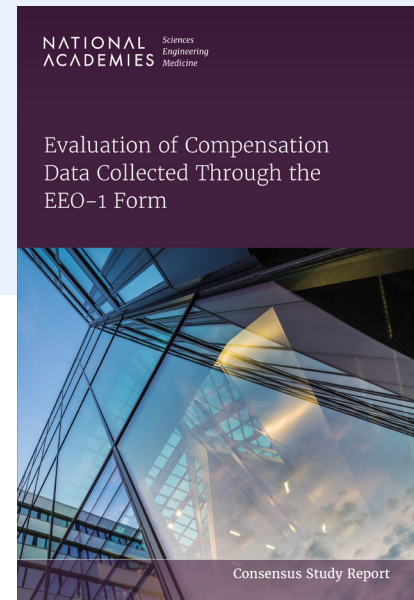


# Evaluation of Compensation Data Collected Through the EEO-1 Form



Pay disparities—inequalities in earnings between women and men and between different racial and ethnic groups—are well documented in national statistics. In 2020, the median earnings of women who worked full time, year-round were 82 percent of those earned by men over the same period.<sup>1</sup> In comparison with non-Hispanic White workers, Black workers earned 76 percent as much and Hispanic workers earned 68 percent as much. Perhaps most strikingly, Black and Hispanic women earned only 63 percent and 55 percent as much, respectively, of what non-Hispanic White men earned.<sup>2</sup> Although differences in pay can often be attributed to differences in workers' education, skills, work experience, or occupation, these factors fail to fully explain these significant sex and race/ethnicity pay gaps.

Differences in pay on the basis of sex and race/ethnicity have been outlawed by the federal government for almost 60 years, and the U.S. Equal Employment Opportunity Commission (EEOC) has statutory authority to enforce pay equity. To achieve its mission, EEOC implements and communicates policies developed to protect workers and investigates charges of discrimination.

A starting point for EEOC investigations into discrimination is the suite of data known as EEO.<sup>3</sup> Particular EEO data collections differ by employer type and filing requirements. The EEO-1 collects information related to establishment and employee characteristics from both private employers with 100 or more employees and

<sup>1</sup> U.S. BLS. (2022). *Women in the Labor Force: A Databook*. Available: <https://www.bls.gov/opub/reports/womens-databook/2021/home.htm>.

<sup>2</sup> U.S. Census Bureau. (2021). *PINC-10. Wage and Salary Workers—People 15 Years Old and Over by Total Wage and Salary Income, Work Experience, Race, Hispanic Origin, and Sex*. Available: <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-10.2019.html>.

<sup>3</sup> The full suite (EEO-1, EEO-2, EEO-3, EEO-4, and EEO-5) collects information from different types of employers regarding the characteristics of establishments and employees.

federal contractors that meet certain criteria. Although initial findings are followed by more in-depth review of individual establishments, EEO data have an important and unique role in initial investigations. EEO data are the only federal source of occupation and demographic data linked to place of establishment.

To improve its ability to investigate and address pay disparities, EEOC expanded the EEO-1 data collection for two years to include measures of pay and hours worked in addition to data on sex, race/ethnicity, count of employees, and employer characteristics from private employers. The new data cover reporting years 2017 and 2018.

In 2020, EEOC asked the National Academies of Sciences, Engineering, and Medicine to examine the quality of the newly collected data (known as Component 2<sup>4</sup>) and provide recommendations for future data collection efforts. The National Academies convened a panel of experts comprised of economists, sociologists, statisticians, survey methodologists, lawyers, and employer advisors with expertise in measuring data quality, pay gaps, pay discrimination, and pay equity to address this task. The panel's report, *Evaluation of Compensation Data Collected Through the EEO-1 Form*, offers recommendations for the use of Component 2<sup>4</sup> data as collected and improvements for future collections.

## FINDINGS

Overall, the panel found that the Component 2<sup>4</sup> data collected for reporting years 2017–2018 have value, but both short-term and longer-term improvements are needed to address significant concerns in employer coverage, data collection protocols, measurement implementation, and conceptual definitions. If implemented, these recommendations could improve the breadth and strength of EEOC data for the purposes of monitoring and addressing pay equity, in some ways reduce employer burden, and respond to employer concerns about the precision of the Component 2<sup>4</sup> data collection instrument used for reporting years 2017–2018.

### Value of Component 2<sup>4</sup> Data and Issues of Concern

The panel found that the Component 2<sup>4</sup> data collected for reporting years 2017–2018 are a valuable resource:

<sup>4</sup> The original survey, which did not include pay or hours worked data, is known as Component 1.

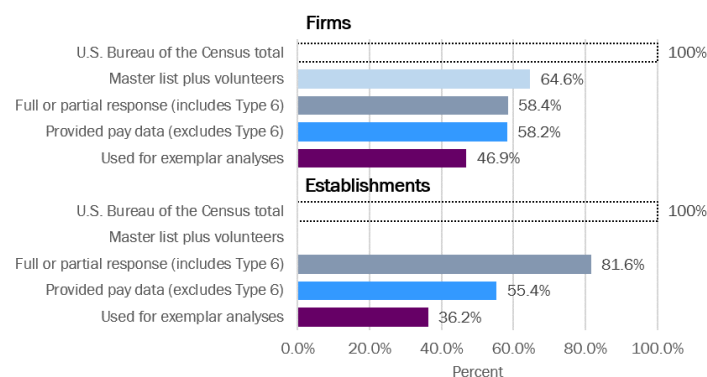
they represent the only federal data source for pay, hours worked, occupation, and demographic characteristics collected at the employer level from the private sector. As collected for reporting years 2017–2018, the Component 2<sup>4</sup> data are helpful for enforcement efforts, for employers' self-assessments, and for providing a broad description of pay gaps. However, there is a strong need to improve the utility of future pay data for these purposes by addressing issues related to coverage, unit response rates, missing data, and extreme values.

### Incomplete Data

Overall, the panel found the data to be incomplete: see Figure 1. Although all eligible firms and establishments were required by law to complete the survey, EEOC only actively reached out to 65 percent of them to request completion of the survey. In total, 58 percent of eligible firms, covering 82 percent of establishments, completed the survey (see Figure 1). However, EEOC did not require firms to provide pay data for establishments with fewer than 50 employees, resulting in pay data for only 58 percent for firms and 55 percent for establishments.

### Reliability

The panel also had concerns about data reliability. Although most of the reported numbers of employees and hours worked appeared to be reliable, the panel found issues in some of the reported numbers of employees or hours that could lead to misleading results if not



**FIGURE 1** Pay data coverage as compared with total eligible, 2018.

NOTES: "Full or partial response" excludes firms reporting more employees than the largest U.S. employer (1.4 million). "Provided pay data" also excludes establishments for whom no pay data were submitted (using the Type 6 form for small establishments). "Used for exemplar analysis" further filters out potentially unreliable data.

SOURCES: Data from panel generated from Component 2<sup>4</sup> employer, establishment, and employee files for 2018; [https://data.bls.gov/cew/apps/data\\_views/data\\_views.htm#tab=Tables](https://data.bls.gov/cew/apps/data_views/data_views.htm#tab=Tables); and [https://www2.census.gov/programs-surveys/bds/tables/time-series/bds2019\\_ez.csv](https://www2.census.gov/programs-surveys/bds/tables/time-series/bds2019_ez.csv).

addressed prior to analysis. After excluding data that were considered less trustworthy, 47 percent of eligible firms and 36 percent of eligible establishments remained for analyzing pay differences.

### Measurement

The panel identified several measurement concerns in the 2017–2018 Component 2<sup>4</sup> data collection, some of which are listed here. First, the measure of pay only partially reflects total compensation and may mask compensation differences; Box 5 from the W-2 form would provide a better measure than the use of Box 1. Second, switching from the method of collecting pay and demographic characteristics in aggregate form (12 pay bands<sup>5</sup>) to collecting this information for individual workers (pay rates) could substantially reduce respondent burden, increase precision in estimating pay gaps, and still protect confidentiality. Third, the 10 job categories<sup>6</sup> in the 2017–2018 data were too broad for describing the workforce.

The collected Component 2<sup>4</sup> data also do not fully measure race/ethnicity or sex. For race/ethnicity, the data do not provide a way to identify more than one specific race for an individual, and they do not support distinguishing Hispanic persons by race. For sex, data are collected on binary sex, but do not include measures of non-binary sex, gender identity, nor sexual orientation, which are part of EEOC’s mandates. Data are also not collected about other groups that EEOC is charged to protect: persons over age 40, persons with disabilities, and veterans.

The collected Component 2<sup>4</sup> data do not include measures of legitimate causes of pay differences, such as educational attainment and job experience. Such information would assist both EEOC enforcement efforts and employers’ self-assessments. Employers’ self-assessments could contribute to improved employment equity.

<sup>5</sup> The 12 pay bands used in the Component 2<sup>4</sup> instrument are \$19,239 and under; \$19,240–\$24,439; \$24,440–\$30,679; \$30,680–\$38,999; \$39,000–\$49,919; \$49,920–\$62,919; \$62,920–\$80,079; \$80,080–\$101,919; \$101,920–\$128,959; \$128,960–\$163,799; \$163,800–\$207,999; and \$208,000 and over.

<sup>6</sup> The 10 EEO-1 job categories used in the Component 2<sup>4</sup> instrument are executive/senior level officials and managers; first- or midlevel officials and managers; professionals; technicians; sales workers; administrative support workers; craft workers; operatives; laborers and helpers; and service workers.

### Appropriate Use of Collected Component 2<sup>4</sup> Data

Despite the errors and measurement concerns in the 2017–2018 Component 2<sup>4</sup> data, the panel found that, with appropriate editing or filtering, the data could be used in two ways:

- They could be used as an initial step in prioritizing investigations and the allocation of resources, including for calculating pay gaps within an establishment under investigation, making comparisons across similar establishments, and for systemic investigations. However, these uses depend on the availability and quality of data for the particular establishments under investigation.
- They could be used to obtain estimates of raw pay gaps at the national level by sex, race/ethnicity, and occupation.

However, the panel concluded that some intended uses are not suitable for the Component 2<sup>4</sup> data:

- Without extensive cleaning, the 2017–2018 data on hours worked are unsuitable for calculating hourly wages.
- The data are unsuitable for direct determinations of bias or reasonable cause for enforcement purposes.

The panel also found that the Component 2<sup>4</sup> data have limited utility to assist employers’ self-assessments because they do not contain measures of legitimate causes of pay differences, such as educational attainment and job experience. Furthermore, the occupation and pay band categories used in the Component 2<sup>4</sup> collection are much broader than the detailed data available to employers. Employers typically use these sources of information to conduct their internal assessments of pay equity.

## RECOMMENDATIONS

### Address Sources of Error in the Data

To address likely sources of error related to the significant lack of coverage of firms and establishments in the data, the panel recommends improving both the master frame and outreach to newly eligible firms. EEOC might use an

interagency agreement with the Bureau of Labor Statistics as a tool for appropriately maintaining business registers.

#### **Improve Data Quality and Ability to Examine Trends**

EEOC's approach to assigning identification numbers made it difficult to match establishments filing 2017 and 2018 reports. The panel understands this was done to protect confidentiality, but there are several better ways to protect confidentiality that do not prevent authorized users from matching records and thereby assessing data quality and trends over time. To address this issue, EEOC should use consistent and unique firm and establishment identifiers, facilitating data merges and data checking.

#### **Improve Ability to Collect More Complete Data**

EEOC should update its instructions to filers to conform to the federal standard on measuring race/ethnicity, which offers solutions for reporting race/ethnicity data in a combined format. EEOC should work with other federal agencies to develop and test ways to measure employees' sex, gender identity, and sexual orientation in a manner appropriate for EEOC data collections. Data on the status of other protected groups might also be collected using established measures, after appropriate development work and taking into account employer burden. Collecting additional types of data, such as education, job experience, and tenure, would inform EEOC's initial investigation of pay disparities and also support employers' self-assessments.

#### **Test, Develop, and Implement the Collection of Individual Worker Data**

EEOC's summary approach used for aggregate pay and hours worked data in reporting years 2017–2018 severely limits the utility of the data collected, unnecessarily increases employer burden, and complicates the collection of additional key information. Collecting data from employers at the level of individual workers may be less burdensome and would markedly increase the utility of the pay data. The Bureau of Labor Statistics' Occupational Employee Wage Survey is a model. Other federal and state agencies routinely collect individual data for tax, social security, and unemployment insurance purposes. After appropriate field testing, if found appropriate, EEOC should transition to the collection of individual-level employee pay data.

#### **Improve Appropriate Access to Pay Data**

EEOC should strengthen consultation and data sharing with the public and with federal and state employment data collection agencies, which will require the implementation of appropriate privacy protections. Doing so would assist employers' self-assessments. Federal interagency collaborations, such as the Federal Committee on Statistical Methodology's Data Access and Confidentiality Committee, are resources that EEOC might consider. Similarly, improvements in data quality and appropriate access may be available by partnering with the U.S. Census Bureau and the Bureau of Labor Statistics. EEOC also should provide employers with benchmark comparison data to industry and locality peers.

#### **PANEL ON THE EVALUATE THE QUALITY OF COMPENSATION DATA COLLECTED FROM U.S. EMPLOYERS BY THE U.S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION THROUGH THE EEO-1 FORM**

**William M. Rodgers III** (Chair), Federal Reserve Bank of St. Louis; **Paul P. Biemer**, RTI International; **Lisa Catanzarite**, UNITE-LA and Associate of Chamber of Commerce Executives; **Siwei Cheng**, New York University; **Rebecca Dixon**, National Employment Law Project; **Lisette Garcia**, The Pennsylvania State University; **Claudia Goldin**, Harvard University; **Judith K. Hellerstein**, University of Maryland; **Elizabeth Hirsh**, University of British Columbia; **M.V. Lee Badgett**, University of Massachusetts; **Kristen M. Olson**, University of Nebraska–Lincoln; **Donald Tomaskovic-Devey**, University of Massachusetts; **Valerie Rawlston Wilson**, Economic Policy Institute

**STUDY STAFF** **Jennifer Park**, Study Director; **Bradford Chaney**, Senior Program Officer; **Eric Grimes**, Senior Program Assistant

#### **FOR MORE INFORMATION**

The Report Highlights was prepared based on the Consensus Study Report *Evaluation of Compensation Data Collected Through the EEO-1 Form* (2022). Copies of the Consensus Study Report are available from the National Academies Press at [www.nap.edu](http://www.nap.edu). Recordings of the workshops organized by the committee are available on the National Academies website at <https://www.nationalacademies.org/our-work/panel-to-evaluate-the-quality-of-compensation-data-collected-from-us-employers-by-the-equal-employment-opportunity-commission-through-the-eeo-1-form>.

The study was sponsored by the U.S. Equal Employment Opportunity Commission. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

#### **Division of Behavioral and Social Sciences and Education**

**NATIONAL ACADEMIES** Sciences  
Engineering  
Medicine

Copyright 2022 by the National Academy of Sciences. All rights reserved.