# 2020 Federal Employee Viewpoint Survey

# Public Release Data File, April 2020 Update

**Compressed File Contents:**

* FEVS\_2020\_PRDF.csv – Comma-delimited (CSV) Federal Employee Viewpoint Survey (FEVS) data file.
* 2020 PRDF Codebook.xlsx – Excel file containing information about the data.
* Read Me.docx – Information about the 2020 data file

## What’s New in 2020?

There were a great number of changes to the 2020 OPM FEVS. The content of the survey received a drastic overhaul, while the context of the survey changed in when it was administered and under unique conditions.

First, the survey itself was dramatically shorted to just 38 core items, down from 71. These are items that have appeared on the survey for some time and were felt to be essential for survey. A host of new items were added to address questions surrounding the COVID-19 pandemic and its effect on the federal workforce. Additionally, some items that have appeared in the past were modified to address the context of the pandemic. Please refer to the codebook for specific information regarding the items and their response choices.

Second, the survey was administered later in the year than usual – mid-September into October as opposed to the May-June administration of years past. The survey was delayed four months in order to make the adjustments necessary to address the unique conditions surrounding the COVID-19 pandemic. This context is important to remember when examining any results in the survey.

As for changes to our Public Release Data File practices, the most substantial changes are (1) more demographic information is included, and (2) we are only providing identifiers for agencies that have at least 750 respondents as opposed to previous years when we provided results one level down. A few of the COVID-19 related items have also been collapsed from their original breakouts due to the sensitive nature of their response choices. (There was a potential for disclosure risk related to very small groups being identifiable.)

## Data File Summary

The public release data file (PRDF) is broken down into five categories:

1. Core Survey Items: The “core” survey items which have appeared in previous surveys, variables Q1-Q38. In addition, items Q58-Q64 are similar to previous years but have been altered in the context of COVID-19.
2. COVID-19 Items: Variables starting with a “V” address issues related to the COVID-19 pandemic. It is important to remember the context of the survey’s administration in mid-September through October 2020 when interpreting the results, especially for items that refer to time (i.e., “before the pandemic).
3. Demographics (all variables starting with “D”): This year’s public use file includes nine demographics. Details of each can be found in the codebook.
   1. DRNO – Race/National Origin
   2. DHISP – Ethnicity
   3. DDIS – Disability Status
   4. DAGEGRP – Age Group
   5. DSUPER – Supervisory Status
   6. DFEDTEN – Federal Tenure
   7. DSEX – Sex
   8. DMIL – Veteran Status
   9. DLEAVING\* – Intent to Leave, with modifications related to COVID
4. Work Unit Identifier: Used to identify where an employee works. For 2020, this is limited to only agency level, and only for agencies with a minimum respondent count of 750 or more.
5. Other: Includes the statistical weighting variable (POSTWT) and a randomly generated ID number (RandomID).

*\*For missing demographic data, a dummy value “X” is used.*

## Data Masking Methodology for Disclosure Avoidance

Starting in 2016, the FEVS PRDF uses a new method to identify at-risk individuals and an optimized masking process to greatly reduce the risk of re-identification and disclosure of confidential survey responses while maximizing the amount of demographic data that can be kept intact.

When it comes to re-identifying individuals, there are two key pieces of information: where they work and their demographic profile.

The first task is to limit identifiable work units. Only agency and one level below the agency were included, and only for components with a minimum respondent count of 300 were considered. Testing showed this number was an acceptable medium between being able to report more work units while keeping most of the demographic data intact.

The second task in the disclosure avoidance process is to limit the demographic information by reducing the number of demographic variables included in the file and collapsing response choices of those that remain. The fewer distinctions in the demographic information, the less masking that must be performed to hide small groups that are particularly at risk for disclosure. This is accomplished by collapsing response choices together in a logical way, such as combining the original supervisory status categories into a more simplified Non-supervisor/Supervisor-type response.

The third task is to identify people who are at-risk of being identified. Individuals are stratified into groups by combining their demographic responses together into a string of characters\*. Example:

**Example Demographic Profile**

|  |  |
| --- | --- |
| SEX | (**B**) Female |
| EDUCATION | (**B**) Bachelor’s Degree |
| MINORITY | (**X**) Missing |
| SUPERVISOR | (**B**) Supervisor/Manager/Executive |

Combined: **B B X B**

Everyone in the same work unit who has a profile of **BBXB** would be part of what is called a “cell” that identifies them as having a unique combination identifying characteristics. The FEVS uses a Rule of Ten to protect respondent confidentiality – at least 10 responses are required to produce a report for any work unit. This same rule is applied to the public release data file – any cell with fewer than 10 respondents is considered at risk of disclosure.

The fourth task involves masking the demographic data in an attempt to roll the at-risk cells into larger cells that aren’t at-risk. This is accomplished by systematically setting demographic values (such A or B) to missing (using the dummy value “X”). A demonstration of this masking/substitution procedure is provided on page 4.

*\*For missing demographic data, a dummy value “X” is used.*

## Masking Procedure Demonstration

In the first pass three at-risk cells are identified (marked in red). Four possible substitutions are presented by replacing one of the demographic values in sequence. For the first at-risk cell (AAAA), changing the fourth “A” value to the “X” value matches the sequence of the AAAX cell which is not at-risk. Everyone in cell AAAA will be reassigned to cell AAAX at the end of this pass through the data. For the at-risk cells ABAB and BABA, a single substitution will not move either into a not-at-risk cell, so not treatment is applied.

Pass 1 (Single Substitution)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CELL | COUNT | x--- | -x-- | --x- | ---x | SOLUTION |
| AAAA | 3 | XAAA | AXAA | AAXA | AAAX | **AAAX** |
| AAAX | 13 | -- | -- | -- | -- | -- |
| ABAB | 6 | XBAB | AXAB | ABXB | ABAX | -- |
| AXXB | 24 | -- | -- | -- | -- | -- |
| BABA | 3 | XABA | BXBA | BAXA | BABX | -- |

In the second pass two substitutions are performed simultaneously. Changing the two middle values of at-risk cell ABAB will allow them to be merged with the cell AXXB which is not at risk. Also note that cell AAAX’s count went from 13 to 16 because the 3 people who formerly had AAAA were combined with the 16 that have AAAX in the first pass.

Pass 2 (Double Substitution)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CELL | COUNT | xx-- | x-x- | x--x | -xx- | -x-x | --xx | SOLUTION |
| AAAX | 16 | -- | -- | -- | -- | -- | -- | -- |
| ABAB | 6 | XXAB | XBXB | XBAX | AXXB | AXAX | ABXX | **AXXB** |
| AXXB | 24 | -- | -- | -- | -- | -- | -- | -- |
| BABA | 3 | XXBA | XAXA | XABX | BXXA | BXBX | BAXX | -- |

The third pass performs three substitutions. This does help move BABA into a not-at-risk cell. No treatment is applied.

Pass 3 (Triple Substitution)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CELL | COUNT | -xxx | x-xx | xx-x | xxx- | SOLUTION |
| AAAX | 16 | -- | -- | -- | -- | -- |
| AXXB | 30 | -- | -- | -- | -- | -- |
| BABA | 3 | BXXX | XAXX | XXBX | XXXA | -- |

In the fourth and final pass, because the at-risk cell BABA hasn’t moved into a not-at-risk cell, the only solution is to remove all the demographic information of those 3 respondents. The combination of no demographic data and a work unit of at least 300 respondents greatly reduce their risk of being disclosed.

Pass 4 (Full Substitution)

|  |  |  |
| --- | --- | --- |
| CELL | COUNT | END SOLUTION |
| AAAX | 16 | AAAX |
| AXXB | 30 | AXXB |
| BABA | 3 | **XXXX** |