Don't Count on Love: Discrepant Historical Measures of American Marriages

Alexander F. Roehrkasse[1]
University of California, Berkeley
18 February 2019

Keywords: census; vital statistics; marriage; historical methods; state formation

DRAFT: Please do not cite or circulate without author's permission

[1] Ph.D. candidate, Department of Sociology, University of California, Berkeley. Address: 410 Barrows Hall, Berkeley, CA 94720-198. Telephone: (510) 642-4766. Fax: (510) 642-0659. Email: roehrkasse@berkeley.edu.

Abstract

Rapid advances in the availability of historical census data are greatly improving historical social research, but much remains unknown about the quality of these data. Full count census microdata are compared to new dataset of county-level vital records of marriages. National census counts of marriage events in 1900 are shown to be 36% lower than counts based on vital records. Multi-year comparisons show discrepancies that are smaller but still greatly in excess of known rates of overall census undercounting. Analysis of exogenous indicators of data quality suggests that both census and vital records quality varied widely across counties, and largely corroborates prior research on the political and demographic correlates of measurement error in official statistics. Implications for historical analysis of marriage and divorce and for the sociology of official knowledge are discussed.

Introduction

In contrast to most other advanced countries that have established national compulsory civil registration systems, the United States has been a laggard in the systematic collection of detailed vital statistics of marriage and divorce. A national system for collecting detailed marriage and divorce statistics was first implemented only in the 1950s, and never achieved complete participation by states before being discontinued in 1996. The poor quality and paucity of U.S. marriage and divorce data has led, even in very recent years, to basic misunderstandings about trends in contemporary family life (Kennedy and Ruggles 2012).

Somewhat in contrast to this data shortfall, the integration and publication of large-scale historical census microdata is quickly increasing historical researchers' ability to leverage "big data" to investigate nuptial life. The bibliography of the Integrated Public Use Microdata Series (IPUMS) project (Ruggles et al. 2018) lists more than 1,200 studies of marriage and the family relying on such data from the United States. The ease with which such data can quickly be put to use, however, risks allowing researchers to overlook or ignore nontrivial data quality issues. A growing but widely overlooked literature explores the validity and reliability of historical census microdata, examining the undercounting of individuals (Ruggles 1991; King and Magnussen 1995; Hacker 2013), the misreporting of individual attributes (Conk 1981; Preston, Lim, and Morgan 1992; A'Hearn, Baten and Crayen 2009; Crayen and Baten 2010), and the consequences of measurement error for analytic results (Clogg, Massagli, and Eliason 1989; Raley 2002).

This study examines the accuracy of historical census measures of marriage events by comparing them to a new national dataset of vital records of marriage. Specifically, I examine discrepancies between marriage events enumerated in the 1900 U.S. decennial census of the population and those recorded in county-level vital records. Both datasets are assumed to be

error-prone, but simple assumptions about overcounting allow me to examine the relative performance of each type of data (Kennedy and Ruggles 2012).

I find significant and pervasive discrepancies between counts of marriages from alternative sources. Only 64% of marriages recorded in vital records during census year 1900 appear in census data. Census-estimated marriage rates per 1,000 unmarried men (women) 15 years and older are 38.2 (42.4) compared to 57.0 (63.4) using vitals data.[2] I demonstrate that the misreporting of marital duration in the 1900 census was nearly three times more common than misreporting of age, but that such misreporting does not explain persistent count discrepancies in excess of a known 5.2% overall census undercount (Hacker 2013). Remaining discrepancies must result either from the frequent misreporting of marital status or the disproportionate underenumeration of married people.

In what follows, I discuss prior research on historical census and vital records quality before outlining the data and methods used to compare historical U.S. marriage data quality. After describing record discrepancies and their administrative and population correlates, I conclude by discussing implications for historical research on marriage and divorce.

Official knowledge problems

What Scott (1998) calls states' "projects of legibility" fall into three broad categories. Censuses are usually universal and compulsory cross-sections of populations undertaken directly by centralized governments. Civil registration systems organize local bureaucrats' regular recording of vital events in populations and their systematic reporting of such records to centralized authorities. With contemporary advances in statistical methodology, surveys perform

[2] Vital records-based estimates of the 1900 marriage rate published by the Centers for Disease Control are slightly higher than my estimates: 61.3 for men and 68.2 for women (https://www.cdc.gov/nchs/data/series/sr_21/sr21_024.pdf).

an increasingly large proportion of the work once done by both censuses and civil registration systems, but because the focus of this paper is on historical data quality, I focus only on censuses and vital records.

Censuses

Census taking is a central activity of modern statecraft. It is, fundamentally, an effort by state actors to develop knowledge of populations as such, to facilitate their governance (Foucualt 2003). Although the enumeration of persons for purposes of political and fiscal apportionment was originally foremost among the mandates of census takers, over the nineteenth century state actors' biopolitical projects expanded to include the measurement of marital status, which they saw as central to the management of fertility and social order.

Census takers' ability to measure populations and their characteristics is not a foregone conclusion, and numerous political sociological studies examine the political and administrative struggles within and between state bureaucracies and civil society that characterize efforts to mount censuses (Loveman 2014; Emigh, Riley, Ahmed 2016a, 2016b). A separate, methodologically focused literature examines measurement error in censuses, generally categorizing it into two broad classes: failures to capture individuals in censuses, and failures to record accurately the attributes of observed individuals.

Undercounting in censuses is a near-universal phenomenon, and results from a combination of enumerator error, foreign migration, language problems, and complex living arrangements. Although some people are often double-counted in censuses, it is usually a much larger number that are never counted. U.S. decennial censuses are thought to have undercounted the population by 6.0% in 1850, 5.2% in 1900, 5.4% in 1940, 1.2% in 1980, and 0.1% in 2000 (Robinson 2001:23; Hacker 2013:88).

When individuals are observed, a variety of factors affect the likelihood that information about them is accurately recorded. Many of these factors mirror the measurement challenges characteristic of survey research. On the one hand, census enumerators may make errors. In the United States, before a permanent census bureau was established in 1902, each decennial census was conducted ad hoc, and census administrators had very little control over enumerators, whose recruitment by federal marshals was characterized by political patronage more than professionalism (Anderson 1990).

On the other hand, census respondents may report inaccurate information about themselves. Misreporting may result from respondents' lack of self-knowledge (Tollnek and Baten 2016), weak incentives to provide accurate information. Respondents may also misreport information when census questions and categories are unintelligible (Emigh, Riley, and Ahmed 2016a, 2016b). And of course, respondents may purposely misreport information, a tendency that is particularly relevant for the case of marriage. For example, black women in the 1910 census overreported their incidence of widowhood and the length of their marital unions, in part to conceal separations and children born out of wedlock (Preston, Lim, and Morgan 1992). Such misreporting may result from social desirability bias, in which respondents seek to manage their social interactions with enumerators, or from the perceived downstream consequences of honest responses. When the public salience of a census is high, the latter set of motivations can also affect undercounting (Kaneshiru 2013; American Sociological Association 2018).

Civil registration systems

Like censuses, civil registration systems seek to develop measures of the population. Differently than censuses, though, civil registration systems do so by locally recording vital events (i.e. births, deaths, marriages, divorces) as they occur, and aggregating local records. The

two steps in this process—local recording and centralized reporting—highlight the two-fold origin and function of civil registration systems. On the one hand, vital recording itself has its origin *not* in the measurement of *populations* but rather in the official establishment of legal statuses for the purpose of clarifying rights and obligations (Dunn 1954). States may therefore promote vital recording not for biopolitical reasons, but rather in pursuit of the rule of law and the stabilization of property rights. On the other hand, the standardization of vital recording and the centralized aggregation of vital records have a later, nineteenth century provenance, one explicitly linked to the calculation of population parameters.

Political sociological research also treats struggles to erect civil registration systems as a case of historical state building, with national state actors struggle to mobilize local ones, and in turn, local bureaucrats vying to register political subjects (Emery 1993; Loveman 2005, 2007). Whereas measurement error in census data consists of both misrecording and underenumeration, problems in civil registration systems are almost exclusively one of missing records. The causes of underrecording in vital records depend on what is being measured. Most literature focuses on the undercounting of births and infant deaths, particularly among marginalized populations (Patterson 1980; Setel et al. 2007). With respect to marriage, several factors are particularly relevant to historical record quality. On the one hand are factors relating to marrying persons and their agents. Marrying people may have insufficient incentives to formalize their unions, particularly where the rule of law is weak or when personal wealth is minimal (Dubler 1998). This is especially true when populations are dispersed and transportation costs to official centers are nontrivial. Marrying people may also and actively resist registering unions if such unions are socially unsanctioned or if individuals are suspicious of incursions by modern officials into traditional institutions like marriage (Outhwaite 1995; Diamont 2001). On the other hand are

factors relating to local officials. Because civil registration systems are often built upon preexisting local vital records systems, the political histories of localities can exert path-dependent influence on the administration and recording of marriages, with consequences for their completeness (Deporte 1926; Richmond and Hall 1929).

Measuring official knowledge problems

How can the accuracy of official statistics be evaluated? Information about populations collected by contemporary states, especially advanced industrial ones, is usually collected using methods designed to allow for tests of validity, reliability, and missingness. Less so for historical censuses. In the United States, the 1870 census was the first to stimulate widespread concern about record quality, and the 1900 U.S. census was the first to employ modern strategies for evaluating data quality, such as by asking respondents to report their age in years and their month and year of birth. Where such precautions were not taken at the time of enumeration, historical researchers must rely on more creative techniques.

With respect to the misrecording of individuals' attributes, record quality can be evaluated by comparing observed distributions to distributions strongly expected for theoretical or empirical reasons. One common example is the analysis of "age heaping." The true distribution of age in populations is relatively continuous, and the distribution of terminal digits in age-in-years is roughly uniform. However, observed distributions of age-in-years in many censuses exhibit sharp discontinuities at cognitively "preferred" terminal digits such as 0 and 5 (Zelnik 1961). Such discontinuities can be analyzed to quantify the misreporting of age.

Evaluating the undercounting of individuals is more difficult. Historical data quality can be assessed through contemporary accounts. Although such accounts can offer helpful clues and guiding details, they are usually anecdotal and often reflect motivated reasoning (King and

Magnuson 1995). Demographic methods use alternative sources of data, such as records of births, deaths, and migration to generate independent estimates of populations which can then be compared to census counts. These methods provide the most convincing estimates of overall census undercounts, but they also have drawbacks. First, they rely on accurate mortality and age data, when both are known to be subject to nontrivial measurement error (Preston, Lim and Morgan 1992; Hacker 2013). Second, they are often only feasible for subpopulations, and rarely provide geographic or demographic detail about measurement error.

A third approach attempts to compare the same measure across multiple sources. Most common is to estimate undercounting by seeking to locate specific individuals from other records in census data. This approach has been particularly useful for identifying which types of people have been most likely to be excluded from historical censuses—the foreign-born; residents of large cities or frontier areas; infants and young children; the indigent; and itinerate groups such as borders, lodgers, and servants—groups whose undercounting is largely corroborated by contemporaneous accounts (Steckel 1991). But record linking studies have also tended to generate inflated and unreliable estimates of overall undercounts.

This study blends the logic of the demographic and record linking approaches, but also simplifies them. Rather than seeking to identify individual records in alternative sources, I compare aggregate counts, but over thousands of geographic subunits, namely, U.S. counties. And rather than using life tables to project population estimates onto census years, I derive flow data from the census which can then be directly compared to vital records. This approach has the merit of relying on many fewer assumptions than alternative methods, but is also subject to unique forms of bias, discussed below. It bears strong similarity to a recent effort by Kennedy

and Ruggles (2012) to evaluate the quality of contemporary divorce data by comparing alternative sources under simple assumptions.

Data and methods

Discrepancies in marriage counts

The primary quantity of interest is the discrepancy between census and vital records counts of marriages. Census data are full-count microdata from the 1900 decennial census, published by the Integrated Public Use Microdata Series, or IPUMS (Ruggles et al. 2018). The 1900 census asked married respondents about the duration of their current marriage.[3] Marriages occurring within the census year (June 1, 1899, to May 31, 1900) were coded "0." Of 27,812,405 married people, 34 men and 118 women have missing data on marital duration. Counts of marriages with durations less than one year therefore offer a measure of the number of legal marriages contracted in the census year. Domestic migration will cause measurement error at the state and county level but not at the national level. Net foreign immigration and mortality will cause, respectively, upward and downward bias in census measures of marriage events at all geographic levels. Notably, however, these sources of bias are confined to individuals whose legal marriage occurred during the census year *and* who migrated or died after their legal marriage and before census enumeration. Rates of legal divorce within the first year of marriage were trivial (Unites States Bureau of the Census 1909b). I calculate census counts of marriages separately using male and female responses, and compare them as a robustness check.

[3] Instructions to census enumerators were: "Number of years married. — Enter in this column for all persons reported as married (column 9) the number of years married (to present husband or wife), as 5, 9, 29, etc.; for persons married during the census year, that is, from June 1, 1899, to May 31, 1900, write '0;' for all other persons leave the column blank. Notice that this question can not be answered for single persons and need not be for widowed or divorced persons." (Barrows 1976)

I digitized vital records of marriage published by the United States Bureau of the Census (1909b). In 1887 and again in 1907, the Census Bureau dispatched agents to most counties to compile the previous twenty years of local marriage and divorce records. Some states with civil registration systems reported county-level statistics directly, and records were collected by mail from local authorities in approximately 765 "sparsely settled or distant counties," and by agricultural surveyors in 206 Southern counties in the second survey. Data were patchy in the first survey, with 59% of counties missing marriage data. By 1900, though, outside of South Carolina, which did issue marriage licenses or record returned licenses, marriage counts were reported for 2,704 of 2,766 counties, or 97.8%. The data are annual flows of marriage "returns" at the county level. Marriage returns are certificates of marriage returned by officiants of celebrated marriages to local record keepers.[4] Vital records tabulate marriages by calendar year, so for comparison with census data I assign vital records counts of marriages to census year 1900 using vital records data from 1899 and 1900 and the monthly distributions of marriages in Massachusetts and Michigan in those years.[5] All subsequent references to years refer to census years—that is, to the twelve-month period preceding June 1 of the referenced year.

At national, state, and county levels I calculate absolute differences in counts of marriage d_a by subtracting vital records counts c_v from census counts c_c. Variation in true population of married people across geographies complicates comparison of discrepancies. I therefore construct the normalized difference:

[4] The Census Bureau also collected information on marriage licenses, but many states issued licenses irregularly or not at all, and so these figures were never reported.

[5] Results are robust to alternative methods of assigning marriages to census years, including using a larger sample of states' monthly marriage distributions in alternative years.

$$d_n = \begin{cases} \frac{d_a}{c_v} * 100, & c_c < c_v\,; \\ 0, & c_c = c_v\,; \\ \frac{d_a}{c_v} * 100, & c_c > c_v\,. \end{cases}$$

Using a piecewise denominator to calculate d_n avoids creating outlying values in sparsely populated geographies and yields a continuous measure with range [-100, 100] that can be interpreted as the *percentage difference in marriage counts*, with negative values indicating census undercounting relative to vital records data, and positive values indicating vital records undercounting relative to census data.

Because vital records are annual and the census measures marital duration in integer values, it is possible to compare census and vital records counts in prior census years. Importantly, however, older estimates will experience increasing measurement error resulting from migration and mortality, as well as from legal divorce and spouses' misreporting of separations as divorces, deaths, or non-existent unions (Preston, Lim and Morgan 1992). Counties with missing vital records of marriage are listwise deleted.

Marriage data quality

Discrepancies between census and vital records counts of marriage are a joint function of the quality of both data sources. Interpreting discrepancies is not, therefore, straightforward. For example, a moderate discrepancy could indicate that one source is valid and the other is moderately biased, or that one source is moderately biased and the other severely so. To help address this problem I develop several measures of data quality that are independent of count discrepancies, and compare them to discrepancies. This helps to better understand which source of data is to blame.

To measure the quality of census data, I examine the misreporting of marital duration. I use a strategy identical to that used to analyze age misreporting. Absent the ability or incentive to provide accurate information about one's age, respondents tend to supply cognitively convenient numbers, usually ones ending with digits 0 or 5 (Zelnik 1961).[6] The aggregate result of misreporting due to digit preference is "heaping," in which response frequencies cluster around predictable responses. Because marital duration, like age, is measured in years, it can be expected to exhibit a similarly discontinuous distribution.

To measure heaping in marital duration I use the Whipple Index, a widely used and validated measure of heaping (A'Hearn, Baten and Crayen 2006). The index takes the form:

$$W = \frac{\sum(n_5 + n_{10} \ldots + n_{55} + n_{50})}{\frac{1}{5}\sum_{i=2}^{51} n_i} \times 100$$

where n_i represents the frequency of respondents reporting marriages of length i. I restrict the range of values over which I evaluate heaping to 2 to 51 years. This balances the frequency of terminal digits, minimizes the number of counties with married populations too small to evaluate heaping (56 counties, for which $\sum_{i=2}^{51} n_i < 250$), and avoids misreporting issues specific to marital durations of less than two years, discussed below. In reporting results, I use a simple transformation of the Whipple index:

$$\widetilde{W} = \begin{cases} \frac{W - 100}{4}, & W > 100; \\ 0, & W \leq 100; \end{cases}$$

[6] It is noteworthy that U.S. Census officials appear by 1900 to have been aware of this behavior, alerting enumerators that "[a]n answer given in round numbers, such as 'about 30,' 'about 45,' is likely to be wrong" and encouraging them in such cases to "endeavor to get the exact" value (Barrows 1976:206). The 1900 census was the first to ask respondents about their month and year of birth as well as their age in years, to evaluate the accuracy of the latter.

that can be interpreted as the *percentage of respondents heaping marital duration* (cf. Crayen and Baten 2010).

To measure vital records quality, I examine legal and administrative characteristics of counties relevant to marriage recordkeeping, as tabulated in United States Bureau of the Census (1909a). I create three indicator variables. The first measures whether a county belonged to a state that required the *centralized reporting* of marriage records to the state. By 1900, ten states comprising 448 counties had such requirements. Because state-level reporting systems encouraged the institutionalization of local marriage recording, counties within these systems can be expected to have produced higher quality marriage counts. (Indeed, such counties were less likely to have completely missing marriage data.) The second measure indicates whether a county belonged to a state requiring that marriage licenses be returned to and recorded by a *non-judicial officer*. By 1900, twenty-nine states comprising 1,726 counties had such provisions, suggesting that they dedicated administrative resources to the recording of marriages rather than relying on existing officers such as judges and court clerks to perform the task. The third measure indicates whether a county belonged to a state specifying penalties—financial, criminal, or otherwise—for marriage recorders' non-performance of duties, including specifying that parties whose marriages were not recorded had a right to bring personal legal suit against the negligent official. In 1900, twenty-five states comprising 1,405 counties allowed for such penalties. I treat the presence of such a provision as a negative indicator of bureaucratization. To the extent that marital recordkeeping was bureaucratized, administrative duties should have inhered to the office and the officeholders themselves should have been absolved of personal liability for state functions (Weber 1976). Hence, penalties should be associated with poorer quality records. It should be noted that because these measures are state-level policy indicators,

they fail to capture variation across states in policy details, or variation across counties within states in the quality and intensity of their implementation.

Other explanatory variables

I also use IPUMS microdata to generate several other measures of counties' demographic characteristics. Because literacy can be expected to bey positively correlated with accurate census reports, I measure the percentage of the population that is *literate*, defined as being able both to read and to write. Because historical censuses are known to have undercounted individuals who were *foreign born*, *urban* residents, or *black* (Hacker 2013), I include measures of the percentages of the population at least 16 years old comprising each of these groups, where urban places are defined as having 2,500 or more people. I also include a measure of the percentage of the population *born in another state* to control for bias resulting from internal migration. Because census quality is also known to have varied by *region*, I also include dummies for the Census Bureau's four geographic regions.

Missing marriages?

How exactly to census- and vital records-based counts of marriages compare? Figure 1 shows the distribution of county-level discrepancies in counts of marriages occurring in census year 1900, with negative values indicating census undercounts as a percentage of vital records and positive values indicating the opposite. Discrepancies calculated using male and female census responses are displayed separately. For both sexes, 90% of counties had a census undercount, with the median (mean) county having a 31% (29%) census undercount.

An examination of the distribution of marital duration in the census, displayed in Figure 2, indicates a possible explanation for such severe discrepancies. Considerably more people

reported being married 1–2 years (1,217,733) than being married less than one year (887,233). Explaining such a discrepancy is difficult except in terms of measurement error: no historical evidence indicates a negative shock to marriage rates in 1900, and neither immigration, mortality, nor divorce patterns offer a plausible explanation for the gap. A much more likely scenario is that individuals married less than 1 year were recorded as having been married 1–2 years. This could result from rounding—it would be unsurprising if nontrivial numbers of newlyweds identified as having been married "one year." It is even more likely that such misrecording resulted from respondents' misunderstanding of the marital duration question. Census enumeration began on June 1, 1900, and the reference date for marital duration was that day. On the one hand, it is possible that responses given at later enumeration dates were systematically upwardly biased—for example, an individual married June 15, 1899 and enumerated July 1, 1900 should have been recorded as having been married less than one year (as of *census* day), but might instead have been misrecorded as having been married for more than one year (which, as of the date of *enumeration*, they in fact had been). On the other hand, respondents may simply have interpreted enumerators' questions to refer to calendar years, upwardly biasing measurement of marital duration for all marriages contracted between June 1, 1899 and December 31, 1899.

Other evidence suggests that marital duration was widely misrecorded in the census. Figure 2 also clearly indicates heaping at terminal digits 0 and 5. Transformed Whipple indices indicate that 4.7% of married individuals expressed digit preference in reporting marital duration in the 1900 U.S. census, compared to 1.8% who heaped their self-reported age.[7] But equally

[7] The unweighted average of percentage of county residence heaping marital duration was 4.2%, compared to 1.5% for age. County-level heaping in marital duration and age were strongly correlated: a bivariate OLS regression model of marital duration heaping on age heaping yields a correlation coefficient of 1.025 ($p < 0.0000$, state-

compelling evidence indicates that misrecording of marital duration cannot explain away count discrepancies. If apparent undercounting of marriages contracted in 1900 was in fact an artifact of misrecorded marital duration, this would result in excessively high counts of marriages lasting 1–2 years. In fact, Figure 2 shows that slightly *fewer* such marriages were recorded than those lasting 2–3 years. It is considerably less plausible to suppose that newlyweds would report having been married two or more years.

Figure 3 gives additional evidence of missing marriages, displaying the total discrepancies in counted marriages by year of occurrence. Discrepancies are calculating by totaling only counties with non-missing marriage data from both sources, and should be interpreted with increasing caution in earlier years because potential bias from foreign migration, mortality, and divorce increases. Known overall census undercounts of the population were approximately 5.2% in 1900 (Hacker 2013), plotted as a dashed line. If apparent count discrepancies were simply the result of wrongly recorded marital duration, the census should have overcounted marriages in years prior to 1900 (that is, relative to the -5.2% baseline), and approximated a net count discrepancy of -5.2% over multiple years. To the contrary, census undercounts of marriages exceeded the baseline in all prior years excepting 1890, an outlier clearly attributable to heaping (cf. Figure 2). Comparing counties with non-missing census and vital records data, marriages contracted in 1900 were undercounted in the census by 35.7%; 21.4% contracted in the period 1899–1900 were undercounted; 12.0% contracted 1894–1900 were undercounted; and 13.7% in the period 1889–1900.

Therefore, although misreporting of marital duration was widespread, accounting for this fact appears to explain at most about one third of census net undercounts of marriages occurring

clustered standard errors). Demographic correlates of heaping in both marital duration and age corroborate prior research on patterns of census misreporting.

in 1900. On the assumption that vital records do not overcount marriages, one is therefore left to conclude that census–vital records discrepancies in counts of marriages result primarily from the erroneous recording of unmarried status for married individuals who were enumerated, and/or the disproportionate underenumeration of married individuals.

Data quality and bounded estimates

National net census undercounts relative to vital records indicate that the latter measure of marriage has higher overall validity. But Figure 1 evidences wide variation across counties in the severity of count discrepancies. Variation in discrepancies could be attributable to the unreliability of census data, vital records, or both. What do exogenous indicators of data quality suggest about the sources of count discrepancies?

Table 1 gives summary statistics for attributes of U.S. counties in 1900. Table 2 reports estimated coefficients and standard errors for models of county-level count discrepancies measured over three different intervals—census years 1900, 1899–1900, and 1894–1900. Models are estimated separately by whether men or women were used to generate census counts of marriages, with sex differences reported. If variation in count discrepancies was driven by unreliable census data, discrepancies should be correlated with exogenous measures of census data quality, such as the degree of heaping in continuous self-reported variables. The first row of Table 2 shows that heaping in marital duration is strongly correlated with census undercounts of marriage in 1900: for male-based estimates, a one standard deviation increase in marital heaping led to a .15 standard deviation widening of the gap between census and vital records counts of marriages. The strength of this effect decays considerably, however, when discrepancies are measured over multiple years. In other words, counties whose residents more frequently

misreported marital durations between years 2 and 51 were also more likely to undercount newlyweds, but not necessarily marriages overall. This pattern offers additional evidence in support of the conclusion that discrepancies in 1900 resulted in part from misreports of marital duration, but that such misreports cannot explain overall net undercounts.

Two pieces of evidence, however indicate that vital records do not themselves provide reliable counts of marriages, and contributed as well to inter-county variation in count discrepancies. First, Table 2 shows that census undercounts relative to vital records were larger in counties that had dedicated personnel for the recording of marriages. This is consistent with the expectation that counties' devotion of more administrative resources to marital recordkeeping led to more accurate—in this case, higher—counts of marriages, and therefore larger census undercounts relative to vital records. Estimated coefficients for centralized reporting and personal penalties for recorders have the expected sign, but are statistically insignificant.

Second, net census *overcounts* of marriage in some 236 counties, visualized in Figure 1, suggest that in some places, census data outperformed vital records outright. Because these counties had disproportionately small populations and were overrepresented in Western states, it is difficult to draw general implications from cases of vital records undercounts. But they may indicate that vital records overall contained nontrivial degrees of marital undercounting, and that observed net census undercounts of vital records understate the true undercount of U.S. marriages.

Population correlates of measurement error

Table 2 further reports estimates of the association between county-level marital count discrepancies and the characteristics of county populations. Less literate counties might be

expected to have higher rates of census misreporting, but any such effect appears to be captured by the parameter for marital duration heaping.

By contrast, counties with greater proportions of urban residents exhibit substantially larger census undercounts of marriages. For male-based estimates, a one standard deviation increase in urban residents leads to a .15 standard deviation increase in the census undercount of marriages. The strength of this effect diminishes by about a third when considering multi-year discrepancies, but remains statistically significant. Counties with larger proportions of black residents also had larger census undercounts of marriage in 1900: a one standard deviation increase in the former is associated with a .09 standard deviation increase in the latter. No such association, however, is evidenced in multi-year counts.

As discussed above, national census estimates of marital events occurring in the United States will be biased upward by foreign migration, and local estimates will be similarly biased by domestic migration. Such bias should increase with multi-year estimates. As expected, larger foreign-born populations are associated with larger census counts of marriage relative to vital records, but only in multi-year estimates and increasingly so with larger periods—almost certainly indicating the inclusion of foreign contracted marriages in census data. Domestic migration, however, seems to contribute no meaningful bias to estimates of count discrepancies.

The most striking correlates of disparities between census and vital records counts of marriage are geographic. Figure 4 plots normalized discrepancies at the state level. Discrepancies appear to be most concentrated in the South, corroborating prior historical research documenting problems of census enumeration in that region. However, estimated models described in Table 2 include region dummies (the reference category is the Northeast), alongside administrative and population covariates. This affords a better understanding of the marginal effects of place itself.

The results complicate the conventional understanding of Southern underdevelopment. The influence of geography on count discrepancies in 1900 alone is ambiguous, but considering multi-year discrepancies shows clear patterns. Census undercounts were indeed severe in the South, but even more so in the Midwest, where, net of other factors, some 13 to 19% of men's marriages went uncounted in the census.

Conclusion

Comparison of U.S. census and vital records data for marriages occurring in 1900 indicates serious flaws in census-based measures of marriage. Analysis of the distribution of marital duration in the census gives evidence of enumeration error consistent with respondents' expression of digit preference and possible misunderstanding of census reference periods. But marital duration misreporting cannot account for multi-year census shortfalls of recorded marriages relative to vital records. These shortfalls indicate either the inaccurate enumeration of large numbers of married individuals as unmarried, or the disproportionate underenumeration of married people, particularly newlyweds.

The large number of marriages missing from the census—and the unevenness of their missingness across groups and geographic areas—has significant consequences for scientific analysis of historical family patterns in the United States. This study compared measures of marital *events* in census data and vital records, but the findings can only be explained if the marital *status* of large numbers of individuals is misrecorded, or individuals are missing not at random from the census, specifically on the basis of their marital status. Marital status measures in historical census microdata have been used to calculate national historical trends in nuptiality (Haines 1996) and marital dissolution (Ruggles 1997; Cvrcek 2009), and are even more

commonly relied upon in the analysis of the causes and consequences of marriage and divorce (Bloome and Muller 2015; Bloome, Feigenbaum and Muller 2017). The results of this study recommend caution when using census-based measures of marital status before the institutionalization of the Census Bureau in 1902.

This study extended the exercise carried out by Kennedy and Ruggles (2012), who compared alternative sources of contemporary marriage and divorce data and, using a simple set of assumptions, were able to show that conventional sources of data on family patterns were plagued by gross undercounts. Fortunately, their critique pointed to a preferred alternative in new census data. The historical situation is less auspicious because vital records, which provided a heuristic for evaluating census data, seem to provide a substitute in only a very limited sense. Vital records are in the first place limited in the information they contain: they measure marital flows but not stocks, and for early periods do not decompose marital events by the characteristics of marrying persons. Nevertheless, such exercises continue to provide helpful background knowledge to those using newly available marriage and divorce data—contemporary or historical—in new ways.[8]

The study also has implications for the sociology of state formation and the sociology of official statistics. Contemporary political sociology takes great interest in how states "see" (Scott 1998), but also in how states are "many-handed" (Orloff and Morgan 2017). This is to say, states construct political subjects and governed populations through their "principles of vision and division" (Bourdieu 1985), but because they are complex, heterogenous, and semi-autonomous organizations, they do so in variegated and sometimes internally inconsistent ways. Synthesizing

[8] Indeed, I've already done most of the analysis for the 1880 census, and will include it in the next draft of this chapter.

these two physical metaphors, we might say that the findings of this study provide evidence that the state requires different prescriptions for each lens.

References

A'Hearn, Brian, Jörg Baten, and Dorothee Crayen. 2006. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." Working paper.

———. 2009. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." *The Journal of Economic History* 69(03):783.

American Sociological Association. 2018. "National Scientific Organizations File Amicus Brief Objecting to Inclusion of Citizenship Question in the 2020 Census." Retrieved February 12, 2019 (http://www.asanet.org/citizenshipamicus).

Anderson, Margo J. 1990. *The American Census: A Social History*. New Haven: Yale University Press.

Barrows, Robert G. 1976. "Instructions to Enumerators for Completing the 1900 Census Population Schedule." *Historical Methods Newsletter* 9(4):201–12.

Bloome, Deirdre, James Feigenbaum, and Christopher Muller. 2017. "Tenancy, Marriage, and the Boll Weevil Infestation, 1892–1930." *Demography* 54(3):1029–49.

Bloome, Deirdre and Christopher Muller. 2015. "Tenancy and African American Marriage in the Postbellum South." *Demography* 52(5):1409–30.

Bourdieu, Pierre. 1985. "The Social Space and the Genesis of Groups." *Theory and Society* 14(6):723–44.

Clogg, Clifford C., Michael P. Massagli, and Scott R. Eliason. 1989. "Population Undercount and Social Science Research." *Social Indicators Research* 21(6):559–98.

Conk, Margo A. 1981. "Accuracy, Efficiency and Bias: The Interpretation of Women's Work in the U.S. Census of Occupations, 1890–1940." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 14(2):65–72.

Crayen, Dorothee and Joerg Baten. 2010. "Global Trends in Numeracy 1820–1949 and Its Implications for Long-Term Growth." *Explorations in Economic History* 47(1):82–99.

Cvrcek, Tomas. 2009. "When Harry Left Sally: A New Estimate of Marital Disruption in the U.S., 1860 - 1948." *Demographic Research* 21:719–58.

Deporte, J. V. 1926. "Development of Statistics of Marriage and Divorce in New York State." *American Journal of Public Health* 16(11):1109–13.

Diamant, Neil J. 2001. "Making Love 'Legible' in China: Politics and Society during the Enforcement of Civil Marriage Registration, 1950-66." *Politics & Society* 29(3):447–80.

Dubler, Ariela R. 1998. "Governing Through Contract: Common Law Marriage in the Nineteenth Century." *The Yale Law Journal* 107(6):1885–1920.

Dunn, Halbert L. 1954. *Vital Statistics of the United States 1950*. Vol. 1. Government Printing Office: Washington, D.C.

Emery, George. 1993. *Facts of Life: The Social Construction of Vital Statistics, Ontario 1869-1952*. Montreal and Kingston: McGill-Queen's University Press.

Emigh, Rebecca Jean, Dylan Riley, and Patricia Ahmed. 2015a. *Antecedents of Censuses from Medieval to Nation States: How Societies and States Count*. New York: Palgrave Macmillan.

———. 2015b. *Changes in Censuses from Imperialist to Welfare States: How Societies and States Count*. Houndmills, Basingstoke, Hampshire ; New York: Palgrave Macmillan.

Foucault, Michel. 2003. *"Society Must Be Defended": Lectures at the Collège de France, 1975-1976*. New York: Picador.

Hacker, J. David. 2012. "New Estimates of Census Coverage in the United States, 1850–1930." *Social Science History* 37(1):71–101.

Haines, Michael R. 1996. "Long-Term Marriage Patterns in the United States from Colonial Times to the Present." *The History of the Family* 1(1):15–39.

Kaneshiro, Matheu. 2013. "Missing Minorities? The Phases of IRCA Legislation and Relative Net Undercounts of the 1990 Vis-à-Vis 2000 Decennial Census for Foreign-Born Cohorts." *Demography* 50(5):1897–1919.

Kennedy, Sheela and Steven Ruggles. 2014. "Breaking Up Is Hard to Count: The Rise of Divorce in the United States, 1980–2010." *Demography* 51(2):587–98.

King, Miriam L. and Diana L. Magnuson. 1995. "Perspectives on Historical U.S. Census Undercounts." *Social Science History* 19(4):455–66.

Loveman, Mara. 2005. "The Modern State and the Primitive Accumulation of Symbolic Power." *American Journal of Sociology* 110(6):1651–83.

———. 2007. "Blinded Like a State: The Revolt against Civil Registration in Nineteenth-Century Brazil." *Comparative Studies in Society and History* 49(1):5–39.

———. 2014. *National Colors: Racial Classification and the State in Latin America*. New York: Oxford University Press.

Morgan, Kimberly J. and Ann Shola Orloff. 2017. *The Many Hands of the State: Theorizing Political Authority and Social Control*. New York: Cambridge University Press.

Outhwaite, R. B. 1995. *Clandestine Marriage in England, 1500-1850*. London and Rio Grande: The Hambledon Press.

Patterson, John E. 1980. "Assessing the Quality of Vital Statistics." *The American Journal of Public Health* 70(9):944–45.

Preston, Samuel H., Suet Lim, and S. Philip Morgan. 1992. "African-American Marriage in 1910: Beneath the Surface of Census Data." *Demography* 29(1):1.

Raley, R. Kelly. 2002. "The Effects of the Differential Undercount on Survey Estimates of Race Differences in Marriage." *Journal of Marriage and Family* 64(3):774–79.

Richmond, Mary E. and Fred S. Hall. 1929. *Marriage and The State*. New York: Russell Sage Foundation.

Robinson, J. G. 2001. "ESCAP II: Demographic Analysis Results." Executive Steering Committee for A.C.E. Policy II, Report no. 1, US Census Bureau, October 13.

Ruggles, Steven. 1991. "Comparability of the Public Use Files of the U.S. Census of Population, 1880-1980." *Social Science History* 15(1):123–58.

———. 1997. “The Rise of Divorce and Separation in the United States, 1880–1990.” *Demography* 34(4):455–466.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2018. *IPUMS USA: Version 8.0 [Dataset]*. Minneapolis: University of Minnesota.

Scott, James C. 1999. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.

Setel, Philip W., Sarah B. Macfarlane, Simon Szreter, Lene Mikkelsen, Prabhat Jha, Susan Stout, and Carla AbouZahr. 2007. “A Scandal of Invisibility: Making Everyone Count by Counting Everyone.” *The Lancet* 370(9598):1569–77.

Steckel, Richard H. 1991. “The Quality of Census Data for Historical Inquiry: A Research Agenda.” *Social Science History* 15(4):579–99.

Tollnek, Franziska and Joerg Baten. 2016. “Age-Heaping-Based Human Capital Estimates.” Pp. 131–54 in *Handbook of Cliometrics*, edited by C. Diebolt and M. Haupert. Berlin, Heidelberg: Springer Berlin Heidelberg.

U.S. Bureau of the Census. 1909a. *Marriage and Divorce 1867-1906, Part 1: Summary, Laws, and Foreign Statistics*. Washington, D.C.: Government Printing Office.

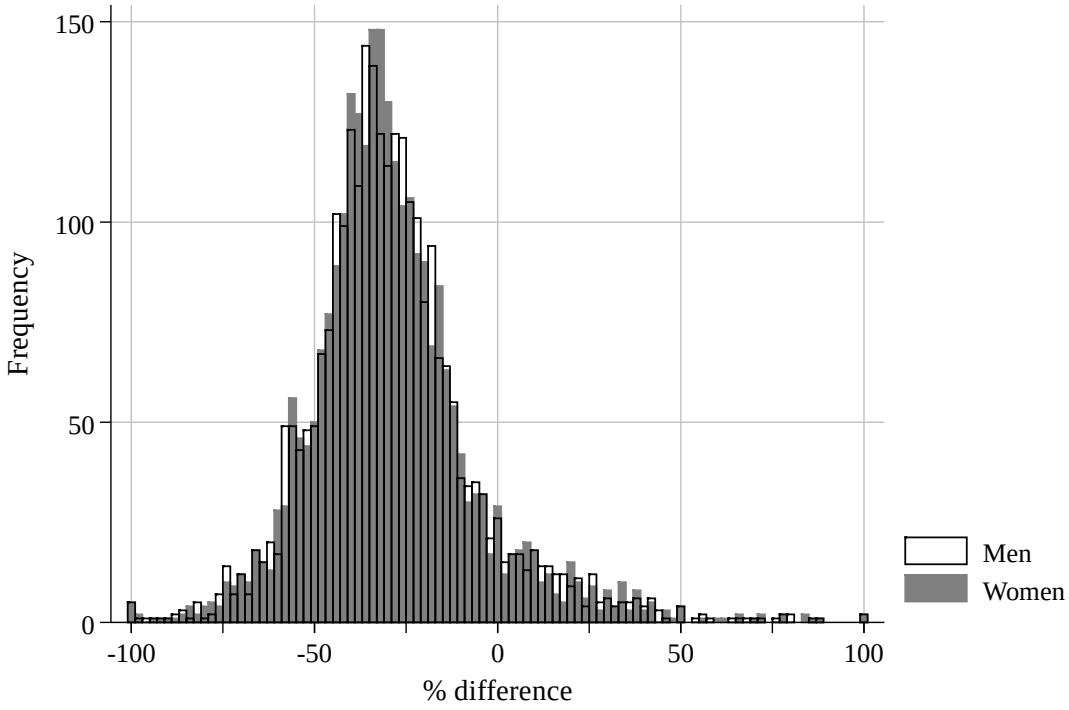
———. 1909b. *Marriage and Divorce 1867-1906, Part 2: General Tables*. Washington, D.C.: Government Printing Office.

Weber, Max. 1978. *Economy and Society*. Berkeley, CA: University of California Press.

Zelnik, Melvin. 1961. “Age Heaping in the United States Census: 1880-1950.” *The Milbank Memorial Fund Quarterly* 39(3):540–73.

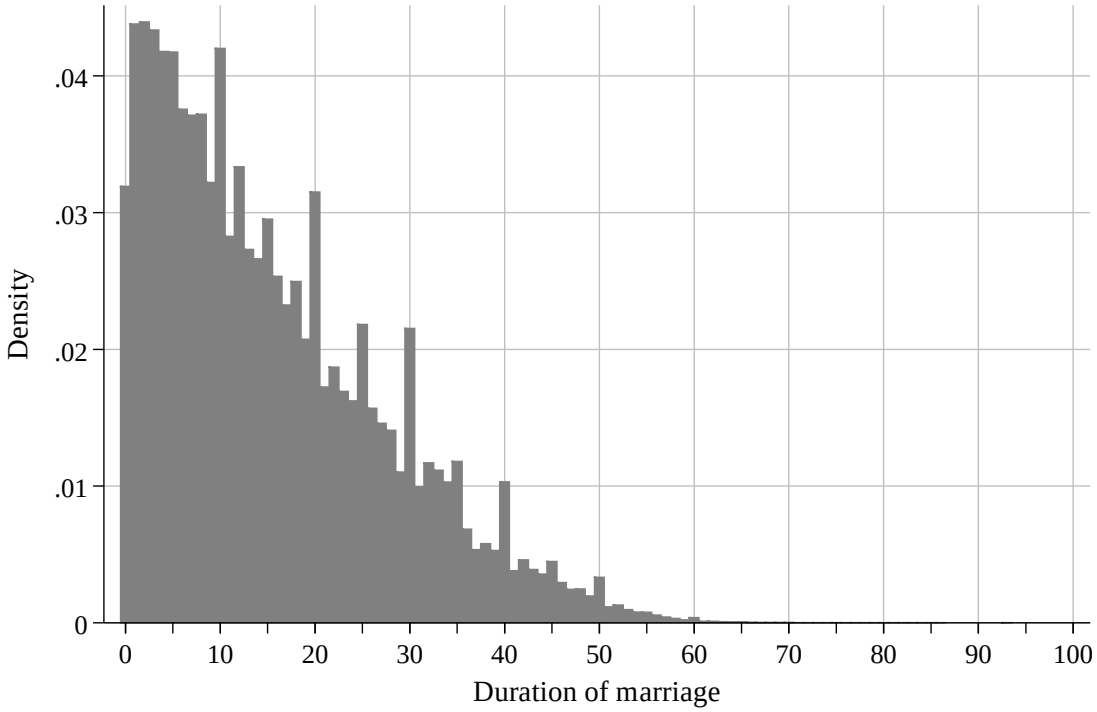
Figures

Figure 1. Percent difference, census and vital records counts of marriages, U.S. counties, 1900.



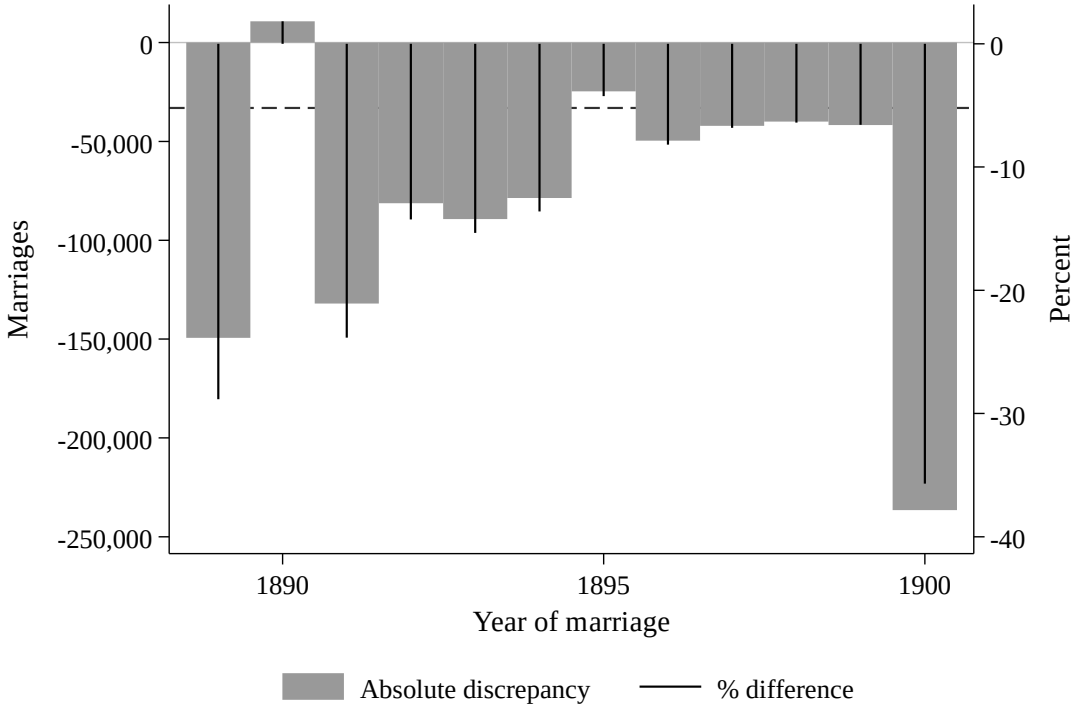
Sources: U.S. Bureau of the Census (1909); Ruggles et al. (2018).

Figure 2. Distribution of marital duration in the 1900 U.S. census.



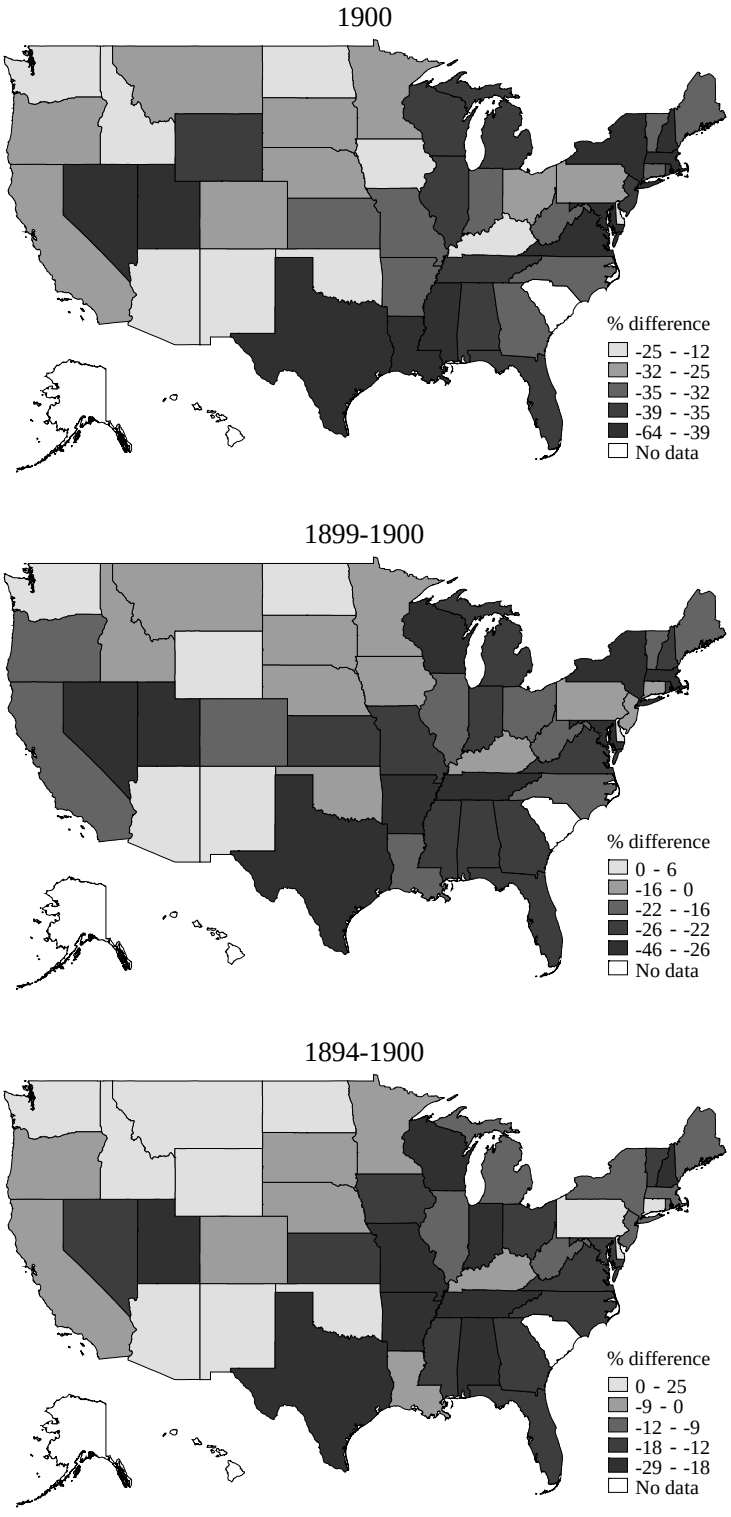
Source: Ruggles et al. (2018).

Figure 3. Annual discrepancies, census and vital records counts of marriages, U.S., 1889–1900.



Sources: U.S. Bureau of the Census (1909); Ruggles et al. (2018).

Figure 4. Percent difference, census and vital records counts of marriages, U.S. states.



Sources: U.S. Bureau of the Census (1909); Ruggles et al. (2018).

Tables

Table 1. Summary statistics, U.S. counties, 1900.

	N	Mean	S.d.	Min.	Max.
% difference in counts					
1900, men	2,698	-28.6	22.8	-100.0	100.0
1900, women	2,698	-28.8	22.6	-100.0	100.0
1900, sex difference	2,698	0.2	4.0	-50.0	38.5
1899–1900, men	2,691	-16.5	20.8	-90.3	89.1
1899–1900, women	2,691	-17.9	19.8	-90.3	88.2
1899–1900, sex difference	2,691	1.4	5.4	-24.0	71.4
1894–1900, men	2,647	-9.7	20.7	-89.1	89.8
1894–1900, women	2,647	-10.3	19.5	-89.3	89.2
1894–1900, sex difference	2,647	0.6	3.6	-15.2	32.4
% mar. dur. heaping	2,783	4.2	2.8	0.0	19.0
% age heaping	2,816	1.5	1.9	0.0	16.5
% literate	2,838	84.8	14.0	26.2	100.0
% urban	2,838	14.0	22.2	0.0	100.0
% black	2,838	12.9	20.6	0.0	93.3
% foreign born	2,838	13.7	16.2	0.0	89.0
% born other state	2,839	29.6	22.4	0.0	97.9
Population in 1,000s	2,838	26.8	73.6	0.004	2,067.9

Sources: U.S. Bureau of the Census (1909); Ruggles et al. (2018).

Table 2. Percent difference, census and vital records counts of marriage, U.S. counties, 1900.

	1900			1899–1900			1894–1900		
Basis of census count	Men	Women	Dif.	Men	Women	Dif.	Men	Women	Dif.
% mar. dur. heaping	-1.576***	-1.480***	-0.096*	-1.007**	-0.865**	-0.142*	-0.494	-0.350	-0.144**
	(0.289)	(0.289)	(0.042)	(0.311)	(0.301)	(0.065)	(0.262)	(0.253)	(0.053)
Centralized reporting	-2.204	-1.947	-0.257	-1.366	-1.242	-0.124	-3.624	-3.266	-0.359
	(2.525)	(2.515)	(0.152)	(2.043)	(1.883)	(0.456)	(2.597)	(2.244)	(0.458)
Non-judicial recording	-3.361**	-3.248**	-0.114	-4.502**	-4.404**	-0.098	-5.158*	-5.096*	-0.062
	(1.172)	(1.187)	(0.192)	(1.543)	(1.440)	(0.349)	(2.034)	(1.907)	(0.318)
Non-recording pen.	0.582	0.493	0.088	1.334	1.209	0.126	1.441	1.236	0.204
	(1.253)	(1.232)	(0.177)	(1.637)	(1.509)	(0.286)	(1.854)	(1.775)	(0.239)
% literate	-0.192	-0.199	0.006	-0.239	-0.231	-0.007	-0.196	-0.189	-0.007
	(0.138)	(0.137)	(0.013)	(0.147)	(0.141)	(0.019)	(0.161)	(0.152)	(0.018)
% urban	-0.158***	-0.159***	0.001	-0.110**	-0.128***	0.019*	-0.083**	-0.092**	0.009
	(0.032)	(0.031)	(0.004)	(0.033)	(0.034)	(0.007)	(0.030)	(0.028)	(0.005)
% black	-0.126*	-0.135**	0.009	-0.048	-0.060	0.012	-0.073	-0.076	0.003
	(0.048)	(0.049)	(0.005)	(0.059)	(0.057)	(0.011)	(0.054)	(0.052)	(0.010)
% foreign born	0.058	0.031	0.027***	0.250**	0.181*	0.069***	0.416***	0.352***	0.065***
	(0.070)	(0.068)	(0.006)	(0.075)	(0.068)	(0.015)	(0.072)	(0.065)	(0.016)
% born other state	0.061	0.046	0.014**	0.067	0.052	0.016	0.110	0.095	0.015
	(0.040)	(0.038)	(0.004)	(0.049)	(0.044)	(0.010)	(0.059)	(0.055)	(0.008)

Table 2. *(continued)*

Basis of census count	1900			1899–1900			1894–1900		
	Men	Women	Dif.	Men	Women	Dif.	Men	Women	Dif.
Region									
Midwest	-6.893*	-5.754	-1.139***	-13.344***	-10.985***	-2.359***	-19.080***	-17.007***	-2.073**
	(2.933)	(2.884)	(0.266)	(3.175)	(2.803)	(0.620)	(4.379)	(3.869)	(0.626)
South	-6.507	-5.935	-0.572*	-11.674**	-10.284**	-1.390*	-10.893*	-9.772*	-1.121
	(3.466)	(3.407)	(0.272)	(3.867)	(3.515)	(0.630)	(4.455)	(3.997)	(0.622)
West	5.373	6.002	-0.630	2.031	1.403	0.628	-0.386	-2.070	1.684
	(3.825)	(3.656)	(0.502)	(4.261)	(3.588)	(1.028)	(5.253)	(4.626)	(0.881)
Observations	2,671	2,671	2,671	2,664	2,664	2,664	2,624	2,624	2,624
Adjusted R^2	0.130	0.122	0.019	0.144	0.116	0.091	0.215	0.170	0.221

Sources: U.S. Bureau of the Census (1909); Ruggles et al. (2018).
Notes: Standard errors, clustered at the state level, in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$