# Mortality and Science: A Comment on Two Articles on the Effects of Health Insurance on Mortality

Robert Kaestner[1]

LINK TO ABSTRACT

While health insurance has several benefits, for example, providing financial protection for financial losses associated with illness, it is the greater use of healthcare associated with it, and the putative health benefits of that greater use of care, that is of primary importance. Intuition strongly suggests that having health insurance will improve one's health and decrease the probability of dying. Supporting the intuition is extensive evidence that health insurance increases health care utilization (e.g., Newhouse et al. 1993; Baicker et al. 2013) and clinical evidence that health care treatments are effective, for example, those catalogued by the Choosing Wisely organization, the National Institute of Clinical Evidence (NICE) in the United Kingdom, the U.S. Prevention Services Task Force (USPSTF) and in reviews of the Cochrane Library (link).

Despite the strong intuition, there is surprisingly little high-quality empirical evidence directly linking health insurance to improved health or lower mortality. The question has been investigated in quite a few studies. There are also several reviews of that literature that have reached various conclusions as to whether there is sufficient evidence to support the claim that health insurance improves health (Hadley 2003; Freeman et al. 2008; Levy and Meltzer 2008; McWilliams 2009; Baicker et al. 2013; Sommers et al. 2017; Woolhandler and Himmelstein 2017). What is noteworthy is that the question is still being vigorously debated. Whether health insurance improves health is central to debates over the value of Medicaid,

1. University of Chicago, Chicago, IL 60637.

Medicare, and the Affordable Care Act, the largest government health programs.

The lack of a consensus about the size and significance of the effect of health insurance on health makes the topic of particular salience to academic researchers. And the uncertainty about whether health insurance improves health breeds mischief in policy debates.[2] Recently, two studies on the effect of health insurance on mortality were published in the *Quarterly Journal of Economics* (QJE). The articles deliver findings that can only be characterized as amazing—health insurance entirely eliminates mortality.

I critically review the evidence in each article. I find that the two studies provide little useful information on this important research question. Both studies lack statistical power to detect reasonably sized effects and produce an array of estimates that cannot be explained by clinical evidence or theories of behavior. And one study has a problematic research design and the other has virtually no external validity—its results, even if true, are relevant for only a tiny fraction of the population and unlikely to be relevant for the remaining population.

## The GLM (2021) and MJW (2021) articles

The first study is "Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach," by Jacob Goldin, Ithai Z. Lurie, and Janet McCubbin (2021)—henceforth abbreviated GLM. The article provides an evaluation of a randomized controlled trial (RCT) conducted under the auspices of the U.S. Treasury Department and Internal Revenue Service (IRS) that involved sending a letter in January 2017 to people who had paid a tax penalty for not having health insurance coverage in 2015. (Only about 20 percent of those without health insurance paid such penalty, the others being exempt because they didn't pay taxes, did not have access to insurance that cost less than 8.05 percent of income, or were exempt for a recognized hardship.) The letter informed the recipients of their options for health insurance coverage and the size of the tax penalty in 2017. The letter was sent to a randomly selected portion (86 percent) of the people who had paid the tax penalty for 2015. Tax information from 2017 and 2018 was used to identify health insurance coverage of this sample of people. Information from Social Security Death file was used to identify sample members who died in 2017 or 2018. The outcomes of those who were sent the letter (treatment group) were compared to the outcomes of those (14 percent) who were not sent the letter

---

2. See the following for evidence of the debate and how research in this area has been used by both sides of the political spectrum: Broaddus and Aron-Dine 2019; Pipes 2019; Davidson 2017; Lowrey 2019; West 2018; Moffit 2020.

(control group). The article mainly focuses on the results for persons from both the treatment group and the control group ages 45 to 64 who did not have health insurance in 2016 for at least one month. The authors conclude: "we interpret our results to support the conclusion that the coverage induced by the intervention reduced mortality" (GLM 2021, 44).

The second study is "Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data," by Sarah Miller, Norman Johnson, and Laura R. Wherry (2021)—henceforth abbreviated MJW. This study used data from the American Community Survey (ACS) from 2008 to 2013 to select a sample of persons born 1950 to 1959 and who would be between the ages of 55 and 64 in 2014. The sample was restricted to those with less than a high school education or who had family income less than 138 percent of the Federal Poverty Level (FPL) at the time of the ACS survey. These data were matched to administrative information on date of death from the Census Numident File and information on Medicaid enrollment from the Centers for Medicaid and Medicare Services (CMS). The statistical analysis of these data involved comparing before-to-after changes in mortality (and health insurance coverage) of those living in states that expanded Medicaid eligibility as part of the ACA to those living in states that did not expand Medicaid eligibility. Based on their analysis, the authors concluded: "we show that the ACA Medicaid expansions substantially reduced mortality rates among those who stood to benefit the most" (MJW 2021, 41).

Both of the articles, therefore, purport to provide evidence supportive of expanding health insurance: MJW (2021) is about Medicaid, while GLM (2021) is about health insurance generally. Both articles focus on the reduction in mortality.

# A lack of statistical power

The most important problem with both articles is that they lacked statistical power to detect an effect of health insurance on mortality that is plausible.

To their credit, GLM conducted an analysis of statistical power and reported it in the appendix to their article (**link**). Their power analysis used several unrealistic assumptions, however, assumptions that obscured the fact that the study was grossly underpowered. To understand why their analysis was flawed, some context is necessary.

The intervention of the study—the IRS's sending of the letter to taxpayers who didn't have health insurance in 2015—resulted in only a very small increase in health insurance coverage. In their preferred sample of taxpayers ages 45 to 64 who lacked health insurance in 2016 for at least one month (as well as in 2015), sending the letter to the treatment group increased the probability of having of any

health insurance coverage, during the 24-month period of 2017–2018, by about 2 percentage points, to 51 percent, as compared to the mean of 49 percent in the control group (that is, those who lacked insurance in 2015 and 2016 and were not sent the letter). Sending the letter increased the months of any insurance coverage among the treatment group by 0.4 months, to 8.2 months, as compared to 7.8 months in the control group. In sum, the intervention had a very small effect on health insurance coverage and thus any difference in mortality between the treated and control groups should plausibly be very small and in line with the very small difference in health insurance coverage. Therefore, when conducting the analysis to detect the level of statistical power, reasonably sized estimates of the difference between the treatment and control groups should be used to provide an accurate estimate of the extent of statistical power. Smaller estimates result in less statistical power, all else equal.

In conducting such a power analysis, what would be a reasonable estimate to use? To answer this question, I start with the extreme assumption that gaining a full year of health insurance coverage would reduce the probability of dying by 100 percent. Indeed, such an estimate would be consistent with a conclusion that obtaining health insurance entirely eliminates mortality. In the sample highlighted by GLM (middle aged without insurance for at least one month in 2016), the intervention increased insurance coverage by 0.358 months, or 3 percent of a year. Thus, assuming that gaining insurance for one year reduced mortality by 100 percent implies that the difference in mortality between the treated and control groups should be about 3 percent. This effect size seems to be a good maximum to use and even that large of an effect seems implausible.

What effect sizes did GLM use in their power analysis? They said they used effect sizes of 1 percent, 3 percent, 5 percent and 7 percent of the mean mortality rate. As already noted, an effect size of 3 percent would correspond to a 100 percent reduction in mortality. Thus, GLM's 3-percent effect size implies that a year of health insurance reduces mortality by nearly 100 percent, which any impartial observer would think implausibly high. Using effect sizes of 5 percent and 7 percent is outlandish. It seems prudent to use an effect size smaller than 3 percent, but surely not a value greater than 3 percent.

But GLM use an implausible assumption that bends the results of their analysis in their favor—so that it indicates greater statistical power—while still appearing to use reasonable effect sizes. Specifically, they assume that the mortality rate of the group affected by the intervention is 2, 3, or 4 times the average mortality rate and then apply the 1-, 3-, 5-, and 7-percent rates. This makes the 3-percent effect size 2, 3, or 4 times larger. As noted, the 3-percent effect size using average rate of mortality is already unrealistically high, so increasing it by a factor of 2, 3, or 4 implies effect sizes that are virtually impossible to conceive of as plausible.

Moreover, there is no evidence to support the assumption that the mortality of those affected is double, triple, or even quadruple the average rate. In fact, data reported in the study shows that the control mean mortality rates among those with no prior insurance (1.007, Table IV) is lower than the mortality rate for those with prior insurance (1.15, derived from Table IV and Table A.XVI). Thus, the "affected" group—those with no prior insurance—actually have lower mortality rates than those with prior insurance.

Thus, the power analysis conducted by GLM substantially inflated the extent of the statistical power of their actual analysis.

Even with this inflation, however, the analysis lacked power. Figure A.5 of the article shows the results of the power analysis conducted by GLM. The statistical power associated with the study for effect sizes of 1 percent and 3 percent never rises above 45 percent—in other words, 45 of 100 times, at most, would a study of this type be able to detect a true effect that is 3 percent or less. This is true even when the assumed mortality rate is assumed to be 4 times the average. The standard that is commonly used to judge whether there is adequate statistical power is 80 percent, which is well above 45 percent. In the same figure, results indicate that the statistical power of the analysis is less than 55 percent for all effect sizes (e.g., including 7 percent) when it is assumed that the mortality rate of the affected group is 2 times the average mortality rate. Finally, if attention is restricted to cases where the effect size is 1 percent or 3 percent, and the mortality rate of the affected is 2 times the average, the statistical power of the analysis never reaches 16 percent, is often under 10 percent, and these are the most likely scenarios.

The extreme lack of statistical power (10 percent) of the analysis should have been a red flag. With such a low level of statistical power, the results of the study are uninformative and the reported findings surely misleading. As discussed in Andrew Gelman and John Carlin (2014), in analyses with low statistical power, for example 10 to 20 percent as is likely in the GLM study, there is a high likelihood that a significant estimate will be a substantial exaggeration (e.g., by a factor of 3 to 5 or more) of the true effect and can even lead to wrongly signed estimates with non-trivial probability. To quote Gelman and Carlin (2014):

> There is a common misconception that if you happen to obtain statistical significance with low power, then you have achieved a particularly impressive feat, obtaining scientific success under difficult conditions.
>
> However, that is incorrect if the goal is scientific understanding rather than (say) publication in a top journal. In fact, statistically significant results in a noisy setting are highly likely to be in the wrong direction and invariably overestimate the absolute values of any actual effect sizes, often by a substantial factor. (Gelman and Carlin 2014, 649)

The likelihood that the problem raised by Gelman and Carlin (2014) characterizes the GLM study is supported by the implausibility of the estimate obtained by the authors, which suggests that 5.6 months of additional health insurance coverage would reduce mortality by approximately 100 percent—a year of health insurance would decrease mortality by 214 percent, which doesn't make a lot of sense, since dying during the year is something that happens to someone at most once. The problems that a lack of statistical power causes that are noted by Gelman and Carlin (2014) are also evidenced by the large, but wrongly signed effects of health insurance (it kills you) that GLM obtain for those ages 25 to 44. I discuss this further below.

Instead of acknowledging the obvious, that the low power of the study is the problem, GLM go to great lengths to argue for alternative interpretations of the implausibly large estimates, interpretations that require unverifiable assumptions about the baseline mortality rate of "compliers" and whether the relationship between health insurance coverage and mortality is linear (same for each additional month of coverage) or concave (the first month of coverage is more important than the second and so forth).[3] But it is not even a close call. Even if we halved the estimate of 214 percent assuming a death rate among those affected that is twice the average death rate of the sample, the still absurdly large estimate is implausible. This is not a surprising result once the very low level of statistical power is taken into account. The results of GLM are uninformative of the effect of health insurance on mortality.

As for MJW (2021), they did not provide a power analysis for their study, which is surprising because MJW cite a prominent critique of prior studies examining the effect of Medicaid on mortality demonstrating that most studies suffered from a low level of statistical power (Black et al. 2021). To remedy this oversight, I conducted an analysis to assess the statistical power of their study. Like the authors, I used the ACS surveys from 2008 to 2013 and selected a sample of people born 1950 to 1959 (ages 55 to 64 in 2014) who had less than a high school degree or had family income less than 138 percent of the Federal Poverty Level. I dropped, as did the authors, non-citizens and those receiving Supplemental Security Income. I obtained a sample of 419,696, which is close to the sample size of 421,648 reported by MJW in Table A.1.[4]

---

3. GLM do not consider the case of a convex relationship—where each additional month is more beneficial.

4. Note that these samples are based on the public-use data. For their mortality analysis, MJW used a sample that was larger (33 percent), derived from data not publicly available. The difference will not materially affect the results of the power analysis I used this sample for because the "effective" sample size is the number of states and not the number of individuals in the sample, although the number of individuals in the sample will have a marginal effect on the analysis.

Using this baseline sample, I constructed a simulated data set assuming that people in the sample, who are low-educated or poor, die at 1.6 times the rate as the average person of the same age and gender who live in the same state as those in the sample (see Black et al. 2021 for a similar analysis). The 1.6 figure comes from MJW's appendix (**link**, see p. 10) who note that the mortality rate of the poor (low-educated) is 1.6 times that of the average person. For each person in the sample, I aged them one year until they died or the year was 2017, which is the last year used in the MJW study. The simulated data set has approximately 3 million person-years. To assign an indicator for whether or not a person died, I randomly assigned people to die in the same proportion as people of their age and sex living in the same state as them. For example, for 60-year-old females living in Illinois, I randomly assigned, on average, 0.012 of them to die, because data from the Centers for Disease Control (CDC) indicated that between 1999 and 2013, on average, 0.007 of females aged 60 living in Illinois died per year and then I multiplied this death rate by the 1.6 figure noted earlier (0.012 = 0.007*1.6).[5] So, exactly like MJW, I had a data set that followed the same sample of people that they used until they died or until 2017 if they did not die by then.

I then assumed that the Medicaid expansion had different effects on mortality in periods after the expansion and I decreased the death rate of people living in the expansion states by that amount using several different assumed estimates ranging from −0.0002 to −0.0009. While these amounts seem quite small, it is necessary to put them into the context of the analysis. As reported by MJW, the expansion of Medicaid decreased the rate of uninsured by 4.4 percentage points for people in expansion states (treated) relative to people in non-expansion states (comparison). Thus, the difference in mortality rates between the treated and comparison groups should be commensurate with the increase in the probability of gaining health insurance coverage. According to MJW the mortality rate to use as a counterfactual—the mortality rate in expansion states had there been no expansion—is 1.63 percentage points (0.0163). So, effect sizes of 0.0002 to 0.0009 represent changes in mortality of between 1.2 and 5.5 percent. These are not small changes when measured against the change in health insurance coverage of 4.4 percentage points (i.e., 4.4 percent of the sample obtained health insurance coverage). The different effect sizes I used imply that gaining health insurance coverage decreased mortality by between 28 and 125 percent. Again, note the unrealistic nature (i.e., immortality) of the upper end of this range.

Using these assumed effect sizes, I obtained 500 estimates of the effect of the Medicaid expansion for each assumed effect size (randomly assigning death in

---

5. When assigning the probability that a person died, I allow there to be some randomness in this average death rate using the standard deviation of the death rate observed between 1999 and 2013.

each of the 500 repetitions). I then calculated the proportion of the 500 estimates that were statistically significant, which provides an estimate of the statistical power of the analysis—what proportion of the time would a study exactly like that of MJW find a true effect statistically different from zero using a 95 percent level of confidence (0.05 level of significance). The results are shown below in Table 1.

**TABLE 1. Analysis of statistical power in MJW**

| Assumed effect of Medicaid expansion on mortality | Implied effect of health insurance on mortality | Share of estimates statistically significant | Mean estimate among those statistically significant | Implied effect of health insurance on mortality of mean estimate |
|---|---|---|---|---|
| −0.0002 | −28% | 14% | −0.0008 | −115% |
| −0.0003 | −42% | 23% | −0.0008 | −115% |
| −0.0004 | −56% | 37% | −0.0009 | −125% |
| −0.0005 | −70% | 47% | −0.0009 | −125% |
| −0.0006 | −84% | 64% | −0.0010 | −139% |
| −0.0007 | −98% | 74% | −0.0012 | −167% |
| −0.0008 | −115% | 85% | −0.0011 | −153% |

*Notes*: Estimates in column 2 are equal to the estimate in column 1 divided by 0.044, which is the change in health insurance coverage reported in MJW, and then divided by 0.0163 which is the counterfactual mortality rate used by MJW. Estimates in column 5 are similarly constructed using the estimates in column 4.

Several points related to the results of Table 1 are noteworthy. First, the level of statistical power of the study for all effect sizes smaller than −0.0008 is below the conventional threshold of 80 percent, and for most of them well below. Second, the implied effect of health insurance on mortality would need to be over 100 percent for the analysis to have adequate (80 percent) statistical power. Again, only if health insurance coverage entirely eliminated mortality would this study reliably find such an effect. For more realistic effect sizes, for example, a reduction in mortality due to gaining health insurance of 56 percent or less, which is still substantial, the study has very low levels of statistical power—less than 38 percent. Third, observe the size of the mean estimate among estimates that are significant. They imply huge effects of health insurance on mortality and they grossly overstate the true effect by a factor of 2 to 4. It is not surprising then that MJW (2021, 37) obtain an estimate of the effect of health insurance on mortality of 184 percent. And even if you ignore my power analysis, the standard errors of estimates of the effect of Medicaid expansion on mortality reported by MJW reveal the lack of power directly. The standard errors imply that the analysis is unable to reliably reject anything but a huge effect—that gaining health insurance reduces mortality by more than 139 percent.

Overall, and like GLM, the MJW study lacked statistical power to detect

plausible effect sizes, and the lack of power, unsurprisingly, resulted in a huge estimate that defies common sense. This result is just another example of the problem described in the quote from Gelman and Carlin (2014) and highlighted by Black et al. (2021).

An interesting implication of the lack of statistical power of these studies is that, if redone, then it would be much more likely to find a smaller and statistically insignificant estimate than the large significant estimate that was found in each study. Would studies that found small insignificant estimates have been published? My suspicion is the answer is no because the lack of statistical power and imprecision of the estimates would likely cause reviewers/editors to conclude that the study was uninformative. This should have been the conclusion for these two studies too because the statistically significant estimates are equally non-informative when there is a lack of statistical power. The fact that a study finds a statistically significant effect is not a sufficient scientific basis for understanding the truth (Gelman and Carlin 2014).

# Findings all over the place

The lack of statistical power that characterizes both the GLM and MJW studies is sufficient to consider them uninformative of the question as to whether health insurance affects mortality and also the question of whether the interventions studied (IRS sending a letter and Medicaid expansions, respectively) affected mortality. A careful review of the estimates presented in each study further highlights the questionable nature of the evidence.

GLM report a variety of estimates of the effect of the IRS-letter intervention on health insurance coverage and mortality, and the effect of health insurance coverage on mortality. Table 2 provides a summary of these estimates by age, sex, marital status and whether a person's state of residence expanded Medicaid as part of the ACA.

As can be observed in Table 2, all of the effects of health insurance on mortality are implausibly large, for example, increasing the mortality rate by 100 percent. Only a few estimates are statistically significant, reflecting the lack of statistical power described earlier. There is also little coherence among the estimates that can be explained by the gain in health insurance or by an appeal to clinical evidence.

Consider the difference in the effects by sex. Why would the effect of health insurance coverage on mortality for females be more than 2.5 times greater than that for males? Men have higher rates of mortality and disease, particularly heart disease, and treatment of heart disease is relatively effective (e.g., statins).

**TABLE 2. Summary of heterogeneous estimates reported in GLM**

| Group | Effect of intervention on health insurance coverage | Effect of intervention on mortality | Effect of health insurance on mortality | Effect of one year of health insurance on mortality as a percent of baseline mortality |
|---|---|---|---|---|
| Ages 25–34 | 0.18§ (0.05) | 0.009¶ | 0.05§ (0.12) | +120%± |
| Ages 35–44 | 0.20§ (0.05) | 0.020¶ | 0.10§ (0.20) | +160%± |
| Ages 45–54 | 0.31§ (0.05) | −0.078¶ | −0.25§ (0.20) | −300%± |
| Ages 55–64 | 0.43§ (0.08) | −0.047 -0.047 | −0.10§ (0.21) | −120%± |
| Ages 45–64 | 0.358 (0.026) | −0.063 -0.025 | −0.178 (0.070) | −212% |
| Males | 0.366 (0.031) | −0.052¶ | −0.142 (0.097) | −142% |
| Females | 0.346 (0.037) | −0.079¶ | −0.229 (0.096) | −373% |
| Married | 0.415 (0.044) | −0.071¶ | −0.170 (0.082) | −257% |
| Not married | 0.314 (0.032) | −0.056¶ | −0.181 (0.112) | −186% |
| Medicaid expansion state | 0.410§ | −0.05¶ | −0.127 (0.078) | −167% |
| Non-expansion state | 0.300§ | −0.078¶ | −0.259 (0.131) | −277% |

*Notes*: The symbol § indicates that the value was an estimate based on Figure I, Figure V, and Appendix Table A.VIII in GLM. The symbol ¶ indicates that the value is derived by multiplying column 1 by column 3 values, which were reported by GLM. The symbol ± indicates that the baseline mortality rate was not reported. In its place, I assumed a baseline mortality of 1 per 100 for those aged 45–54 and 55–64; 0.75 for those aged 35–44; and 0.5 for those aged 25–34.

Why would the effect of health insurance *increase* mortality by 160 percent among those ages 35 to 44 and decrease mortality by 300 percent for those ages 45 to 54? It is difficult to think that there is some clinical explanation or behavioral explanation for such a difference.

Another example, not reported in Table 2 but in GLM's appendix Table A.XXIV, relates to income. The effect of health insurance on mortality was 10 times larger (and implausibly large in absolute value) for those with income greater than 138 percent of Federal Poverty Level as compared to those with income under 138 percent of Federal Poverty Level. Why would gaining health insurance decrease mortality for higher-income persons and not have any effect on lower-income persons? The gain in insurance coverage was larger for the lower-income group (Table A.VIII) and the mortality rate was similar (Table A.XXIV). In sum,

the huge, disparate, and seemingly arbitrary estimates when viewed as a whole are not easily explained and are almost surely due to the low statistical power of the study.

Another statistical issue that was not addressed in GLM is multiple testing bias. As John List, Azeem Shaikh, and Yang Xu (2019) show, it is necessary to adjust for multiple testing bias when examining treatment effects for different subgroups. GLM (2021) estimate scores of models by age, sex, income, prior insurance coverage, and state. The multiplicity of subgroups suggests that if one were to adjust for multiple testing bias to GLM's estimates, it would substantially reduce their statistical significance and exacerbate the lack of power.

The variation in the effect of health insurance on mortality clearly points up just how weak is the relationship between the effect of the intervention on gains in insurance and the effect of the intervention on mortality. For example, those ages 55 to 64 had a gain in health insurance coverage that is approximately 40 percent greater the gain experienced by those ages 45 to 54, but the effect of the intervention on mortality was 66 percent greater for the younger age group than the older age group. Again, there are no obvious behavioral or clinical explanations for these findings. Other evidence also shows a similarly weak relationship between the effect of the intervention on health insurance coverage and the effect of the intervention on mortality. In Figure A.IX (by treatment arm) and Figure A.X (by state) in GLM, the correlation between the effect of the intervention on insurance and the effect of the intervention on mortality are not statistically significant. It is notable that only one of three figures examining the correlation between the effect of the intervention on insurance and the effect of the intervention on mortality was reported in the GLM text (Figure IV). The figure reported in the text was the only one of the three that showed a statistically significant correlation and also the only one of the three analyses to include individuals with prior coverage in 2016, which is not the sample used in the main analyses (which excludes those previously covered by insurance).

A similarly weak relationship between the effect of the intervention on health insurance and the effect of the intervention on mortality is found by treatment arm as shown in Table 3. There is no obvious reason why the effect of health insurance on mortality (last column of Table 3) among a group similarly situated and of the same age should differ by treatment arm after adjusting for the different effects of the different treatment arms on health insurance coverage, as is done in the last two columns of Table 3. Perhaps there is some behavioral or clinical reason to expect differences in the effect of health insurance across treatment arms, but it is unknowable from GLM's study and would be purely an ad-hoc explanation. A more likely explanation is the lack of statistical power of the analysis.

**TABLE 3. Heterogeneous effects of intervention by treatment arm, as reported in GLM**

| Treatment arm | Effect of intervention on health insurance coverage | Effect of intervention on mortality | Effect of health insurance on mortality | Effect of one year of health insurance on mortality as a percent of baseline mortality |
|---|---|---|---|---|
| Basic | 0.389 (0.032) | −0.033 (0.030) | −0.084 | −100% |
| Early | 0.524 | −0.076 | −0.145 | −174% |
| Non-personalized | 0.251 | −0.047 | −0.187 | −224% |
| Exemption info | 0.367 | −0.060 | −0.163 | −196% |
| Spanish | 0.355 | −0.055 | −0.154 | −185% |
| *Notes*: Estimates in columns 1–3 come from GLM. Estimates in column 4 are author's calculations. | | | | |

Here too, List et al. (2019) argue that it is necessary to adjust for multiple testing bias when examining the impact of different treatment arms. In this case, there are eight treatment arms, although only five categories are reported (Tables A.XIII and A.XX). Standard errors of the effect of each treatment arm on mortality are not reported in the text (only the significance of the difference in a treatment arm from the "basic" arm are reported). Only for the "basic" arm is there a standard error, and for this arm, the intervention had no significant effect. It is likely that adjusting for the multiple testing bias would render the other estimates insignificant, too.

A final anomaly is found in Figure III of GLM, which shows the effect of the intervention on mortality for those ages 45 to 64 with no prior coverage. The figure shows that the effect of the intervention grows with time, which is surprising because the effect of the intervention on health insurance coverage is declining with time. One possibility, which I offer speculatively and was not something the authors considered, is that cumulative health insurance coverage and the care it implies has an increasingly beneficial effect (see footnote 2). However, this explanation is inconsistent with the argument the authors make to justify the implausibly large estimate they obtain of the effect of health insurance on mortality that it is the first month or two of coverage that matters and not the cumulative effect. My point is the authors can't have it both ways. And again, the large headline estimate and growing estimates over time in Figure III are surely due to a lack of statistical power.

MJW also reported many estimates for different demographic and socio-economic groups. Table 4 summarizes their estimates. Again, there is no obvious explanation for the disparate effects of health insurance on mortality by age, race/ethnicity and other characteristics. For example, why would obtaining health insurance *increase* mortality by 61 percent among those ages 40 to 49 and decrease

mortality by 43 percent among those ages 50 to 54? Similarly, why would obtaining health insurance *increase* mortality of black persons by 50 percent, which is not a trivial number, and decrease mortality by an impossible 208 percent among white people?

TABLE 4. Summary of heterogeneous estimates reported in MJW

| Group | Effect of Medicaid expansion on uninsured | Effect of Medicaid expansion on mortality | Effect of health insurance on mortality | Effect of health insurance on mortality as a percent of baseline mortality | Comparison to GLM |
|---|---|---|---|---|---|
| Ages 19–29 | −0.095 (0.011) | 0.00007 (0.00005) | 0.0007 | 67% | |
| Ages 30–39 | −0.084 (0.011) | −0.00005 (0.00017) | −0.0006 | −23% | |
| Ages 40–49 | −0.082 (0.012) | 0.00023 (0.00022) | 0.0028 | 61% | |
| Ages 50–54 | −0.079 (0.011) | −0.00032 (0.00048) | −0.0041 | −42% | −300% (Ages 45–54) |
| White | −0.044 (0.010) | −0.00169 (0.00041) | −0.038 | −208% | |
| Black | −0.050 (0.015) | 0.00045 (0.00097) | 0.009 | 50% | |
| Hispanic | −0.035 (0.014) | −0.00072 (0.00044) | −0.021 | −231% | |
| Other | −0.045 (0.013) | −0.00047 (0.00149) | −0.010 | −110% | |
| Males | −0.040 (0.011) | −0.00184 (0.00063) | −0.046 | −230% | −142% |
| Females | −0.048 (0.011) | −0.00085 (0.00058) | −0.018 | −140% | −373% |
| Married | −0.026 (0.012) | −0.00133 (0.00075) | −0.051 | −426% | −257% |
| Not married | −0.055 (0.011) | −0.00132 (0.00052) | −0.024 | −124% | −186% |
| Less than HS | −0.032 (0.013) | −0.00163 (0.00080) | −0.051 | −334% | |
| Less than 138% FPL | −0.055 (0.011) | −0.00131 (0.00047) | −0.024 | −132% | −28% |
| *Notes*: Estimates in column 3 were obtained by dividing column 2 by column 1. | | | | | |

A comparison of estimates in MJW with those in GLM (last column) underscores that both studies lack statistical power, produce estimates that are implausibly large, and, what's more disconcerting, produce estimates that are quantitatively and qualitatively different despite estimating a similar parameter— the effect of health insurance on mortality. The comparison underscores that both

studies provide little reliable information about the effect of health insurance on mortality.

Other evidence reported in MJW also resists easy explanation. For example, the Medicaid expansion had a beneficial effect on external causes of death, and an impossibly large effect on deaths due to internal causes. Medicaid expansions were also associated with a non-trivial decrease in mortality among those with incomes of 400 percent or more of the FPL despite almost no change in health insurance coverage among this group.

Overall, both the GLM and MJW studies report hugely disparate results across demographic and socioeconomic groups, and across the two studies, that are not easily explained by any behavioral model or clinical evidence. The value of these estimates in terms of revealing information about the effect of health insurance on health is nil.

## The Medicaid expansion is not an experiment

While GLM is based on a true experiment, the MJW study is based on a 'natural' experiment and a difference-in-differences design that assumes that states that did not expand Medicaid were a valid comparison group for states that did expand Medicaid. The validity of this assumption is always difficult to assess empirically, but it is particularly difficult when the statistical power of a study is very low. For example, to bolster the case for interpreting their results as causal estimates, MJW lean heavily on analyses that show that trends in mortality in expansion and non-expansion states did not diverge *significantly* in years prior to the Medicaid expansion. However, when there is a lack of statistical power, it is highly likely that there will be no significant differences in trends in mortality prior to the Medicaid expansion because the analysis cannot reliably detect reasonable differences in mortality between expansion and non-expansion states. Thus, finding no differences is expected because of the lack of statistical power—it is not a validation of the research design.

To provide independent evidence related to the validity of the research design that does not suffer from a lack of statistical power, I used the initial ACS sample of MJW to assess whether trends in employment and wages conditional on being employed differed by whether a state eventually expanded Medicaid or not. The analysis was conducted for the period prior to the Medicaid expansions of 2014 and used data from 2008 to 2013.

The analysis is straightforward. I used Ordinary Least Squares (OLS) regression methods to obtain estimates of the calendar-year pattern of employment and conditional earnings in states that did and did not expand Medicaid between 2014

and 2017 (as in MJW). The regression model adjusted for separate state and year effects, and separate effects for each year of age, race/ethnic category, marital status category and sex. In some models I also adjusted for state-by-race/ethnicity and year-by-race/ethnicity and results were very similar. I then tested whether the year effects differed between expansion and non-expansion states. For both model specifications, statistical tests rejected common year effects (test of joint hypothesis that year-by-expansion indicators were zero) at the 0.05 level of significance. This result suggests that employment among the sample used by MJW was trending differently by whether or not a state eventually expanded Medicaid. In fact, estimates showed a relative increasing trend in employment in expansion states, as shown in Figure 1a. Similarly, conditional on being employed, estimates from identical regressions, but using log earnings as the dependent variable, indicated that earnings were growing more slowly in expansion states than non-expansion states and that these differential year effects were significant at the 0.12 and 0.09 level depending on the model (Figure 1b).

**Figures 1a and 1b**. Estimates of pre-trend differences—in employment (Fig. 1a) and log earnings (Fig. 1b)—between expansion and non-expansion states
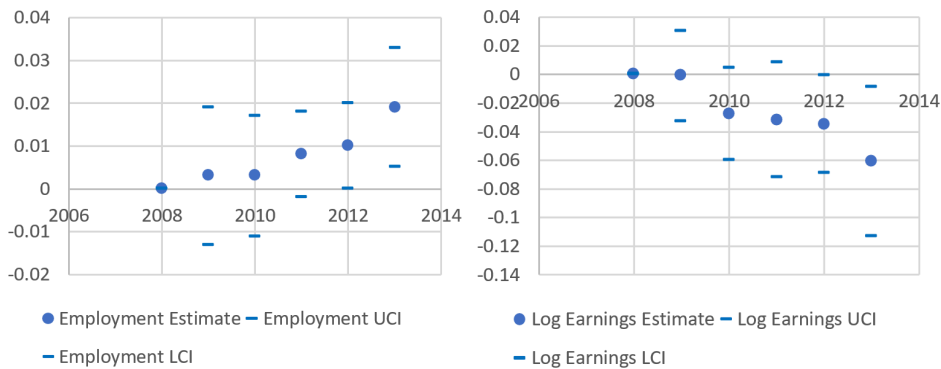


*Notes*: Estimates in Figures 1a and 1b come from a model that adjusted for state-by-race/ethnicity and year-by-race/ethnicity (see text for other controls).

The evidence of dissimilar trends in employment and wages conditional on being employed between expansion and non-expansion states raises questions about the validity of the difference-in-difference research design used by MJW to examine the effect of the Medicaid expansions on mortality. The supporting evidence provided by MJW with respect to mortality is not very persuasive once the lack of statistical power is considered. While a rejection of the identifying assumption for one outcome (e.g., employment) does not necessarily mean that the difference-in-differences method is invalid for another outcome (e.g., mortality),

it should raise alarm bells. Mortality is determined by many factors including socioeconomic factors such as employment and earnings and macroeconomic forces that underlie changes in individual employment and earnings. The lack of credible support of the research design specifically related to mortality and the direct evidence of an invalid design for outcomes related to mortality suggests that the MJW study is unlikely to produce causal estimates. This exacerbates the lack of statistical power already noted.

# A lack of external validity

External validity is an important aspect of scientific inquiry, although it is usually ignored in much economics research. It refers to whether a study that finds that an intervention, such as expanding health insurance coverage, "works somewhere" or for someone will work somewhere else or for someone else (Cartwright 2011).

The experimental intervention (receipt of a letter) of GLM was targeted at people who were without health insurance in 2015 and who paid a tax penalty. This group consisted of approximately 9 million persons and represents only 20 percent of taxpayers who were uninsured at some point in 2015. Notably, among those who were uninsured in 2015, the experimental sample had income that was 40 percent higher than the larger sample of uninsured persons. Thus, the targeted sample was unlike the average uninsured person. In addition, the primary analysis was limited to those without health insurance for at least one month in 2016, which is a subset of approximately 5 million people, or 55 percent of the 9 million in the experimental sample and about 10 percent of all uninsured persons in 2015.

Finally, and remarkably, only about 48,000 people of those who had received the letter obtained insurance coverage as a result. When the sample is limited to those ages 40 to 64 who lacked insurance for at least one month in 2016, which is the primary sample highlighted by GLM, the number of people who gained health insurance coverage for any period of time was only about 21,000. It is this tiny fraction of people from the larger samples that the results of the study are applicable.

The following quotation from GLM seems to agree with my conclusion of little external validity:

> Although the baseline mortality rate for extensive-margin compliers may be higher or lower than for other complier groups, and although a difference in the baseline mortality rates is only one reason that the ACR may differ from the average effect of coverage on mortality for the overall uninsured

population, *this result suggests that the effect of coverage on mortality may be particularly large among the individuals induced into coverage from the intervention we study, as compared with other policies that reduce uninsurance.* (GLM 2021, 41, emphasis added)

To emphasize, when the authors write "*among the individuals induced into coverage from the intervention*" they are referring to about 21,000 people among whom only a small fraction died. The group of 21,000 is a remarkably small number of people, among the 7.6 million who received the letter,[6] demonstrating the hoped-for effect (that is, getting health insurance). Findings about this sliver of people, irrespective of the nature of the findings, should not carry much weight in policy debates. The references to millions of people in the experiment belies the actual tiny number of people that the estimate of the effect of health insurance on mortality is based. There is no reason to believe that the estimate of the effect of health insurance on mortality that pertains to this small group of people would be applicable to the vast majority of uninsured or the vast majority of people with current insurance. Any pretense otherwise is disingenuous. Moreover, estimates in GLM differ in qualitative and quantitative ways from those reported in MJW. While I believe I have provided a good case that estimates from either study are not reliable and not informative, the contradictory evidence surely implies that at least one of these has no external validity, and quite possibly neither does.

# Conclusion

My review of GLM (2021) and MJW (2021) reveals that both studies were severely under-powered to detect a reasonably sized effect of health insurance (or treatment as measured by receiving a letter or Medicaid expansion) on mortality and, because of that, were prone to grossly overestimating the effect of interest if not get the direction of the effect wrong.[7] The lack of statistical power would be heightened by appropriate adjustment for multiple testing bias. And when results are viewed in a comprehensive manner, inconsistencies and anomalies arise that seem inexplicable from a behavioral or clinical perspective. Adding to these problems is the fact that the MJW study was an observational study that is likely

---

6. As reported in GLM (2021), 8.9 million people paid the IRS tax in 2015 for being uninsured. Of these, 86 percent or 7.6 million received the IRS letter. GLM (2021), however, focused the analysis on a sample of those aged 45–64 who were uninsured for at least one month in 2016, or 1.4 million people, and approximately 1.1 million of these received the IRS letter. Of this 1.1 million people, approximately 21,000 obtained health insurance because of the IRS letter.

7. A similar lack of power with correspondingly implausibly large estimates characterizes a paper that examines the effect of Medicaid expansion on mortality, that of Borgschulte and Vogler (2020).

to be biased by unmeasured confounding, and that the GLM study has virtually no external validity. The flaws of the two studies leads me to conclude that we learned little about the effect of health insurance on mortality from them.

I return to the question of why these articles were published in such a prestigious journal despite all of the red flags—e.g., the implausible size of estimates and the lack of statistical power, and wildly contradictory estimates by demographic and socioeconomic characteristics—that render estimates in both papers uninformative.

Here I speculate, but from my experience.

I have found the economics profession to be so narrowly focused on research design and internal validity that it often fails to consider the statistical power of an analysis, the scientific plausibility of the findings, and external validity. Because the GLM study was based on an experiment, then, to many economists, estimates had to be valid and informative. Similarly, because the MJW study passed a series of "robustness checks" of questionable value, estimates had to be valid and useful. The fact that the estimates in both papers are anything but informative because of the flaws that were there for all to see if bothering to look exemplifies a growing problem in economics—a preoccupation with research design that comes at the expense of all other ingredients that make an analysis scientifically valid.

Second, I raise the possibility that the articles were published because the results, however unreliable, align with political objectives, namely, support for greater government intervention in healthcare. It is difficult to imagine that similar studies that found small, insignificant effects of health insurance on mortality, which as I noted would be the much more likely outcome given the lack of statistical power of the two articles, would be published.

# Appendix

Data and code related to this research is available from the journal website ([link](#)).

# References

**Baicker, Katherine, Sarah L. Taubman, Heidi L. Allen, Mira Bernstein, Jonathan H. Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein**. 2013. The Oregon Experiment—Effects of Medicaid on Clinical Outcomes. *New England Journal of Medicine* 368(18): 1713–1722.

**Black, Bernard S., Alex Hollingsworth, Leticia Nunes, and Kosali Ilayperuma**

**Simon**. 2021. Simulated Power Analyses for Observational Studies: An Application to the Affordable Care Act Medicaid Expansion. *NBER Working Paper* 25568. National Bureau of Economic Research (Cambridge, Mass.). **Link**

**Borgschulte, Mark, and Jacob Vogler**. 2020. Did the ACA Medicaid Expansion Save Lives? *Journal of Health Economics* 72: 102333.

**Broaddus, Matt, and Aviva Aron-Dine**. 2019. Medicaid Expansion Has Saved at Least 19,000 Lives, New Research Finds. November 6. Center on Budget and Policy Priorities (Washington, D.C.). **Link**

**Cartwright, Nancy**. 2011. Predicting 'It Will Work for Us': (Way) Beyond Statistics. In *Causality in the Sciences,* eds. Phyllis McKay Illari, Federica Russo, and Jon Williamson, ch. 35. New York: Oxford University Press. **Link**

**Davidson, John Daniel**. 2017. Let's Stop Pretending Medicaid Saves Lives. *The Federalist* (FDRLST Media, Chicago), June 29. **Link**

**Freeman, Joseph D., Srikanth Kadiyala, Janice F. Bell, and Diane Martin**. 2008. The Causal Effect of Health Insurance on Utilization and Outcomes in Adults: A Systematic Review of US Studies. *Medical Care* 46(10): 1023–1032.

**Gelman, Andrew, and John Carlin**. 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science* 9(6): 641–651.

**Goldin, Jacob, Ithai Z. Lurie, and Janet McCubbin** (GLM). 2021. Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach. *Quarterly Journal of Economics* 136(1): 1–49. **Link**

**Hadley, Jack**. 2003. Sicker and Poorer—The Consequences of Being Uninsured: A Review of the Research on the Relationship between Health Insurance, Medical Care Use, Health, Work, and Income. *Medical Care Research and Review* 60(2_suppl): 3S–75S.

**Levy, Helen, and David Meltzer**. 2008. The Impact of Health Insurance on Health. *Annual Review of Public Health* 29: 399–409.

**List, John A., Azeem M. Shaikh, and Yang Xu**. 2019. Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics* 22: 773–793. **Link**

**Lowrey, Annie**. 2019. The Supreme Court Is Bad for Your Health. *The Atlantic*, July 31. **Link**

**McWilliams, J. Michael**. 2009. Health Consequences of Uninsurance Among Adults in the United States: Recent Evidence and Implications. *Milbank Quarterly* 87(2): 443–494. **Link**

**Miller, Sarah, Norman Johnson, and Laura R. Wherry** (MJW). 2021. Medicaid and Mortality: New Evidence From Linked Survey and Administrative Data. *Quarterly Journal of Economics* 136(3): 1783–1829. **Link**

**Moffit, Robert E**. 2020. Government-Controlled Health Care Won't Help Us Live Longer. March 4. Heritage Foundation (Washington, D.C.). **Link**

**Newhouse, Joseph P., and the Insurance Experiment Group**. 1993. *Free for All?: Lessons from the RAND Health Insurance Experiment*. Cambridge, Mass.: Harvard University Press.

**Pipes, Sally**. 2019. Medicare For All Won't Result In Better Health Outcomes. Forbes.com, March 18. **Link**

**Sommers, Benjamin D., Atul A. Gawande, and Katherine Baicker**. 2017. Health

Insurance Coverage and Health—What the Recent Evidence Tells Us. *New England Journal of Medicine* 377(6): 586–593.

**West, Rachel**. 2018. Expanding Medicaid in All States Would Save 14,000 Lives Per Year. October 24. Center for American Progress (Washington, D.C.). **Link**

**Woolhandler, Steffie, and David U. Himmelstein**. 2017. The Relationship of Health Insurance and Mortality: Is Lack of Insurance Deadly? *Annals of Internal Medicine* 167: 424-431. **Link**

# About the Author

**Robert Kaestner** is a Research Professor at the Harris School of Public Policy of the University of Chicago. He is also a Research Associate of the National Bureau of Economic Research, an Affiliated Scholar of the Urban Institute and a Senior Fellow of the Schaeffer Center for Health Policy of USC. Prior to joining Harris, Kaestner was on the faculty of the University of Illinois, University of Illinois at Chicago, University of California, Riverside, the CUNY Graduate Center and Baruch College (CUNY). He received his Ph.D. in Economics from the City University of New York. He received his BA and MA from Binghamton University (SUNY). His research interests include health, demography, labor, and social policy evaluation. He has published over 125 articles in academic journals. Recent studies have been awarded Article of the Year by AcademyHealth in 2011 and the 2012 Frank R. Breul Memorial Prize for the best publication in *Social Services Review*. Dr. Kaestner has also been the Principal Investigator on several NIH grants focused on Medicare and Medicaid policy. Kaestner is an Associate Editor of the *Journal of Health Economics* and the *American Journal of Health Economics*, and on the Editorial Board of *Demography* and *Journal of Policy Analysis & Management*. His email is kaestner.robert@gmail.com.

Discuss this article at Journaltalk:
**https://journaltalk.net/articles/6034**