# EXHIBIT B

Declaration of Stacy Woods

**IN THE UNITED STATES DISTRICT COURT**
**FOR THE EASTERN DISTRICT OF MICHIGAN**
**SOUTHERN DIVISION**

CONCERNED PASTORS FOR SOCIAL ACTION, et al.,

Plaintiffs,

v.

NICK A. KHOURI, et al.,

Defendants.
_______________________________________/

Case No. 16-10277

Hon. David M. Lawson

Mag. J. Stephanie Dawkins Davis

**DECLARATION OF STACY WOODS, Ph.D., M.P.H.**

I, Stacy Woods, declare as follows:

**QUALIFICATIONS**

1. I am a data scientist, with focused expertise in data visualization and spatial and non-spatial data analysis (statistics).

2. I am presently employed as a scientist by the Natural Resources Defense Council, Inc. (NRDC), a nonprofit health and environmental organization, and a plaintiff in this litigation. My work for NRDC focuses on data analysis and data visualization, including data mapping. My day-to-day work includes locating and analyzing (and if appropriate mapping) data, including data gathered by NRDC and data obtained from other sources. In my current position, I have applied advanced statistical methods and models as appropriate to data and research questions to support scientific research, policy analysis, and litigation at NRDC.

3. I received my Ph.D. in Environmental Health from the Johns Hopkins Bloomberg School of Public Health (JHSPH). My doctoral dissertation utilized spatial statistics and GIS (geographic information systems) to assess and visualize changes in air quality and the influence of federal and state regulations on air quality over time. During my graduate training, I taught Master's and doctoral students as the lead teaching assistant for the spatial statistics and GIS courses at JHSPH for several years. I also received a Master of Public Health degree from JHSPH. My Master's thesis utilized statistical cluster analysis and GIS to identify and visualize hotspots of Lyme disease transmission in Howard County, Maryland. I received my Bachelor of Science degree in Entomology and Nematology from the University of Florida.

4. Before joining NRDC, I worked as an epidemiologist in the Health Effects Division, Office of Pesticide Programs, U.S. Environmental Protection Agency. Before that, I held positions as a Mirzayan Science and Technology Fellow at the National Academies of Science, Engineering, and Medicine; as a Brown Scholar in Community Health at Johns Hopkins; and as a Senior Research Assistant with the Johns Hopkins Center to Reduce Cancer Disparities. In these positions and during my graduate school training, I used statistics to analyze data, and the results of my analyses appear in a number of reports and published papers.

5. My *curriculum vitae* is attached as Exhibit 1 to this Declaration.

6. By virtue of my training, research, prior experience, and knowledge of pertinent scientific literature, I consider myself an expert on spatial and non-spatial data analysis (statistics).

## ANALYSIS AND CONCLUSIONS

7. I was asked by counsel for NRDC to evaluate the data and information provided by the City of Flint in support of the conclusions in its report submitted pursuant to Paragraph 30 of the Settlement Agreement in this case. *See* Chaudhary Decl.[1] Ex. 8 [hereinafter, Paragraph 30 Report]. Attached as Exhibit 2 to this Declaration is a list of the data and other documents from the City that I reviewed. All of the information in this Declaration is based upon my education, personal knowledge, and experience, as well as my personal review of the documents listed in Exhibit 2.

8. The City's Paragraph 30 Report concluded that: "It is not reasonably likely that there were more than 18,000 lead or galvanized steel service lines at replacement eligible households" in Flint as of March 28, 2017. Paragraph 30 Report at 1; *see* Settlement Agmt. ¶ 29, ECF No. 147-1.

9. As used by the City, I understand the term "hit rate" to mean the proportion of excavations conducted in a particular area that uncover lead or

---

[1] Declaration of Dimple Chaudhary (dated June 20, 2018), submitted in support of Plaintiffs' Motion to Enforce Paragraphs 29 and 30 of Settlement Agreement.

galvanized steel service lines. For excavations covered by the Agreement, I understand that the City must replace those service lines that the City confirms are lead or galvanized steel. In other words, if the City discovers lead or galvanized steel service lines at 10 out of 100 addresses where it conducts excavations, the hit rate would equal 10%, indicating that 10% of those 100 households require service line replacement.

<u>The City's predicted hit rate (80%) and observed hit rate (70%) in 2017</u>

10. In the document with file name, "2018.03.06 Followup Summary.pdf," the City states that "[i]nitial estimates" predicted that lead or galvanized steel service lines would be identified at "80% of the addresses" in the areas slated for excavations during Phase IV.[2] Chaudhary Decl. Ex. 13 at 2 [hereinafter, Follow-Up Summary]. The City also states that it confirmed a lead or galvanized steel service line at 70% of the homes at which it conducted excavations during 2017. *Id.* Accordingly, the City reports that its observed hit rate for 2017 (roughly 70%) was lower than its "projected" hit rate (80%). *Id.*

11. Since the observed hit rate for work done in 2017 was lower than the expected (predicted or projected) hit rate, the City extrapolates that "the original

---

[2] I understand that "Phase IV" excavations occurred roughly during 2017, and will use "2017 excavations" and "Phase IV excavations" interchangeably, for simplicity. *See* Paragraph 30 Report at 1 n.2 ("Phase IV [service line replacement] work occurred in 2017 after execution of the [Settlement Agreement].").

18,000 ceiling was likely an overestimation." *Id.* However, as explained below in Paragraphs 12 through 22, this conclusion does not follow from the premise.

12. Prior to the start of the 2017 excavations, it appears that the City assumed that certain factors (explanatory variables), such as "City records of service line composition" and "age of local water-related infrastructure," *see id.*, were appropriate to use in predicting whether a given home would have a lead or galvanized steel service line. Those assumptions led the City to predict that it would find a lead or galvanized steel service line at 80% of the homes where it conducted excavations in 2017.

13. Plaintiffs have asked the City to detail the factors (variables) that informed the initial 80% projected hit rate, including during a conference call on April 6, 2018, in which I participated. In response, the City could not identify who generated the 80% projection or what process he or she used to generate it. Furthermore, the City has not provided a complete and precise list of the data, records, and other factors it considered in constructing the 80% projection, or specific information regarding how these factors informed the 80% projection.

14. For example, the City states that its predictions were "generated by consultants based on the review of available City records of service line composition, age of local water-related infrastructure (fire hydrants/water mains), and other available data etc." *Id.* This list is neither complete nor precise (per the

use of "etc."), nor does it provide any information regarding how these factors were used in calculating the 80% projection (e.g., how close to a home did a fire hydrant have to be for the hydrant's age to be considered a relevant factor in predicting whether that home had a lead or galvanized steel service line?).

15. Because the City has not identified all of the explanatory variables that informed the 80% prediction, it has not provided—and could not provide—any assessment of the explanatory variables based on the outcome of the excavations done to date. That is, the City now has data about the observed outcomes of the 2017 excavations. However, based on the City's responses to Plaintiffs' inquiries concerning its Paragraph 30 Report, I understand that the City has not *used* this new data to assess how the variables that informed the 80% projection (the explanatory variables) actually performed in predicting the presence of lead or galvanized steel service lines. In other words, the City has not used this new data to assess whether the explanatory variables it relied on in making the 80% prediction were in fact appropriate predictors of the presence of lead or galvanized steel service lines in 2017, and thus whether these factors are appropriate to use prospectively to estimate the hit rate in areas where the City will conduct excavations in 2018 and 2019.

16. A hypothetical example illustrates this issue: Consider a scenario in which the City, prior to undertaking any excavations, hypothesized that red brick

buildings more frequently had lead or galvanized steel service lines than buildings made of other materials. In this hypothetical scenario, assume that red brick was a popular building construction material around the same time that lead or galvanized steel was commonly used in service lines. Thus, the City used the presence of red brick buildings as an explanatory variable in predicting hit rates, so that areas with more red brick buildings were predicted to have higher hit rates than areas where buildings were mostly constructed with other materials. Assume that the areas slated for excavation in 2017 had more red brick buildings than the areas slated for excavation in 2018 and 2019. The City's initial model thus predicted a higher hit rate in 2017 than in 2018 and 2019.

17. Now assume that the truth in the context of this hypothetical scenario is that the number of red brick buildings does not, in fact, help predict the number of lead or galvanized steel service lines in a given area. Therefore, a prediction model that projected a lower hit rate in 2018 and 2019 compared to 2017 based, in part or in whole, on the lower number of red brick buildings in the 2018 and 2019 areas compared to the areas excavated in 2017, is unsound. In this scenario, the City could use the outcomes from 2017 to assess the appropriateness of using red brick buildings as an explanatory variable. Once the City found that the presence of red brick buildings did not, in fact, help predict the presence of lead or galvanized steel service lines, the City should then refine the prediction model so

that the presence of red brick buildings does not inform the predicted hit rates in 2018 and 2019.

18. The statistical techniques to assess the appropriateness of an explanatory variable start with an initial prediction model (in the context of the City's Paragraph 30 Report, the considerations that informed the City's original 80% projection for 2017). As service lines are excavated and their compositions are identified, a data file can be built that contains the outcomes (the service line composition discovered) and the values for the explanatory variables for each house (e.g., the age of the fire hydrant closest to the house, the age of the house). The initial prediction model is then re-run using this newly acquired data. The model output will include an assessment of the appropriateness of each explanatory variable given the outcome data. This output is critical, because it provides updated information (based on observed data) about whether the factors the City expected to be reliable predictors of the presence of lead and galvanized steel service lines are actually dependable predictors.

19. For example, one could use regression analysis to assess if and how the explanatory variables were actually related to the observed outcomes of the City's excavations conducted in 2017. In the regression technique, a statistical computer program (e.g., SAS) is used to run the predictive model with the information obtained in 2017. The results are numeric indicators, such as the

regression coefficient,[3] standard error,[4] and p-value,[5] that express the strength, direction, and appropriateness of the relationship between each explanatory variable and the outcome.

20. The City's support for its declaration that "the original 18,000 ceiling was likely an overestimation" seems to rest on the lower-than-expected hit rate for work done in 2017. *See* Follow-Up Summary. Therefore, the City reckons that the reduction from 80% estimated replacements in 2017 to 70% actual replacements in 2017 will reduce the overall number of replacements to less than 18,000. But without any assessment of what drove the discrepancy between the predicted and observed hit rates for work done in 2017, it is not possible to infer how the original predictions for areas to be excavated in 2018 and 2019 will actually perform, as

[3] A regression coefficient is a number that describes the direction (positive or negative) and relative magnitude of the relationship between the explanatory variable and the outcome.

[4] The standard error is a measure of variation that describes the difference between the observed and predicted outcomes. *See* Beth Saunders-Dawson & Robert G. Trapp, Basic and Clinical Biostatistics (2d ed. 1994).

[5] The p-value is a measure of statistical significance of the explanatory variable that can inform whether a specific explanatory variable should remain in the prediction model. The p-value describes the probability that the model outcome could be due to random chance. If the p-value is a small number (e.g., less than 0.05), then the probability that the model output (the observed relationship between the explanatory variable and the outcome) was due to chance is low (e.g., less than 5%), and thus the explanatory variable may be retained in a data-refined prediction model. National Research Council, *Reference Manual on Scientific Evidence* 249-51 (3d ed. 2011), *available at* https://www.nap.edu/read/13163/chapter/7#249.

described in Paragraphs 15 through 17. Specifically, without an assessment of why the 2017 hit rate was overestimated, it is not possible to use the fact of that overestimation to surmise if the original predictions for 2018 and 2019 will likely correctly, over-, or under-estimate the number of lead or galvanized steel service lines.

21. A hypothetical example illustrates this issue: Consider a scenario in which, prior to any excavations, the predicted total hits (lead or galvanized steel pipes) were as follows: 9,000 to be identified in 2017, 6,000 in 2018, and 3,000 in 2019, for an overall prediction of 18,000 lead or galvanized steel pipes. Assume that at the end of 2017, 7,800 lead or galvanized steel lines were identified, far fewer than the 9,000 predicted. However, in 2018, the City finds 7,000 lead or galvanized steel lines, and in 2019, it finds 3,500 lead or galvanized steel lines. In this scenario, the overestimation of hits for excavations in 2017 (from the predicted 9,000 to the observed 7,800) did not mean that the overall prediction of 18,000 was an overestimation. In fact, the underestimation of hits for excavations in 2018 and 2019 (from the predicted 6,000 and 3,000 to the observed 7,000 and 3,500 for 2018 and 2019, respectively) meant that the overall prediction of 18,000 was an *under*estimation; the actual number of lead or galvanized steel service lines identified by the end of 2019 in this example was 18,300. To reiterate, without any assessment of how the observed outcomes compared to the predicted outcomes in

2017 and whether the predictions for 2018 and 2019 should be recalculated in light of the information gathered in 2017, the City could not surmise how the 2017 overestimation would influence the original overall 18,000 hit prediction.

22. Thus and in sum, as outlined in Paragraphs 12 through 21, the fact that there was a lower-than-predicted hit rate observed in 2017 does not support the City's conclusion that, as of March 28, 2017, "[i]t is not reasonably likely that there were more than 18,000 lead or galvanized steel service lines at replacement eligible households," as that term is defined in the Settlement Agreement. *See* Paragraph 30 Report at 1; Settlement Agmt. ¶¶ 11, 12, ECF No. 147-1.

<u>Analysis of projected hit rates of 60% for 2018 and 50% for 2019</u>

23. The City projects that its hit rate will be 60% for work conducted in 2018 and 50% for work conducted in 2019. Paragraph 30 Report at 2. After reviewing the City's Paragraph 30 Report and the other documents listed in Exhibit 2 to this Declaration, I cannot discern what calculations the City used to generate these specific projections.

24. Plaintiffs have asked the City to detail the factors (variables) that informed the projected hit rates (60% for work conducted in 2018 and 50% for work conducted in 2019). Specifically, during a phone call with the City and representatives from AECOM (the City's project management firm) on April 6, 2018, I asked about the basis for the 60% and 50% projected hit rates, and the City

confirmed that the process for developing those projections was "not statistical." *See* Chaudhary Decl. ¶ 19.

25. As explained below in Paragraphs 26 through 51, after reviewing the documents provided by the City in support of their hit rate projections, I conclude that there is no support for the specific projections that the hit rate will be 60% in 2018 and 50% in 2019.

*Map labeled, "Occupied Housing with Lead, Galvanized, and Unknown Connections"*

26. The 2016 map attached as Exhibit 14 to the Chaudhary Declaration appears to show, through solid-colored block-like shapes, the locations (parcels) of a number of service lines in the City. The service lines are color-coded to show compositions of lead (red), galvanized steel (blue), or unknown (green). The map does not purport to show the locations of copper service lines in the City.

27. The map has very little supporting information describing the data it represents. For example, the map does not indicate what data informed the locations and service line compositions of the mapped parcels (color blocks). That is, it is not clear if the service line composition for each parcel shown on the map was inferred from historical city records, explanatory (predictor) variables such as the age of nearby water-related infrastructure, or some other method or

combination of methods.[6] Furthermore, the map provides no details about how many of the total number of occupied households[7] in Flint were included in the data represented on the map (i.e., the map excludes households with copper service lines but does not explicitly state that fact, and it is unclear if the map also excludes other categories of occupied households).

28. Additionally, the map makes no predictions about the service line composition for unknown (green) parcels. As a result, a map reader cannot make informed quantitative predictions of the relative hit rate for different areas of the City based on the solid colored parcels on the map. That is, while one may assume that areas with a high concentration of red or blue parcels will require service line replacement (because red and blue parcels represent households with lead and galvanized steel service lines), one can make no assumptions whatsoever regarding areas with a high concentration of green parcels (those with unknown service line compositions). Therefore, one cannot make comparisons between areas with many red/blue parcels and areas with many green parcels.

---

[6] The map also does not provide an assessment of the relative accuracy of the data that was used to generate the mapped parcels. This is important because, for example, historical records of service line composition in Flint have been shown to be inaccurate in previous analyses. *See* Chaudhary Decl. Ex. 5.

[7] For simplicity, I use the phrase "occupied households" to refer to "replacement eligible households," as that term is defined in the Settlement Agreement in this case. *See* Settlement Agmt. ¶¶ 11, 12.

29. Underlying the colored blocks (parcels) is a shading scheme that appears, based on the map's title and legend, to illustrate the density of parcels for occupied households in Flint for which the map maker[8] had information regarding service line composition (lead, galvanized steel, or unknown). This density shading (density map), appears to have been derived from the relative count of colored blocks (parcels) in a given area. The density shading is shown on a continuous scale, so that areas with few parcels (low information density) are shaded green and areas with many parcels (high information density) are shaded red. There is no indication of the numerical cut points for each category of density. For example, it is unclear how many parcels in a given area constitutes a "high" information density (red shading).

30. First, it is important to note that the density map does not illustrate the total number of occupied houses in Flint, and thus does not illustrate the overall occupied housing density across Flint. This is because an illustration of the density of all occupied housing would need to include households with copper service lines and households with lead, galvanized steel, or unknown service lines. This map, however, does not include homes with copper service lines (Paragraph 26); it counts only households with lead, galvanized steel, or unknown service lines. Thus, the mapped density is calculated from a subset of all Flint households.

[8] University of Michigan Flint Geographic Information Systems Center.

31. The ability to assess the total number of houses (i.e., houses with copper, lead, galvanized steel, and unknown service lines) is an essential element of calculating the hit rate for a given area, since the hit rate is equal to the number of houses with lead or galvanized steel divided by the total number of houses. Thus, one cannot make any reasonable inference about the hit rates for a given area based on the density map (shading), regardless of the density shade (e.g., red versus green shading).

32. Second, as a map reader, I cannot make any conclusions about how many lead or galvanized steel service lines are in a given area based on the density shading of that area on the map. This is because this map does not predict the density of specific service line connection types (e.g., the density of lead or galvanized steel service lines) across the City. Rather, it only shows the density of information across the City, and that information is limited to lead, galvanized steel, or unknown service lines (i.e., the information considered excludes copper service lines). This is illustrated in the examples below (Paragraphs 34 through 37).

33. The hypothetical and alternative scenarios in Paragraphs 34 through 37 illustrate several limitations of the map. Specifically, the two scenarios illustrate that (1) the map reader can make no assumptions regarding the presence of lead or galvanized steel service lines based on the density shading alone, (2) regarding the

solid colored parcels, whether the map reader can draw conclusions about the number of lead or galvanized steel service lines in areas with red and blue parcels (but no green parcels) depends on the quality and completeness of the parcel data, which is unknown to the map reader, and (3) overall, under no circumstances can the map reader draw any conclusions about the hit rate in any areas in Flint based on the information presented in the map.

34. Consider a hypothetical scenario in which the parcels (colored boxes) on the map accurately and precisely capture all lead and galvanized steel service lines in the City, and all other non-copper service lines are included as unknown service lines in the parcel data. In other words, the red, blue, and green boxes show every non-copper service line in Flint—the parcel data is complete. That is, assume that the number of occupied households represented on the map is exactly equal to the number of occupied households in Flint with non-copper service lines, so that the only information missing from the map is the copper service line households. In this scenario, consider a green shaded area with two red, one blue, and no green colored boxes. The area is shaded green because there is service line information for only three households in the area (i.e., the area has low information density). The map reader cannot assess if there are in fact more than three households in the area; e.g., there may be 150 houses in the area with copper service lines, which were excluded from the map. Thus, little is known about the service line

composition for all households in the area. But the map reader does know that, regardless of how many total households are actually in the area, there are *exactly* two households with lead and one household with galvanized steel service lines (and thus exactly three required service line replacements in this area).

35. In this scenario, even though the map reader knows that there are exactly three required service line replacements in this area, the reader cannot predict the hit rate, because the reader does not know the total number of households in the area.[9]

36. Consider an alternate hypothetical scenario in which the parcel data is not complete; that is, assume the parcels (colored boxes) on the map do not accurately and precisely show all lead, galvanized steel, and unknown service lines in the City. In this alternate scenario, a map reader can infer little about a green shaded area with two red, one blue, and no green boxes. As with the prior hypothetical scenario, there is information about only three households in the area, so it is shaded green indicating low information density. And again, the two red and one blue box mean that there are three households that will require

[9] Calculating the hit rate requires knowing the total number of households in the area, including households with copper service lines. If there were 100 total households in the area, the hit rate in this example would be 3% (3 out of 100 households had lead or galvanized steel service lines). If there were 10 total households in the area, the hit rate would be 30% (3 out of 10 households had lead or galvanized steel service lines).

replacement in the green shaded area. However, contrary to the prior hypothetical scenario (Paragraphs 34-35), in this alternate scenario, the reader cannot assume that all households with lead or galvanized steel service lines are represented on the map. Therefore, the reader cannot determine if there are exactly three or more than three households requiring replacement in the green shaded area; e.g., there may be 150 houses in the area with lead or galvanized steel service lines that were not captured in the underlying parcel data and thus were excluded from the map. Thus, the map reader cannot determine whether the green shaded area under consideration with two red, one blue, and no green boxes will require as few as three or many more service line replacements (the reader has no idea how many more) under this alternate scenario in which the underlying parcel data do not accurately and precisely capture all lead and galvanized steel lines in the City.

37. Again, in this alternate hypothetical scenario, the map reader cannot make any reasonable predictions about the hit rate, because again, the map reader does not know how many total households are in the area, an essential element to calculating the hit rate. Furthermore, in this scenario, the map reader also does not know how many households in the area have lead or galvanized steel service lines (due to the incompleteness of the parcel data).

38. It is impossible for the map reader to assess whether the hypothetical scenario in Paragraph 34 or the alternate hypothetical scenario in Paragraph 36 is

appropriate for the City's map (Chaudhary Decl. Ex. 14), because the map includes no discussion of its source data, and no discussion of the data limitations, as described in Paragraph 27 & n.6. Thus, the map reader cannot determine if the map accurately and precisely captures all lead and galvanized steel service lines in the City, as posited in the hypothetical scenario in Paragraph 34, or if it does not, as posited in the alternate hypothetical scenario in Paragraph 36.

39. Regardless of which scenario applies, as a map reader, I cannot draw any conclusions about the likely hit rates without information regarding the total number of households (including households with copper service lines) in each area of the map. In particular, I cannot draw any conclusions about whether the hit rates will be higher in the red shaded areas than the green shaded areas.

40. In short, the map lacks information regarding the source data for service line composition, contains no assessment as to the accuracy of this source data, contains no information regarding copper service lines, does not represent all occupied households in Flint, and does not attempt to predict the density of specific service line connection types (e.g., the density of lead or galvanized steel service lines) across the City. As a result, this map does not provide support for the City's specific predictions regarding the hit rate in areas slated for work in 2018 and 2019, nor does it provide support for the City's inference that the original 18,000 figure was likely an overestimate for the total number of lead and

galvanized steel service lines in Flint.

*Hydrant Dataset and Map*

41. I understand that on June 8, 2018, the City provided Plaintiffs with an Excel spreadsheet containing a list of fire hydrants in Flint, including information about their "date of fabrication." Chaudhary Decl. Ex. 23. The City also provided a map of hydrant locations, color-coded by date of installation[10] (red and orange dots indicate hydrants installed prior to 1950, green dots indicate hydrants installed after 1970). *Id.* Ex. 23-2. From these two documents, I do not know whether all of the hydrants listed in the spreadsheet are included in the map, or how the map deals with the 101 entries in the spreadsheet with a "FabDate" of "0." The hydrant map also does not overlay the hydrant ages on the areas where the City conducted excavations in 2017 or the areas where the City intends to conduct excavations in 2018 or 2019.

42. These documents show that data about the age of hydrants in Flint could be used by the City to inform its projected hit rates. However, the data and

[10] The hydrant map legend specifies that the information illustrated on the map is the date of hydrant installation. *See id.* Ex. 23-2. In the spreadsheet, the column labeled "InstallDat" (presumably, installation date) is empty for every hydrant, but the column labeled "FabDate" (presumably, fabrication date) is populated with a year (although for 102 hydrants, the year fabricated is reported as "0" or "202"). The City provided no definitions for "installation date" and "fabrication date" and no information regarding how these terms are related. For simplicity, I have assumed these terms to be interchangeable for the purposes of this Declaration.

documents provided to me do not detail *how* this information was used to inform the City's estimates. For example, the spreadsheet and map do not show how the hydrant installation dates in the areas where the City conducted excavations in 2017 compare to the hydrant installation dates in the areas where the City plans to conduct excavations in 2018 and 2019. Nor do the documents answer the question of the spatial extent of a hydrant's predicted influence: that is, how close must a fire hydrant be to a given home for its installation date to be relevant to the probability that the service line for the home will be made of lead or galvanized steel? For example, is the presence of an "old" fire hydrant within 1 mile of a home a strong indicator that the home will require service line replacement? How about the presence of an "old" fire hydrant within 5 or 10 miles of a home?

43. In other words, although the City states that the hydrant data was "a major element used in assessing the age of the City's water distribution infrastructure," Chaudhary Decl. Ex. 23, I cannot discern precisely how the City applied the hydrant data in its analysis in the Paragraph 30 Report. Thus, the hydrant data—absent further analysis or explanation—does not provide support for the City's projected hit rates of 60% and 50% for work slated to be done in 2018 and 2019, respectively, nor does the hydrant data provide support for the City's conclusion that the original 18,000 figure was likely an overestimate for the total number of lead and galvanized steel service lines in Flint.

*Excavation Results from April 26, 2018, through June 4, 2018*

44. I understand that on June 8, 2018, the City provided Plaintiffs with an Excel spreadsheet that the City states "covers all hydrovacs [sic] excavations conducted in 2018, through June 4, 2018." *Id*. Ex. 23; *see id.* Ex. 23-1. This Excel spreadsheet was provided in support of the 60% predicted hit rate for work to be done in 2018 and further in support of the City's inference that it is "not reasonable to conclude that there were more than 18,000 lead or galvanized steel service lines in the City." *Id.* Ex. 23. However, the data in this spreadsheet fails to provide any support for either of these two conclusions.

45. The excavation spreadsheet contains the following information for the 954 excavations already completed in 2018: Parcel (household) address, unique identification number, public- and private-side service line composition(s), date of excavation, and whether the parcel had copper service lines on both the public and private lines and therefore did not need replacement. *See id.* Ex. 23-1. The information in the spreadsheet may be used to calculate the hit rate for this subset of 2018 parcels (37.2%). However, it is inappropriate to make predictions regarding the hit rate for the remaining parcels slated for excavation in 2018 based on this limited data.

46. There is no information in the excavation spreadsheet that compares

the 954 households already excavated to the approximately 5,046[11] parcels remaining to be excavated in 2018. Specifically, there is no information regarding the explanatory variables (factors that effectively predict whether or not a parcel will require service line replacement) for these 954 parcels, or how the explanatory variables compare between the 954 parcels already excavated and the 5,046 parcels remaining to be excavated in 2018. For example, if historical service line composition records are a sound predictor of the presence of lead or galvanized steel service lines as suggested by the City, *see* Follow-Up Summary, how do these records compare quantitatively between the 954 parcels excavated thus far in 2018 and the remaining 5,046 parcels to be excavated in 2018? In essence, to assess how the hit rate for this subset of 2018 parcels will effectively predict the hit rate for all work in 2018, the City must first assess whether the parcels in this subset are adequately representative of all parcels slated for work in 2018 with respect to important factors that predict the presence of lead or galvanized steel service lines.

47. Consider, for example, applying the City's logic—that the spreadsheet for the subset of work done to date in 2018 can help predict the overall 2018 hit rate without any additional information or analysis—to the work done through 2017 (Phases I – IV). Recall that, prior to the work being completed in

[11] I estimated that there were 5,046 parcels remaining to be excavated in 2018 based on the 6,000 excavations slated for Phase V (2018), *see* Paragraph 30 Report at 5, minus the 954 excavations done to date as reflected in the Excel spreadsheet.

2017, the predicted hit rate was 80%. Per the City's Report, we know that, in reality, the 8,843 excavations completed through 2017 were found to have a cumulative hit rate of roughly 70%, with 6,256 lead or galvanized steel service lines identified and replaced. *See* Paragraph 30 Report at 2.

48. Consider this hypothetical scenario: In June 2017, the City was challenged to support its predicted hit rate of 80% for all of the work to be done in 2017 based on the work done to date in June 2017. In this scenario, by June 2017, the City had excavated 954 parcels and found that 599 of them did not have lead or galvanized steel service lines, while 355 parcels did, for a hit rate of 37.2%. Thus, in this scenario, the work the City completed through June of 2017 *by chance* selected 355 of the total 6,256 parcels with lead or galvanized steel service lines ultimately discovered during the work completed through the end of 2017. If we consider the subset of work completed by June 2017 to be a sample of the total work completed through the end of 2017, we see that by chance the subset oversampled houses that did not need replacement and undersampled houses with lead or galvanized steel service lines. Thus, the hit rate for this subset of parcels (37.2%) does not provide an appropriate indication of the hit rate for all parcels slated for work through 2017 (70%). It would thus be erroneous to use the hit rate of this subset to refine the predicted hit rate for work yet to be done in 2017 without first assessing if the parcels in this subset were adequately representative

of all parcels slated for work in 2017. Yet this is precisely the logic the City is relying on with respect to the 2018 predictions: the City is basing its overall 2018 hit rate predictions on a subset of 954 parcels excavated to date in 2018 without providing any information about how well those 954 parcels represent the entire set of parcels to be excavated in 2018. Regardless of the actual overall hit rate for all parcels excavated in 2018, the subset of 954 parcels excavated to date could have produced a hit rate of 37.2% based on the chance selection of parcels that mostly do not require service line replacement.

49. Because there is no information tying the results of the excavations conducted from April 26 through June 4, 2018, to the explanatory (predictive) variables the City asserts can effectively predict the hit rate for a given area, the observed hit rate for work done through June 4, 2018 (37.2%) does not provide support for the City's projected hit rate of 60% for all work to be done in 2018.

50. Furthermore, the City has not posited that the data contained in the excavation spreadsheet (Chaudhary Decl. Ex. 23-1) provides any information or support regarding the City's projected hit rate of 50% for work slated to be done in 2019. Finally, because the excavation spreadsheet provides no support for the expected number of service lines needing replacement among the parcels to be excavated in 2018 and 2019, the spreadsheet provides no support for the City's conclusion that the original 18,000 figure was likely an overestimate for the total

number of lead and galvanized steel service lines in Flint.

* * *

51. In sum, none of the documents or data provided by the City that I reviewed provides a sound justification for its predictions that the hit rate will be 60% in the areas slated for excavations in 2018, and 50% in the areas slated for excavations in 2019, or for the City's assertion that the original 18,000 figure was likely an overestimate for the total number of lead and galvanized steel service lines in Flint at replacement eligible households as of March 28, 2017.

Application of prior observed hit rate

52. As described in Paragraphs 13 and 14, the City has not provided a complete list of explanatory variables used to predict the hit rate of lead or galvanized steel service lines in areas of Flint where it has not yet conducted excavations. Nor has the City used the data it has collected to date to provide an assessment of the applicability of those explanatory variables to its prediction of the likely hit rate in the areas where it has not yet conducted excavations.

53. In the absence of such an assessment, the data collected to date itself may be used to predict the outcome (hit rate) for areas where work is left to do. Multiple established and accepted statistical theories support the use of the observed hit rate (here, 70.7%, *see* Paragraph 30 Report at 2 (reported hit rate for work conducted through 2017)) as a sound predictor of the future hit rate.

54. For example, the *frequentist* line of statistical reasoning holds that the probability of an event is approximately equal to the frequency with which the event happens in a large number of trials.[12] Thus, the probability of a hit on lead or galvanized steel for every parcel in Flint covered by the Settlement Agreement (the hit rate for the excavations in Phases I – VI) is approximately equal to the frequency with which the hit happened in the excavations done to date (i.e., 70.7%, the hit rate of the 8,843 excavations completed as of February 2018, *see* Paragraph 30 Report at 2).

55. Another line of statistical reasoning, *spatial statistics*, holds that the applicability of the observed hit rate to date in predicting the hit rate for areas slated for work in 2018 and 2019 is appropriate due to the theory of spatial autocorrelation, i.e., events in close spatial proximity to each other tend to be more similar than events that are more spatially separated.[13] Since all events (hits, i.e., the presence of lead or galvanized steel service lines) in the past (work done to date) and future (work to be done in 2018 and 2019) are in close spatial proximity to each other (in the same city), the observed hit rate for the work done to date is an appropriate predictor of the future hit rate for work to be done in 2018 and

---

[12] Clifford Konold, *Informal Conceptions of Probability*, 6 Cognition and instruction 59-98 (1989).

[13] Oliver Schabenberger & Carol A. Gotway, *Statistical Methods for Spatial Data Analysis* (2005).

2019.

56. A third line of statistical reasoning, Bayes' theorem, holds that new data should be used to refine existing hypotheses.[14] Thus, Bayes' theorem supports the application of the data from the work done to date (the observed hit rate) as a refined predictor of the future hit rate.

57. Applying the City's observed hit rate for work completed through 2017 (70.7%) as the expected hit rate for households slated for excavation in 2018 and 2019, and adding that prediction to the number of households that required replacement in the work done through 2017, results in a conclusion that there were likely more than 18,000 lead and galvanized steel service lines in Flint at replacement eligible households as of March 28, 2017. This result is based on the following calculations:

($a_1$) total number of occupied households covered by the Settlement Agreement as eligible for replacement = 28,400[15]

---

[14] National Research Council, *supra* note 5, at 173; Sharon Bertsch McGrayne, *The Theory that Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines and Emerged Triumphant from Two Centuries of Controversy* (2011).

[15] I have assumed that Flint has approximately 28,400 residential parcels with active service line connections to Flint's water system, based on a letter from Robert Bincsik, Director, Department of Public Works, City of Flint, to Chris Korleski, Director, Water Division, Region 5, U.S. EPA, dated May 16, 2018. Chaudhary Decl. Ex. 19. However, it is unclear whether this assumption accurately reflects the number of replacement eligible households in Flint as of March 2017. Indeed, earlier correspondence from the City suggests that the number of occupied

($b_1$) total households with excavations completed through 2017 = 8,843[16]

($c_1$) total households with excavations completed through 2017 that were found to have lead or galvanized steel service lines = 6,256[17]

($d_1$) total replacement eligible households that have not had service line excavations through 2017 = [total replacement eligible households] – [total households where excavation completed through 2017]

total replacement eligible households yet to be excavated = ($a_1$) – ($b_1$) = 28,400 – 8,843 = **19,557**

($e_1$) number of households yet to be excavated that are predicted to have lead or galvanized steel service lines based on the hit rate of work done to date (70.7%) and rounding to the nearest whole number = [($d_1$) multiplied by 0.707] = 19,557 * 0.707 = 13,826.799 = **13,827**

($f_1$) total replacement eligible households predicted to have lead or galvanized steel service lines, based on excavations done to date plus predictions for households slated for excavation in 2018 and 2019 = ($c_1$) + ($e_1$) = 6,256 + 13,827 = **20,083**[18]

---

residential parcels in Flint may be much higher, roughly 39,525. A letter from Gen. Michael McDaniel to Keith Creagh, Director, Michigan Department of Natural Resources, dated November 1, 2016, states there are roughly 51,000 residential parcels in Flint. *Id.* Ex. 5. Assuming that 22.5% of those parcels are unoccupied (based on the percentage of unoccupied homes reported in a September 2016 letter from Gen. McDaniel to Mr. Creagh, *see id.* Ex. 4) yields an assumption that Flint has 39,525 occupied, i.e., "replacement eligible" households (51,000 multiplied by [1 – 0.225]). I do not know which of these two assumptions is more accurate, but I have used the more recently reported number (28,400) for purposes here.

16 *See* Paragraph 30 Report at 2.

17 *See id.*

18 Again, this estimate assumes 28,400 occupied residential parcels per ($a_1$). However, if the calculations were completed using the higher estimate of 39,525 residential parcels per footnote 15, the total replacement eligible households predicted to have lead or galvanized steel service lines, based on excavations done

($g_1$) total replacement eligible households in Flint predicted to have lead or galvanized steel service lines (20,083) > 18,000

58. As described in Paragraphs 15 through 21, in addition to using the observed hit rate, the City may use other data as explanatory variables to refine the predicted hit rate for the areas of the City slated for work in 2018 and 2019, but only if it undertakes an assessment of the appropriateness of these variables to a prediction model.

59. For example, the City may use historical records that indicate service line composition at households slated for work in 2018 and 2019 to inform their estimated hit rates for 2018 and 2019, but only after an assessment of this potential explanatory variable is conducted. If the City chooses to use historical records as an explanatory variable in predicting the future hit rate, then it should provide an analysis of the relative accuracy of those historical records, based on the excavation work done to date. This is because these records have been shown to be inaccurate in previous analyses. *See* Chaudhary Decl. Ex. 5 attach., at 2. The City should also provide quantitative evidence that historical records were a strong predictor of the outcome in the work done to date, as described in Paragraphs 18 and 19.

---

to date plus predictions for households slated for excavation in 2018 and 2019 = 27,948.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Dated: June 20, 2018

Stacy E. Woods, Ph.D., M.P.H.

# INDEX OF EXHIBITS

| **Exhibit** | **Description** |
|---|---|
| 1 | *Curriculum vitae* of Stacy Woods, Ph.D., M.P.H. |
| 2 | List of reviewed data and other documents from the City of Flint |

# EXHIBIT 1

**Stacy E. Woods, MPH, PhD**
1152 15th Street NW, Suite 300
Washington, DC 20005
(202) 513 - 6260 ext. 2260
Swoods@nrdc.org

## EDUCATION

**Doctor of Philosophy, Environmental Health Sciences**
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
June 2016

**Risk Sciences and Public Policy Certificate** (2013)
**Dissertation:** "Investigating the space-time variation in fine particulate matter pollution in the Northeastern United States, 2000 – 2014"

**Master of Public Health**
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
December 2010

**Capstone:** "Spatial analysis of Lyme disease in Howard County, Maryland"

**Bachelor of Science, Entomology and Nematology**
University of Florida, Gainesville, FL
December 2001

## WORK HISTORY

**Staff Scientist**, Natural Resources Defense Council (NRDC), Washington, DC (November 2017 – present).

- Conduct statistical analysis for scientific research projects
- Create static and interactive maps for various NRDC research projects, policy analyses, and litigation
- Provide advice on use of spatial and non-spatial data, statistics, and visual displays of data including maps and graphs to NRDC colleagues
- Serve as manager of ArcGIS software package and provide trainings on ArcGIS to NRDC staff

**Epidemiologist**, United States Environmental Protection Agency, Office of Pesticide Programs, Washington, DC (July 2016 – November 2017).

- Composed systematic literature reviews for pesticides undergoing registration review

**Mirzayan Science & Technology Policy Fellow**, The National Academies of Science, Engineering, and Medicine, Washington, DC (January 2015 – April 2015).

- Drafted internal reports and study prospectuses for the Board on Environmental Sciences and Toxicology (BEST)

- Attended ad hoc committee meetings at the National Academy of Sciences, which included members of academia, policy makers, nonprofits, and industry

**C. Sylvia and Eddie C. Brown Scholar in Community Health**, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (August 2011 – June 2016).

- Developed innovative approach to evaluate small-scale variability of particulate pollution to identify area-specific pollution trends and represented area of potential public health concern through static and interactive maps
- Applied original small-scale methodology to investigate how the fracking industry has influenced particulate pollution variability across Pennsylvania
- Collected and analyzed ambient air quality samples for benzene in South Baltimore
- Executed spatial and temporal analyses of neighborhood drug markets and sexually transmitted infections in Baltimore City, and created static maps for publication

**Teaching Assistant (Lead)**, Geographic Information Systems (GIS) and Spatial Statistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (January 2012 to April 2015).

- Gave lectures, held office hours, prepared assignments, and graded graduate students of the GIS and spatial statistics courses

**Senior Research Assistant**, Center to Reduce Cancer Disparities, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (January – August, 2011).

- Collaborated on papers for public dissemination to technical audiences regarding longitudinal epidemiology studies of cancer disparities
- Provided spatial assessment of socio-economic and health outcomes in Prince George's County, Maryland, including creating maps for communication of health outcomes to non-technical audiences

**Public Health Intern**, Howard County Health Department, Columbia, MD (October 2009 – March 2010).

- Designed and implemented study to characterize geographic distribution of Lyme disease and identify significant disease clusters in Howard County, Maryland

**Biologist**, University of Florida, Florida Medical Entomology Laboratory, Vero Beach, FL (September 2005 – August 2007).

- Conducted species surveys and characterized changes in relative abundance of Florida mosquitoes in response to control activities

**PUBLICATIONS**

Woods, SE, Waugh, DW, Koehler, KA, Davis, MF, Fox, MA, Rule, AM, and Curriero, FC (2017). Investigating the large scale trends and small scale spatial variation in PM2.5

pollution and the efficacy of federal emissions regulations in reducing fine particulate pollution in the northeastern United States. [In Progress]

Woods, SE, Waugh, DW, Koehler, KA, Davis, MF, Fox, MA, Rule, AM, and Curriero, FC (2017). The association of the fracking industry with small scale variability in PM2.5 pollution in Pennsylvania, 2004 - 2014. [In Progress]

Meyer, WK, . . . Woods, SE, et al. (2017) Ancient convergent loss of PON1 yields deleterious consequences for modern marine mammals. [Manuscript accepted, pending publication date, *Science*]

Jennings JM, Woods SE, Curriero FC (2013). The spatial and temporal association of neighborhood drug markets and rates of sexually transmitted infections in an urban setting. Health & Place Volume 23, September 2013, Pages 128–137.

## DISCIPLINARY SKILLS

**GIS and statistical skills and software:**

- Mapping, data management, and analysis in GIS (ArcGIS including Desktop Suite and Online, qGIS, R)
- Education in and advanced application of spatial, multilevel, and longitudinal data analyses
- Statistical coding and analysis in R, STATA, SAS, and SaTScan

**Additional computer skills and software:**

- Online literature databases including PubMed, Scopus, Web of Science, Science Direct, and Google Scholar
- Citation manager software including Endnote, RefWorks, and Zotero
- Adobe Creative Suite including Photoshop

# EXHIBIT 2

**EXHIBIT 2 TO DECLARATION OF STACY WOODS, Ph.D., M.P.H.**

The following is a list of the data and other documents from the City that I reviewed for my declaration.

| **Exhibit, as numbered in Chaudhary Decl.** | **Description** |
|---|---|
| 4 | Letter from Gen. Michael C.H. McDaniel, FAST Start Coordinator, to Keith Creagh, Michigan Department of Natural Resources (Sept. 28, 2016) |
| 5 | Letter from Gen. Michael C.H. McDaniel, FAST Start Coordinator, to Keith Creagh, Michigan Department of Natural Resources (Nov. 1, 2016) |
| 8 | Paragraph 30 Evaluation, City of Flint (Feb. 8, 2018) |
| 13 | Follow-Up Summary, City of Flint (Mar. 6, 2018) |
| 14 | Map, "Occupied Housing with Lead, Galvanized, and Unknown Connections" (created Aug. 9, 2016) |
| 19 | Letter from Robert Bincsik, City of Flint, to Chris Korleski and Anthony Ross, U.S. EPA (May 16, 2018) |
| 23 | Email from William Kim, Flint City Attorney's Office, to Sarah Tallman and Dimple Chaudhary, NRDC (June 8, 2018) |
| 23-1 | Excel file, "HVI Data 4.26.18-6.4.18.xlsx" |
| n/a | Excel file, "Hydrant_Data_20180607.xlsx" |
| 23-2 | Map, "Hydrant Installation Map_Optimized.pdf" (issued June 5, 2018) |
| 23-3 | Map, "Water Card Results Map 11x17.pdf" (issued June 8, 2018) |
| 23-4 | Maps, "Phase 5 Water Card Reading.pdf" |