STATE OF WASHINGTON

OFFICE OF FINANCIAL MANAGEMENT

Insurance Building, PO Box 43113 □ Olympia, Washington 98504-3113 □ (360) 902-0555

December 21, 2020

Disclosure Avoidance System Team:

Thank you for providing the November 16, 2020, Privacy Protected Microdata File for our review. While some improvements have been made since the first PPMF release, we are still very concerned about the usability of the data and the implications to state and local governments that have come to rely on this data when making policies, funding allocations, and other critical decisions.

This PPMF release is limited in scope to the data necessary to create the P.L. 94-171 redistricting data product. It was our understanding from presentations on differential privacy (DP) that the full scope of tables and information being released has to be taken into consideration when determining the privacy loss budget. Given that, we are extremely concerned that decisions about the privacy loss budget will be made soon without giving stakeholders the opportunity to provide feedback on the detailed data that includes 5-year age, sex race and ethnicity data. Stakeholders need to have access to multiple PPMF runs with different privacy loss budgets. Without runs containing full demographic and housing details, and variations of epsilon (the privacy loss budget statistic), partners' feedback with be limited.

We are very concerned by the prevalence of illogical and implausible values in the most recent PPMF data release. Such values are going to be an issue with the data user community until they are resolved. We believe that making occupancy status invariant, limiting the distribution of household population to blocks with occupied housing units, forcing each occupied unit to have at least one person, and limiting the frequency of outlier household sizes would go a long way towards improving the validity of the data. The 2020 Census data will simply not be credible if illogical and implausible values exist.

We are very concerned that the data shows less racial diversity in some counties and in many sub-county areas.

We believe the root cause of many these issues is the independent processing of the housing and population data that DP currently employs. We urge you to re-think that choice. Processing the data in this way removes all household relationships from the data tables. Those relationships are extremely important and are the source of very valuable information for a large number of census data products. We believe the de-coupling of this relationship will have far-reaching consequences affecting not only the decennial census data, but also the American Community Survey data.

Illogical Values and Related Biases

The illogical distribution of occupied housing units continues to create errors that must be corrected if the data are to be viewed as credible. For example, there are 9,197 blocks in Washington with 174,071 people in households, but no occupied housing units. Similarly, there are 8,978 blocks with occupied units, but no population to occupy them. There are 25,005 people in 500 blocks with an implausible household size greater than 20 people. The largest household size of any block in Washington from the 2010 Census was 16 and it was only one block. Until these illogical and improbable situations are corrected, analysis of this data is limited because the people and occupied housing involved would have to be re-relocated to correct these issues. For the present iteration of the PPMF, we find these inconsistencies very troubling.

There are large biases this process imposes on the occupancy rates. In Washington, occupancy rates at the census block level are on average 0.042 persons lower than they would be without DP. Occupancy rates range from 0 to 1 so this is a 4.2% bias towards lower occupancy. In Washington, there are 119,427 blocks with housing units. If the DP process were truly random, the difference in average occupancy rate should be near zero with such a large number of blocks. The standard deviation of the occupancy rate was 0.166 in the original 2010 Census data. It more than doubles in the PPMF data to 0.314.

When we combine the population and housing data derived statistics from the PPMF files show serious issues. To calculate the household size (PPH), the population in households is divided by the number of occupied housing units. At the block level, the average PPH is 0.482 higher than it was in the original 2010 data. This means there is nearly half a person more per household. As with the occupancy rate, if the effects of the DP process were random one would expect this value to be near zero. The standard deviation for household size was 0.840 in the PL data. In the latest PPMF file it is nearly four times higher at 3.333.

These biases are lessened as the size of geography increases, but they are still noticeable at the city level. City occupancy rates, on average, are 0.012 less than the original 2010 data. City PPH is 0.043 persons higher, and the standard deviation of PPH increases from 0.441 to 0.534.

The noise from the reassignment of vacancy status combined with the decoupling of household and population estimates has far reaching impacts. *This is of great concern because we, like many other states and local governments, rely on the housing unit method to create population estimates in the postcensal period.* The housing unit method uses administrative records to update housing unit counts in an area. It then applies occupancy rates and PPH ratios to those counts to estimate population. This method is extremely sensitive to error in those statistics. We estimated how different our 2020 city estimates would have been if the error in the latest PPMF had been present in the 2010 Census. For cities under 1,000 people, the difference averages 15.6%. One city under 500 people would have a population difference of 93.5%. These errors have real consequences given that cities in Washington receive annual funds based on population.

We discovered other problems in the latest version of the PPMF that defy logic and common sense. There are 8,436 blocks in Washington where the entire population of 94,444 persons was under the age of 18. While there are some living situations where households have no adults, in the original 2010 Census data there were only 11 such blocks with 186 people.

Race and Ethnicity

There appears to be a flaw in the DP process that is systematically decreasing racial diversity. Thirty one percent of Washington blocks saw a reduction in the number of races represented, while only 6% saw an increase. This phenomena continued up to the city and county levels with 47% of cities and 15% of counties having fewer races represented in their data. This is a serious issue for anyone trying to use census data to understand racial representation and diversity.

There is a strong correlation between the percent of people in an area that are of Hispanic origin and the average PPH of that area. This correlation has an impact on the way city PPH's are adjusted in our postcensal county and city population estimates. In the original 2010 Census data, a regression analysis of the percent Hispanic predicting PPH at the city level has an R-squared value of .600 and an effect of 1.865. It is a very strong predictor in a model. The impact of DP reduces the strength of this to an R-squared of 0.394 but increases the effect to 2.038. That change is particularly disconcerting, as an increase in PPH of that magnitude would have a large impact on sub county population estimates. The degradation

of this relationship at the city level is troubling, but at the block level, the relationship disappears. The block level relationship goes from having a R-squared value of 0.124 in the published data to a minuscule value of 0.002 in the latest PPMF. The relationship between percent Hispanic population and household size should be much stronger.

Redistricting

A complete and accurate Census is critical for a functioning democracy. This a primary reason conducting the census was in our constitution. Much attention has been given to the quality of the count at the state level, as this is how congressional house seats are apportioned. The accuracy of state population counts is of such import, that it has been designated as invariant in the DP process. However, very little attention has been given to the quality of the data needed for federal and state legislative redistricting efforts. Even less attention has been given to local voting districts or precincts. In Washington the redistricting commission responsible for delineating these boundaries is temporary and will not be formed until January of 2021. We feel it is important that this critical use case be brought to light during the PPMF review process and we offer the following information.

Washington designated 49 legislative districts in 2010. They averaged 137,235.5 people each and no district varied from that value by more than 50 people. This small variance was the result of careful crafting by the redistricting commission. Using the PPMF data the average population of the resulting districts does not change at all. However, the maximum variation from that average increases over fourfold to 228 people. Had this differential privacy process been in place in 2010, the commission would have defined the districts differently reflecting the erroneous values. A simple proxy for the demographic modeling used in redistricting is the percent of a district's adult population that is in a minority group (not non-Hispanic white). We looked at how this value was changed by the differential privacy process. We were pleased to find that these values were not modified in a way that would likely affect the redistricting process. The percent minority in any districts was different by no more than 0.81 percent.

Voting precincts are much smaller units than legislative districts. Washington had 6,696 precincts in 2012. Precincts are key component of all voting jurisdictions in the state. They are defined by county, city and town boundaries. They are further divided by county legislative authority, port commissioner, PUD commissioner, school, school director and fire districts. Delineating these boundaries is done at the start of the redistricting process. These boundaries are then allocated and divided into congressional and legislative districts. Detailed accurate block level population counts are critical to this process. The Washington State Guide to Redistricting (https://www.sos.wa.gov/_assets/elections/redistrictingguide.pdf) list several steps that require high accuracy. For example, one step is identifying existing precincts with no population so they can be incorporated into neighboring precincts without affecting any voters. This recent PPMF file would have flagged 22 precincts as not having voting age population when they in fact did. Conversely, the PPMF data has 2 precincts with voting population which in fact, had none.

Another step-in redistricting is identifying precincts that exceed the maximum number of registered voters. Washington law (RCW 29A.16.040) sets this maximum to 1,500 voters (although some did exceed this size), and counties can set it to a lower value. The error introduced by the most recent PPMF file would make this determination difficult. Precinct voting age populations were altered by as much as 11% by the DP process. Additionally, the PPMF changed 10 small precincts, with 84 people, to having no voting age adults.

The racial makeup of voting precincts also changes. This would very likely affect the political negotiations as to which precincts belong in which legislative or congressional districts. The PPMF bias towards the reduction of racial diversity noted in an earlier section is particularly problematic at the

precinct level. There were 4,512 precincts with fewer races represented after DP, while 318 showed an increase. Furthermore, 136 precincts switched from majority white non-Hispanic to a majority non-white or Hispanic. Only 73 precincts went the other way, switching from a minority majority to a majority white non-Hispanic. When looking at the voting age population by race this is phenomena is more pronounced, with 162 precincts changing to majority non-white or Hispanic with only 60 moving the other direction.

Washington law (RCW 44.05.140) also requires state agencies to supply the redistricting commission with identifiable information on incarcerated inmates, persons committed to receive involuntary behavioral health treatments, and persons residing or placed in juvenile justice facilities. This inmate information is reported as of April 1 of the census year. The commission is then to determine the last known address of these individuals and reallocate them by race into those precincts. In the past, this was possible because the population counts in the blocks containing these facilities were extremely precise and could be closely correlated with the administrative counts. This task would be difficult using the recent PPMF.

There is a bias with the group quarter allocation (see Figure 4). Blocks with group quarters and an average group quarter size less than 10 (in the original 2010 data) have on average 1.83 more people per group quarter due to DP. Blocks with group quarters from 10 to 500 people in SF1 have on average 3.71 fewer people per group quarter in the PPMF data. The error introduced in the 10 to 500 range is from -65 to +144 with a standard deviation of 19.2. Error this large is likely to present a challenge to reallocation. This table shows the effect DP had on the blocks containing correctional facilities in Washington state. The table includes the black population as an example of how the issue affects an individual race group. Additional race group, which are not shown in the table, are impacted as well.

	2010 PL			2010 PPMF			Error Introduced by DP		
Block	Group Quarters Population	Under 18	Black	Group Quarters Population	Under 18	Black	Group Quarters Population	Under 18	Black
530090002001051	896	0	244	911	3	252	15	3	8
530110405041069	219	0	39	223	1	35	4	1	-4
530210208003025	1,445	0	266	1,457	0	266	12	0	0
530270016002013	1,967	5	372	2,004	0	340	37	-5	-32
530319507021027	377	0	72	335	0	0	-42	0	-72
530459604001012	178	17	20	142	0	2	-36	-17	-18
530459606001008	1,673	3	315	1,671	1	293	-2	-2	-22
530530725042008	827	0	144	838	0	143	11	0	-1
530610522091002	2,463	2	489	2,459	14	466	-4	12	-23
530630104011011	2,184	0	320	2,216	0	309	32	0	-11
530719204001001	2,300	4	552	2,264	0	594	-36	-4	42
530330326023045	126	120	25	90	89	24	-36	-31	-1
530419710001027	193	103	60	168	90	0	-25	-13	-60
530499504002046	98	94	17	54	74	0	-44	-20	-17
530670127205045	203	156	35	181	167	0	-22	11	-35

The noise infused into the census data will make Washington's redistricting process more difficult and potentially less accurate. It is unlikely the commission will be able to follow the intent of the law in their relocation of selected incarcerates unless this issue is addressed. We believe other states will face similar challenges with the data.

Conclusion

Thank you again for the opportunity to review this data. As stated earlier we believe that making occupancy status invariant, limiting the distribution of household population to blocks with occupied housing units, forcing each occupied unit to have at least one person and limiting the frequency of outlier household sizes would go a long way towards improving the validity of the data. Even if those changes are made we would still have serious concerns about changes in the race and ethnicity distributions and the accuracy of the voting age population data. We urge you to reduce the overall amount of noise infused in the data and increase the accuracy of the age and race data as it is crucial for accurate redistricting, health analysis, planning, fund distribution and many other applications. We urge you to take your time and fully engage data users in the review process by providing more detailed PPMF data sets with varying privacy loss budgets. It is extremely important to address these issues, as the use of this data will have serious and long lasting impacts on Washington's people.

Sincerely,

Mike Mohrman, State Demographer
Forecasting and Research Division, Office of Financial Management

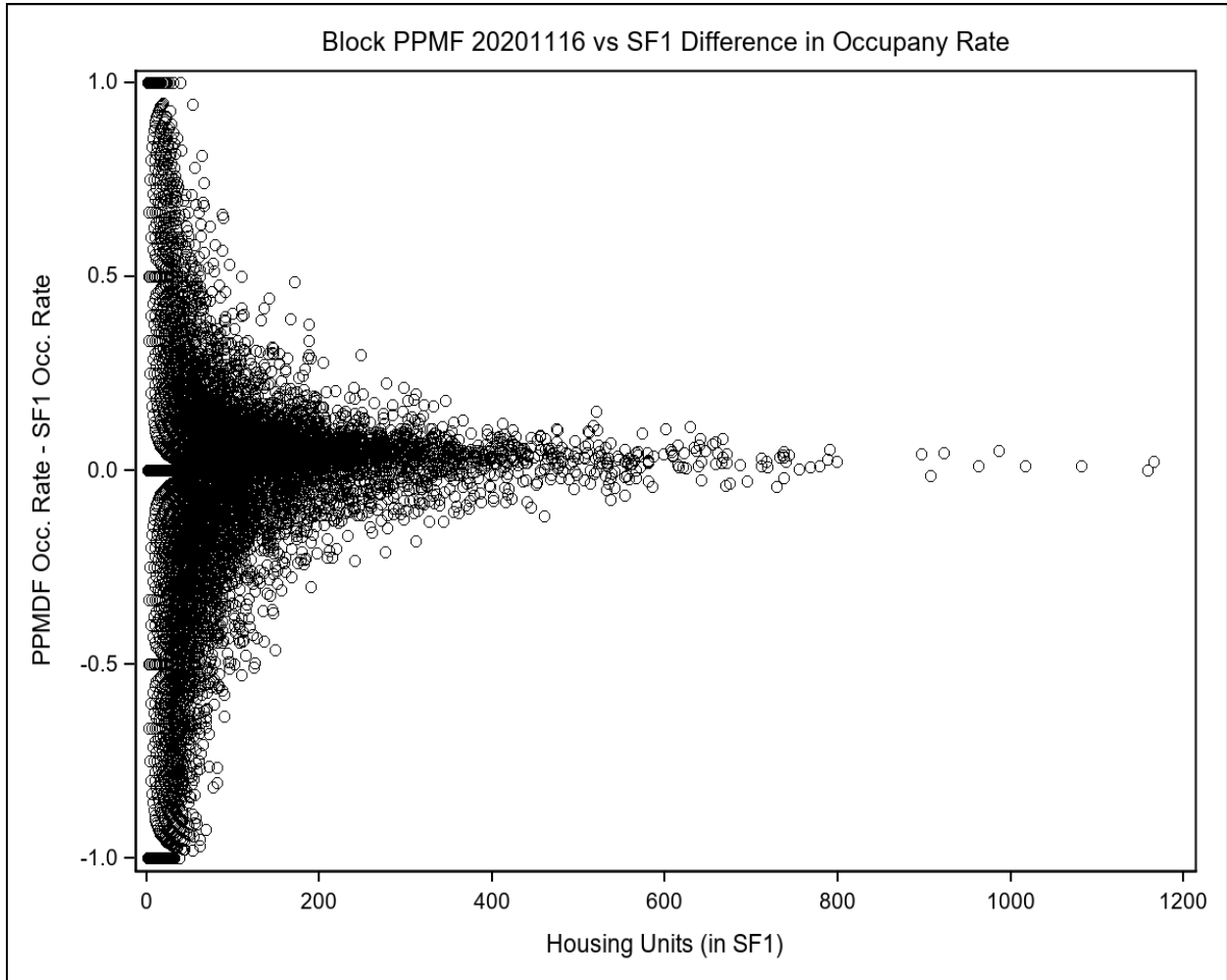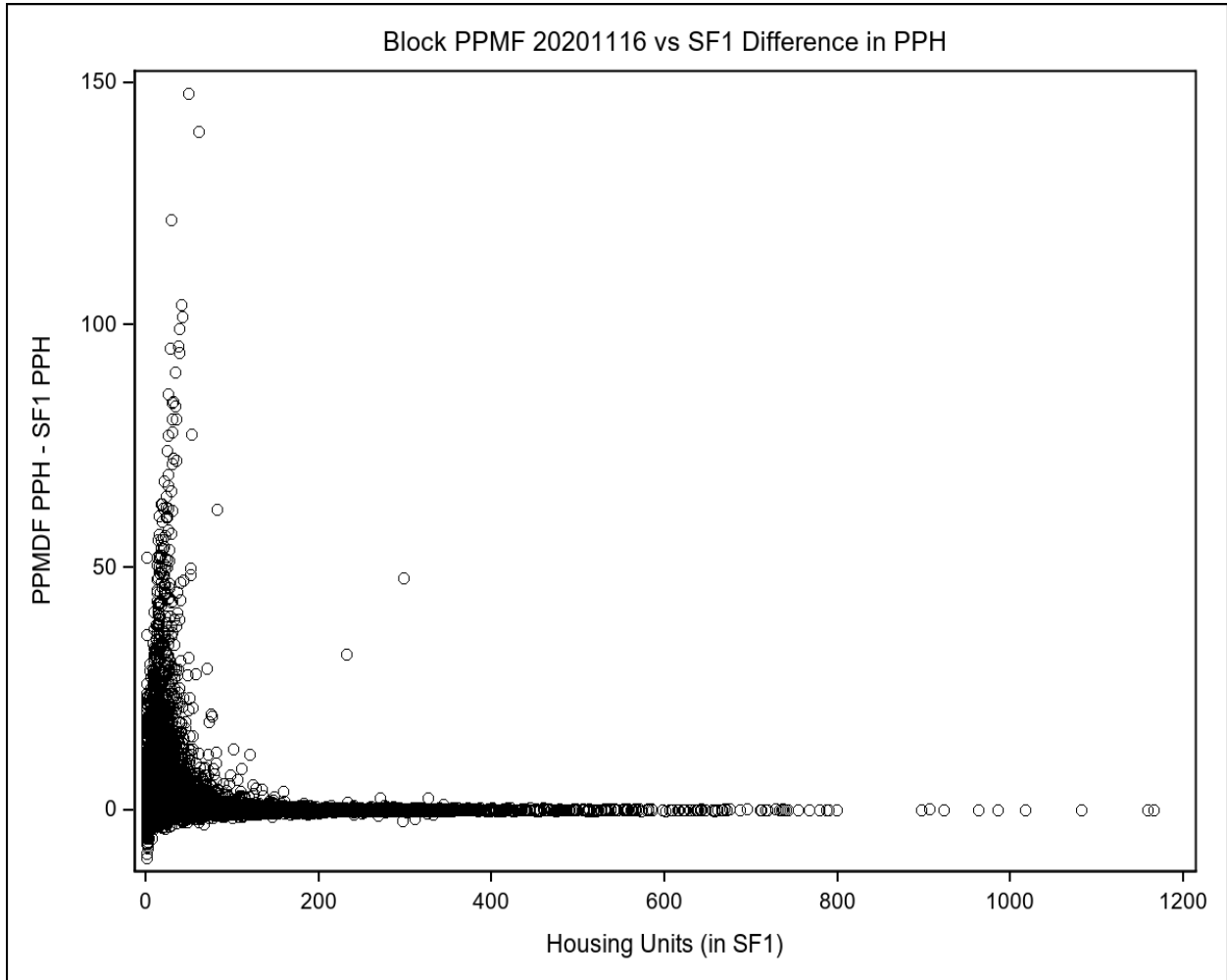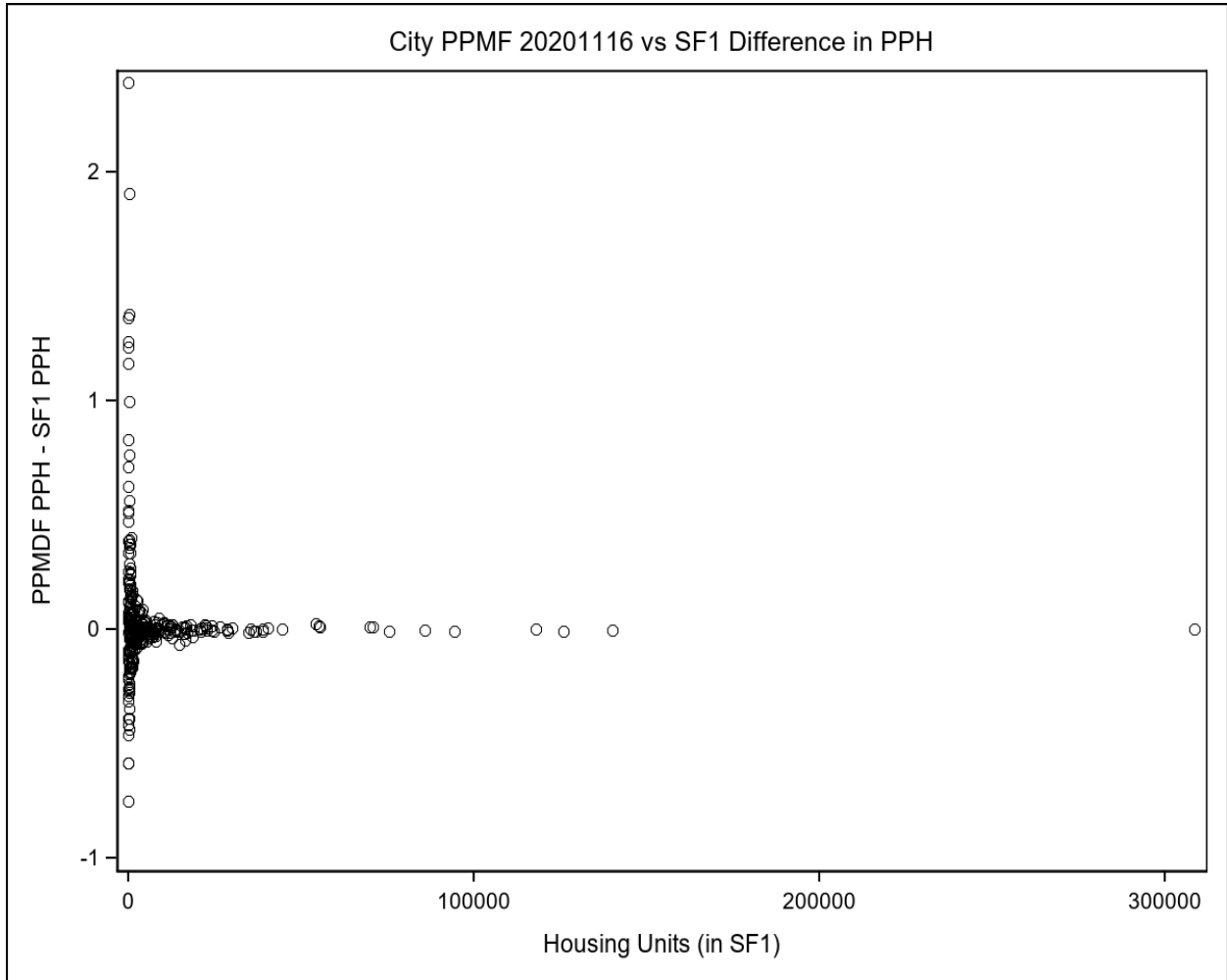Cc: Marc Baldwin, Assistant Director

Appendix

Figure 1

Block PPMF 20201116 vs SF1 Difference in Occupany Rate

Figure 2

Block PPMF 20201116 vs SF1 Difference in PPH

Figure 3

City PPMF 20201116 vs SF1 Difference in PPH

Figure 4

Relationship of Block GQ size to DP error in GQ Pop