

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

## Overall Quality of the New EEO-1 Pay Data

Differences in pay based on sex and race/ethnicity have been illegal for almost 60 years, with the U.S. Equal Employment Opportunity Commission (EEOC) having statutory authority to enforce pay equity. To achieve its mission, EEOC investigates charges of discrimination and seeks to resolve them through such methods as conciliation and mediation, as well as litigation. A starting point of EEOC investigations into discrimination is the suite of data known as EEO.<sup>1</sup> Particular EEO data collections differ by employer type and filing requirements. The EEO-1 collects information from private employers with 100 or more employees (and federal contractors meeting certain criteria) regarding establishment and employee characteristics.

Although review of the EEO-1 findings is followed by collecting more detailed data from the establishments being investigated, EEO data have an important and unique role. They are the only federal data source that gives EEOC a look at the degree to which there are disparities in work and demographic characteristics within an individual establishment. Yet, until recently, these data did not include pay or hours worked data for private-sector employees, which limited EEOC's ability to investigate pay disparities. To address this, EEOC expanded its EEO-1 data collection (called Component 1 data) to include measures of

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

pay and hours worked (called Component 2 data) in 2019 for reporting years 2017 and 2018.

In 2020, EEOC asked the National Academies of Sciences, Engineering, and Medicine to examine the quality of Component 2 data and provide recommendations for future data collection efforts. This brief describes the panel's findings and recommendations regarding the overall quality of the EEO-1 Component 2 pay data as collected to inform investigations of pay differences among protected groups.<sup>2</sup>

### INITIAL INVESTIGATIONS

Under law, EEOC initial investigations must be completed by staff analysts. After an initial assessment, EEOC prioritizes cases for further analysis and discussion with employers. In the context of the EEO-1 pay data collection, staff analysts would compare pay differences by sex, race/ethnicity, occupation, and hours worked across employers:

- Pay data are taken from Box 1 on W-2 forms, categorized in 12 pay bands: (1) \$19,239 and under; (2) \$19,240–\$24,439; (3) \$24,440–\$30,679; (4) \$30,680–\$38,999; (5) \$39,000–\$49,919; (6) \$49,920–\$62,919; (7) \$62,920–\$80,079; (8) \$80,080–\$101,919; (9) \$101,920–\$128,959;

<sup>2</sup> The panel's report also covers the data's utility for employers' self-assessments, for national comparisons, and for protecting vulnerable groups.

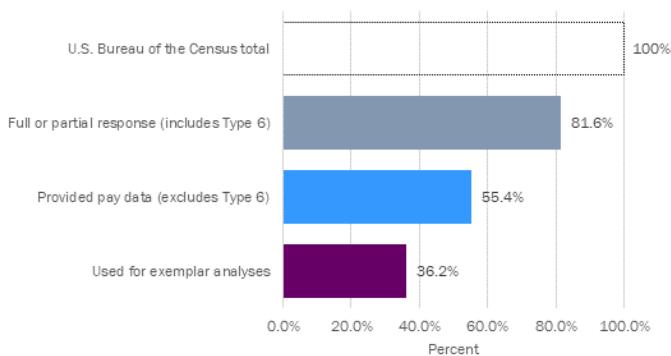
(10) \$128,960–\$163,799; (11) \$163,800–\$207,999; and (12) \$208,000 and over.

- Sex is reported as male or female; non-binary can be reported in the remarks section.
- Race/ethnicity can be reported as Hispanic or Latino, White, Black or African American, Native Hawaiian or Other Pacific Islander, Asian, American Indian or Alaska Native, and two or more races. The measures do not allow distinguishing Hispanic persons by race.
- Occupation covers 10 categories: executive/senior level officials and manager; first- or mid-level official and manager; professional; technician; sales worker; administrative support worker; craft worker; operative; laborer and helper; and service worker.

Employers can be organized as single-establishment firms or multiple-establishment firms. For initial investigations using EEO-1 data, the staff analysts focus their analysis at the establishment level.

## KEY FINDINGS

The panel found that after cleaning, the EEO-1 data could be useful for comparing raw pay gaps within industry and across geography. However, as shown in Figure 1, there are significant limitations.



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

NOTES: “Full or partial response” excludes firms that reported more than 1.4 million employees. “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.

SOURCE: Data from panel generated from Component 2 employer, establishment, and employee files for 2018; [https://www2.census.gov/programs-surveys/bds/tables/time-series/bds2019\\_fz.csv](https://www2.census.gov/programs-surveys/bds/tables/time-series/bds2019_fz.csv).

## Incomplete Data

The panel found EEOC received data for 82 percent of eligible establishments (see Figure 1). Filers had the option of providing only broad summary data (no pay data) for establishments with fewer than 50 employees. Overall, the survey data contained pay data for 55 percent of eligible establishments. After excluding values that appeared unreliable or could not be verified, however, only about 36 percent of establishments eligible for the EEO-1 pay data collection had data that the panel analyzed to examine pay differences. The limitation is especially an issue for small establishments: it affects the analysis of individual establishments under review, as well as comparisons with peer establishments.

## Reliability

The panel found what appeared to be some large errors in some of the employee counts and hours worked. Without extensive cleaning, employee counts and hours worked data could contribute to misleading results when describing employer size and pay differences. Reliability concerns were more difficult to address in the hours worked measure, in part because there were no Component 1 data for checking the Component 2 responses.

## Measurement

The panel found measures of pay, work status, and occupation to be incomplete and imprecise. Specifically, Box 1 of IRS form W-2 does not reflect total compensation. For both the hours worked and pay data, the absence of data on part-time, full-time, and part-year status can produce misleading results. Occupation categories are overly inclusive and do not reflect the modern workforce. These limitations mean only very simple pay analyses can be conducted.

## KEY RECOMMENDATIONS

The panel’s report offers several recommendations to improve the overall quality of EEO-1 pay data; several of these are listed below.

EEOC should improve the coverage of its master list, perhaps using an interagency agreement between EEOC and Bureau of Labor Statistics to appropriately maintain business registers. [Recommendation 4-1]

EEOC should eliminate Type 6 reports and mandate Type 8 reports for all establishments in multi-establishment firms of 100 or more employees. Consolidated reports (Type 2) could then be eliminated, and firm-level data created by summing across establishment reports. These actions would increase coverage, simplify reporting, and reduce respondent burden. [Recommendation 2-2]

Before 2017–2018 Component 2 data are used to assist initial investigations of charges, for employer self-assessment, or for research on pay differences more generally, the data should be carefully reviewed and cleaned. Filtering on employee counts and on hours worked would be beneficial, but some issues would be best addressed by modifying the basic data collection methodology. [Recommendation 5-1]

Before future collection of Component 2 data, EEOC should conduct a field test to investigate issues of burden, data availability, and questionnaire design. The field test should examine the sources of errors in the hours worked and employee count data, and should assess the functioning of new survey questions.

**PANEL ON THE EVALUATE THE QUALITY OF COMPENSATION DATA COLLECTED FROM U.S. EMPLOYERS BY THE U.S. EQUAL 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 Issue Brief 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.

Solutions to be tested may include placing employee-count and hours worked data side by side, as in the data-upload mode. Cognitive interviews may inform EEOC of employers' interpretations of survey questions, difficulties faced in answering, and strategies used to obtain the reported data. [Recommendation 5-2]

EEOC should collect W-2 Box 5 data to measure total compensation, instead of W-2 Box 1 data. [Recommendation 3-3]

EEOC should adopt the Standard Occupational Classification system for classifying occupations to provide greater precision for comparisons of similarly situated employees. To limit respondent burden, EEOC should explore established, improved data systems for occupational coding of individual-level job titles, such as those used by the Bureau of Labor Statistics' Occupational Employment Wages Statistics collection. [Recommendation 3-8]

EEOC should distinguish between Fair Labor Standards Act-exempt and non-exempt workers, and between part-time and full-time workers. A measure for the number of weeks worked should be included to account for part-year employment. EEOC should only collect hours worked for non-exempt employees. [Recommendation 3-9]

Division of Behavioral and Social Sciences and Education

NATIONAL  
ACADEMIES  
Sciences  
Engineering  
Medicine

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