## DATA USERS' NOTE: WRAAK PUBLIC USE DATA FILE

#### Introduction

This document summarizes the revisions to the data set and accompanying documentation from the Worker's Rights- Access, Assertion, and Knowledge (WRAAK) survey.

## **Deleted variables**

Several variables were deleted from the dataset either to reduce disclosure risk or to eliminate redundancy and increase usability of the data.

# **Aggregated variables**

Many variables were aggregated to reduce disclosure risk due to small cell counts. The following aggregations were performed:

- The hourly wage and industry variables were each aggregated to reduce disclosure risk, producing the hourly\_wage\_recoded and D4\_RECODED variables. Hourly wage is recoded in large intervals with more detail at lower wage levels. The industry variable is recoded into 13 major industry groups using the two-digit industry code classifications.
- The values "Don't know," "Refused," and "Does not apply" were aggregated into one value for all variables. To keep variables consistent within the dataset, these values were also aggregated for variables that did not exhibit small cell counts for these observations.
- Thirty other variables had values with small cell counts. For each variable, the value in question was combined with other, similar values until the resultant aggregated value had at least three observations.

#### Added variables

- A binary variable was added (low\_hourly\_wage) that indicates whether hourly wage was reported to be less than \$1.00. This information provides documentation within the dataset of the recoding of these observations to missing before aggregating the wage variable.
- A variable was added that indicates the stratum for each observation. This information allows data users to correctly calculate measures of error for estimates produced by the survey using standard statistical software.

All of these revisions to the variables and values were incorporated into the codebook.

<sup>&</sup>lt;sup>1</sup> Small cell count is defined as values within variables that have fewer than three observations.