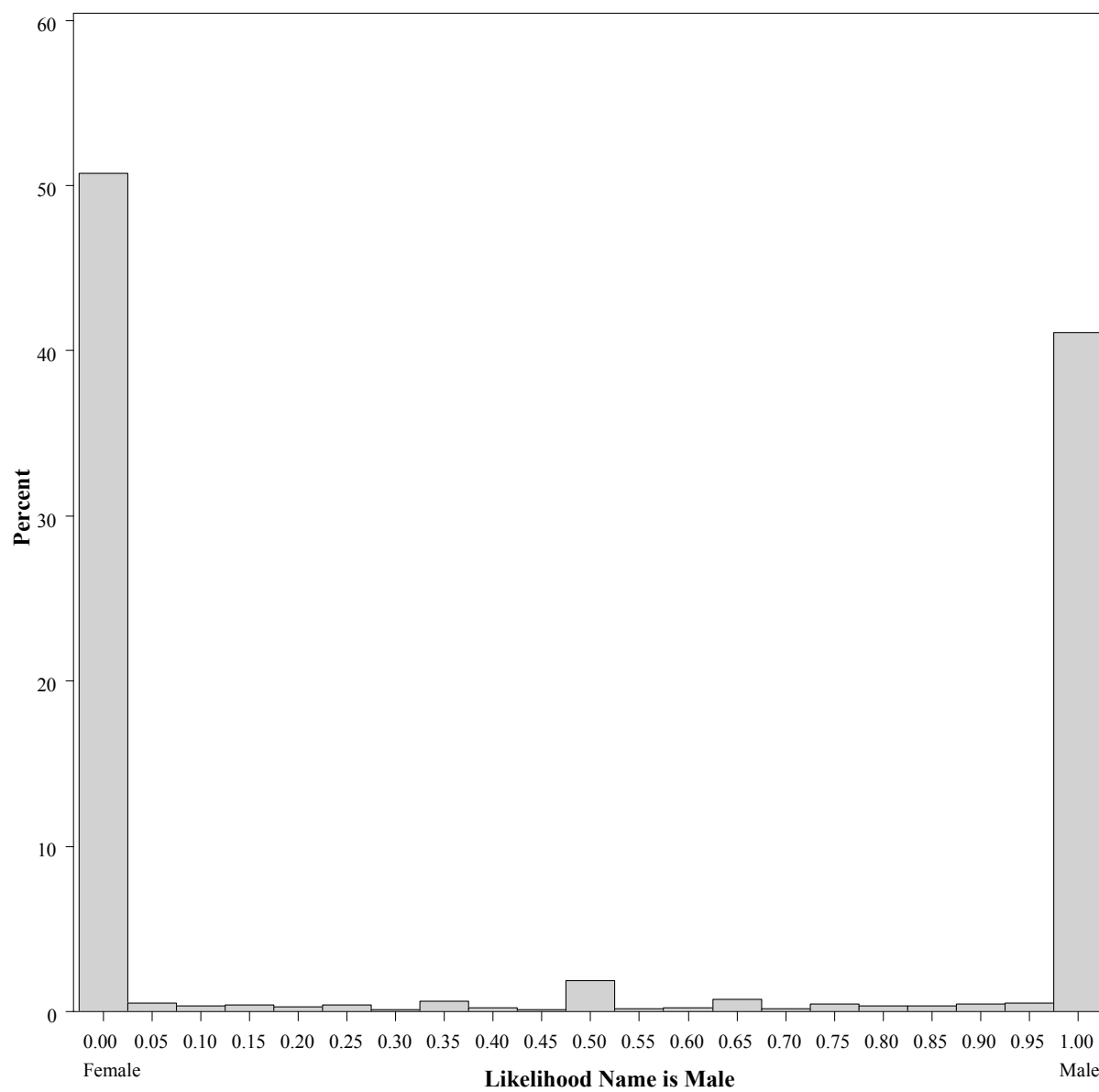
Table 8: Census Responses to the Sex Question, 99% Threshold

	Count	Reported Male	Did not report sex	Reported both sexes
Census Population over 16	207,241,145	48.16	1.13	0.02
Panel A. Total				
Name	26,754	28.33	1.95	0.14
Name and sex-coding	10,279	38.71	1.85	0.13
Panel B. Male to Female				
Name	12,193	28.79	2.32	0.20
Name and sex-coding	7,106	16.64	1.90	0.13
Panel C. Female to Male				
Name	14,561	27.95	1.65	0.09
Name and sex-coding	3,173	88.05	1.73	0.13

Source: SSA Numident, 1936–2010 linked to 2010 Census.

Notes: Proxy responses are not included. Individuals must be alive on 1 April 2010, at least 16 years old at the point of the name change or sex-coding change, and both changes must be permanent (within the SSA files) to be included. Name changes must meet the 99 percent confidence threshold. Corresponding tables for the 90 and 95 percent thresholds are available upon request.

Appendix 2 Figures



Source: SSA Numident, 1936-2010

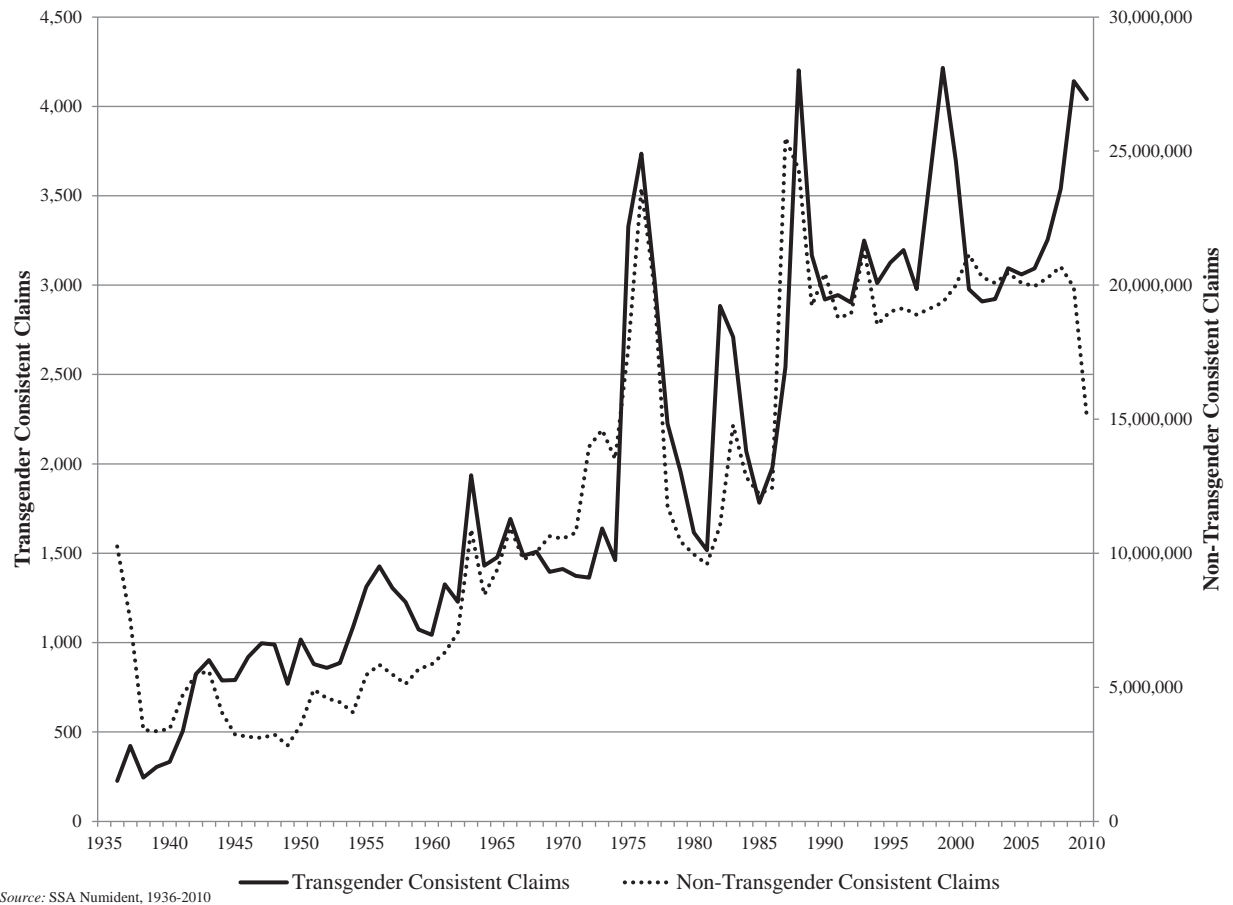


Figure 2: Frequency of Transgender-Consistent and All Other Claims



Figure 3: De-trended Annual Percent Change in Transgender-Consistent Claims

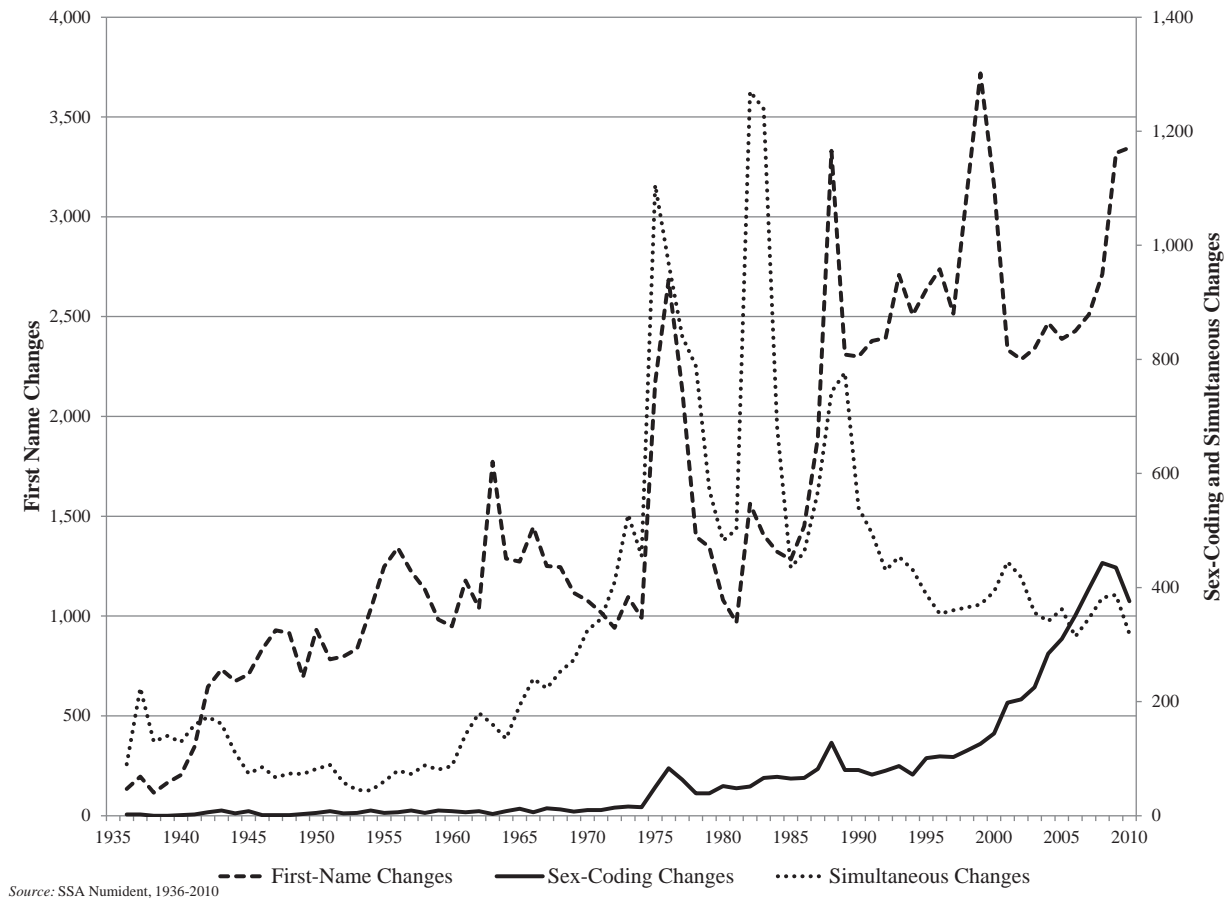


Figure 4: Frequency of Likely Transgender-Related Claims, by Claim Type

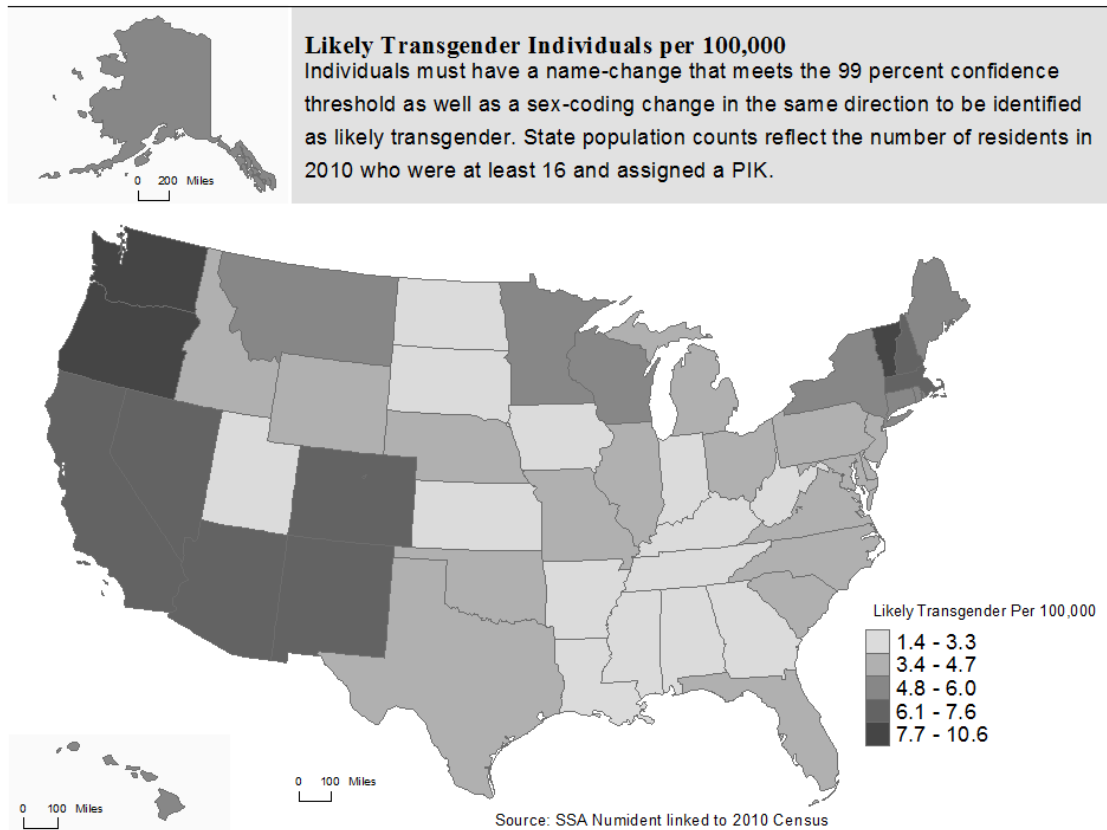


Figure 5: Likely Transgender Individuals per 100,000