Review of: Grossman DS, Slusky DJG: *The Effect of an Increase in Lead in the Water System on Fertility and Birth Outcomes: The Case of Flint, Michigan.*

Nigel Paneth MD MPH, Michigan State University

For the past few months, MDHHS epidemiologists studying perinatal outcomes in Flint have asked me to consult on some of the data analyses they have performed intended to shed light on whether the episode of water contamination may have affected outcomes such as infant mortality, low birthweight and preterm delivery in Flint. In the course of this work, I was recently asked to review the above paper and provide my views on its methods, results and conclusions.

Focus of this review: I will focus my comments on the adverse outcomes described in the paper that are generally dealt with in epidemiology – the fetal death rate, birthweight, gestational age, and abnormal conditions of pregnancy. General fertility in the population, as contrasted to fertility problems that pose clinical issues, is not strictly a health problem, and is determined by many complex behavioral factors and personal choices. Thus I will discuss the fertility findings in this paper only in relation to the authors' conclusion that the change in fertility they describe in Flint can safely be attributed to a higher risk of fetal loss. The question of whether in fact there was or was not a change in fertility in Flint requires inspection of both birth rates in Flint and elsewhere in our state, and the population estimates that underlie calculation of fertility rates, and those data are not immediately available to me.

Since this paper has apparently yet to be peer-reviewed, I will provide the kind of review I expect this paper would receive from a journal that publishes studies concerning perinatal health outcomes in populations such as the ones addressed in this paper. If this review is shared with the authors, I hope it will be helpful to them in addressing the questions they raise in the most scientifically rigorous manner possible.

Study Outcomes

1. **Fertility rate**

As noted above, the general fertility rate, defined as the number of live births in a population divided by an estimate of the number of women aged 15-44 in that population, is determined by many complex behavioral factors. But changes in rates of fetal loss are rarely sufficient to alter general fertility in an entire population. I know no instance, including the literature on famines and other disasters, which are often accompanied by substantial drops in fertility, in which a change in fertility in a general population has been clearly linked to a change in the rate of spontaneous fetal loss. It would be useful if the authors were to provide examples of any episodic toxic exposures with established effects on fetal loss sufficiently large to alter measured fertility in an entire population.

Recognized fetal losses fall into two categories - losses before 20 weeks of gestation, generally referred to as miscarriages or early fetal losses, and losses after 20 weeks of gestation, generally referred to as late fetal losses or stillbirths. The first category is far larger, since an estimated 15% of recognized pregnancies end in miscarriage, while late fetal deaths rates (denominatored to live births) are currently reported in the US to be about 6/1,000. Thus reported fetal deaths rates account for only about 4% of spontaneous pregnancy losses.

Only late fetal deaths are reportable in Michigan, and the completeness of reporting of late fetal deaths is a problem in all jurisdictions. The claim in the paper is that fertility declined by about 8.5 births per thousand women of child-bearing age. At the same time, the fetal death rate, as described in their paper, rose by 0.1 deaths per thousand women of child-bearing age. Thus the reported fetal death rate can only account for a little over 1% of the fertility decline.

Not discussed in this paper is another form of fetal loss – induced abortion. The abortion rate in Michigan is currently about 15 per thousand women of child-bearing age. With general fertility rates of between 60 and 70 per thousand women of child-bearing age, changes in abortion rates have a much greater potential to affect fertility than changes in late fetal losses. To put this in perspective, and using the figure of 67/1,000 for fertility (the figure provided in Table 1 for non-Flint births in the pre-water change period), among 1,000 births in Michigan, 150 would be expected to miscarry spontaneously, 180 to be aborted, 6 to be a late fetal loss and 664 to result in a live birth. The ratio of the two large forms of fetal loss not enumerated in this paper – spontaneous and induced abortions – to reported late fetal deaths, the only category of loss examined in this paper, is thus greater than 50 to 1. It should be noted that induced abortions are a reportable outcome of pregnancy, and if one is to make a full case for a change in fertility in Flint or elsewhere in our state, it would be advisable to examine these data.

The authors then make the case that a change in the sex ratio at birth, with a reduction of the proportion of births that are male, is evidence of fetal toxicity, because males are more susceptible to environmental influences on fetal survival. The sex ratio change was such that 1,009 males were born in Flint in the post-water change period they examine, whereas, if the sex ratio had remained constant, 1026 males should have been born, a difference of 17 births. These 17 fewer males represent about 0.1% of the fertility rate.

Thus the two changes affecting fertility directly examined in this paper can together account for not much more than 1% of the change in fertility reported, while larger factors that affect the rate of fetal loss are unexamined. And yet changes in these two minor parameters – late fetal deaths and sex ratio - are the only data provided before the paper concludes (section 6.3, page 24) that "*the results in this section provide direct support for the Flint water change causing a culling of the weaker fetuses*".

In addition to the common forms of fetal loss on which the authors provide no data, the other factors that affect fertility directly, which is to say behaviors affecting the conception rate such as economic conditions, availability of contraception and many other factors, are considered in only one way. The single insight into this complex set of behaviors is based on survey data from the American Time Use Survey (ATUS) that suggests to the authors that sexual activity increased in Flint after the water change, and therefore cannot have affected fertility. The variable requested by ATUS is "*time spent in personal or private activities*", which the authors note, includes having sex among several other activities. The authors also mention another limitation, which is that the survey unit is Genesee County, only 25% of whose residents live in the City of Flint. Not noted by the authors is that ATUS, a national study, conducts a total of just 16,000 annual interviews across the nation, that the samples are stratified by race and other factors to represent Michigan and the US rather than the sampled county itself, and has a response rate of less than 60%. Given the size of Genesee, the sample size for these estimates must have been very small and the relevance of this survey to Flint, very uncertain.

In most of the authors' tabulations of fertility rates in Flint and in comparison populations (Table 1, Figures 3, B1, B2, panel A of B3), fertility rates in Flint were quite consistently *lower* than the comparison population, *before t*he Flint water change. In Table 1, the pre-water change fertility rate of the 25 largest cities in Michigan is given as 67 per thousand women of childbearing age and for Flint it is 62 per thousand. Moreover, Figure 3 shows that at times before the water switch, the extent of lower fertility in Flint was even greater, reaching a difference of about 10 per thousand women of childbearing age in late 2010/early 2011. This substantially lower fertility in Flint *before* the water change is left unexplained.

It is impossible to conclude, from the data provided, that even if fertility did drop in Flint as the authors claim, that the drop was due to an increase in spontaneous fetal loss.

2. **Late Fetal Death Rate**

Unlike most states, data on fetal deaths are recorded in Michigan differently from data on live births. Live births are reported on certificates, as are fetal deaths in most states. Fetal deaths (and incidentally, induced abortions as well) are considered *reports* in Michigan. This distinction was made by the state to retain fetal death data as "*confidential statistical reports to be used only for medical and health purposes and shall not be incorporated into the permanent official records of the system of vital statistics*." Unlike all other vital data elements, identifying information on the mother is not requested for late fetal deaths.

For this reason and others, even though the state requires reporting of a fetal deaths over 20 weeks and weighing more than 400 grams at birth, the pressure to report is less than for infant deaths, or even live births, both of which are part of the state's vital statistics system, and which, in the case of live birth, produces a certificate that is required for many legal purposes. The birthweight requirement of 400 grams, which is unique to Michigan (most states do not require a weight threshold, and those that do use a different threshold) also probably reduces reporting of all late fetal deaths, some of whom are below the weight threshold, even though over 20 weeks.

Studies have documented substantial under-reporting (Greb et al, 1987) and over-reporting (Makelarski et al 2010) of late fetal deaths in state reporting systems, even in locations where certificates are required. For this reason, state-reported fetal death rates are not nearly as widely used as indicators of perinatal health as are infant mortality data, which are built on a much stronger foundation, despite the fact that the number of reported late fetal deaths is very similar to the number of infant deaths.

Turning to the trends described in this paper, one encounters at least three anomalies. First, the fetal deaths are denominatored to women of child-bearing age and not to live births. All epidemiologic studies of which I am aware denominator fetal death either to live births (in which case the statistic is called a fetal death *ratio*), to all births, or to an estimate of the number of pregnancies in utero at the gestational week in which the fetal death was recognized. The reason for this is straightforward. A fetal death requires a pregnancy; including women not pregnant in the denominator can produce misleading results because the fetal death rate then becomes nearly entirely a function of the fertility rate.

The second anomaly is that the authors' data showing that the reported fetal death rate (per thousand women of child-bearing age) in Flint after the water change is *lower* than found in the rest of Michigan either before or after the water change. Their reported fetal death rates (Table 1, page 18) in the rest of Michigan ranged from 0.34 to 0.36, while in Flint, post-water change, it was 0.32. If lead was involved in causing fetal deaths in Flint, why is the fetal death rate *in the post-water change period* 6% lower than in the rest of Michigan? The authors emphasize only that the latter figure in Flint represented an increase in the fetal death rate in Flint from a pre-change figure of 0.18. But they do not ask themselves why, in a city with socio-economic circumstances worse than the rest of the state, with more women with characteristics that put them at higher risk of late fetal death than elsewhere in the state, the fetal death rate they take as the baseline in Flint was half (0.18 compared to 0.36) that of the fetal death rate in the rest of the state. Common sense would have urged them to take a longer view of the period prior to the water change, in case the surprisingly low rate in Flint just before the water change was an anomaly. They might also wish to calculate a conventional fetal death rate (i.e. denominatored to all births), and examine more closely the quality of fetal death reporting in Flint and in Michigan, as well as the validity of their denominator data.

The third anomaly, is that although the increase in the fetal death rate from 0.18 to 0.32 is highlighted in the abstract, Table 1 shows that it is not statistically significantly different from the change in the non-Flint areas. This not surprising since the absolute difference was just 0.16 fetal deaths per 1,000 women of child bearing age. What is surprising is that the authors take a non-significant finding and highlight it as the largest health effect included in the abstract.

Birthweight and Gestational Age:

The authors find no meaningful change in mean birthweight or mean gestational age over time in Flint, although they note that Flint births, both before and after the water change weigh some 200g less than births elsewhere in Michigan, testifying to the longstanding higher risk for adverse outcomes for newborns in Flint. They do not examine changes in the clinically important extremes of the distribution over time, most usually described as the percent of births <2,500 g and/or born before 37 weeks.

The index they choose to describe gestational growth, grams/week, is a flawed statistic, since fetal growth is not constant by gestational week. A more useful and valid measure is the fetal growth ratio, which compares each infant's birthweight to the remainder of infants born at the same gestational week. Again, it is surprising to read the authors claim in their conclusion (page 35) that the data provide *"suggestive evidence of …. a decrease in birthweight and gestational age"* when the change they report in Table 1 - a 40 gram decline in birthweight (when other Michigan cities experienced a decline of 25 grams) and a decline in gestational duration of approximately one day – are both too small to even approach statistical significance.

3. **Abnormal Conditions of Pregnancy:**

The authors finding is that these conditions, as reported on birth certificates, actually *decreased* in frequency in Flint, from 18.5% to 17.7% of live births. While the authors acknowledge that this represents a "*positive health outcome*" (page 25), they make no attempt to explain it, and the meaning it may have for their hypothesis.

Comparison populations and adjustments for risk

The authors state "…*only the city of Flint switched their water sources at this time, while the rest of Michigan did not. These areas provide a natural control group for Flint in that they are economical similar areas, and with the exception of the change in water supply, followed similar trends in fertility and birth outcomes over this time period."*

The claim that the rest of Michigan is economically similar to Flint is, of course, unsupportable, Although the wording in the paper is not precise on this point, the term "non-Flint births" as used in Table 1 appears to refer to the 25 largest cities in MI other than Flint. In any case, the comparison population shows, in Table 1, which describes the main findings of the study, that mothers of the non-Flint births used as the central comparison population were more than five times as likely as mothers in Flint to have a college education.

The appendices contain analyses which attempt to control for some factors that affect the fetal death rate such as maternal education, race and age of mother, as well as contrasts to a variety of different sub-populations in Michigan. Since no change in fertility occurred in any of the comparison areas, and the composition of the Flint population did not change during the brief time period examined, it is not really surprising that the reported decline in fertility remained significant in all of these contrasts.

The importance of the correct comparison population is highlighted when the authors turn to using the 29 most populous counties in Michigan as the comparison. Puzzlingly, in this comparison, Flint, which as noted above, in all other contrasts has *lower* fertility than the comparison regions, both before and after the water switch, now is found to have consistently *higher* fertility than the control population, which, from Appendix figure C2, is nearly always below 50 per thousand women of childbearing age. How it is possible for the 25 largest cities in Michigan to have a fertility rate about a third higher than that of the 29 largest counties in Michigan (two groupings which must surely overlap considerably) is not explained. These shifts in fertility data, depending upon which comparison population is used, indicate that the authors need to describe and verify their information base both for births and for women of childbearing age very carefully.

Data Sources

An unfortunate difficulty in the presentation of this paper is the absence of any critical consideration of the limitations of the several data sources they use, especially vital data and fetal death reports. It is standard practice in perinatal epidemiology to carefully describe and weigh the strengths and limitations of the data sources in use. As expressed by one noted perinatal epidemiologist, "*I issue a call to editors of peer-reviewed journals publishing studies using vital statistics as primary sources: Require authors either to document or otherwise evaluate the quality of their data for primary research variables or to comprehensively assess the potential impact of mismeasurement on their findings*" (Kirby 2002).

Focus on Lead

It has become apparent that although lead contamination was a major component of the change in the water supply of Flint to the Flint River in 2014, there may well have been other chemical changes involving other toxic substances. To focus solely on lead is to unnecessarily limit the scope of scientific thought on these important issues. If effects on fertility or fetal death were to be found, it would be important to keep an open mind about the many possible reasons for these effects, and not to isolate one particular chemical in a study in which, by its nature, no direct assessment of the specific causal agent is possible.

Conclusion of this Review

Careful analyses of available data to try to assess any possible health effects of the water contamination episode in Flint are very welcome. However, it is essential to examine carefully the underlying data sources and their limitations, to examine the meaning of findings contrary to the hypothesis, and to endeavor to explain internally contradictory findings such as have been highlighted in this review. Time trends in health outcomes are one way to explore the health effects of time-delimited exposures, but always present the difficulty that it is rare indeed when only one thing is changing over time.

I conclude that the data in this paper are an insufficient basis to establish that an adverse effect of the water contamination occurred on rates of fetal death, or on any other adverse birth outcome examined, i.e. birthweight, gestational age or abnormal conditions of pregnancy.

As to whether fertility declined in Flint relative to the rest of Michigan, I would recommend that the authors recheck the underlying data very carefully, paying attention especially to estimates of the relevant denominator populations and making sure that a full picture of long-term fertility in Flint is explored, and not just data from a small number of years. In any case, none of the data on fetal death rates described in this paper can come close to explaining the trends in fertility in Flint that this paper reports.

REFERENCES

Greb AE, Pauli RM, Kirby RS. Accuracy of fetal death reports: comparison with data from an independent stillbirth assessment program. Am J Public Health 1987;77:1202–6.

Kirby RS: Re: Case-cohort analysis of agricultural pesticide applications near maternal residence and selected causes of fetal death. Am J Epidemiol. 2002 Apr 15;155:779-80;

Makelarski JA, Romitti PA, Caspers K et al: Use of active surveillance methodologies to examine over-reporting of stillbirths on fetal death certificates. Birth Defects Research A 2010; 91: 1004-1010.