Methodology Appendix

By Isaac Maddow-Zimet, Kathryn Kost and Sean Finn

© 2020 Guttmacher Institute
Acknowledgements

This report was prepared by Isaac Maddow-Zimet, Kathryn Kost and Sean Finn, all of the Guttmacher Institute. It was edited by Jenny Sherman. Research support was provided by Ashley Little. Aletha Akers, Haley Ball, Joerg Drewke, Liza Fuentes, Kimberly Lufkin and Elizabeth Nash reviewed the report and provided comments and suggestions. The Guttmacher Institute gratefully acknowledges the unrestricted funding it receives from many individuals and foundations—including major grants from the William and Flora Hewlett Foundation and the David and Lucile Packard Foundation—which undergirds all of the Institute’s work.

Suggested Citation

This appendix describes the methodology and data sources used to calculate the estimates presented in *Pregnancies, Births and Abortions in the United States, 1973–2016: National and State Trends by Age*, as well as in the accompanying data set and tables. These estimates encompass pregnancy, birth and abortion counts and rates over time and by age-group, for the 50 states and the District of Columbia. Our estimation procedure involves combining multiple data sources, each with its own strengths and limitations. We describe each data source and how we assessed data quality. We conclude with a discussion of how and why our estimates may differ from other published estimates, including prior estimates published by the Guttmacher Institute.

**Definitions of terms**

We define pregnancies as the sum of births, abortions and fetal losses (i.e., miscarriages, ectopic pregnancies and stillbirths). Demographic rates are defined as the number of events (in this case, pregnancies, births or abortions) divided by the number of individuals who could experience the event. The accuracy of demographic rates depends on having accurate counts of both events and the population of people who are able to become pregnant. In reality, that population includes some proportions of cisgender women, of transgender men and of people whose gender is nonbinary. To our knowledge, however, there are no comprehensive estimates of the number or proportion of the population that is able to become pregnant. As a proxy, the population we use for calculation of rates is the number of women in a given age-group (the denominator), as reported by the Census Bureau. Consequently, we describe rates as being among women, although counts of events (the numerator: births, abortions, fetal losses or all pregnancies) include outcomes among all people able to become pregnant, regardless of their gender.

Rates in the tables, text and accompanying data are calculated as the number of events per 1,000 women of a given age-group. Rates among women younger than 15 are calculated as the number of events per 1,000 women aged 14; rates among women aged 40 or older are calculated as the number of events per 1,000 women aged 40–44. Abortion ratios are the number of abortions per 1,000 live births within a given age-group.

“Age” refers to an individual’s age when their pregnancy ended. This can lead to some measurement error in pregnancy estimates, particularly for age-groups where there are large differences in pregnancy rates between the highest age in one age-group and the lowest age in the subsequent group. For example, most individuals who conceive at age 19 and whose pregnancy results in a birth will give birth approximately nine months later, after they turn 20. Thus, their pregnancies would not be counted among the 15–19-year-old or 18–19-year-old age-groups. Because pregnancies are concentrated at the highest ages among 15–19-year-olds, the pregnancy rates among these groups are likely higher than the estimates in this report. Similarly, “year” refers to the calendar year in which the birth, abortion or fetal loss occurred, not when the conception occurred.
Data and methods

The estimates produced for this report rely on multiple data sources, and differ slightly for national- and state-level statistics. Citations are available in the Data Sources section.

National-level estimates are based on data from the following sources:

- Number of births, by age: National Center for Health Statistics (NCHS) National Vital Statistics System (NVSS)
- Number of abortions: Guttmacher Institute Abortion Provider Census (APC)
- Distribution of abortions, by age: Centers for Disease Control and Prevention (CDC) Abortion Surveillance Reports
- Population of women: Census Bureau

State-level estimates are based on data from the following sources:

- Number of births, by age: NCHS NVSS
- Number of abortions, by state of residence: Guttmacher Institute APC, CDC Abortion Surveillance Reports and data from state health departments
- Distribution of abortions, by age: CDC Abortion Surveillance Reports and data from state health departments
- Population of women: Census Bureau

Births

We obtained the numbers of births occurring each calendar year in each state and in the District of Columbia from the NCHS’s National Vital Statistics System. These data are based on information abstracted from birth certificates and include the age of the mother at delivery.

In instances of a multiple birth (i.e., twins and higher order births), we count each infant delivered as a single pregnancy. This results in a slight upward bias in pregnancy rates—a bias that is more prominent for older women compared to younger women, as the incidence of multiple births increases with maternal age. The impact is also likely to vary over the time period represented in our data set. In the most recent year for which estimates are available (2016), 6% of births to people aged 40–54 were multiple-birth deliveries. This is a substantial increase from the earliest year (1973), in which only 2% were multiple in the same age-group. In contrast, fewer than 2% of births were multiple among women aged 15–19 in 2016, a proportion that has been relatively stable over time. As a result, there is likely only a small impact on pregnancy rates estimated for this age-group.

*National Vital Statistics Reports refer to births from a single pregnancy with multiple fetuses as “multiple births.” However, this does not necessarily mean that there were multiple live births from a single pregnancy. Birth certificates in the United States define multiple births to include fetal loss. An individual birth certificate of a multiple live birth does not indicate the outcome of the other fetus(es).
Abortions: national-level


Our estimates include only reported legal abortions. There is evidence that illegal abortions were performed for some time after the U.S. Supreme Court effectively legalized the procedure in 1973, but there are no accurate records or estimates of the number. Thus, our estimates of abortion and pregnancy in the early- to mid-1970s may be underestimated.

Distributions of abortions by age

While we obtained total counts of abortions occurring each year in the United States from the Guttmacher Institute’s APC, we obtained the proportion occurring among each age-group using data on the characteristics of abortion patients from the CDC’s Abortion Surveillance Reports. For each year since 1969, the CDC has compiled abortion data provided by state health departments, most of which maintain an abortion reporting system with state-specific standard certificates that providers use to record abortions and associated patient characteristics. The CDC’s estimates are subject to some limitations. Not all states require providers to report abortions, and states that do have reporting systems differ in the completeness of reporting, the types of providers that must report and the information obtained on the abortion certificate. In addition, reporting to the CDC is voluntary, and not all states that collect abortion data report the results to the CDC each year. Finally, in given years, state-specific data sometimes have high proportions of cases that are missing data on the age of the patient.

Issues of data quality and methods to ensure consistency in estimation across years

To produce comparative estimates from one year to the next, we use a methodology developed several decades ago by Guttmacher Institute research staff to minimize the effect of year-to-year differences in which states report data to the CDC, as well as year-to-year fluctuations in the quality or completeness of data reported. To do so, we calculate yearly changes in the proportion of abortions to each age-group using only data from those states that have comparable data for the previous year. We consider the data to be comparable if a state reported data on abortions by age to the CDC in both years, if there was no more than a 4% change in the proportion of abortions to those of unknown age and if the population for which
abortion information by age was collected was identical between years (e.g., only residents of the state). If these conditions are met, we then apply these changes to an ongoing historical trend anchored by a “base” year (2010).

Appendix Figure 1 shows the impact of this approach for one age-group: 20–24-year-olds. The red line shows the proportion of all abortions to individuals in this group as reported in the CDC report for a given year (as a proportion of abortions with known age). The blue line shows the proportion after being adjusted using our described methodology. The unadjusted and adjusted proportions are extremely close for most of the reporting period, particularly in more recent years; however, the data diverge at several points, illustrating the error in the trend that can be introduced when relying solely on the unadjusted CDC reports.

Two such divergences are labeled in the figure. The first is the spike in 1984, the year that the CDC stopped reporting data by age for California. Because the proportion of abortions to individuals aged 20–24 in California was lower than the national average in the early 1980s, excluding California from the CDC’s surveillance in 1984 led to a sharp increase in the proportion of abortions to 20–24 year olds in the CDC report. Our applied adjustments to the data account for this and other changes in reporting. Although there was, in fact, a small increase in the proportion of abortions to 20–24-year-olds in states with data for both years, it was much smaller in scale than the spike seen in the unadjusted CDC data.
In some cases, data artifacts can occur in the CDC report’s distributions because of changes in data quality in specific states. For example, from 2005 to 2006, Tennessee had a large increase in the proportion of abortions had by patients with unknown age, from 0.3% of abortions to 11.4%; meanwhile, among abortions had by patients with known age, the proportion had by 20–24-year-olds dropped from 34% to 30%. This decrease was likely a reflection of the higher proportion of missing data in 2006 as compared to 2005 (both the proportion of abortions had by patients with unknown age and the proportion of abortions had by 20–24-year-olds returned to approximately their prior levels in 2007); our adjustment methodology explicitly accounts for this kind of variation in data quality or completeness.

Representativeness of CDC abortion data

The adjustments described above account for year-to-year changes in the state data that are reported to the CDC. However, these adjustments cannot ensure that the data on the proportion of abortion patients with specific characteristics, including age, accurately reflect the true distribution nationally. In particular, because several states do not regularly report abortion data with patient characteristics to the CDC (e.g., California, Florida, Maryland, New Hampshire and Wyoming), the CDC distributions may be biased to the extent that the distribution of abortion patients by characteristic from those missing states differs from the national distribution.

To examine the extent of this potential bias, we systematically check the adjusted and unadjusted distributions of abortions by age against a secondary source of data: the Abortion Patient Survey (APS), a nationally representative survey of abortion patients fielded periodically by the Guttmacher Institute. Estimates from the APS, which is a sample-based survey, have some degree of sampling error and are only available for selected years. However, because the sampling design is constructed to be nationally representative, we can use the estimates to assess to what degree CDC distributions may be biased by excluding states that do not report data. Appendix Figure 2 shows the result of this comparison for each age-group.
In almost all cases, the confidence intervals of estimates from the 1994, 2000, 2008 and 2014 APS surveys overlap with the raw and adjusted CDC estimates. This suggests that the absence of some state data in CDC reports does not have a large impact on estimates of abortion for each age-group. The lack of impact may reflect minimal variation across the states in the distribution of abortions by age.

In contrast, missing data from states may have a considerable impact on CDC distributions of abortions by race and ethnicity. The CDC published tables for abortions by Hispanic origin starting in 1991, and tables for abortions by race and ethnicity combined starting in 2007. Appendix Figure 3 shows large differences in distributions of abortions by race and ethnicity.

†The CDC has also published estimates by race alone since the early 1990s; however, because of differences in how the data were collected, there are no comparable estimates from the APS.
between the unadjusted CDC data and comparable estimates from the APS, for years in which data was available.

Estimates from the CDC of the proportion of abortion patients who are of Hispanic origin, as well as the proportion of abortion patients who are Black and not of Hispanic origin, are substantially different from corresponding estimates from the APS—especially in recent years. In 2014, the CDC estimate of the proportion of abortion patients of Hispanic origin was more than 10 percentage points lower than the corresponding estimate from the APS; that same year, the proportion of abortion patients who were Black and not of Hispanic origin was more than 8 percentage points higher in the CDC than in the APS.

These discrepancies are much larger than those by age, in part because fewer states reported data by Hispanic ethnicity than reported data by age (in 2014, 36 and 46 states, respectively).

Methodology: Pregnancies, Births andAbortions in the United States
Consequently, it is likely that the states that do report data on ethnicity are not representative of the United States as a whole; in addition, the distribution of abortion patients by combined race and ethnicity likely varies more between states than the distribution of abortion patients by age. In particular, the CDC estimates do not include California or Florida, two states that have large Hispanic populations (with state residents accounting for 35% of all Hispanic women of reproductive age in the country in 2016) and which together account for almost a third of the abortions occurring in the United States (31% in 2016).

For these reasons, we believe that the CDC data on Hispanic ethnicity is not nationally representative, and we do not calculate pregnancy rates by race and Hispanic ethnicity in this report because we lack confidence in the reliability of the data currently available. Although APS estimates are designed to be nationally representative, they are only available for a few years and have wide confidence intervals around their estimates (the 95% confidence interval for the proportion of abortion patients who were Hispanic in 2014 ranges from 21% to 29%). The APS therefore has a limited ability to capture changes over time, or even between racial and ethnic population groups.

In prior reports focusing on pregnancy among adolescents and young adults, we published rates for 15–19-year-olds by race and ethnicity combined; on the basis of the data-quality investigations presented here, we see a need to obtain better quality data to produce estimates that would accurately represent trends or levels in national pregnancy rates by race and ethnicity. To provide full access to all estimates and ensure comparability for those who may use previously published estimates of pregnancy rates by race and ethnicity, we include statistics by race and ethnicity in our public data set, applying the same sources and methodology as in all other updated historical estimates. We do not publish them in the report’s tables, however, and they should be used with caution, given the variability around the estimates and the conflicting estimates from different data sources highlighted here.

### Abortions: state-level

Our estimates of the incidence of abortion by age-group at the state level are based on estimates of abortions by state of residence. The methodology for estimating abortions by state of residence requires data collected from numerous sources, in addition to abortion counts from the Guttmacher Institute’s Abortion Provider Censuses (APCs). Abortions by state of residence are only available for selected years in the period 1988–2005; starting in 2005, they are estimated for every year.

The APCs provide the annual number of abortions occurring in each state for 1988, 1992, 1996, 2000, 2005, 2008, 2010, 2011, 2013, 2014 and 2016. Starting with the number of abortions performed in each state for patients of all ages (residents and nonresidents), we reassign abortions to the patients’ state of residence on the basis of information collected by state abortion reporting agencies. The Centers for Disease Control and Prevention (CDC) compiles this information in an annual table showing, for each state, the state of residence for all individuals who had had an abortion in that state. The CDC table suppresses information on any combination of state of residence and state of occurrence with fewer than 50 abortions, so we use more-complete data available from many state health department websites and, where
necessary, contacted the states directly for this information. Some states do not collect this information; for these states, we use data collected directly from abortion facilities on patients’ state of residence. For the years 2006, 2007, 2009, 2012 and 2015, we estimate the total number of abortions among residents of each state using Stineman interpolation and the stinepack package in R version 3.6.1.\textsuperscript{21,22}

From the total of each state’s residents who had had an abortion, we use the proportions who were in each age-group from the CDC Abortion Surveillance Report for the year of the estimate. In a few cases where states did not provide data to the CDC, or where significant corrections had been made to the data subsequent to its publication in the CDC report, we use data collected directly from state health departments. For states with incomplete or no information on the age of patients who had had an abortion, we estimate the proportion of abortions obtained by individuals in each age-group, or in selected age-groups, using the national distribution or the distribution from neighboring states (Methodology Appendix Table). This estimation is more common in earlier years of the data (11 states in 1988 vs. five in each year from 2010 to 2016). Note that state-specific abortion counts do not sum to national occurrence totals; this is in part due to the small number of abortions in the United States had by nonresidents (<1% in 2016), and in part due to the methodological differences in the way national and state estimates are calculated.\textsuperscript{‡} This discrepancy is small (<1%) for most years, but it can be larger in years in which total counts of abortions occurring in the United States and the counts of abortions among residents of each state were both estimated through interpolation. In particular, in 2006, the method of interpolation for abortions nationally is able to capture a spike in national counts that our state-specific interpolation methods do not, which leads to a slightly larger discrepancy in totals (by 42,520 abortions, or 3.4% of the national total).

Our calculation methods assume that women in each age-group travel outside their home state for abortion services in roughly similar proportions. This assumption may not be valid for states with parental involvement requirements, or in surrounding states to which individuals younger than 18 may travel to obtain an abortion. The extent to which young people travel out of state because of these requirements is not known; further research to measure how frequently this occurs and whether it varies by age could contribute to the accuracy of the estimates.

Fetal losses

The incidence of pregnancy includes conceptions that do not result in induced abortion or live birth. The majority of these pregnancies end in spontaneous abortion (i.e., miscarriage); smaller numbers end in ectopic pregnancy (implantation of a fertilized ovum outside of the main uterine cavity) or stillbirth (most often defined as pregnancy loss after the 20th week of pregnancy). Stillbirths are generally reported by hospitals in death records. Many states collect miscarriage data, but reporting is incomplete, particularly for miscarriages that occur prior to 20 weeks’ gestation. In fact, many women can experience a very early miscarriage without

\textsuperscript{‡}In prior releases of this report, a small adjustment factor was applied to the state-specific estimates so that they would add to the total. We no longer include this adjustment.

Methodology: Pregnancies, Births and Abortions in the United States
realizing it, perhaps experiencing what they believe to be a late menstrual period. Our estimates of pregnancy are intended to include only fetal losses after six or seven weeks.

We estimate fetal loss as the sum of a proportion of births (20%) and of abortions (10%). While these proportions are rough approximations for the numbers of fetal losses, they account for the fact that the incidence of fetal loss in a population is dependent on the ratio of abortions to births. Pregnancies terminated by abortion are exposed to the risk of fetal loss for a shorter duration of time than pregnancies that are continued to term, especially if the abortion occurs in the early weeks of the pregnancy. Most miscarriages occur by the 14th week of pregnancy, and more than 90% of abortions occur before that time. Estimation of fetal loss as a proportion of births and abortions accounts for these different gestation-dependent risks associated with pregnancies that end early (abortions) and those that end later (births). Put simply, the greater the proportion of pregnancies terminated by abortion, the less opportunity for a miscarriage or other fetal loss to occur and the lower the overall incidence.

It is possible to obtain other estimates of fetal loss. In particular, the National Center for Health Statistics (NCHS) reports of pregnancy statistics have used respondents’ reports of pregnancy loss from the nationally representative National Survey of Family Growth (NSFG) to estimate national-level incidence of recognized fetal loss. Estimates from the NSFG are obtained for a specific period of time, covering fertility experiences from five to seven years before the year of the survey. However, evidence suggests that fetal loss is underreported in the NSFG, with substantial variation in underreporting according to respondents’ sociodemographic characteristics. Prior work has also found substantial variation between age-groups in how early a pregnancy is recognized; this could, in turn, drive apparent variation in reports of fetal loss that are not representative of real underlying patterns.

For this report, we calculate annual rates over a time period of more than 40 years. Instead of using NSFG self-reports, which would require us to make assumptions about changes in the degree of underreporting over time and between age-groups, we used the proportionate approximation—20% of births and 10% of abortions. Although this approximation is constant over time, it is sensitive to changes in the relative distribution of births and abortions over this longer period. This is important because in years when abortion rates are high, relatively fewer fetal losses can occur. Failure to account for changes in the relative distribution of births and abortions (which have changed dramatically for some age-groups since the early 1970s) would mean that the incidence of fetal losses—and consequently the incidence of pregnancy—could be over- or underestimated.

Finally, these proportions—20% and 10%—are intended only as approximations based on what little data are available on the incidence of fetal loss. Because we use fixed proportions for all age-groups, we likely underestimate pregnancy rates among older individuals, for whom fetal loss may be more likely. We use the estimates of fetal loss to provide comparable estimates of pregnancy over time that account for variations in the relative incidence of live births and abortions. Because of the limitations of these approximations of fetal loss incidence, these estimates should not be used to study trends in rates of fetal loss over time or for comparisons of incidence between age-groups.
Other sources of pregnancy statistics

The historical estimates in this report may differ from those found in other sources. First, for national-level estimates provided in reports published by the NCHS, estimation of fetal loss differs significantly, as previously described.\textsuperscript{19} Second, the denominator we use in calculating rates among women younger than 15 is the female population aged 14, because most pregnancies occurring among these very young women are likely among 14-year-olds (this was true for 81% of births in 2016). Estimates in existing NCHS reports use the number of women aged 10–14 for the denominator.\textsuperscript{2} We use the number of women aged 15–19 in the denominator for rates among all women younger than 20 because from 1973 to 2016, nearly all (95–97%) pregnancies among women younger than 20 were among those aged 15–19. Finally, our denominators are based on population estimates produced by the Census Bureau for July 1 of each year and revised periodically. Our rates may differ slightly from those published elsewhere if the population estimates are obtained from differing “vintages” (estimates are updated yearly between decennial censuses) and if the rates have been updated using the final intercensal population estimates available after each national census. For the years 1980, 1990, 2000 and 2010, the NCHS uses the April 1 census counts; we use the July 1 estimates for these and all other years. We update our rates when the final intercensal estimates are released.

Our state-level estimates for adolescents and young adults differ slightly from prior reports published by the Guttmacher Institute, as we updated our methodology and applied it retroactively to data for all previous years to ensure year-to-year comparability. In particular, we no longer adjust our state-level estimates to add to national totals. We use additional data for specific states; in particular, we recalculated estimates for Texas for the years 2006–2013 because of an error in the calculation of age at outcome,\textsuperscript{32} which was included in the data that that state provided to the CDC. With that correction to the data, estimated pregnancy rates for 15–19-year-olds in Texas changed substantially for 2006–2013, compared to Guttmacher’s prior published estimates. However, in comparing this report’s current estimates for each state for every year to the estimates in Guttmacher’s prior publications, the estimated pregnancy rates for 15–19-year-olds changed by less than 4% in 90% of the state-year pairs.

In prior Guttmacher reports focusing on pregnancy among adolescents and young adults, we estimated pregnancy rates by combined race and ethnicity for 15–19-year-olds nationally and at the state level.\textsuperscript{33} In this report, we drop those estimates at the national level because of the data quality issues described. At the state level, we do not calculate estimates by a cross of age with race and ethnicity because we believe that a key assumption—that patients in each age-group and race and ethnicity population group travel outside their home state for abortion services in roughly the same proportions—is violated for these data. Although no studies to our knowledge have examined the characteristics of patients travelling to other states (as compared with patients who have abortions in their state of residence), recent research has found strong differentials in distance travelled according to patient race,\textsuperscript{34} likely reflecting structural differences in the substantial material, informational and social resources required to travel across state lines.\textsuperscript{35}
Data sources

Population


Births


Abortions


**METHODOLOGY APPENDIX TABLE: Methods for estimating percent distribution of abortions by age of patients for states with missing data**

<table>
<thead>
<tr>
<th>State</th>
<th>Years estimated</th>
<th>Estimation methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>1988</td>
<td>Weighted average of Georgia and Mississippi</td>
</tr>
<tr>
<td>Connecticut</td>
<td>1992</td>
<td>Age distribution of abortions in Connecticut in 1993</td>
</tr>
<tr>
<td>Delaware</td>
<td>1988, 1992, 1996</td>
<td>Weighted average of Maryland, North Carolina and Virginia</td>
</tr>
<tr>
<td>Iowa</td>
<td>1988, 1992, 1996</td>
<td>Weighted average of Kansas, Minnesota, Missouri, Nebraska and South Dakota</td>
</tr>
<tr>
<td>Illinois</td>
<td>1988, 1992</td>
<td>Weighted average of Indiana, Missouri and Ohio</td>
</tr>
<tr>
<td>Louisiana</td>
<td>2005–2009</td>
<td>The proportion of abortions had by people aged 15–17 and 18–19 among abortions had by 15–19-year-olds, distributed according to weighted average of Texas, Mississippi and Arkansas</td>
</tr>
<tr>
<td>State</td>
<td>Years estimated</td>
<td>Estimation methods (continued)</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
2011–2016: Simple average of Maine, Massachusetts and Vermont |
| New Jersey    | 1988            | The proportion of abortions had by people aged 15–17 and 18–19 among abortions had by 15–19-year-olds, distributed according to weighted average of New York and Pennsylvania |
| West Virginia | 1988            | Weighted average of Tennessee and Maryland                                                   |
2011–2016: Simple average of Colorado, Idaho, Montana, Nebraska, South Dakota and Utah |
References


Methodology: Pregnancies, Births and Abortions in the United States


