

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde, and Dawn Phelps  
**Date:** December 4, 2019  
**Subject:** Disability Analysis File (DAF) processing error and implications for historical data

---

The purpose of this memo is to document an error that we discovered during processing the 2017 version of the Disability Analysis File (DAF17). This error was present in all earlier versions of the DAF and its predecessor, the Ticket Research File (TRF). As a result of the error, measures of earnings and benefits due for SSI beneficiaries were too low, with the effect being larger for historical data than for more recent data. Variables derived from earnings, including measures of suspense or termination for work (STW) and benefits forgone due to work (BFW) were also affected, also with larger effects on historical data.

In what follows, we outline the nature of the error, the variables it affected, and how those variables differ under the old (incorrect) and new (correct) processing.

## Description of the error

The coding error affected records in the DAF that were derived from the Supplemental Security Income (SSI) Longitudinal file (SSI-LF), during the process of converting multiple SSI records into the one-record-per-beneficiary structure of the DAF. Each record in the SSI-LF contains monthly data on benefits due, benefits paid, and earnings information (used to calculate SSI benefits due) covering a certain number of months. While each record can contain data for many months, in certain instances, SSI beneficiaries may have more than one record. A beneficiary can have multiple records for several reasons. For example, a new record is often established if it is a new period of SSI eligibility, if a child SSI recipient becomes an adult SSI recipient, if there are changes in the individual's status that affect benefits, if the record becomes too "complicated" over time due to recording of earnings or benefits data, or if the record simply runs out of space after many months of benefit receipt. Because of the various reasons for which a new SSI record can be established, more than half of SSI beneficiaries in the DAF have more than one SSI record.

Each SSI record has an "establishment date" that indicates when it was started. Though one might expect that multiple SSI records for the same person would not contain overlapping data for the same calendar months, this is not always the case in practice. In fact, it is common for a record to have data populated in months that precede the record establishment date, and thus, overlap with monthly data from an earlier record. Moreover, data for the same month from two different records do not always align. In some cases, the later-established record may have populated data from months before the establishment date with zeros, rather than leaving the

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 2

Mathematica

values missing or populating the record with the same information as the earlier record. To address this, we designed the intended DAF construction logic to allow us to create a single DAF record from a combination of multiple SSI-LF records for cases with multiple SSI records with inconsistent values for a given month is as follows:

- The monthly value from the latest established record will overlay the value from the earlier-established record, provided that the later-established value is not zero or missing.
- If the value on the later-established record is zero or missing then the value from the earlier-established record will be used.

The processing error represented a departure from the intended logic. Specifically, in some instances, the erroneous processing allowed zero values after the first record to overwrite non-zero values in the first established record. Affected records include those for which the first established record had a populated monthly value but *all subsequent records* had a zero amount for that same month. In those cases, the zero value incorrectly replaced a valid, populated value from the first established record. In practice, this meant that older positive benefits due and earnings values were set to zero based on newer records when they should not have been.

To demonstrate this, we show a hypothetical example in Table 1. In this example the values for January, February, and April of 2005 (0501, 0502, and 0504) were incorrect using the old DAF processing because these records had only zero values for these months in all records after the first record. Though March 2005 (0503) had a 0 in the third record, it was correctly processed because there was a populated value of \$25 on the second record.

**Table 1. Example of combining SSI-LF records in the old (incorrect) and new (corrected) DAF processing algorithm**

	Record establishment date	EICM 0501	EICM 0502	EICM 0503	EICM 0504	EICM 0505	EICM0 506
<b>Raw data from the SSI-LF</b>							
Record 1	January 2002	150	50	200	100	.	.
Record 2	April 2005	0	0	25	0	0	0
Record 3	June 2005	0	0	0	0	50	100
<b>Old DAF processing</b>	--	0	0	25	0	50	100
<b>New DAF processing</b>	--	150	50	25	100	50	100

### Variables affected by the error

The error affected variables related to SSI benefits due and earnings; it did not affect any character variables. It also did not affect benefits paid variables, as those data were correctly processed by summing together information from all established SSI records. It also did not

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 3

Mathematica

affect any variables specific to the SSDI program. The variables that were directly affected by the error are:

- Countable earned income, *EICMyymm*
- Unearned income, *UINCyymm*
- Federal benefit amount due, *FAMTyymm*
- State benefit amount due, *SAMTyymm*
- Total amount due (the sum of FAMT and SAMT), *DUESyymm*

Several additional variables were indirectly affected by the error because they were derived from the variables listed above. These are:

- An indicator for 1619a status, *PROAyymm*
- An indicator for 1619b status, *PROByymm*
- An indicator for concurrently being due an SSI and Social Security Disability Insurance (SSDI) benefit, *CONCyymm*
- Variables indicating suspense or termination of cash benefits for work; *STWSSIyymm* for SSI alone, and *STWCMyymm* that considers both SSI and SSDI
- Variables indicating the dollar value of cash benefits suspended for work; *BFWSSI\_DRAFTyymm* for SSI alone, and *BFWCMyymm* for combined SSI and SSDI

## Implications of the error for snapshot statistics

There are two general implications of the error on benefits due and earnings. First, because the error caused a subset of earnings (or benefits due) records with positive values to be replaced with zeros, correcting the error resulted in the incorrect observations being changed from zeros to positive values. Therefore, aggregate measures of earnings (or benefits due) derived from the affected variables are expected to increase as a result of the correction. Similarly, measures of the number of beneficiaries with non-zero earnings (or benefits due) based on the affected variables are expected to increase. However, average values calculated over the set of non-zero earnings (or benefits due) values could increase, decrease, or stay the same. The direction of the change depends on the distribution of the incorrectly omitted non-zero observations compared to the distribution of the remaining observations. For example, if the correct values of the observations that were incorrectly set to zero tend to be higher than the unaffected observations, after the correction, the average observed non-zero value for an affected variable will increase.

The second implication is that the effect on current year values is expected to be small or non-existent, because the probability of multiple records in a single year is small, but the magnitude of the effect of the error will increase the farther back in time (relative to the current DAF year) one observes. This is because only beneficiaries with multiple SSI record establishments could

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 4

Mathematica

have overwritten values, and the likelihood of that is greater the farther back in time one moves from the DAF year. For example, in the DAF17, the error would affect a larger share of observations in 2010 than in 2017. Furthermore, we would expect that the error in the 2010 observations in the DAF17 would be of a larger magnitude than the error in the same observations in the DAF13.

## Magnitude of the effect of the error

### 1. Directly-affected variables

Figures 1, 2, and 3 show summary measures of the magnitude of the effect of error at three points in time for the directly-affected variables. The three points in time, January of 1995,<sup>1</sup> 2005, and 2015; are arbitrary but provide a snapshot at evenly-spaced intervals over the time period covered by the DAF17. The appendix to this memo contains graphs of these measures that span all months in DAF17 (January 1994 through December 2017). In the remainder of the memo, we discuss variables using only their prefix, omitting the *yymm* suffix of the variable name; we instead reference the calendar month of interest.

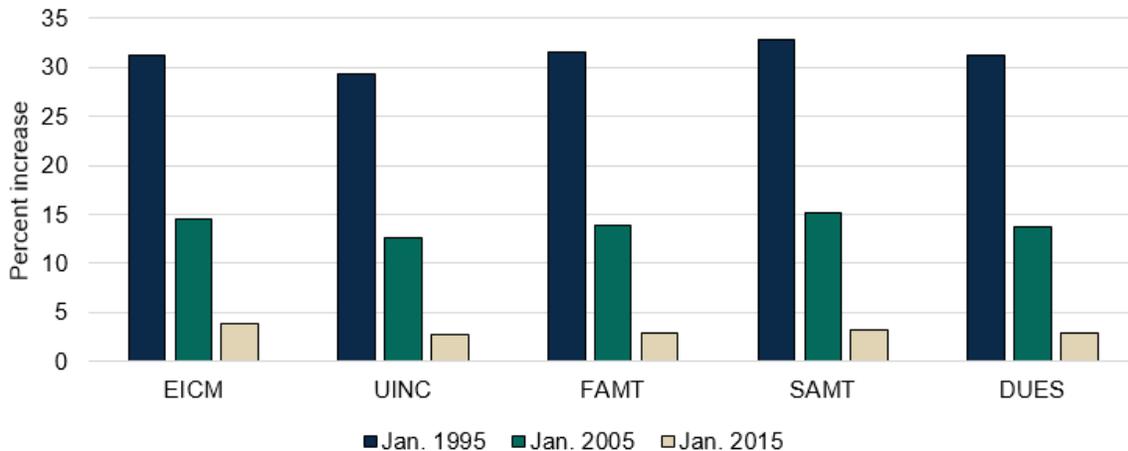
As discussed above, correcting the processing error resulted in an increase in populated positive values in almost all months, with the change getting larger for months farther in the past. Figure 1 shows the percentage increase in the number of beneficiaries with a positive value as a result of the correction.<sup>2</sup> For all of the directly-affected variables, the correction resulted in an increase in positive values of approximately 30 percent for observations in January 1995. Positive observations in January 2005 increased by about 15 percent, and in January 2015, the increase was about 3 percent.

---

<sup>1</sup> DAF17 contains data on beneficiaries who participated in SSI in one or more months starting in March 1996 through December 2017. The earliest observations for these beneficiaries go back to January of 1994. Therefore, observations from January 1995 do not include beneficiaries who had stopped participating before March 1996. In contrast, the January 2005 and 2015 observations include all SSI participants in those months. We confirmed that the patterns we show for 1995 look similar to those in 1996 and 1997, so differences across time are not related to DAF selection criteria.

<sup>2</sup> In what follows, we consider positive values versus zero values because we expect that the variables of interest should be positive (earnings, benefits due). In actuality, correcting the processing error also could have replaced zero values with negative values, had those been in the data.

**Figure 1. Percentage increase in the number of beneficiaries with a positive value after correcting the SSI-LF processing error, as a share of beneficiaries with a positive value before correcting the error**



Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

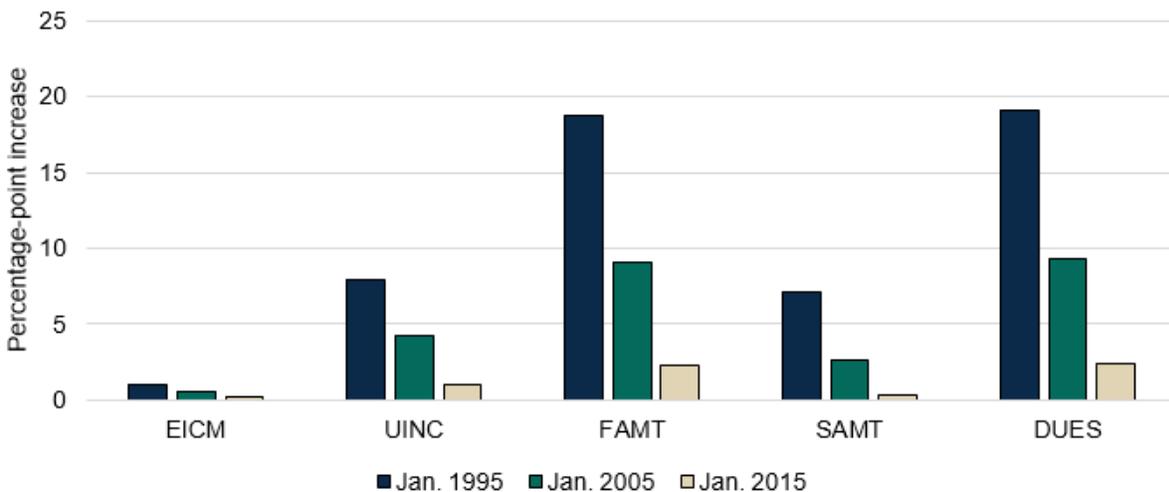
Note: Positive values summed across all beneficiaries in the DAF in the month shown, with the percentage change calculated by comparing the DAF17 corrected value to the share with positive values before reprocessing.

The records that changed as a share of positive records before the correction, shown in Figure 1, may not give a complete picture of the magnitude of the effect on beneficiary-level statistics, as some variables in the administrative records apply to virtually all beneficiaries, while others apply to very few. To consider this, we calculated the number of beneficiaries whose value on a given variable changed from a zero to a positive value after reprocessing, as a share of the number of beneficiaries receiving benefits in a given month (Figure 2). We calculated this value before and after reprocessing. We present the difference in those values—the percentage-point change—that results from the correction in the share of beneficiaries with a positive value among those who are also in non-terminated SSI payment status (*PSTA<sub>yymm</sub>*). For convenience we will refer to the set of beneficiaries with non-terminated SSI payment status as “SSI beneficiaries.”

Although the increase in the percentage increase in positive values does not vary greatly across affected variables (Figure 1), the change in the share of beneficiaries varies widely, as shown in Figure 2. This variation is a function of the number of SSI beneficiaries who had a positive value to begin with. For example, relatively few beneficiaries have positive values of EICM, so a 31 percent increase in the number of records with a positive value translates to a one percentage-point increase in the share of SSI beneficiaries with a positive value (from 3 to 4 percent). In contrast, the 32 percent increase in the number of records with a positive value of FAMT in January 1995 represents a 19 percentage-point increase in the share of SSI beneficiaries with a positive value (from 59 to 78 percent). The correction also resulted in a similar increase in SSI beneficiaries with a positive values for DUES in January 1995 (consistent with DUES being the

sum of FAMT and SAMT). In January 2005, the increases for these variables were about nine percent, while in January of 2015 the increase was two percent as a share of all SSI beneficiaries. The percentage-point increases for UINC and SAMT were approximately eight, four and one in January 1995, 2005, and 2015, respectively.

**Figure 2. Percentage-point increase in the share of SSI beneficiaries with positive values after correcting the SSI-LF processing error**



Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

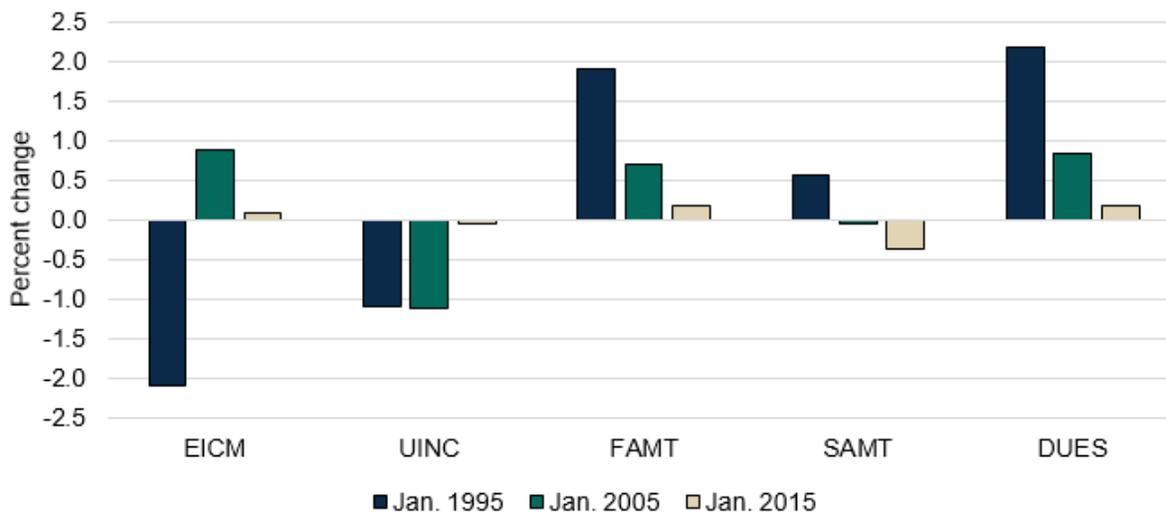
Note: Percentage point change is calculated by subtracting the share of SSI beneficiaries with a positive value before the correction from the share of SSI beneficiaries with a positive value after the correction.

Average dollar values did not change much after correcting the processing error. Figure 3 shows the percent change in the average dollar value of positive observations as a result of the correction. The largest change in the average positive value of the directly-affected variables across all months of the DAF17 (January 1994 through December 2017) was less than 3 percent. In January 1995, the average positive value of the income variables (EICM and UINC) decreased slightly and the average positive value of the benefits due variables (FAMT, SAMT, DUES) increased slightly. These changes indicate that the beneficiaries whose records were affected by the processing error had slightly lower earned and unearned income in January 1995 than beneficiaries whose records were unaffected. Given the inverse relationship between earnings and SSI benefits due, this pattern is consistent with an increased average value of benefits due in the corresponding months.

Figure 3 also shows how average values for the affected variables changed in January 2005 and January 2015; the changes are not necessarily uniform over time. In January of 2005, the average positive value of EICM increased by just under 1 percent after the correction whereas the average positive value of UINC decreased by just over 1 percent. The average positive value of FAMT and DUES increased by about 0.7 percent while the average positive value of SAMT was

essentially unchanged. The percentage change was less than one-half of a percent for all of the earnings and benefits due variables in January 2015.

**Figure 3. Percentage change in the average dollar value among observations with positive values after correcting the SSI-LF processing error**



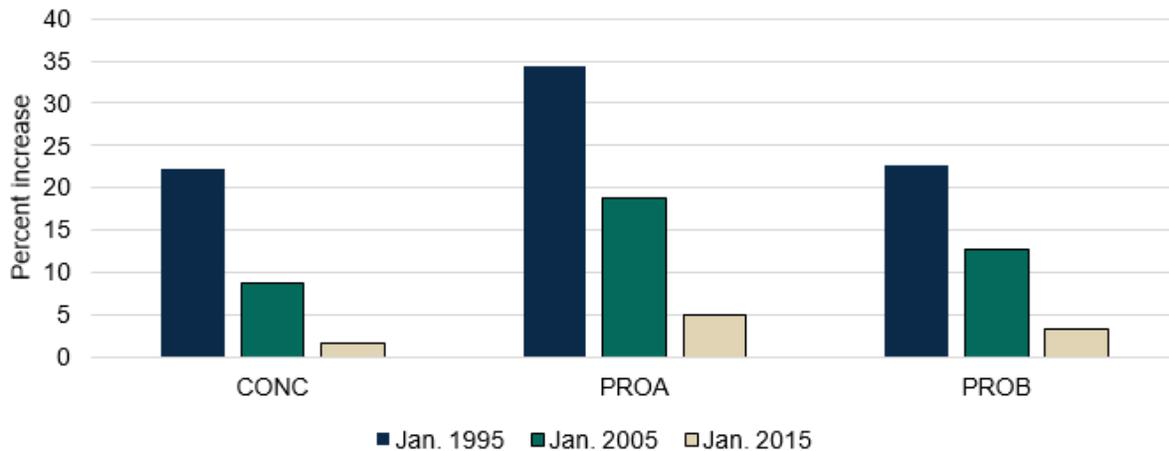
Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

## 2. Indirectly-affected variables

We use a similar set of figures as those shown in the previous section to describe the effects of the error on CONC, PROA and PROB, before moving in a slightly different direction to discuss the effects on STWSSI, STWCM, BFWSSI and BFWCM.

The share of records affected by the error for indirectly affected variables is similar to the share of core variables that were affected directly. Figure 4 shows the percent increase in positive records for the binary variables CONC, PROA, and PROB. In each case, a positive record indicates that the observation has a value of 1 instead of 0. This means that the beneficiary was in the status in a given month—for example, concurrently received SSI and SSDI payments in the month (CONC), was in 1619a status (PROA), or was in 1619b status (PROB). The percentage increase in each time period for the variables shown is of a similar magnitude as the increases shown for the directly-affected variables in Figure 1 from which they were constructed. In January 1995 the increase was 22 percent, 34 percent, and 23 percent for CONC, PROA, and PROB, respectively. In January 2005 the increase was about half of the magnitude in the period ten years prior. By January 2015, the magnitude of the increase was no greater than 5 percent.

**Figure 4. Percentage increase in the number of beneficiaries with value equal to 1 after the SSI-LF processing error was corrected in DAF17, as a share of beneficiaries with a positive value before the error was corrected**

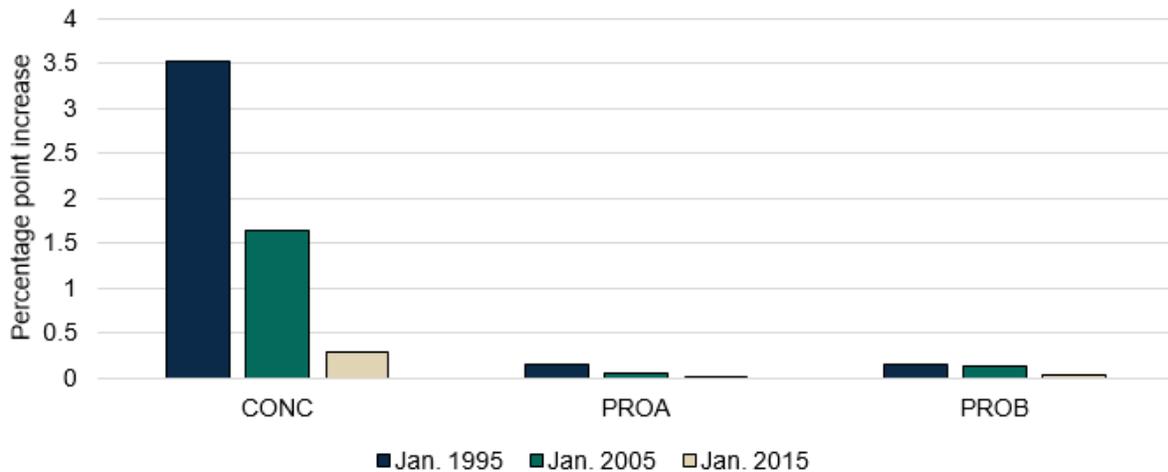


Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

Note: Positive values summed across all beneficiaries in the DAF in the month shown, with the percentage change calculated by comparing the DAF17 corrected value to the share with positive values before reprocessing.

Figure 5 shows the percentage-point change that results from the correction in the share of beneficiaries with a value equal to 1 among those who are also in non-terminated SSI payment status (*PSTA<sub>yymm</sub>*). The share of SSI beneficiaries flagged as being concurrently due SSI and SSDI benefits (CONC equal to 1) increased by 3.5 percentage points in January 1995 as a result of the correction. In January 2005, the increase was about 1.5 percentage points and in January 2015, the increase was about one-third of a percentage point. Relatively few beneficiaries have 1619a (PROA equal to 1) or 1619b (PROB equal to 1) status. As a result, even though there was a non-negligible percentage increase in the number of beneficiaries flagged as having either 1619a or 1619b status (Figure 4), the change in the share of SSI beneficiaries flagged as 1619a or 1619b is less than two tenths of a percentage point.

**Figure 5. Percentage-point increase in the share of non-terminated SSI beneficiaries with value equal to 1 after correcting the SSI-LF processing error**



Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

Note: Percentages calculated by dividing the number of SSI beneficiaries with a positive value by the number of beneficiaries in non-terminated SSI payment status (PSTAyymm) in the month shown.

We now describe the effects of reprocessing on variables that measure the effects of beneficiary work activity—STW and BFW. Because SSI benefit amounts are affected by the amount of earned and unearned income that are recorded, the corrected algorithm for processing SSI-LF data resulted in changes to these variables. For users unfamiliar with the STW and BFW concepts and measures, we suggest reviewing Volume 3- Tips for Conducting Analysis with the DAF17 available at: <https://www.ssa.gov/disabilityresearch/daf.html>. In the case of SSI, there are a number of interactions between earned and unearned income that make it difficult to predict the effect of the error on the STW and BFW measures.

In Table 2, we show the distributions in each month under the old and corrected processing methods. Because most SSI beneficiaries do not have earnings that would result in STW, the overall distribution of STWSSI looks quite similar before and after the correction. Yet, comparing the change between the old (incorrect) and new (correct) processing algorithms highlights patterns that echo the earlier findings: the affected values are among those who had earned or unearned income (STWSSI=1,2,3,4) and the magnitude of the change increases the farther back in time one goes. These statistics also show that at each point, correcting the error resulted in an increase in the number of SSI beneficiaries with an STWSSI value indicating suspense or termination of cash benefits due to work (STWSSI=1,2,3), an increase in the share with earnings, but whose unearned income alone led to benefit suspense or termination (STWSSI=4), and a decrease in the share who were in suspense or termination status for a reason not determined to be work (STW=8).

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 10

Mathematica

We present the effects on BFWSSI in Figures 6 and 7, following a similar structure to earlier figures. Because the error affected earnings and SSI benefits are reduced by \$1 for every \$2 of earnings above an income disregard, we would expect to see BFWSSI increase with the corrected algorithm. However, because STWSSI=4 cases (countable unearned income was sufficient to cause suspension) also increased, and because those cases do not accrue BFWSSI, we would not expect the magnitude to be as large as for earnings alone.

In Figure 6, we show the percentage increase in the number of beneficiaries with a positive BFWSSI value after the correction and the percentage increase when limiting to beneficiaries in STWSSI=0,1,2,3 (the only groups who can accrue BFWSSI).<sup>3</sup> As a share of the beneficiaries with positive BFW in a month, correcting the error increased positive BFWSSI by 33 percent in January 1995, 15 percent in January 2005, and 4 percent in January 2015. Relative to the number of SSI beneficiaries in current pay status or with benefits suspended or terminated for work in a given month, however, the magnitude of the change was 1 percent or less, reflecting the fact that relatively few SSI beneficiaries accrue BFWSSI in a given month.

Figure 7 shows that average BFWSSI increased by 0.5 to 3.5 percent in the months shown, without a clear pattern over time. This suggests that the BFWSSI that was erroneously omitted tended to be slightly higher than the BFWSSI for beneficiaries whose values were included.

In the appendix, we present statistics for both STWSSI and BFWSSI, and the combined version of both variables (STWCM and BFWCM). The magnitude of the change for the latter is smaller than for the former, which is expected based on their construction. The combined variables take into account beneficiaries' status in both programs, when relevant. In the case of STW, STWCM errs toward current pay status if a beneficiary is in current pay status in one program and not the other. In the case of BFWCM, BFWSSI and BFWDI are summed together, with the former generally representing much smaller dollar values than the latter, per program rules.

---

<sup>3</sup> Because the STWSSI values and BFWSSI values changed as a result of correcting the error, we used a denominator for this calculation that limited to beneficiaries who had STWSSI=0,1,2,3 under the old processing method. As shown in Table 2, the share with STW=0,1,2,3 increased overall, so had we instead used the corrected STWSSI values as a denominator, this share would have been smaller.

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 11

Mathematica

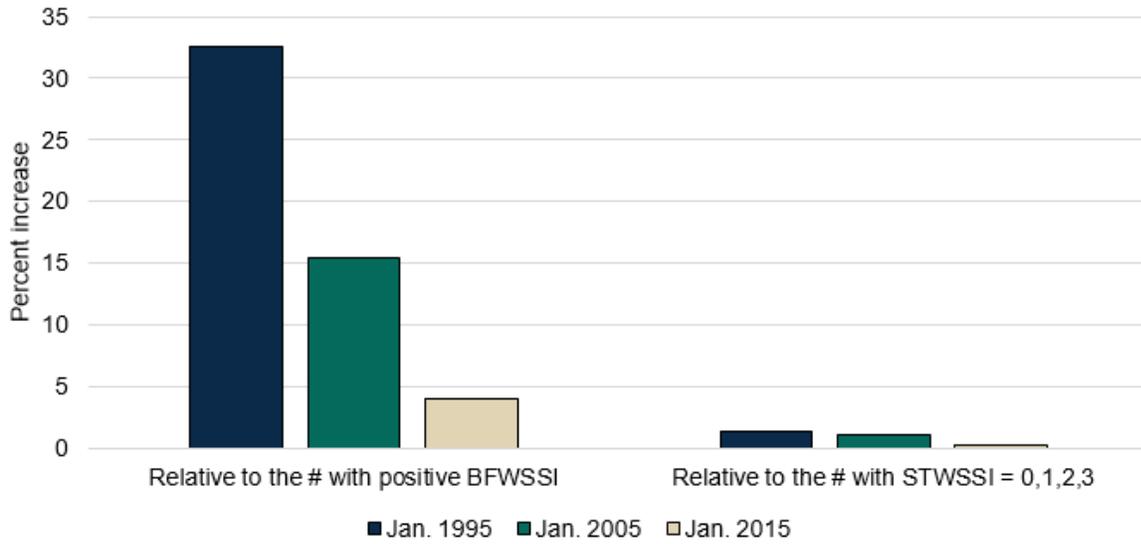
**Table 2. Percentage distribution of STWSSI values before and after correcting the SSI-LF processing error**

	January 1995- Old	January 1995- New	Percent change (New vs. Old).	January 2005- Old	January 2005- New	Percent change (New vs. Old).	January 2015- Old	January 2015- New	Percent change (New vs. Old).
STWSSI=0	74.5	74.5	0%	54.6	54.6	0%	49.9	49.9	0%
STWSSI=1,2,3	0.7	0.9	32%	1.5	1.7	18%	1.3	1.4	4%
STWSSI=4	0.1	0.2	37%	0.3	0.4	8%	0.3	0.3	2%
STWSSI=8	17.1	16.9	-1%	16.4	16.2	-1%	9.8	9.8	0%
STWSSI=9	7.5	7.5	0%	27.2	27.1	0%	38.6	38.6	0%

Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

Note: Percent change is calculated as a share of the pre-correction distribution. The values shown in the table are rounded to the nearest tenth but the percentage change was calculated prior to rounding.

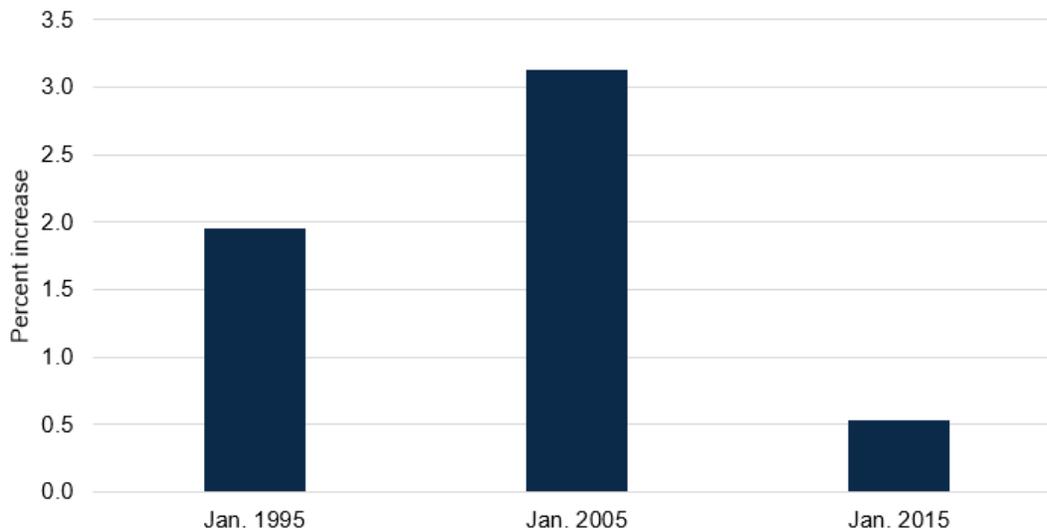
**Figure 6. Percentage increase in the number of beneficiaries with positive BFWSSI after correcting the SSI-LF processing error**



Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

Note: Positive values summed across all beneficiaries in the DAF in the month shown, with the percentage change calculated by comparing the DAF17 corrected value to the share with positive values before reprocessing.

**Figure 7. Percentage change in the average dollar value of BFWSSI among positive observations after correcting the SSI-LF processing error**



Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

## Characteristics of beneficiaries with affected records

Because of the nature of the reprocessing error, beneficiaries with certain characteristics were more likely to have affected records. In this section, we present statistics on the distribution of sex, age, years since SSI initial eligibility, and primary disabling condition for beneficiaries whose FAMT and EICM were affected by the error. Because other affected variables are closely related to FAMT and EICM, we would expect similar patterns for those variables, but did not include comparisons here for ease of presentation. Again, we present statistics from January of 1995, 2005, and 2015.

Table 3 shows the characteristics of beneficiaries whose FAMT records were affected by the processing error compared to those whose records were not affected. The “affected” category consists of beneficiaries for whom the processing error caused a positive value to be set to zero. The “not affected” category consists of beneficiaries with a *positive value* of FAMT both before and after the correction. The largest difference between the two groups is age. On average, beneficiaries whose records were affected are younger than those whose records were not. In particular, a much larger share of beneficiaries whose records were affected are under age 18 (as of January 1995, 2005, or 2015). As noted previously, a new record is often established if a child SSI recipient becomes an adult SSI recipient. Beneficiaries who are under 18 are more likely than older beneficiaries to have been affected by the error because, all else equal, they are more likely to have had a new record established prior to the construction of DAF17. Given the difference in age, it follows that the affected group has fewer years of SSI eligibility on average

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 14

Mathematica

(at the time they are observed in 1995<sup>4</sup>, 2005, or 2015). Other notable differences are that the affected group has a larger share of men (by about 7 percentage points) and a larger share of people with diagnoses in group 1 (autistic disorders; developmental disorders; or childhood and adolescent disorders not elsewhere classified), especially in 2005 and 2015.

Table 4 shows the characteristics of beneficiaries whose EICM records were affected by the processing error compared to those whose records were not affected. As in Table 3, the “affected” category consists of beneficiaries for whom the processing error caused a positive value to be set to zero and the “not affected” category consists of beneficiaries with a positive value both before and after the correction. Because many SSI beneficiaries do not have earned income (and because EICM is an input into FAMT), this table consists of a subset of the beneficiaries described in Table 3. Beneficiaries whose EICM records were affected are younger on average but the difference is smaller than for FAMT and it is primarily in the 18 to 29 age group instead of the under-18 group. This is expected given that earned income is rare for SSI beneficiaries under the age of 18. Beneficiaries whose records were affected have slightly fewer years of SSI eligibility on average and there is little difference in the gender distribution in the two groups. For both groups of beneficiaries, the modal diagnosis group is intellectual disability but the share of beneficiaries with an intellectual disability is larger for beneficiaries whose records were affected by the error than for those whose records were not.

---

<sup>4</sup> The share with missing values of SSIELIG\_FIRST in January 1995 is higher than in January 2005 and 2015, and is higher for the group whose records were not affected than for the group whose records were affected. An investigation across other years (not shown) showed that the pattern changed in 2004. In 2004, there were changes in the source files used for SSI and in the PSTA codes that resulted in selection into the DAF. Additionally, in late 2003, there were changes in SSA’s use of some PSTA codes; together these changes mean that the variable we used to define years since SSI eligibility was not as frequently populated for DAF beneficiaries prior to 2004 as it has been since that time. As a result, the distribution on SSIELIG\_FIRST in January 1995 is not directly comparable to the later years.

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 15

Mathematica

**Table 3. Comparison of beneficiary characteristics for those whose FAMT records were or were not affected by the SSI-LF processing error**

	January 1995		January 2005		January 2015	
	Not affected	Affected	Not affected	Affected	Not affected	Affected
<b>Observations</b>	3,201,186	1,012,129	4,647,120	643,166	6,322,528	185,418
<b>Percent female</b>	52.8	46.0	53.3	45.9	49.2	43.4
<b>Age</b>						
Mean	37.5	27.4	39.0	26.9	38.7	26.4
18-29 (%)	16.6	41.7	17.6	45.6	20.5	49.9
30-39 (%)	15.8	14.0	13.4	14.8	14.6	13.6
40-49 (%)	20.0	15.8	14.0	10.2	11.7	9.6
50-59 (%)	19.0	14.8	22.3	13.0	14.8	8.2
60-FRA (%)	20.6	10.3	22.7	11.2	26.2	11.9
<b>Years since SSI award</b>						
Missing (%)	6.1	0.0	0.0	0.0	0.0	0.0
Mean	7.0	6.0	10.0	7.6	11.7	9.1
<1 (%)	10.6	15.2	9.9	13.1	7.7	8.1
1-4 (%)	37.2	50.0	25.5	35.4	24.7	33.7
5-9 (%)	21.3	14.0	21.1	18.9	21.4	24.8
10-19 (%)	19.0	11.7	30.3	26.1	24.8	16.6
20+ (%)	5.9	9.1	13.2	6.6	21.5	16.8
<b>Disabling condition (%)</b>						
Missing	31.6	26.6	13.4	10.2	13.0	8.0

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 16

Mathematica

	January 1995		January 2005		January 2015	
	Not affected	Affected	Not affected	Affected	Not affected	Affected
1. Autistic disorders; developmental disorders; or childhood and adolescent disorders not elsewhere classified	1.8	5.2	6.1	16.1	12.1	26.4
2. Intellectual disability	17.3	24.6	19.2	20.8	15.6	14.8
3. Mood disorders; organic mental disorders; schizophrenic and other psychotic disorders; or other mental disorders	19.3	18.5	26.9	23.7	26.2	20.0
4. Endocrine, nutritional, and metabolic diseases; circulatory system; digestive system; genitourinary system; nervous system and sense organs; or respiratory system	15.7	11.0	17.6	13.6	15.7	14.0
5. Musculoskeletal system and connective tissue	5.3	2.6	8.0	4.6	9.3	5.6
6. Infectious and parasitic diseases or injuries	3.7	2.6	4.2	3.2	3.9	3.1
7. Congenital anomalies; blood and blood-forming organs; skin and subcutaneous tissue; or other	1.9	3.8	2.9	6.2	3.5	7.4
8. Unknown value	3.5	5.1	1.7	1.5	0.8	0.7

Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 17

Mathematica

**Table 4. Comparison of beneficiary characteristics for those whose EICM records were or were not affected by the SSI-LF processing error**

	January 1995		January 2005		January 2015	
	Not affected	Affected	Not affected	Affected	Not affected	Affected
<b>Observations</b>	175,859	55,017	284,067	41,260	333,376	12,690
<b>Percent female</b>	44.8	45.9	49.7	51.4	49.1	48.7
<b>Age</b>						
Mean	35.3	34.3	38.4	34.9	37.9	34.3
18-29 (%)	2.2	3.9	2.0	3.8	2.1	3.3
30-39 (%)	36.0	36.8	29.0	39.6	35.4	44.9
0-49 (%)	31.4	29.7	24.3	22.1	21.5	24.2
50-59 (%)	17.2	19.3	24.2	20.6	16.0	12.4
60-FRA (%)	10.5	8.4	16.0	10.8	18.8	11.5
<b>Years since SSI award</b>						
Missing (%)	19.0	0.2	0.5	0.5	0.4	0.8
Mean	9.4	8.6	12.2	9.7	13.6	13.1
<1 (%)	6.4	7.2	9.6	9.3	8.1	5.7
1-4 (%)	22.2	35.6	17.6	21.0	18.4	14.3
5-9 (%)	18.5	21.7	17.0	21.1	17.8	20.2
10-19 (%)	26.0	22.9	33.8	39.2	26.3	29.3
20+ (%)	8.0	12.5	21.5	8.9	29.0	29.8
<b>Disabling condition (%)</b>						
Missing	23.4	21.2	7.8	8.4	7.9	7.0

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 18

Mathematica

	January 1995		January 2005		January 2015	
	Not affected	Affected	Not affected	Affected	Not affected	Affected
1. Autistic disorders; developmental disorders; or childhood and adolescent disorders not elsewhere classified	0.5	0.6	2.1	3.0	6.7	8.4
2. Intellectual disability	30.6	36.5	32.6	33.8	28.1	34.1
3. Mood disorders; organic mental disorders; schizophrenic and other psychotic disorders; or other mental disorders	20.2	22.2	27.6	31.2	27.5	27.0
4. Endocrine, nutritional, and metabolic diseases; circulatory system; digestive system; genitourinary system; nervous system and sense organs; or respiratory system	13.9	9.5	16.0	12.3	14.5	11.8
5. Musculoskeletal system and connective tissue	3.6	2.1	6.4	4.5	8.0	4.8
6. Infectious and parasitic diseases or injuries	3.2	2.9	4.1	4.1	3.9	3.7
7. Congenital anomalies; blood and blood-forming organs; skin and subcutaneous tissue; or other	1.2	0.9	1.8	1.5	2.5	2.8
8. Unknown value	3.5	4.1	1.8	1.3	0.8	0.5

Source: DAF17, using the corrected method of processing SSI-LF records compared to the method in place in DAF16 and earlier.

**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 19

Mathematica

## Discussion

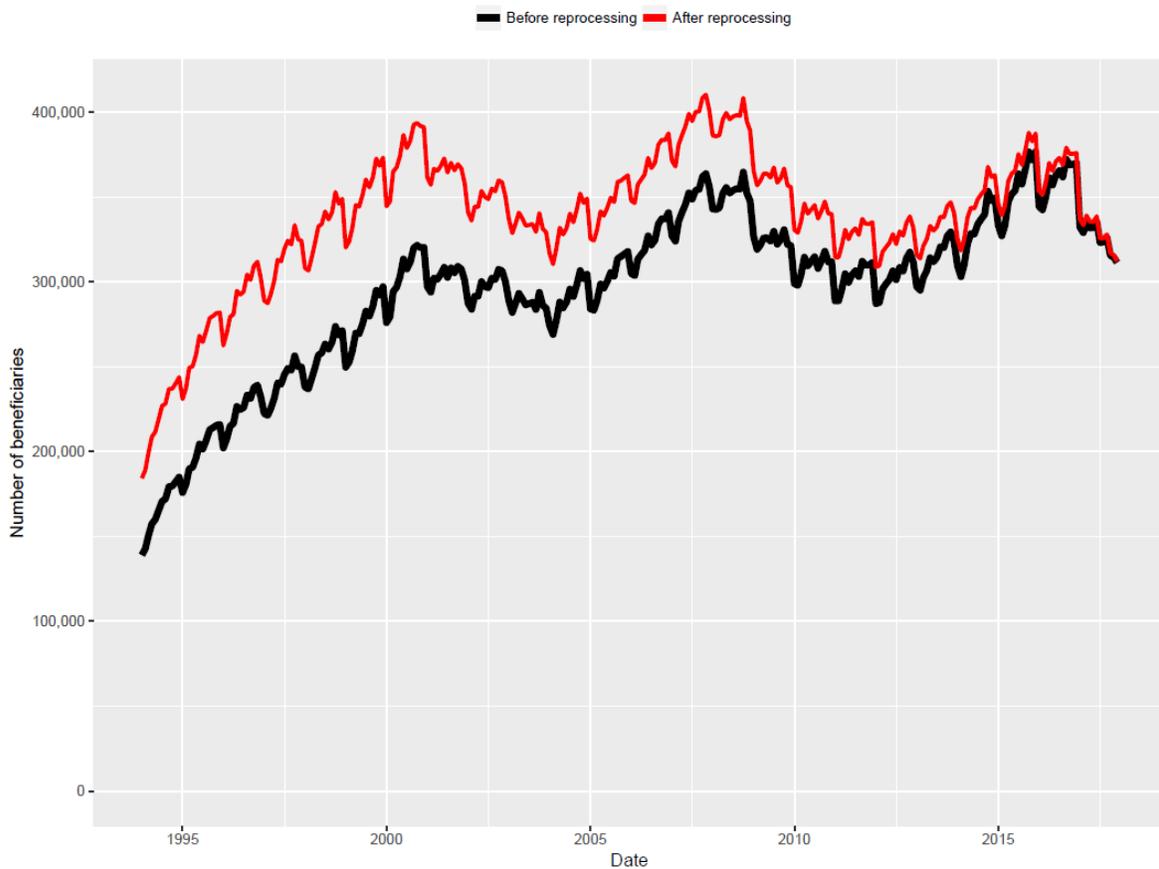
Our analysis of the impact of the processing error suggests that across all of the affected variables, the magnitude of the effect was smaller in the years closer to the current DAF year than in years farther in the past. Our checks were limited to comparing DAF17 under the old (incorrect) and new (corrected) processing algorithm, though we would expect to find similar patterns had we performed similar comparisons in earlier versions of the DAF.

Because earlier versions of the DAF with the incorrectly processed SSI-LF data have been used in research and analysis, we think it is important to get a sense of the implications of the change in a broader way than what we have presented here. We continue to work on that analysis and will provide additional details as they become available. We expect to produce: (1) statistics similar to those in this memo for beneficiaries more likely to have had their records affected by the error or of strong policy interest, including young adults ages 18-29, (2) cross-sectional and cohort statistics for groups defined by age, and (3) statistics that look at cohorts of new awardees over time.

## Appendix: Monthly graphs of measures of change due to the error

This appendix contains graphs showing various measures of the magnitude of the change in the affected variables due to the error in each month contained in DAF17, from January 1994 through December 2017. There are three sets of graphs: 1) Figures A1-A15 for the continuous variables directly affected as a result of the error (EICMyymm, UINCyymm, FAMTyymm, SAMTyymm, and DUESyymm), 2) Figures A16-A24 for the binary variables that were affected (CONCyymm, PROAyymm, PROByymm), and 3) Figures A25-29 for variables related to STW and BFW (STWSSIyymm, STWCMyymm, BFWSSI\_DRAFTyymm, BFWCM\_DRAFTyymm).

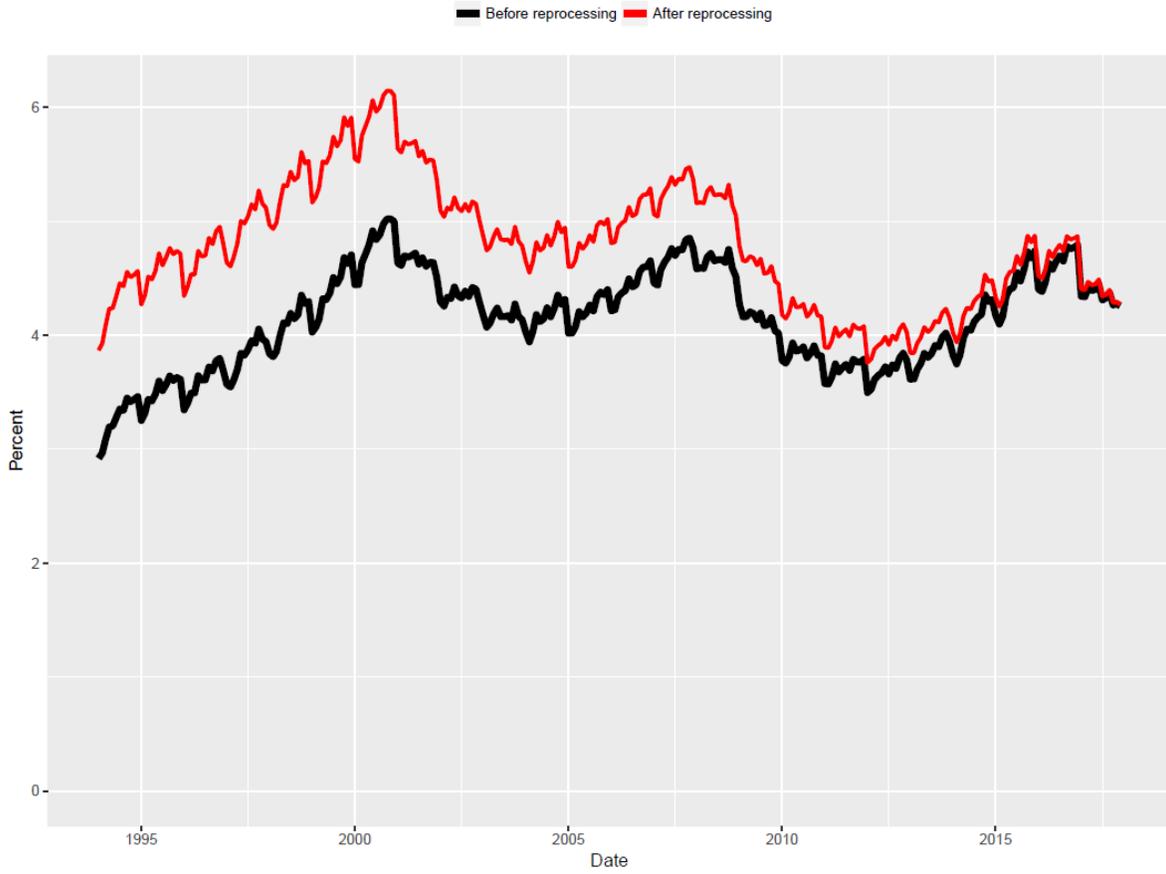
**Figure A1. EICMyymm: Number of beneficiaries with value > 0**



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 21

Mathematica

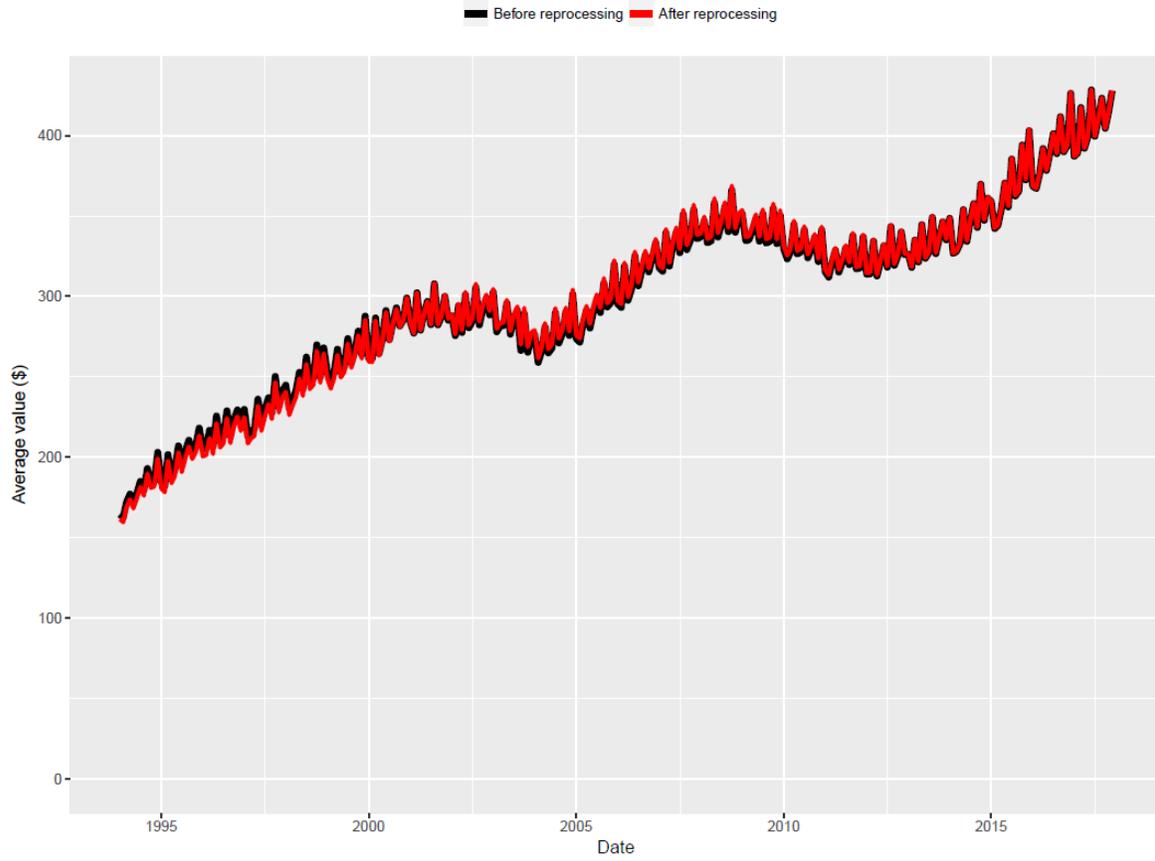
Figure A2. EICMyymm: Percent of SSI beneficiaries with value > 0



**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 22

Mathematica

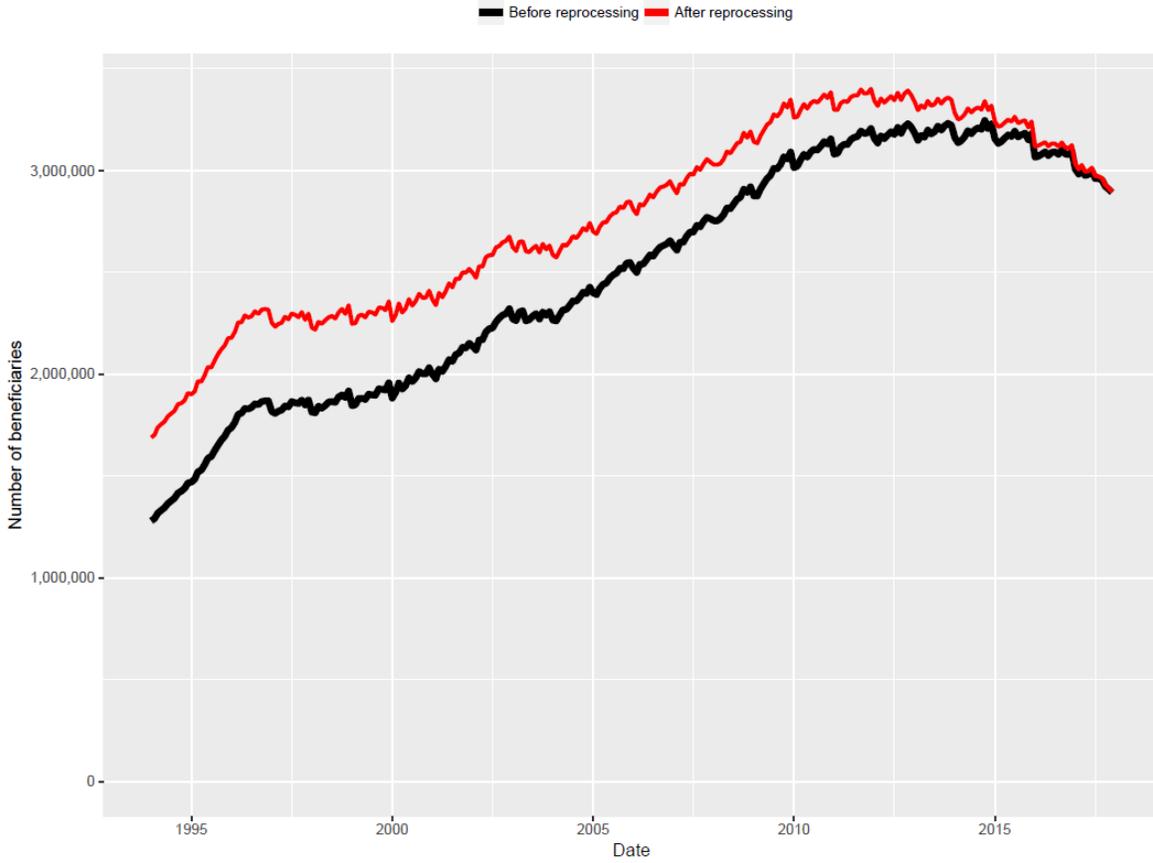
**Figure A3. EICMyymm: Average value among beneficiaries with value > 0**



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 23

Mathematica

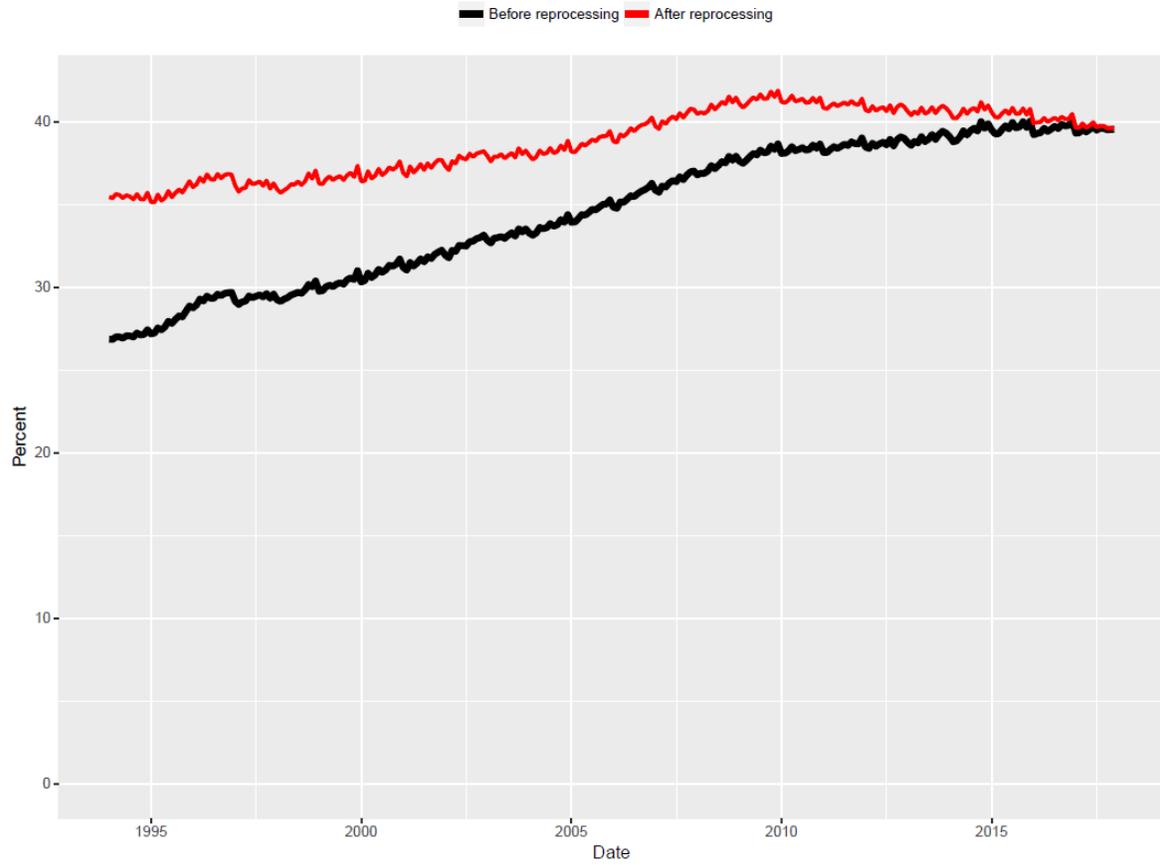
Figure A4. UINCyymm: Number of beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 24

Mathematica

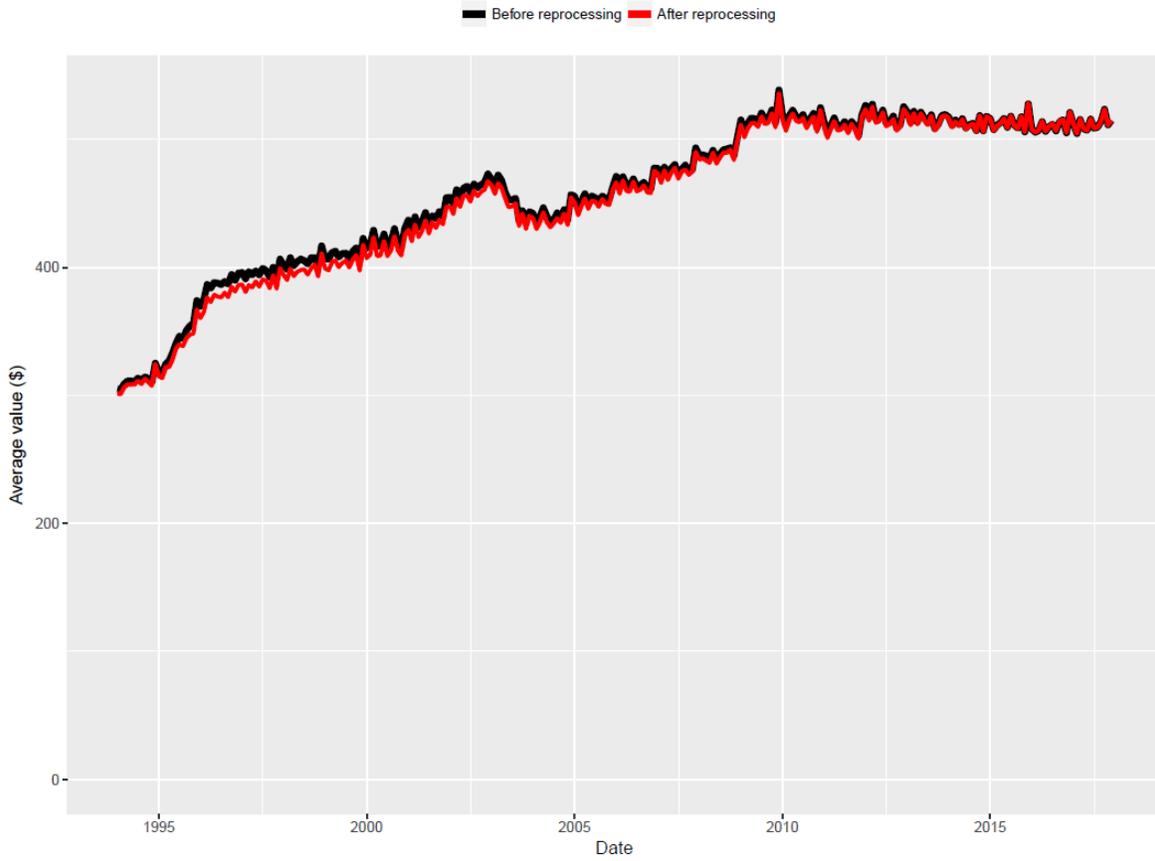
Figure A5. UINCyymm: Percent of SSI beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 25

Mathematica

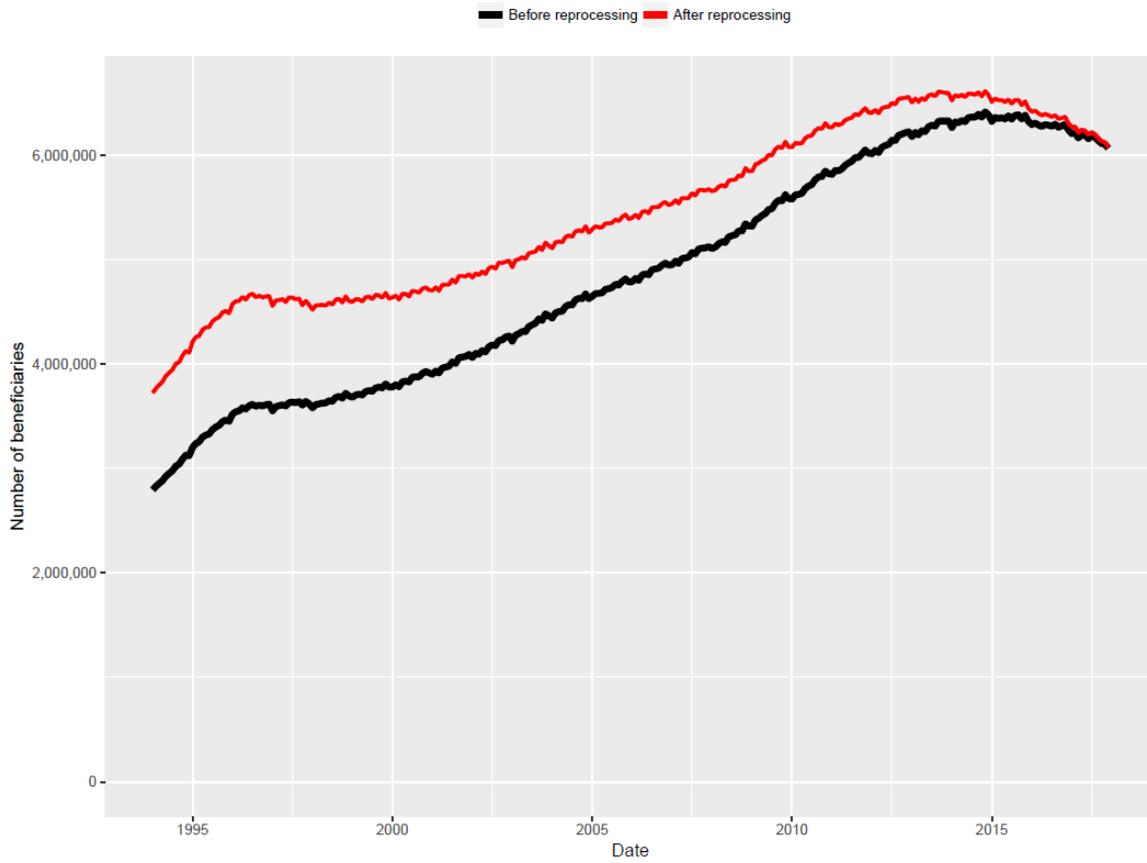
Figure A6. UINCyymm: Average value among beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 26

Mathematica

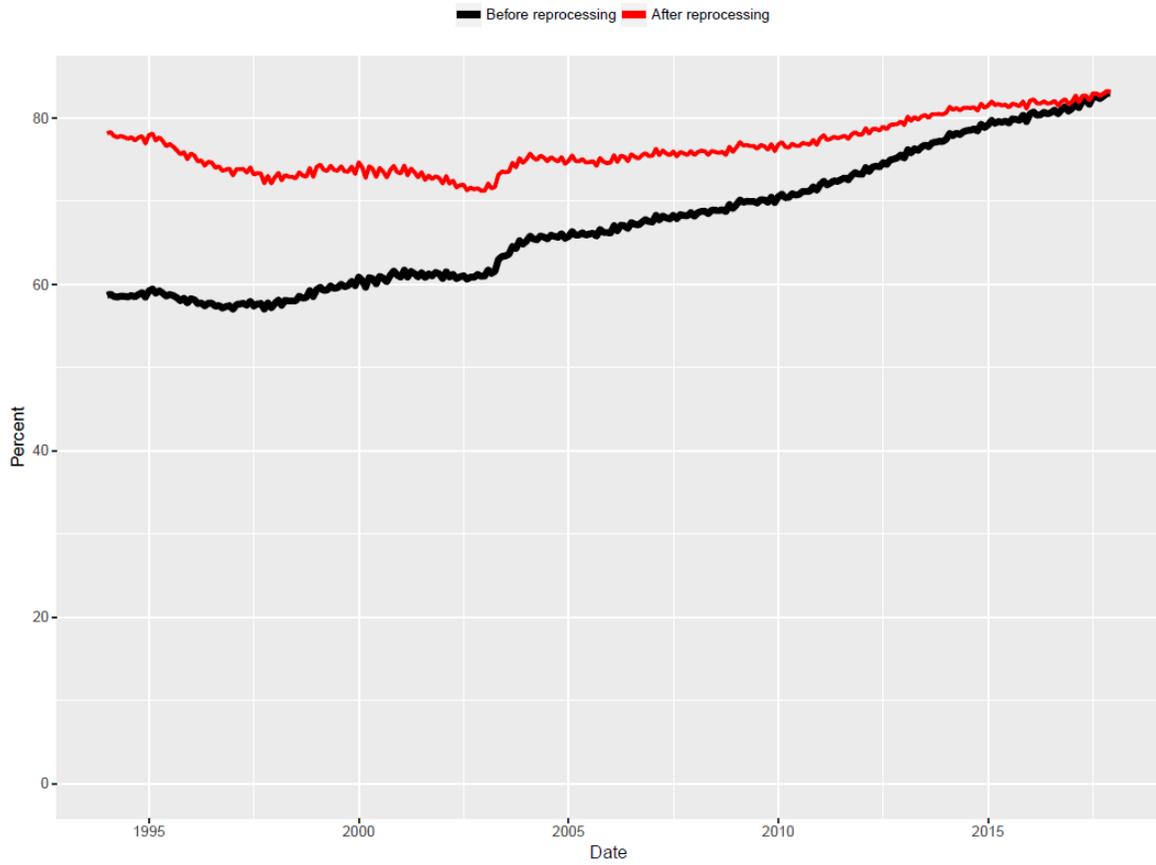
Figure A7. FAMTyymm: Number of beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 27

Mathematica

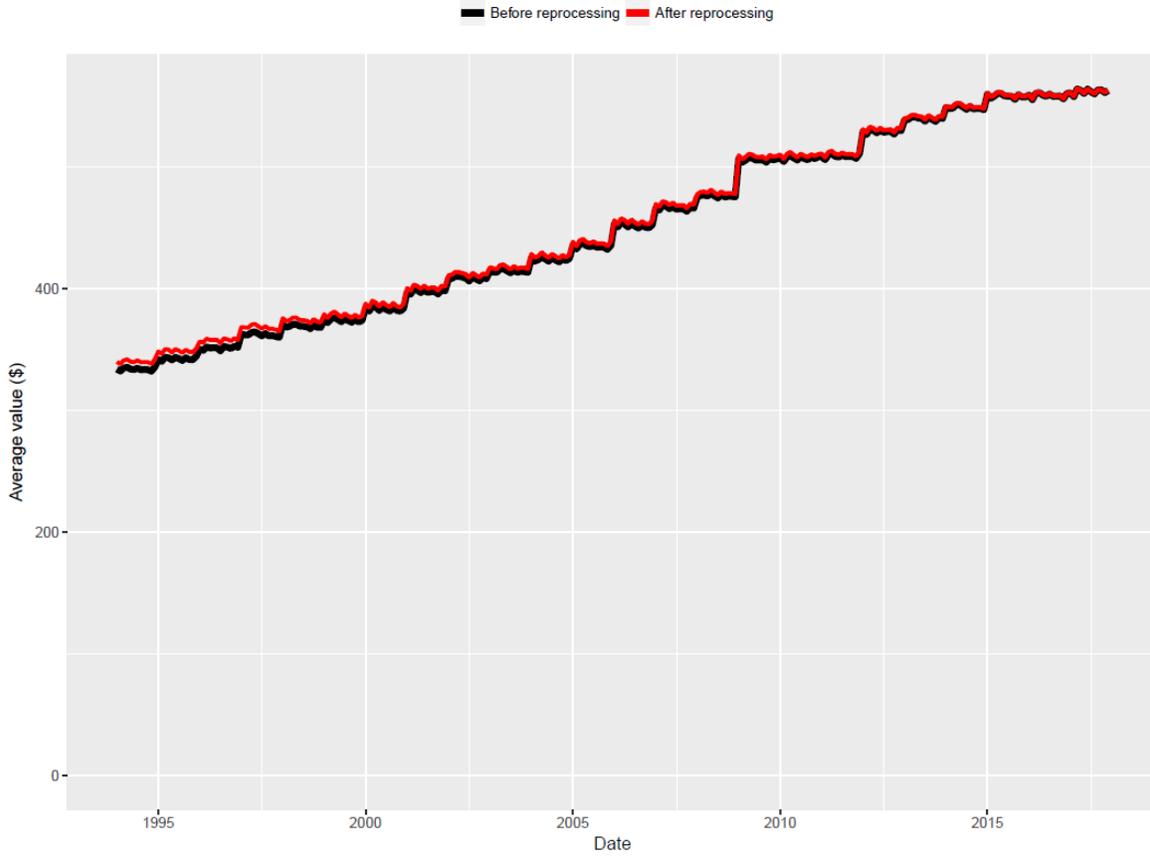
Figure A8. FAMTyymm: Percent of SSI beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 28

Mathematica

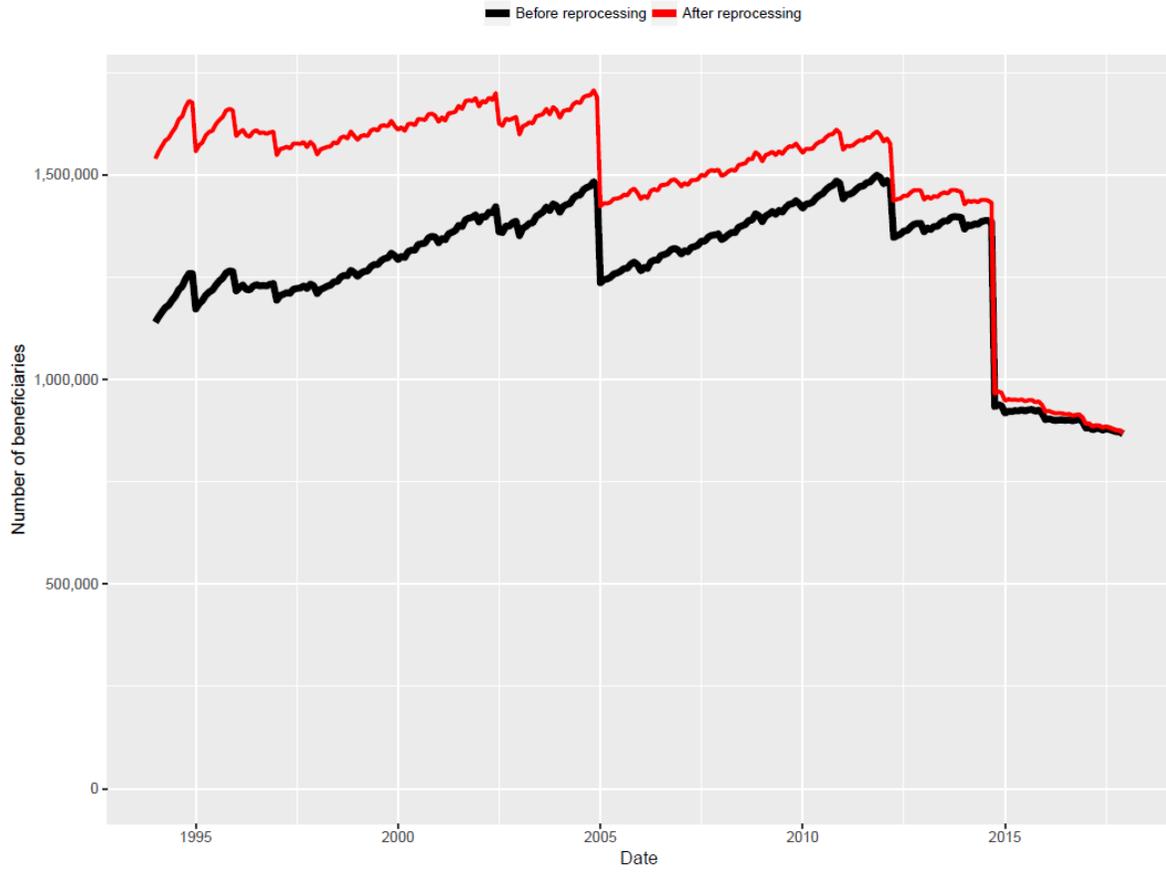
Figure A9. FAMTyymm: Average value among beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 29

Mathematica

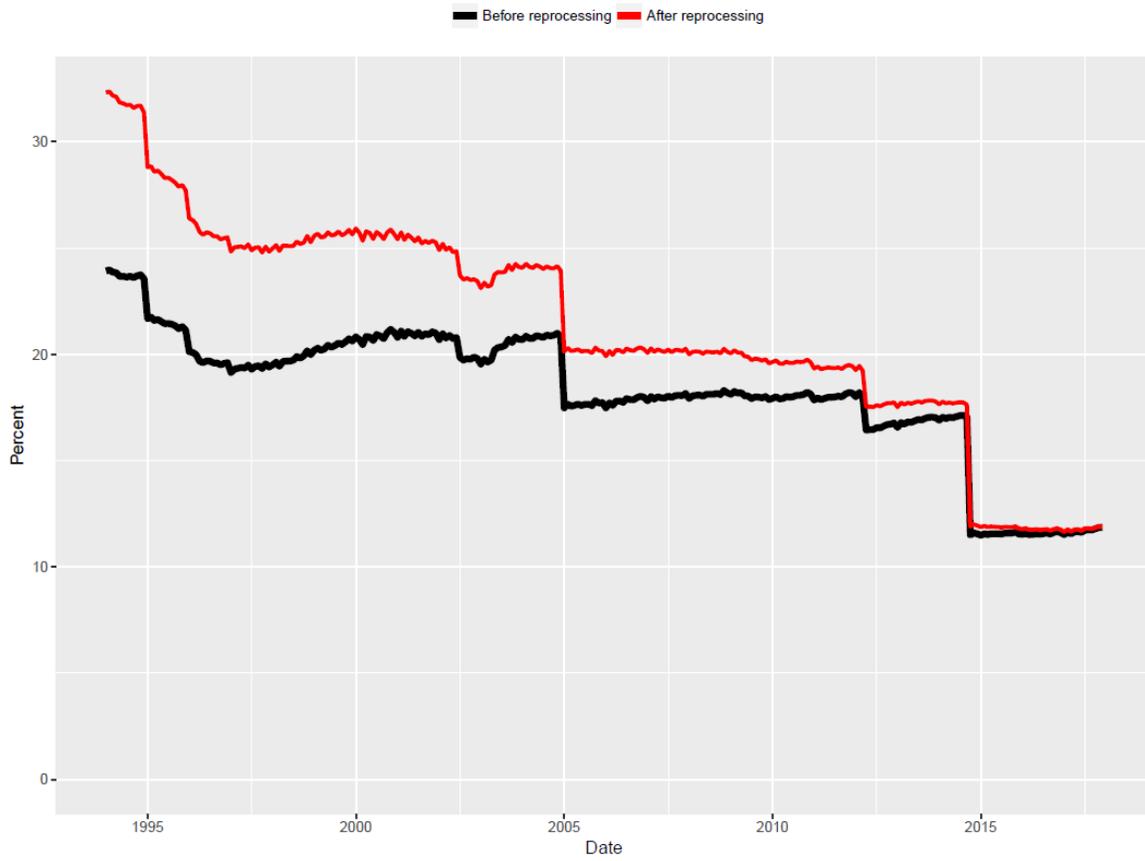
Figure A10. SAMTyymm: Number of beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 30

Mathematica

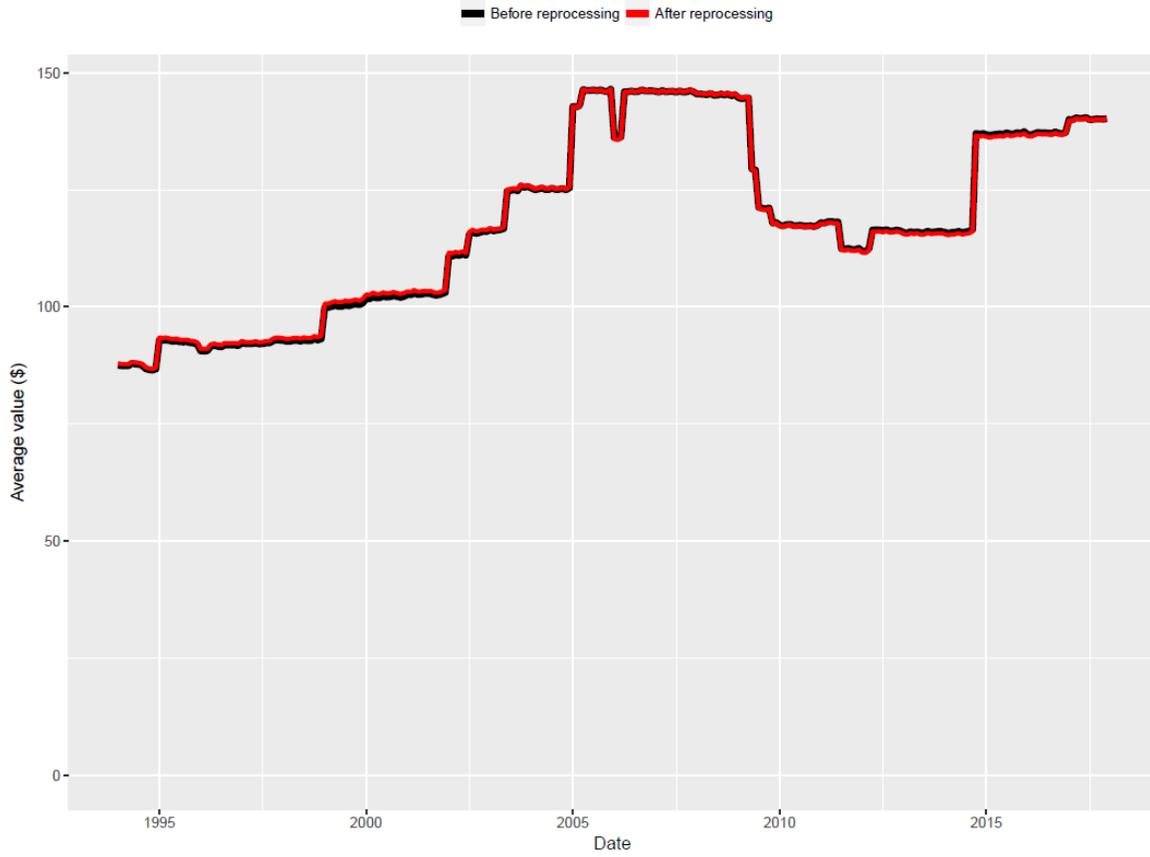
Figure A11. SAMTyymm: Percent of SSI beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 31

Mathematica

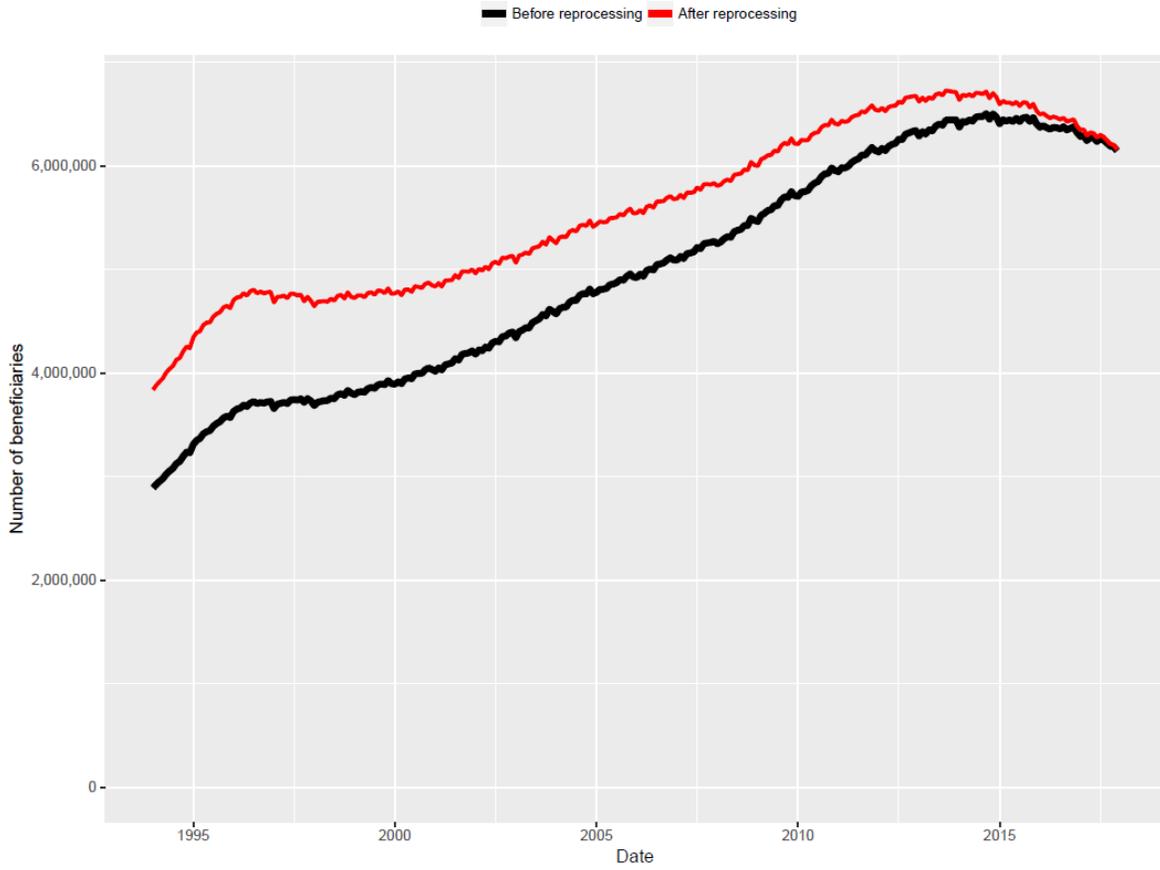
Figure A12. SAMTyymm: Average value among beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 32

Mathematica

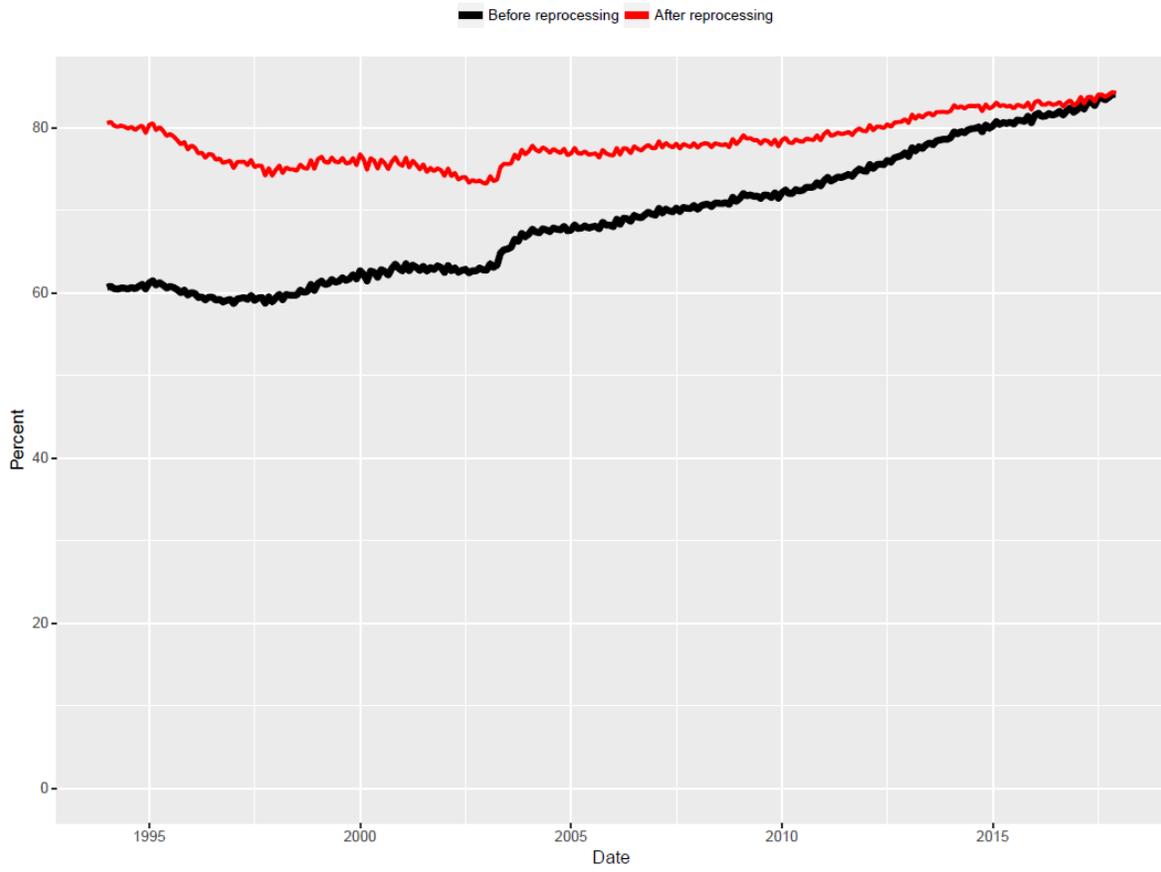
Figure A13. DUESymm: Number of beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 33

Mathematica

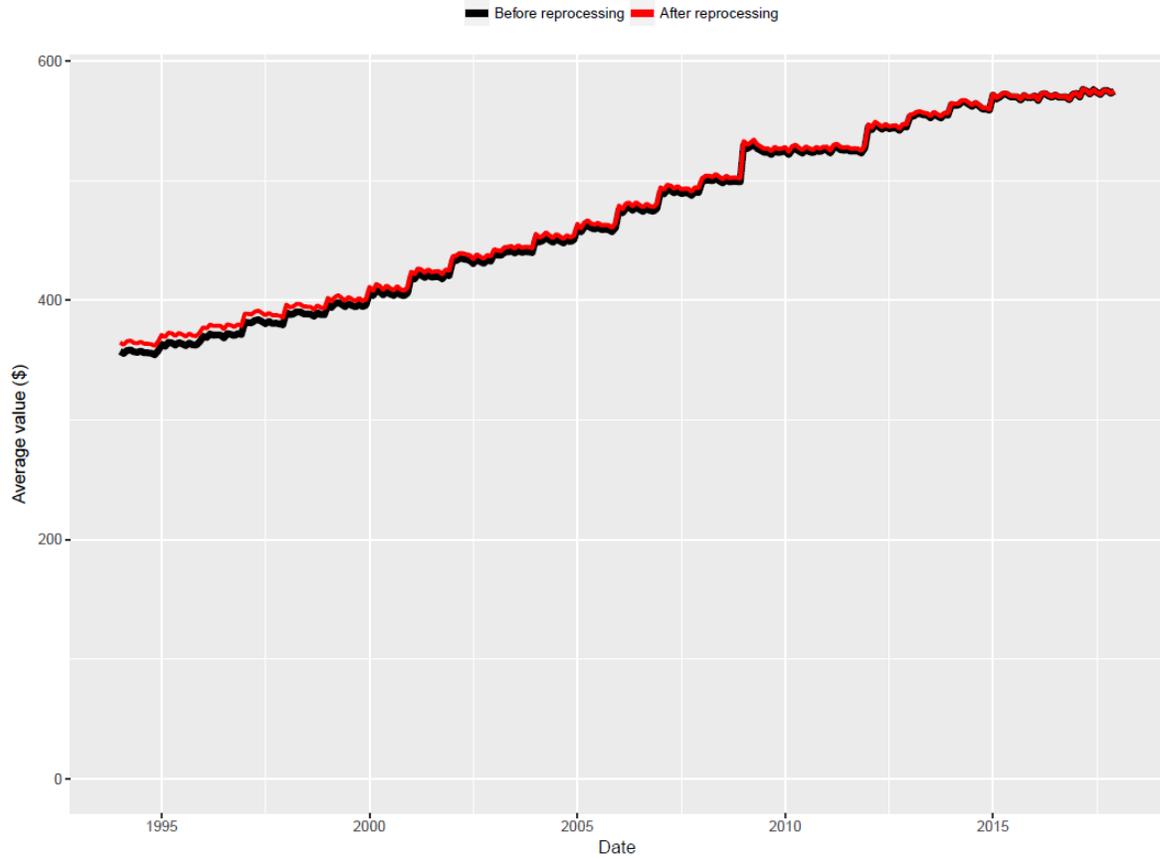
Figure A14. DUESymm: Percent of SSI beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 34

Mathematica

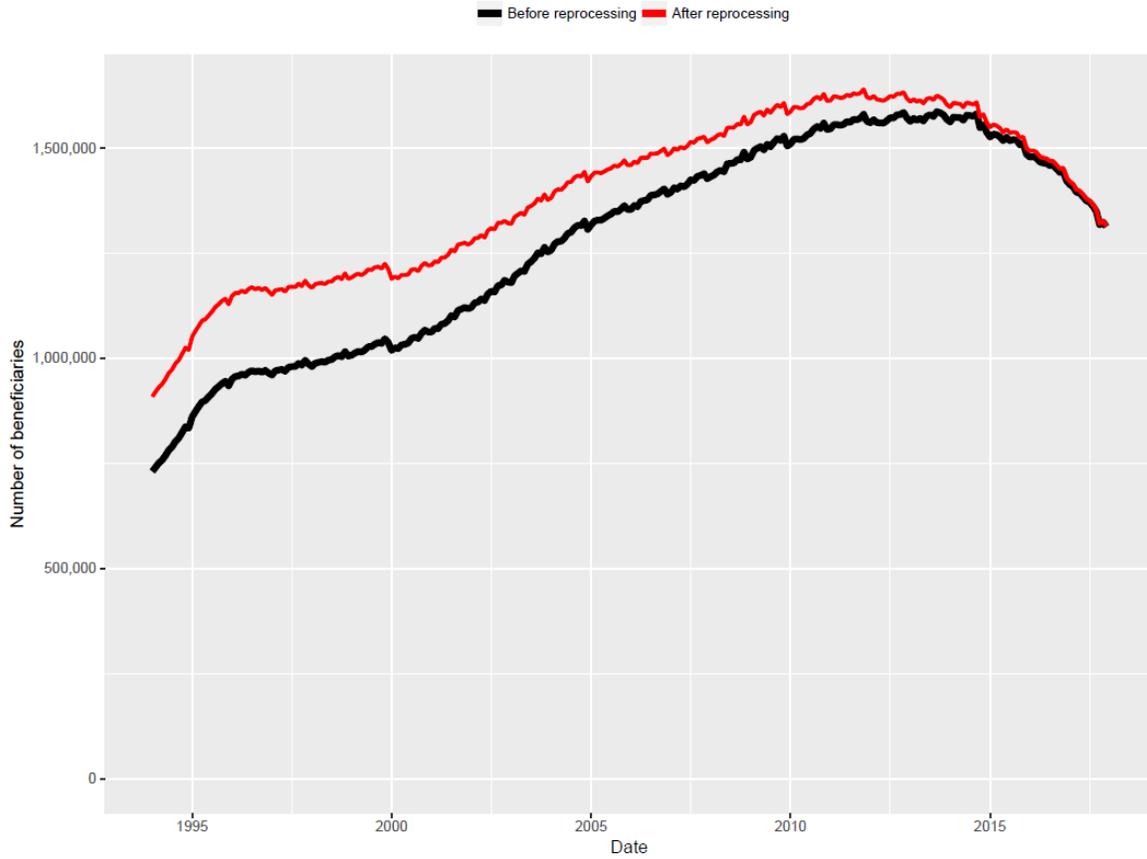
Figure A15. DUESymm: Average value among beneficiaries with value > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 35

Mathematica

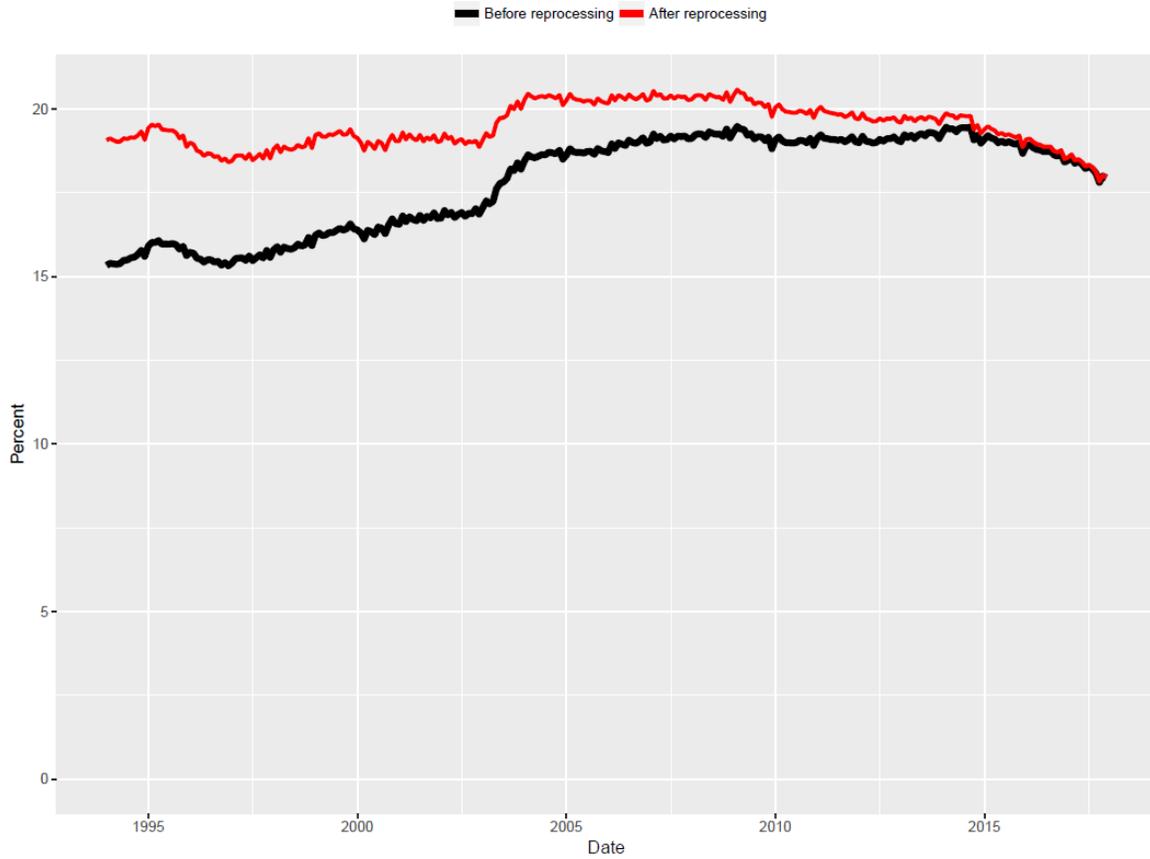
Figure A16. CONCyymm: Number of beneficiaries with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 36

Mathematica

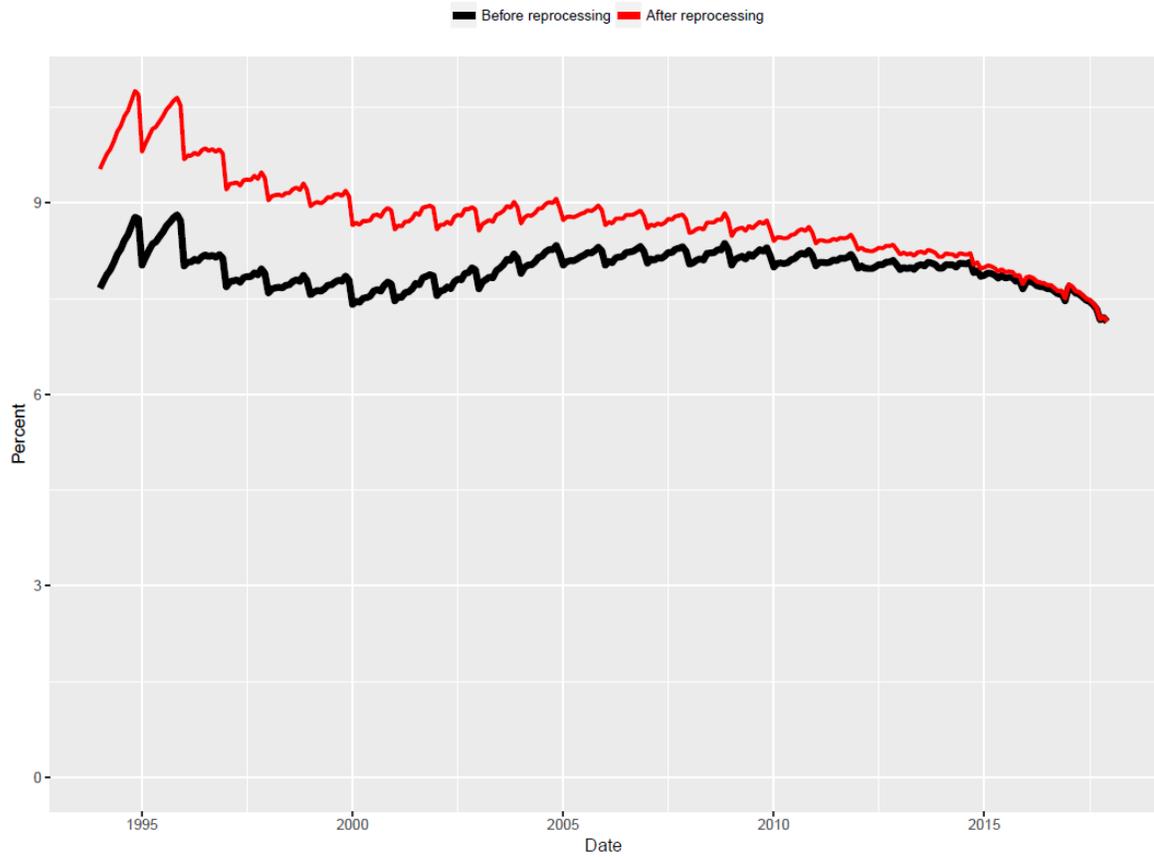
Figure A17. CONCyymm: Percent of SSI beneficiaries with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 37

Mathematica

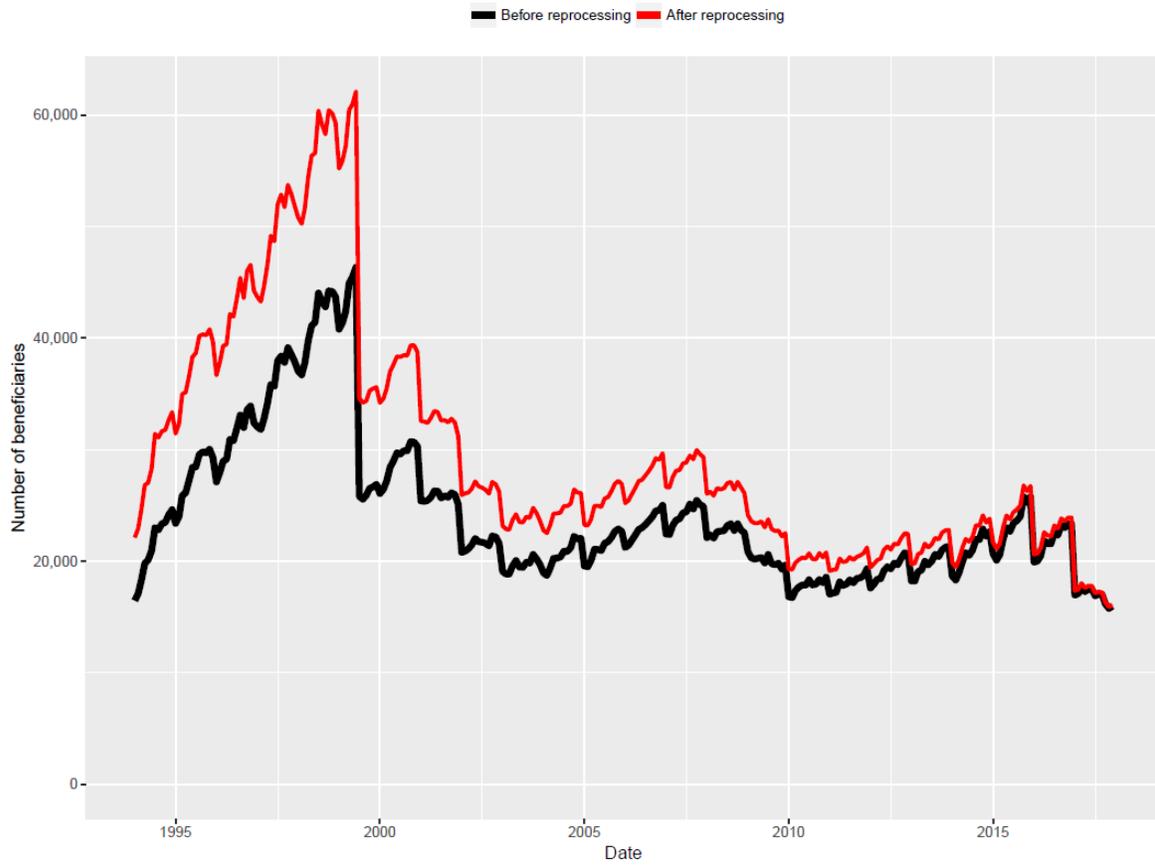
Figure A18. CONCyymm: Among beneficiaries with populated values, percent with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 38

Mathematica

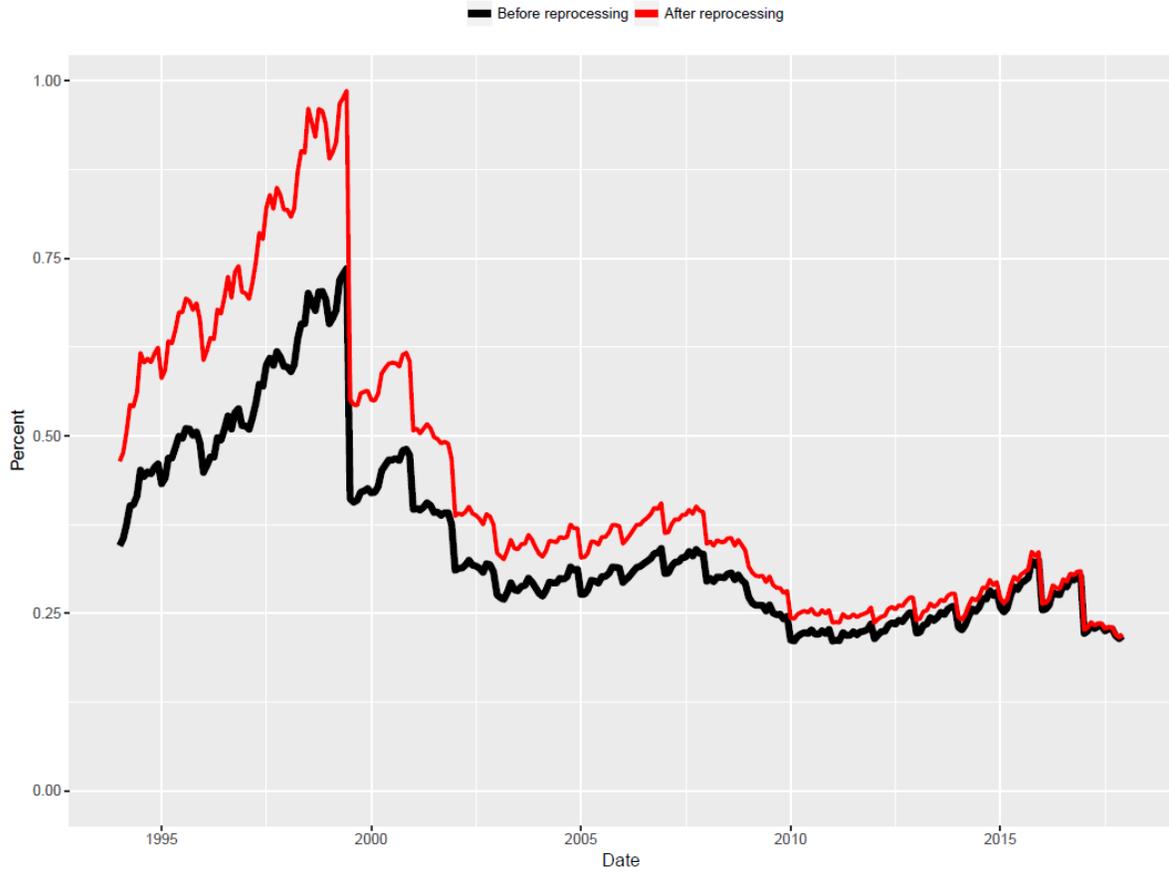
Figure A19. PROAymm: Number of beneficiaries with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 39

Mathematica

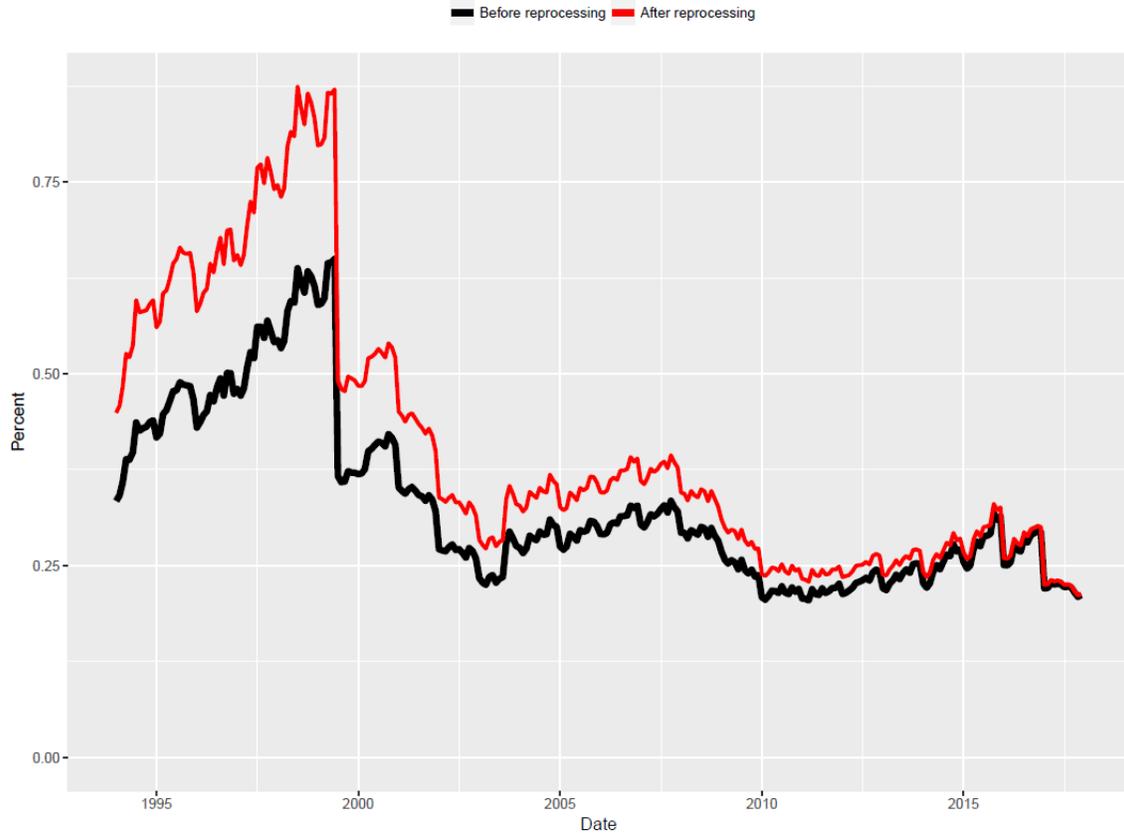
Figure A20. PROAyymm: Percent of SSI beneficiaries with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 40

Mathematica

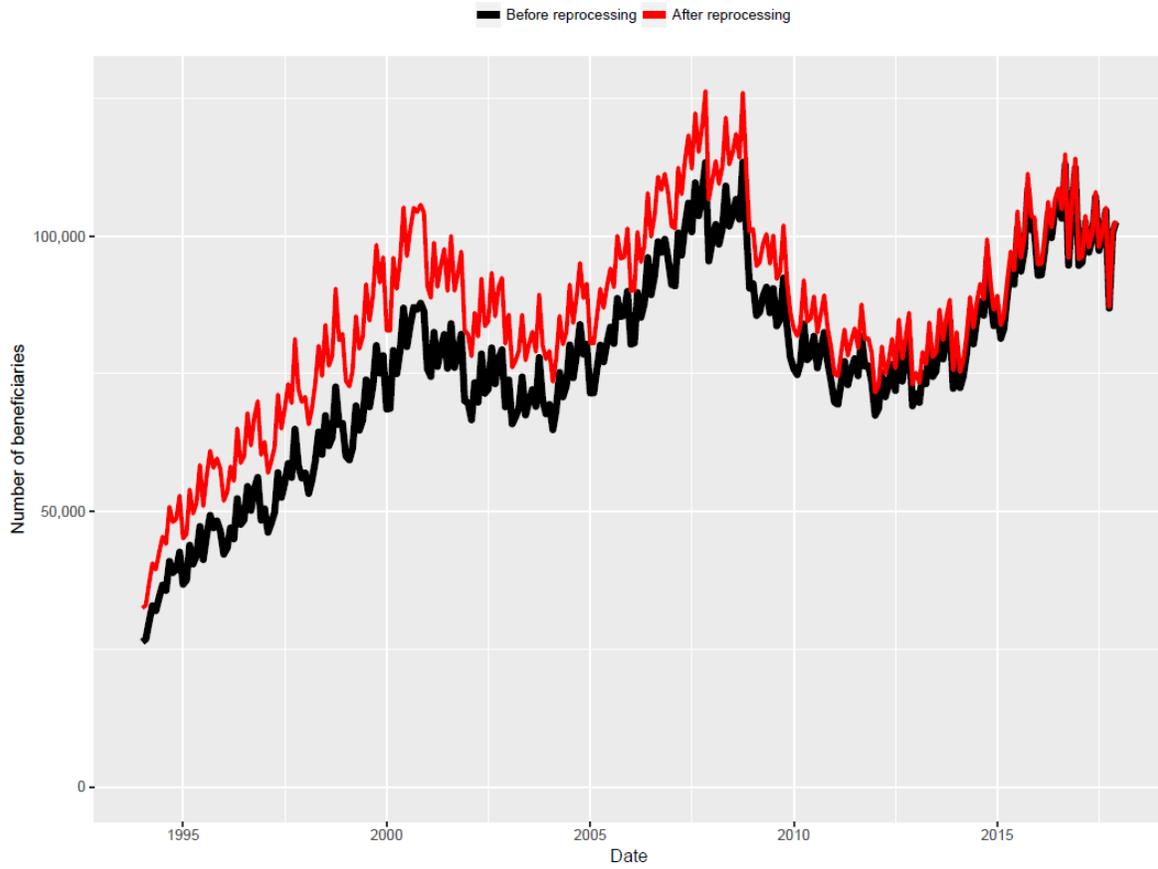
Figure A21. PROAymm: Among beneficiaries with populated values, percent with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 41

Mathematica

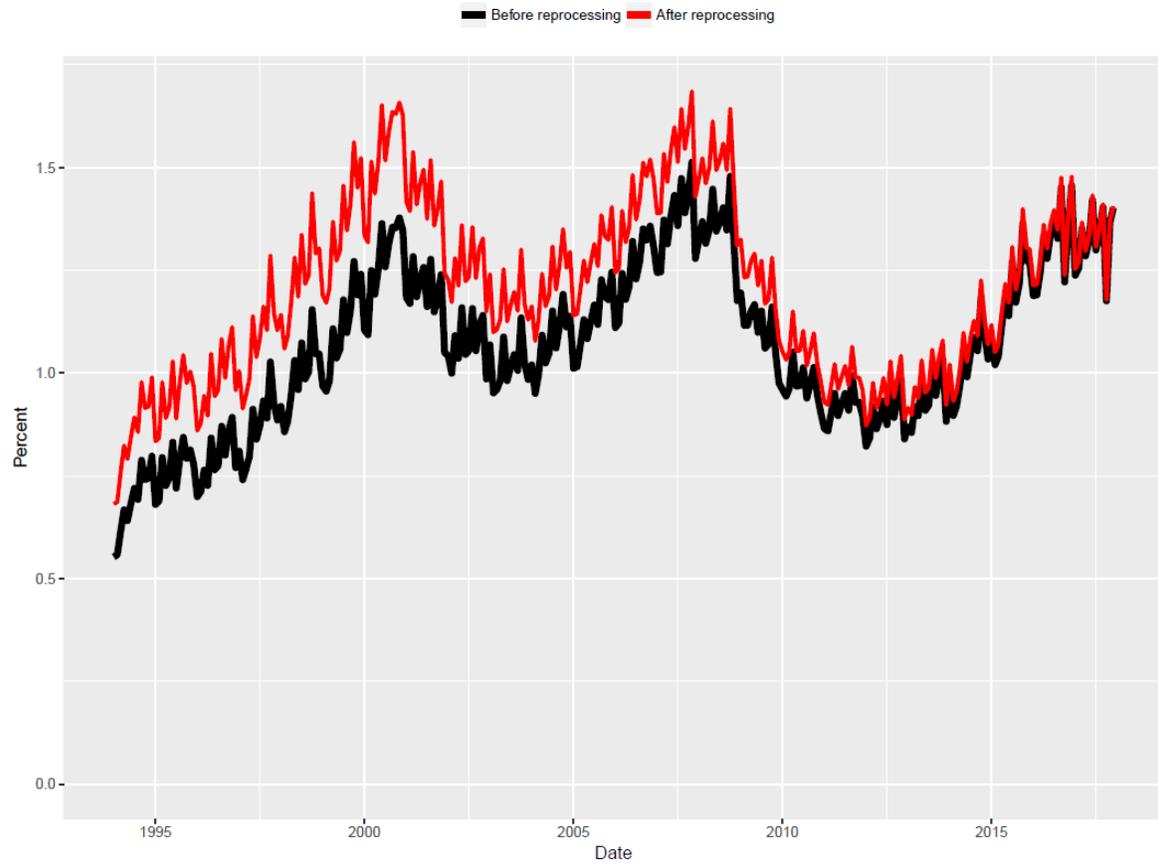
Figure A22. PROByymm: Number of beneficiaries with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 42

Mathematica

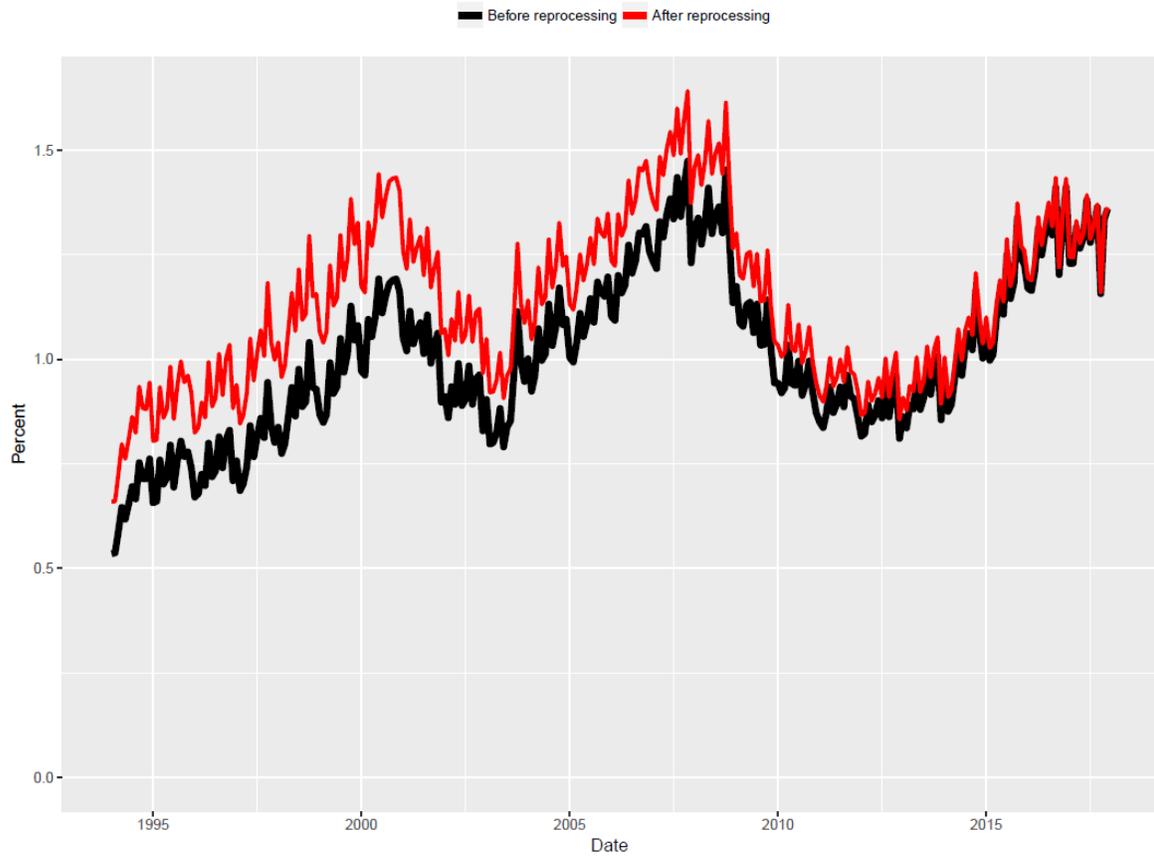
Figure A23. PROByymm: Percent of SSI beneficiaries with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 43

Mathematica

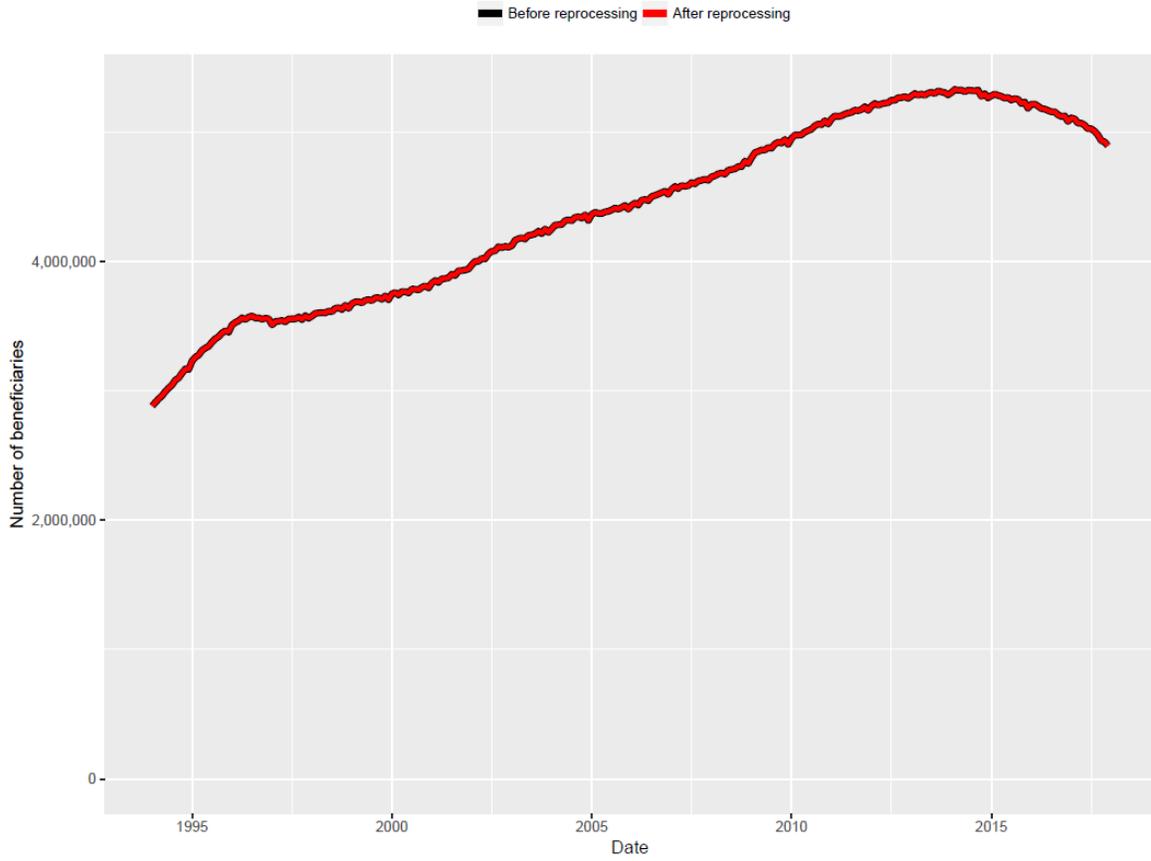
Figure A24. PROByymm: Among beneficiaries with populated values, percent with value = 1



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 44

Mathematica

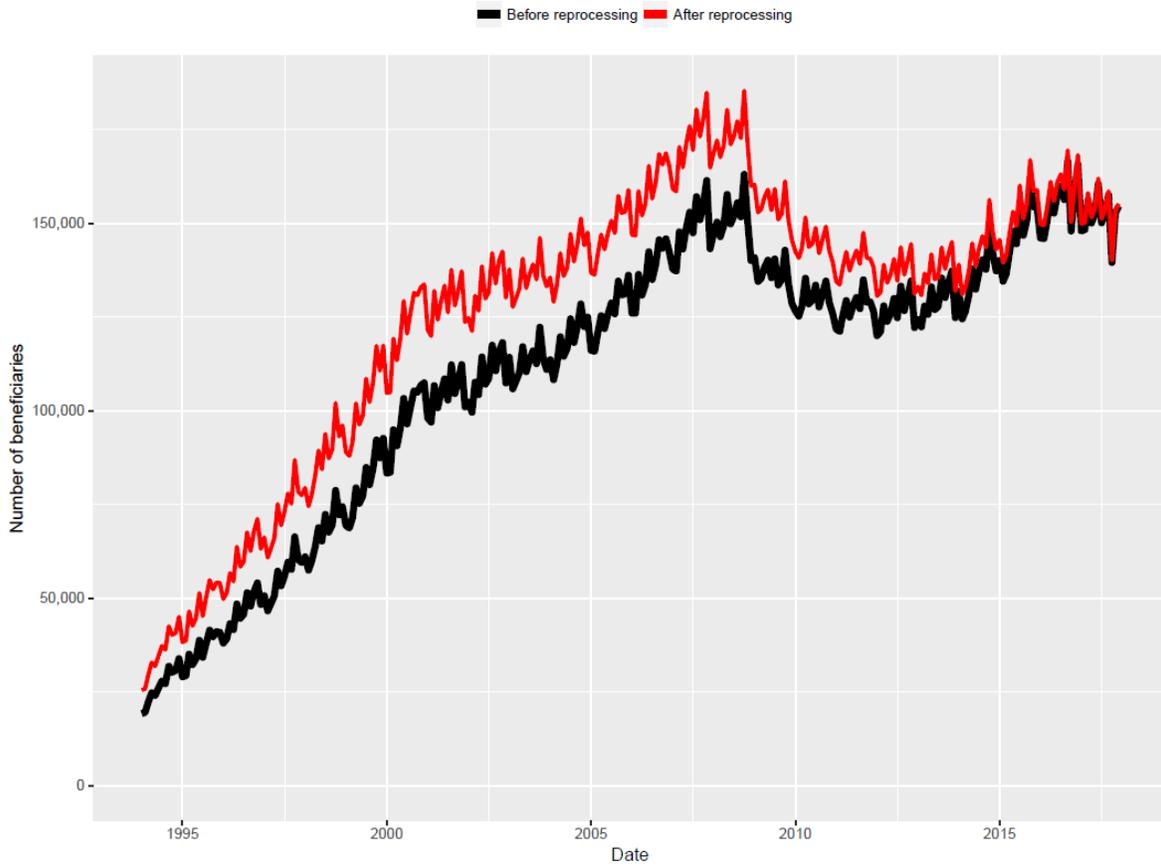
Figure A25. STWSSlyymm: Number of beneficiaries where STWSSI = 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 45

Mathematica

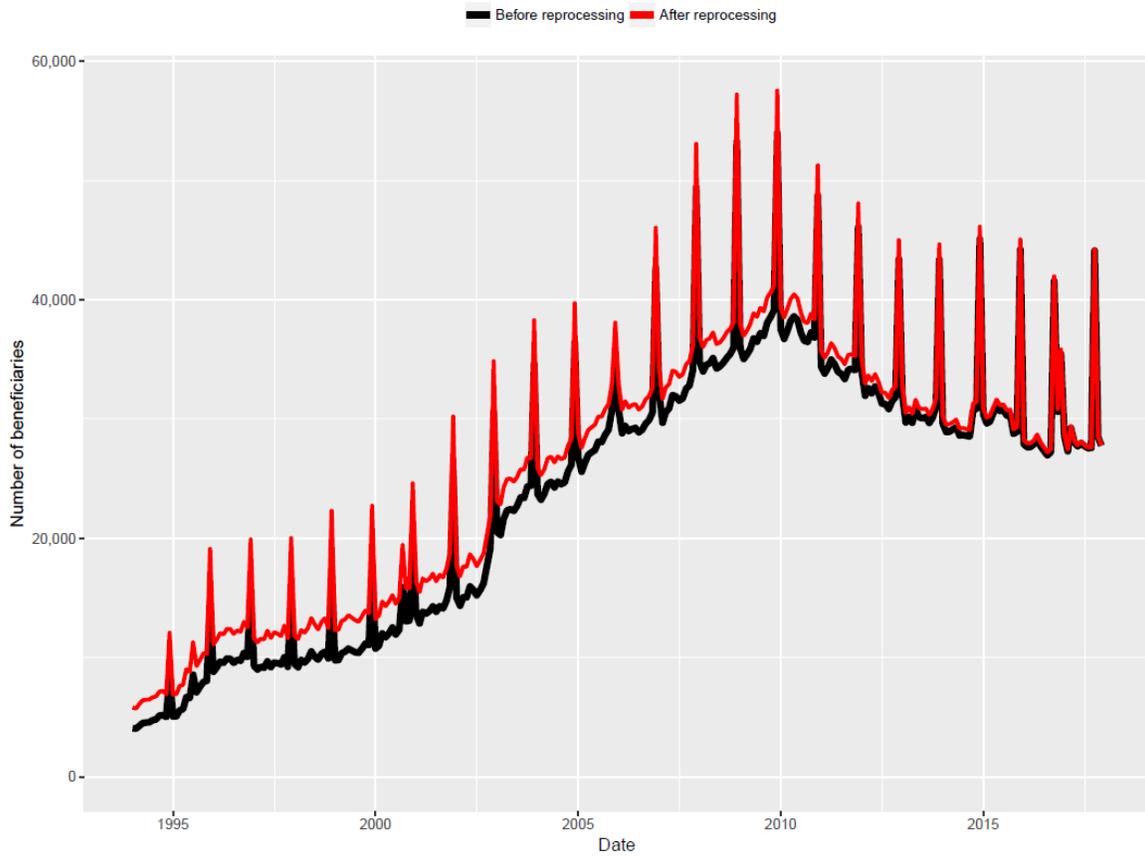
Figure A26. STWSSlyymm: Number of beneficiaries where STWSSI = 1, 2, or 3



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 46

Mathematica

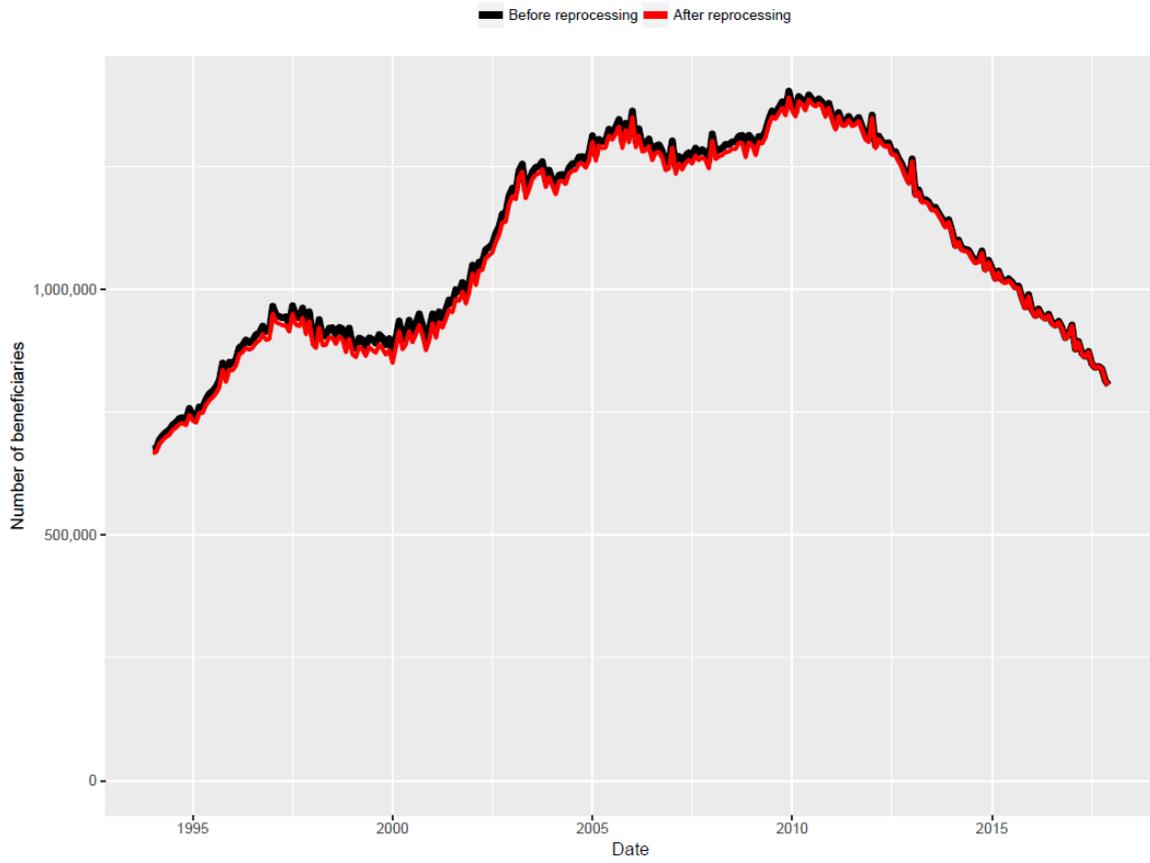
Figure A27. STWSSlyymm: Number of beneficiaries where STWSSI = 4



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 47

Mathematica

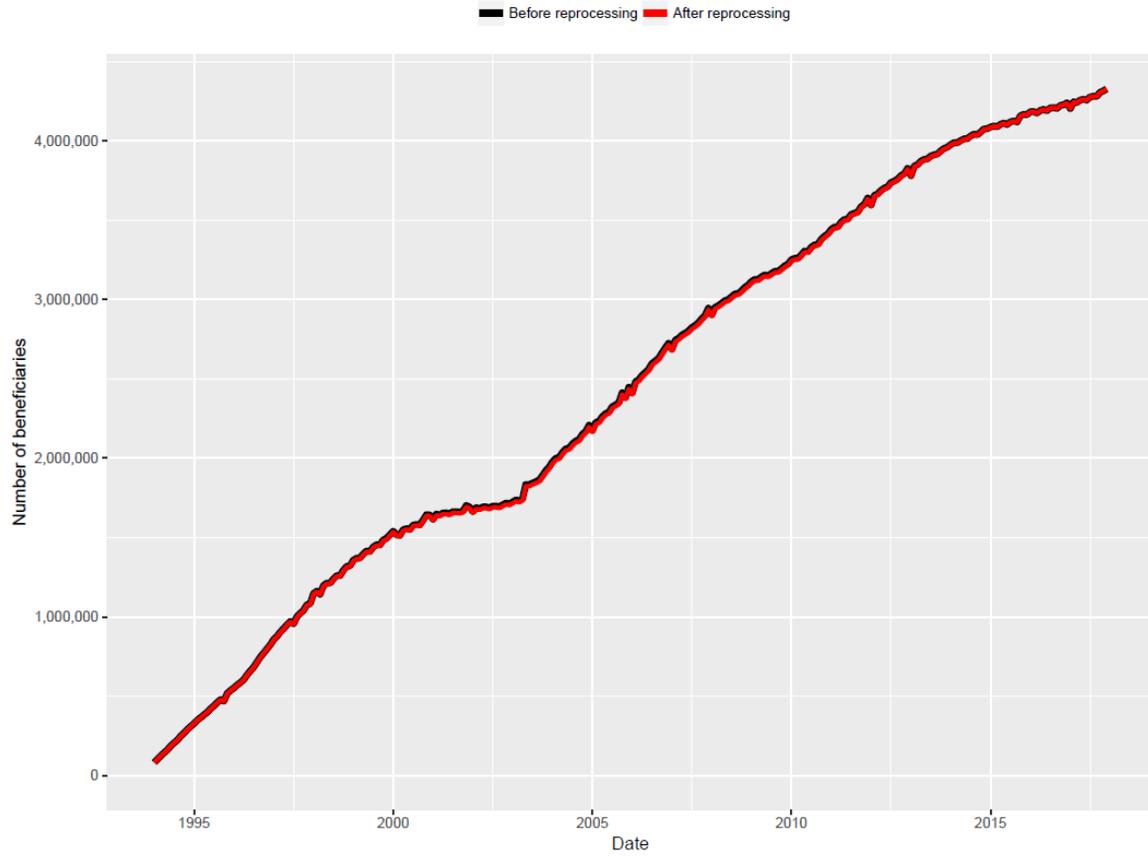
Figure A28. STWSSlyymm: Number of beneficiaries where STWSSI = 8



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 48

Mathematica

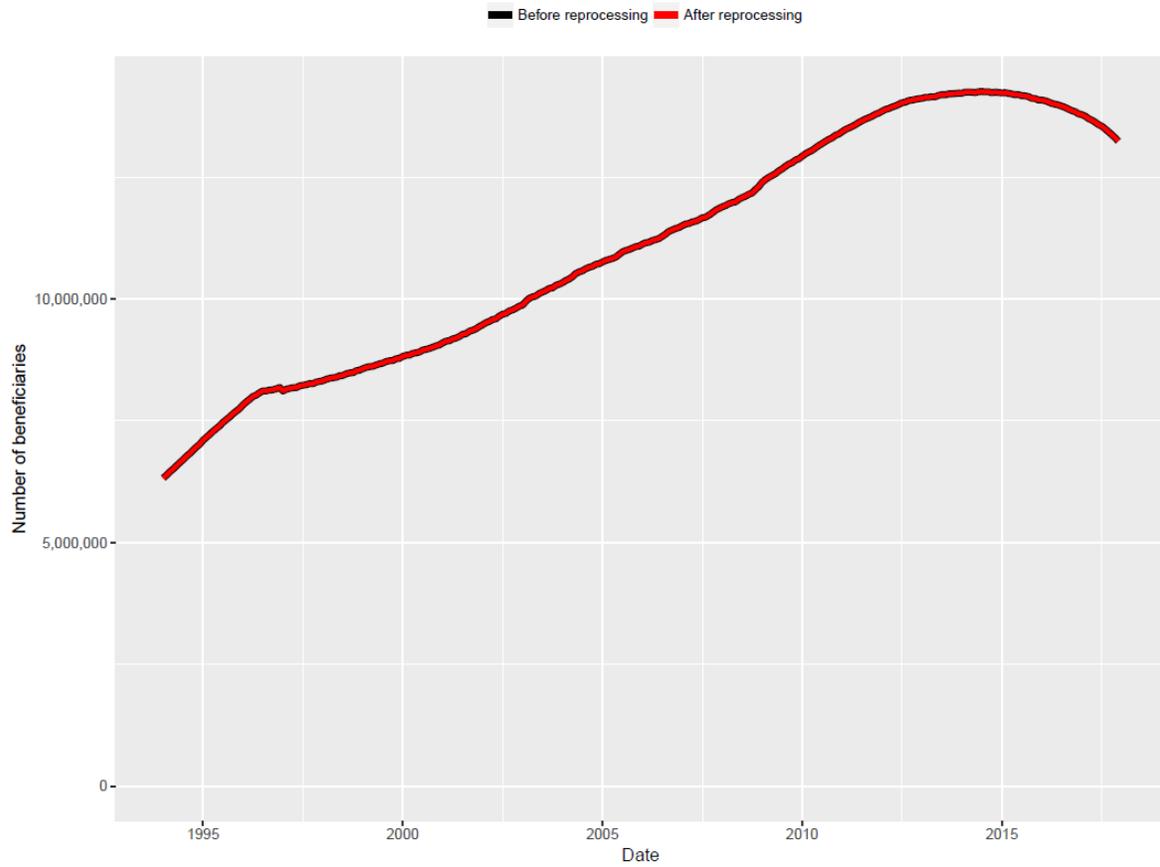
Figure A29. STWSSlyymm: Number of beneficiaries where STWSSI = 9



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 49

Mathematica

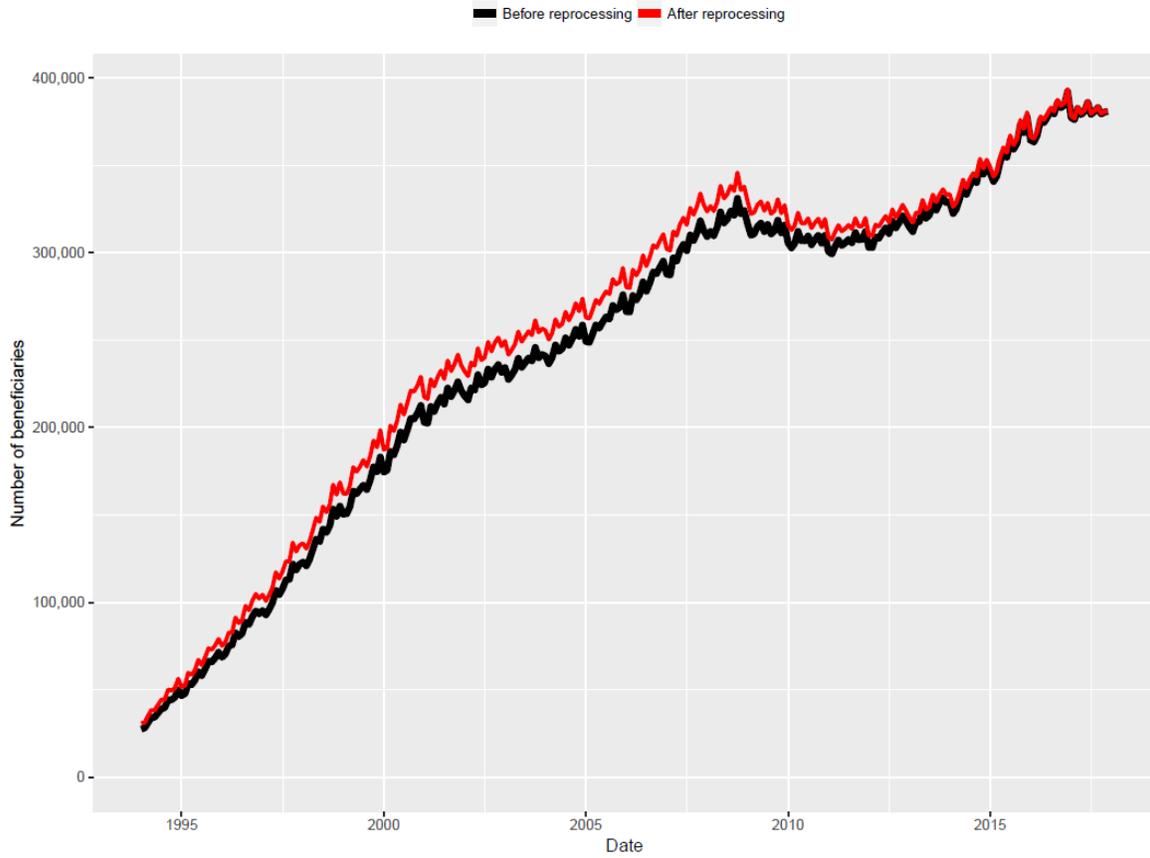
Figure A30. STWCMymm: Number of beneficiaries where STWCM = 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 50

Mathematica

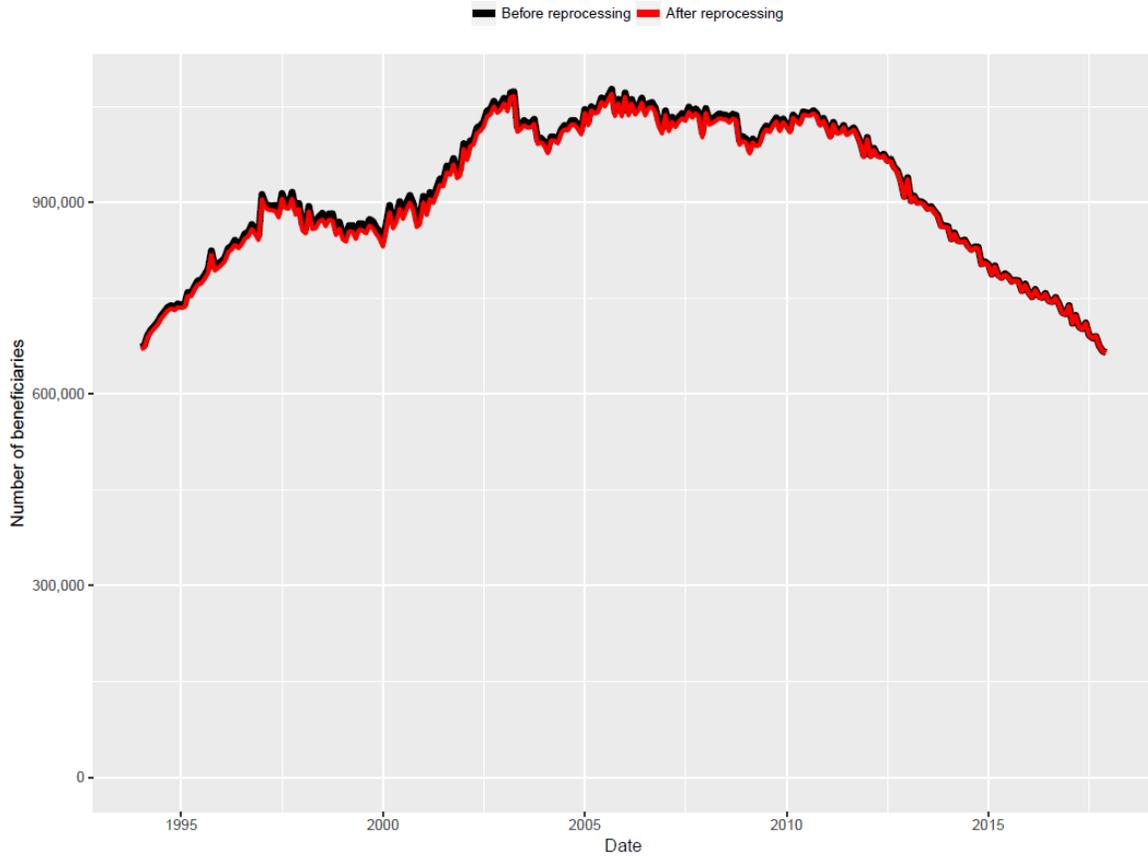
Figure A31. STWCMymm: Number of beneficiaries where STWCM = 1, 2, or 3



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 51

Mathematica

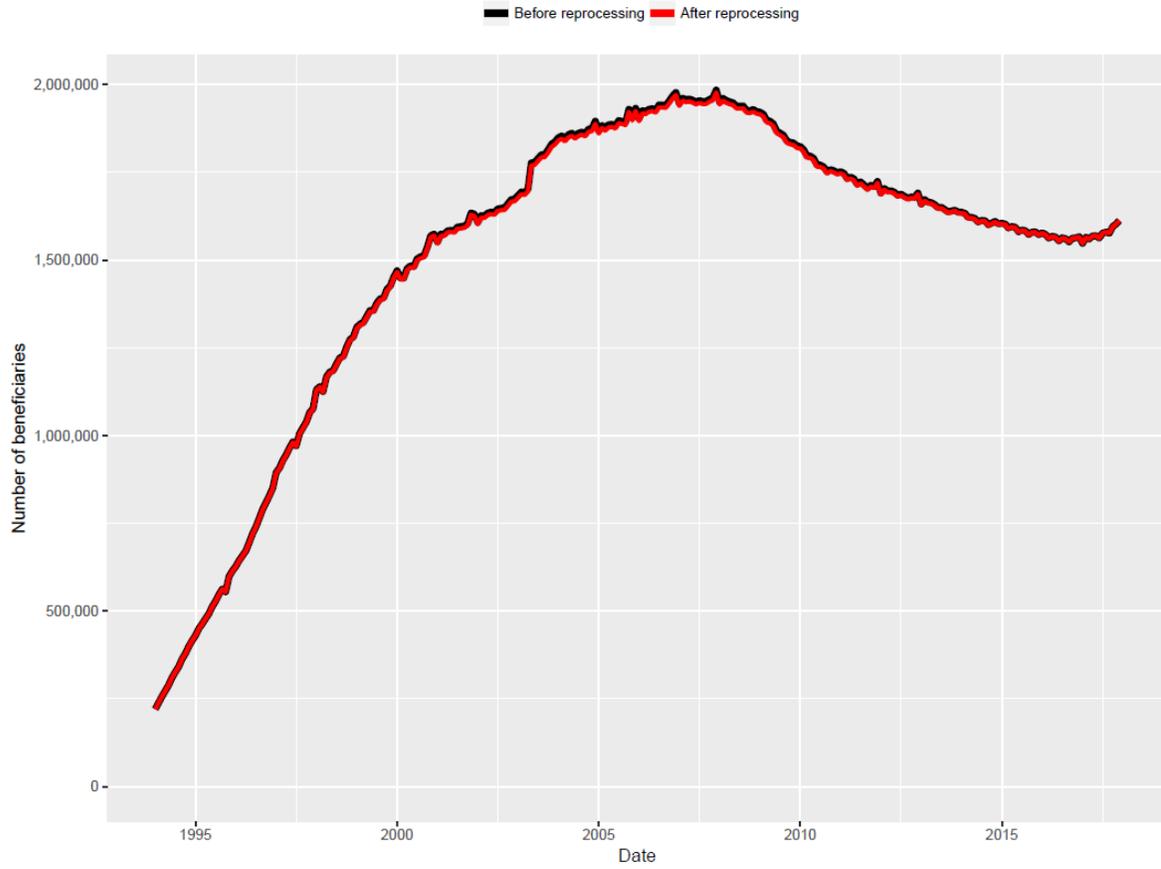
Figure A32. STWCMymm: Number of beneficiaries where STWCM = 8



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 52

Mathematica

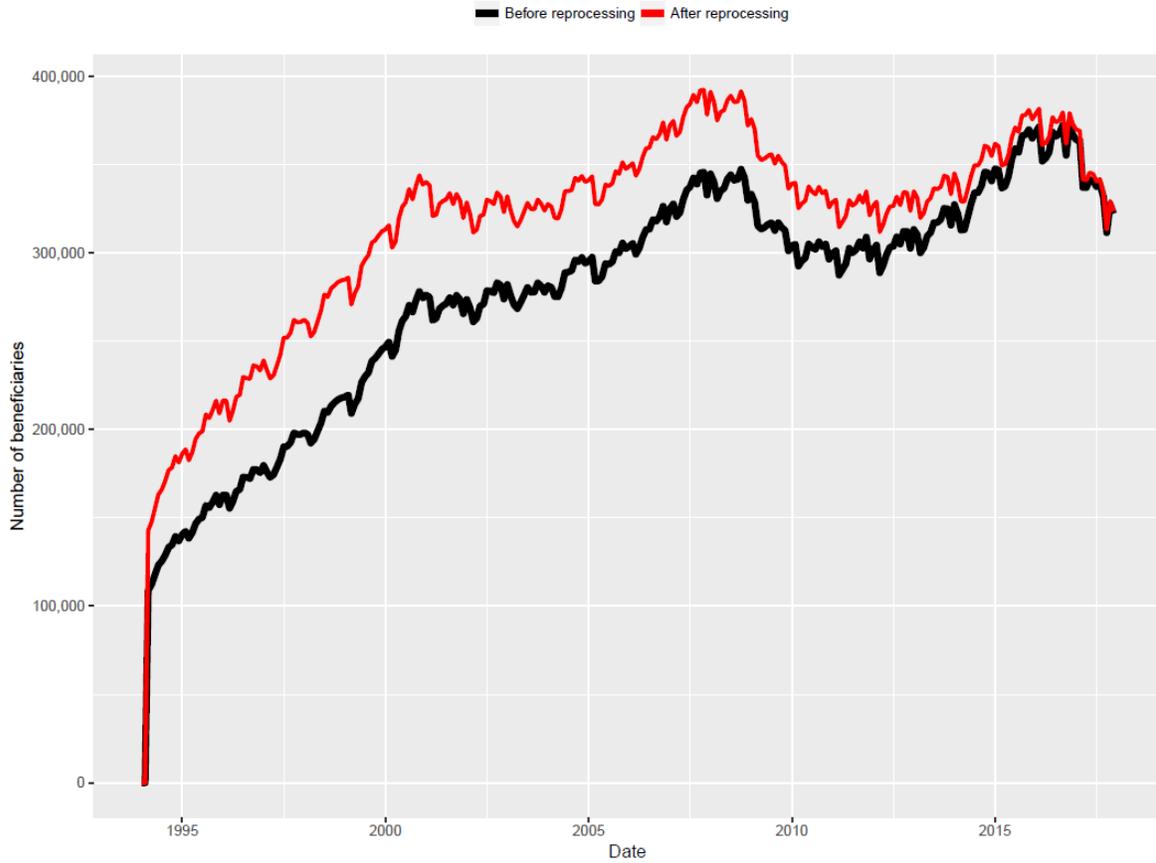
Figure A33. STWCMymm: Number of beneficiaries where STWCM = 9



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 53

Mathematica

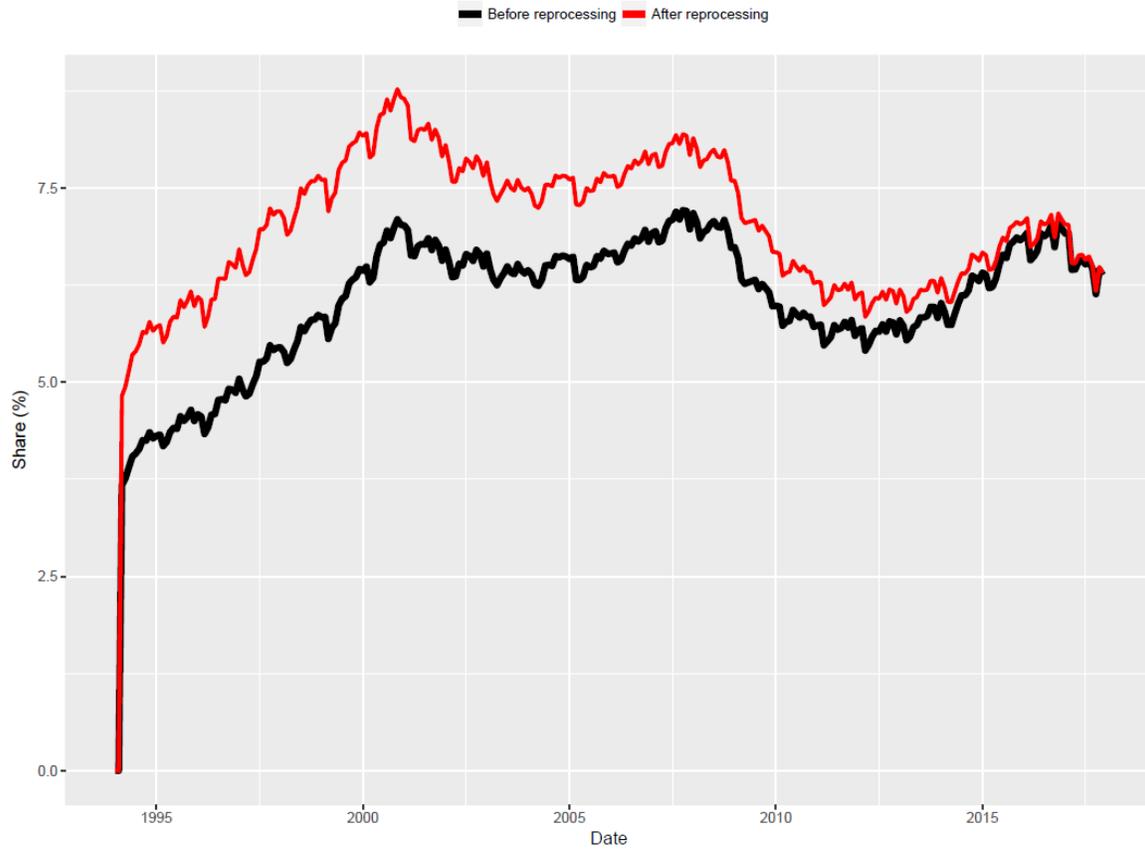
Figure A34: BFWSSI\_DRAFTyymm: Number of beneficiaries with values > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 54

Mathematica

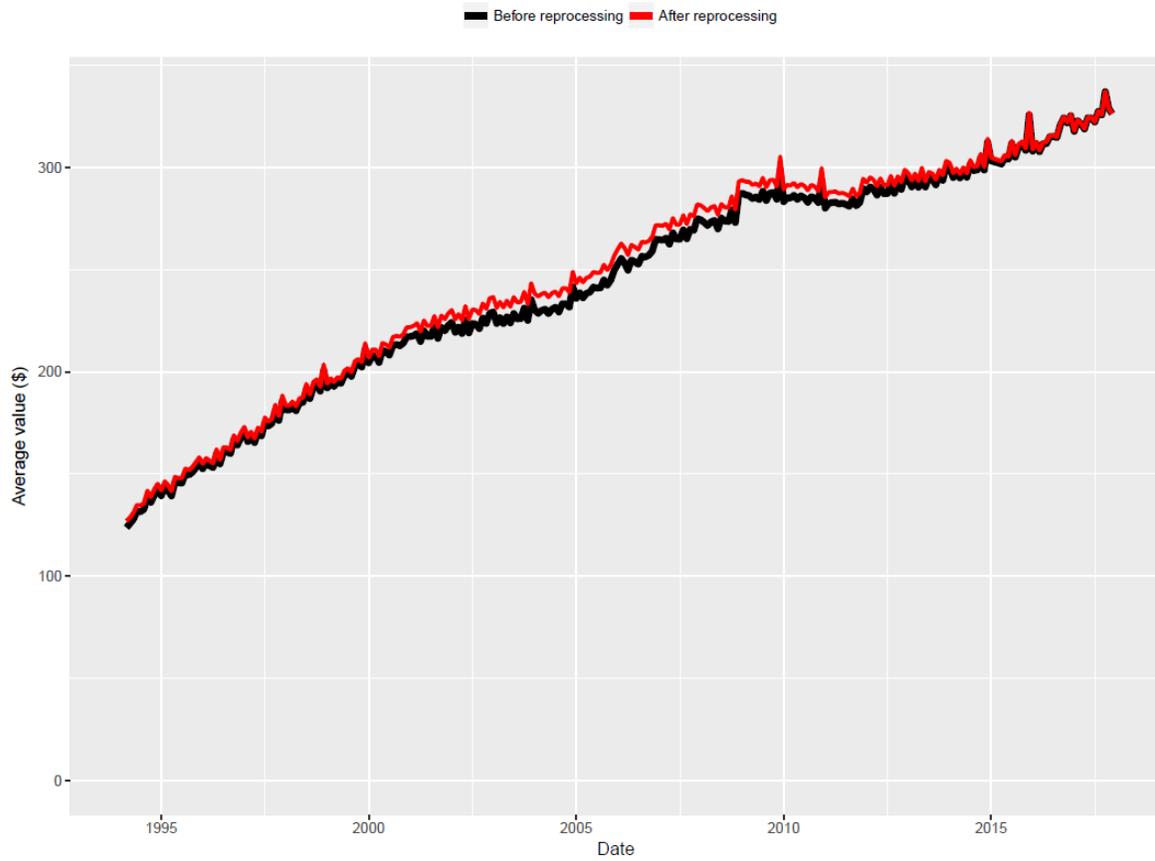
**Figure A35: BFWSSI\_DRAFTyymm: Among beneficiaries with STW = 0, 1, 2, or 3, share (%) of beneficiaries with values > 0**



**To:** Paul O'Leary, Social Security Administration  
**From:** Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
**Date:** 12/04/2019  
**Page:** 55

Mathematica

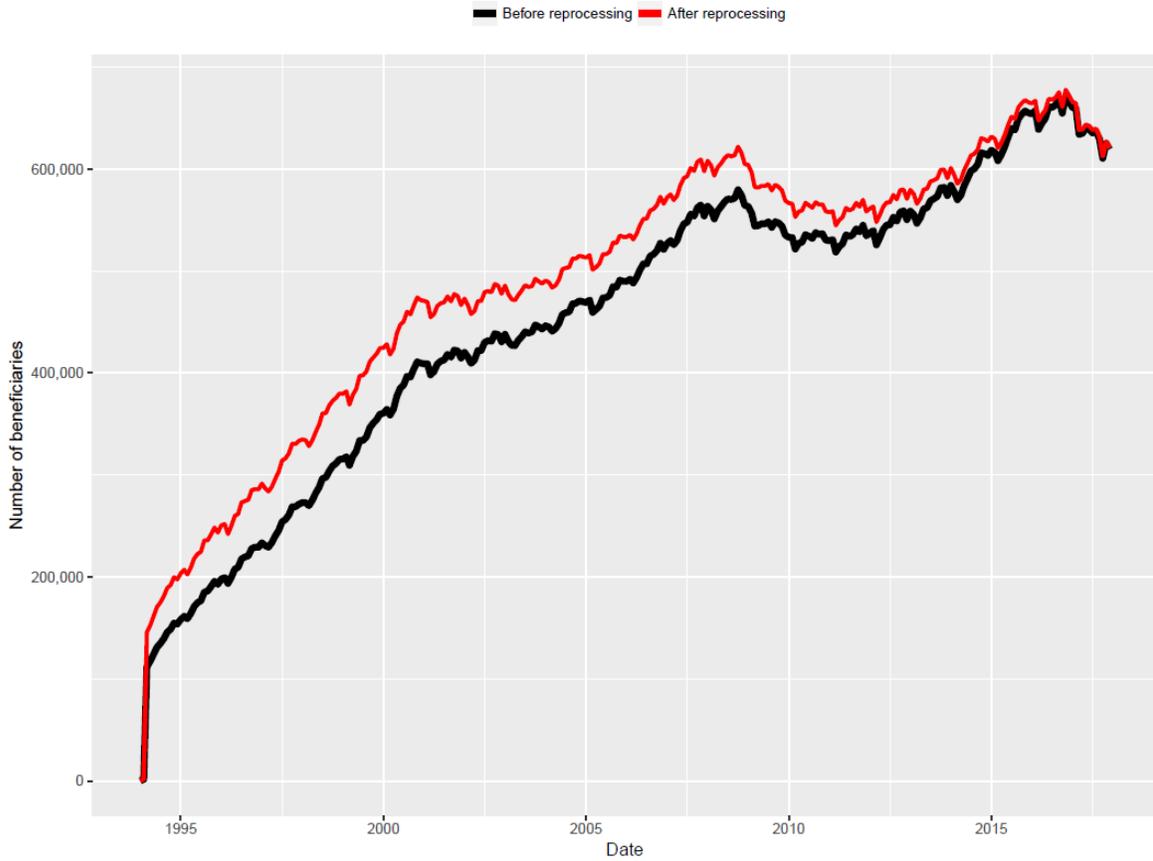
**Figure A36: BFWSSI\_DRAFTyymm: Average of values > 0**



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 56

Mathematica

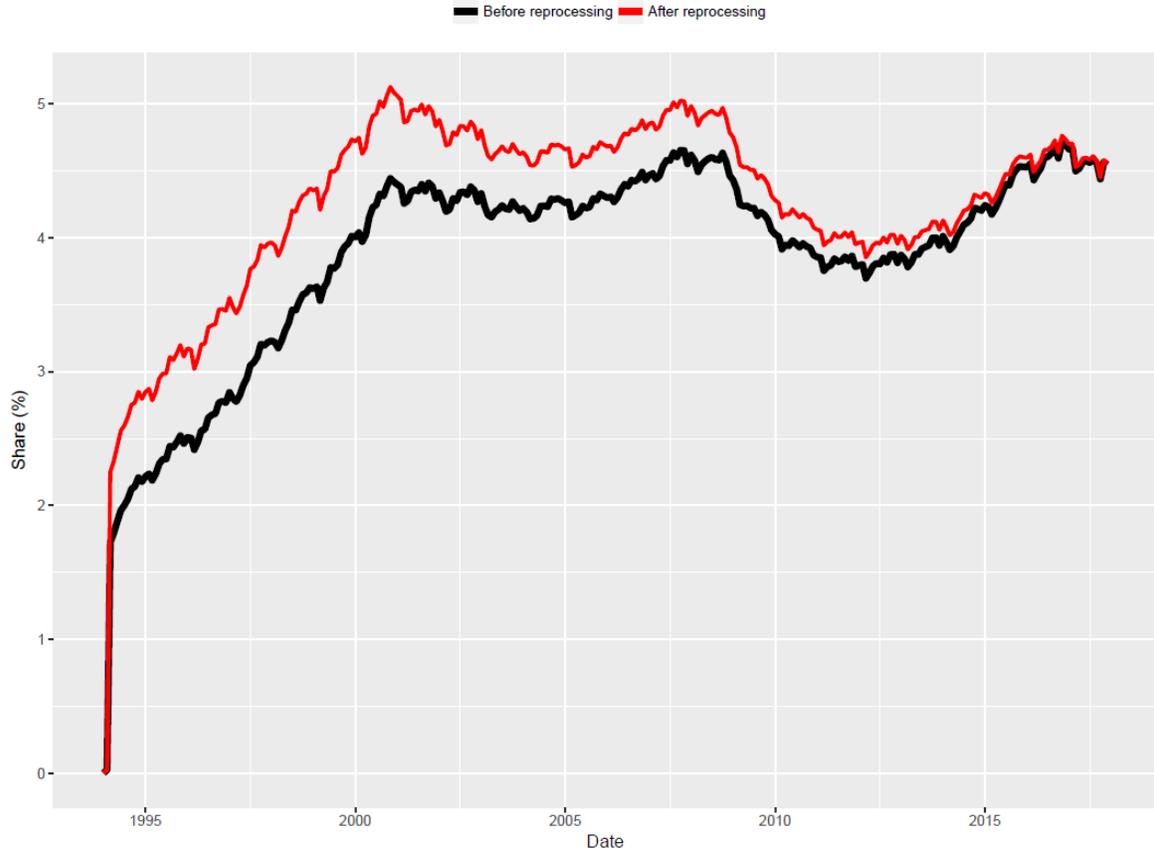
Figure A37: BFWCM\_DRAFTyymm: Number of beneficiaries with values > 0



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 57

Mathematica

**Figure A38: BFWCM\_DRAFTyymm: Among beneficiaries with STW = 0, 1, 2, or 3, share (%) of beneficiaries with values > 0**



To: Paul O'Leary, Social Security Administration  
From: Michael Anderson, Jody Schimmel Hyde and Dawn Phelps  
Date: 12/04/2019  
Page: 58

Mathematica

Figure A39: BFWCM\_DRAFTyymm: Average of values > 0

