Reexamining the Effect of Clean Water.

A Replication Study of Anderson, Charles, and Rees (*AEJ: Applied Economics*, 2022)

Ivan Kozlov [*] Carson Jones [†]

Journal of Comments and Replications in Economics, Volume 5, 2026-11, DOI:10.18718/81781.62

JEL: H75, I12, I18, J13, Q53
Keywords: Water filtration, Infant mortality, Typhoid mortality, Difference-in-differences, Two-way fixed effects, Variation in treatment timing treatment effect heterogeneity, Doubly-robust estimator, Public health infrastructure

Data Availability: The R code and the data to reproduce the results of this replication can be downloaded at JCRE's data archive (DOI: 10.15456/j1.2025190.2321978684).

Please Cite As: Kozlov, I., and Jones, C. (2026). Reexamining the Effect of Clean Water. Findings from 24 European Countries. A Replication Study of Anderson, Charles, and Rees (*AEJ: Applied Economics*, 2022). *Journal of Comments and Replications in Economics*, Vol. 5 (2026-11). DOI: 10.18718/81781.62

Abstract
We replicate and extend Anderson et al. (2022), who reassessed the effects of early 20th-century public health interventions on urban mortality found by Cutler and Miller (2005). Their study, using two-way fixed effects, found significant reductions in infant and typhoid mortality. We replicate their main results but extend the analysis to address concerns about treatment effect heterogeneity in difference-in-differences with multiple time periods by applying the estimator proposed by Callaway and Sant'Anna (2021b). Our reanalysis confirms their results for typhoid mortality but finds heterogeneous and largely statistically insignificant effects on infant mortality, suggesting that prior estimates may overstate the impact of water filtration.

[*]Department of Economics, Michigan State University
[†]Department of Economics, University of Maryland cjones68@umd.edu

Received July 10, 2025; Revised April 22, 2026; Accepted June 17, 2026; Published July 29, 2026.

1 Introduction

The decline in urban mortality during the early 20th century is widely regarded as one of the most significant public health achievements in U.S. history. Work by Cutler and Miller (2005) attributed much of this improvement to investments in urban public health and clean water infrastructure, particularly the filtration and chlorination of municipal water supplies. More recently, Anderson et al. (2022) revisited this question, correcting and expanding the dataset used by Cutler and Miller. Their findings reaffirmed the importance of water filtration, particularly in reducing infant and typhoid mortality, although with more modest effect sizes than previously reported, and de-emphasized other measures such as chlorination.

In this paper, we replicate the main analyses of Anderson et al. (2022) and extend their work by addressing a key methodological concern: the use of traditional fixed effects methods leaves their estimates vulnerable to bias in the presence of heterogeneity. To address this, we apply the multiple-period difference-in-differences estimator developed by Callaway and Sant'Anna (2021b), which allows for the unbiased estimation of average treatment effects even when treatment effects vary across groups and time periods.

Using their dataset, we successfully replicate the findings of Anderson et al. However, our alternative analysis of infant mortality reveals substantial heterogeneity in treatment effects across cities, and we find that the average treatment effect on the treated of water filtration is not statistically significant. In contrast, the results for typhoid mortality remain significant and show a greater effect size than the original estimates. These findings suggest that the broader mortality benefits of water filtration may have been more limited and less precisely identified than previously thought, while its impact on typhoid mortality remains robust.

2 Summary of Anderson et al.

Reexamining the Contribution of Public Health Efforts to the Decline in Urban Mortality by Anderson et al. (2022) investigates the impact of several public health interventions implemented across major U.S. cities in the early 20th century. The researchers modify the dataset and build on the analyses from an influential study by Cutler and Miller (2005), which examined the effects of clean water technologies on mortality. Using a panel of 25 major cities, Anderson et al. examine both general and disease-specific mortality rates, relying on U.S. Census data. Their analysis employs a fixed-effects model, controlling for city-specific demographic factors, year and city fixed effects, and city-specific time trends in mortality to isolate the effects of public health measures.

A key contribution of the paper is its correction of transcription errors in Cutler and Miller's original data. Anderson et al. identified that 79 out of 410 infant mortality counts in Cutler and Miller's study were incorrectly transcribed, leading to an overestimation of the impact of water filtration on infant mortality. After correcting these errors, Anderson et al. found that the estimated effect of water filtration on infant mortality diminished by two-thirds, and the effect on total mortality decreased by half. These adjustments significantly changed the conclusions about the importance of water filtration in reducing mortality during this period.

Despite the adjustments, Anderson et al. still find that water filtration had a measurable, though

Table 1: Descriptive Statistics for Key Variables

	Mean	SD	Description
Filtration	0.428	0.493	= 1 if city had a water filtration plant
Chlorination	0.625	0.480	= 1 if city treated water supply chemically
Clean water project	0.236	0.423	= 1 if city had completed a clean water project
Sewage treatment or diversion	0.248	0.430	= 1 if city had a sewage treatment plant or diverted sewage away from water supply
Bacteriological standard	0.641	0.477	= 1 if city set a bacteriological standard for milk
TB test	0.540	0.496	= 1 if city required the tuberculin testing of cows
% female	0.503	0.016	Female percent of city population
% non-white	0.090	0.100	Non-white percent of city population
% foreign	0.193	0.109	Foreign-born percent of city population
% under 15	0.255	0.032	Percent of city population $<$ 15 years old
% 15 to 44	0.529	0.025	Percent of city population 15–44 years old
% 45 and older	0.216	0.038	Percent of city population $\geq$ 45 years old

reduced, effect on urban mortality rates, particularly in reducing infant mortality and typhoid-related deaths. However, the authors show that the effects were limited, and other public health interventions, such as chlorination, sewage treatment, and milk quality regulations, did not have significant effects on mortality. This contrasts with earlier studies, suggesting that while water filtration played a role in improving public health, it explains only part of the overall decline in urban mortality during the early 20th century.

3 *Original Findings and Replication*

The main analysis is presented in a few tables and figures, with Tables 2 and 3 (6 and 7 in Anderson et al.) and Figure 1 (Figure 5 in Anderson et al.) being the most important. Table 2 presents the results of the fixed-effects regression analysis, examining the impact of public health interventions such as water filtration, chlorination, clean water projects, sewage treatment, and milk regulations on total mortality. The main finding is that water filtration is associated with a modest but statistically insignificant reduction in total mortality, but the effect is smaller than previously estimated by Cutler and Miller. The coefficients for other interventions, such as chlorination and sewage treatment, are either insignificant or positive, suggesting that these measures had no meaningful impact on reducing overall mortality. This finding supports the paper's thesis that, while water filtration contributed to the decline in mortality, its effect has been overstated and other public health interventions had limited or no measurable impact.

Infants and children saw the greatest declines in mortality during the epidemiological transition, largely due to reductions in infectious disease (Costa 2015). Because of this, both Cutler and Miller and Anderson et al. are concerned with infant mortality and not just total mortality. Table 3 runs the same regressions as Table 2, but with infant mortality as the dependent variable. Unlike the total mortality estimates, the effect of water filtration on infant mortality is statistically significant and corresponds to a decrease of approximately 11% in the full model specification.

One concern about the robustness of these estimates is that the effect of the treatment may change over time. In the case of filtration specifically, Anderson et al. note that after a filtration plant was built it often took time to switch water supply networks over to the

Table 2: The Effects of Water Quality, Sewage Treatment/Diversion, and Clean Milk on Total Mortality

	(1)	(2)	(3)	(4)	(5)
Filtration	-0.020			-0.019	-0.011
	(0.016)			(0.016)	(0.016)
Chlorination	0.015			0.012	0.009
	(0.014)			(0.013)	(0.013)
Clean water project	-0.017			-0.014	-0.011
	(0.025)			(0.024)	(0.023)
Sewage treatment/diversion		0.025		0.022	0.018
		(0.025)		(0.024)	(0.022)
Bacteriological standard			0.016		0.015
			(0.017)		(0.018)
TB test			0.019		0.015
			(0.014)		(0.013)
Observations	1024	1024	1024	1024	1024
R^2 (within)	0.930	0.930	0.931	0.930	0.931

Table 3: The Effects of Water Quality, Sewage Treatment/Diversion, and Clean Milk on Infant Mortality

	(1)	(2)	(3)	(4)	(5)
Filtration	-0.131			-0.128	-0.111
	(0.045)			(0.044)	(0.041)
Chlorination	0.093			0.088	0.082
	(0.044)			(0.038)	(0.039)
Clean water project	-0.069			-0.061	-0.055
	(0.066)			(0.061)	(0.058)
Sewage treatment/diversion		0.067		0.048	0.040
		(0.067)		(0.056)	(0.049)
Bacteriological standard			0.040		0.024
			(0.036)		(0.037)
TB test			0.061		0.040
			(0.042)		(0.031)
Observations	1024	1024	1024	1024	1024
R^2 (within)	0.981	0.979	0.980	0.981	0.981

new source. More generally, Cutler and Miller speculate that the effect of treatment could increase over time as individuals learn to more effectively use the intervention. They also suggest that the opposite is possible—the effect may decline over time if filtration crowds out investment in other public health interventions, or if those who survive due to the intervention are sickly and have higher mortality rates than those who would have survived without treatment. To investigate this question, Anderson et al. decompose their filtration dummy into a series of leads and lags (indicating up to five years before or after the construction of a filtration plant). The estimated treatment effects on infant mortality for each lead and lag are presented with 90% confidence intervals in Figure 1. There is no evidence of pre-treatment trends, and all of the confidence intervals before filtration cross zero. The effect of filtration is significant a year after filtration begins, and it becomes larger over time. This supports the results in Table 3, which suggest that water filtration had a significant effect on infant mortality.

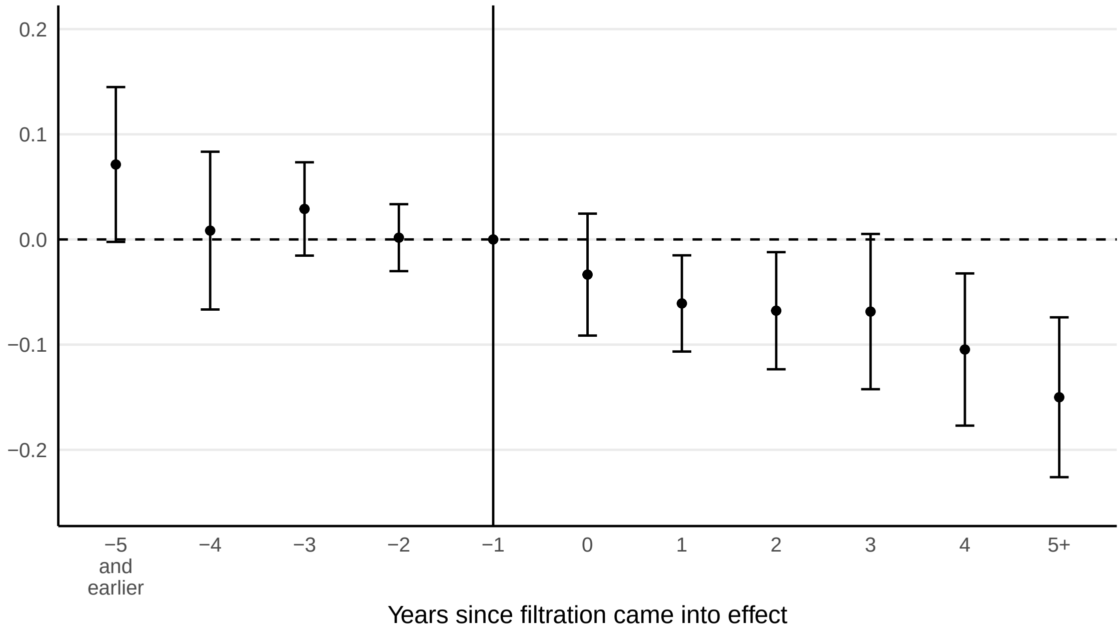

Figure 1: Pre- and Post-filtration Trends in Infant Mortality

In addition to total mortality, Anderson et al. examine cause-specific mortality and find that water filtration led to large and statistically significant reductions in typhoid mortality (Table 4 in this paper and Table 10 in Anderson et al.). Because typhoid is primarily a waterborne disease, this result provides strong evidence of a direct causal effect of improvements in water quality on health outcomes. These effects stand in contrast to the generally modest and often statistically insignificant estimates for other interventions, including chlorination, sewage treatment, bacteriological standards, and tuberculin testing.

Table 4: The Effects of Water Quality, Sewage Treatment/Diversion, and Clean Milk on Mortality by Cause

	Typhoid mortality	Diarrhea/ enteritis mortality	Diarrhea/ enteritis mortality (under age 2)	Non-pulmonary TB mortality	Non-pulmonary TB mortality (under age 2)
Filtration	-0.170	-0.154	-0.158	0.038	0.024
	(0.070)	(0.085)	(0.099)	(0.057)	(0.028)
Chlorination	0.005	0.089	0.107	0.008	-0.007
	(0.040)	(0.079)	(0.096)	(0.054)	(0.033)
Clean water project	0.054	-0.158	-0.159	0.073	0.035
	(0.075)	(0.147)	(0.198)	(0.045)	(0.028)
Sewage treatment/diversion	-0.025	0.190	0.196	-0.067	-0.026
	(0.057)	(0.120)	(0.148)	(0.042)	(0.027)
Bacteriological standard	-0.052	0.028	0.048	0.028	0.008
	(0.037)	(0.055)	(0.066)	(0.042)	(0.027)
TB test	0.040	0.160	0.159	-0.002	0.004
	(0.063)	(0.078)	(0.097)	(0.041)	(0.030)
Observations	1024	974	1024	1024	924
R^2 (within)	0.939	0.971	0.964	0.937	0.843

4 *Concerns with the Analysis*

Our main concern with the estimates presented by Anderson et al. is treatment effect heterogeneity. When there are more than two time periods,[1] the two-way fixed effects estimator will produce biased estimates if treatment effects are heterogeneous across groups, time, or exposure (Callaway and Sant'Anna 2021b). Anderson et al. and Cutler and Miller recognize that heterogeneity may be a problem, and are concerned that the treatment effect may vary with the length of exposure—hence their use of lagged treatment dummies as a robustness check (see Figure 1). But, as Callaway and Sant'Anna show, the estimates that two-way fixed effects models with binned endpoints (e.g., the "5+ years" lag in Figure 1) produce in these situations can be biased, even if we assume that treatment effects are homogeneous across groups.

But we have reason to believe that there may be heterogeneity in treatment effects across groups. As Cutler and Miller admit, the effects of clean water likely differed between socioeconomic groups. The poor suffered more from waterborne illness and would have therefore experienced greater benefits from filtration. But we have no data on how access to the newly filtered water differed by socioeconomic status within cities. Furthermore, the methods used to treat water were different between cities (slow sand vs. mechanical filtration), and the way water was distributed could affect its quality as well (e.g., in Cleveland, filtered water was initially mixed back into the unfiltered water supply). The crowding-out or learning effects that would cause the treatment effect to change with exposure time could also differ between cities. These are all reasons to suspect that the effect of filtration on infant mortality was heterogeneous across cities.

[1] The two-period case of two-way fixed effects being equivalent to the classic difference-in-differences setup with two time periods and a single group that is treated in the second period.

A Goodman-Bacon (2021) decomposition displays the individual comparisons that the two-way fixed effects estimator is averaging, and can show potential issues with heterogeneity or problematic (late vs early) comparisons. The decompositions for the effect of water filtration on infant mortality and on typhoid mortality (Figures 2 and 3) show a great deal of heterogeneity in the 2x2 difference-in-differences estimates. In the case of infant mortality, the problematic comparisons are concentrated in lower part of the graph (i.e. they may be biasing the weighted average in a more strongly negative direction) as compared to the case of typhoid mortality, where the late-versus-early comparisons appear more randomly distributed. This suggests that the TWFE estimate may be biased for infant mortality, but not for typhoid.

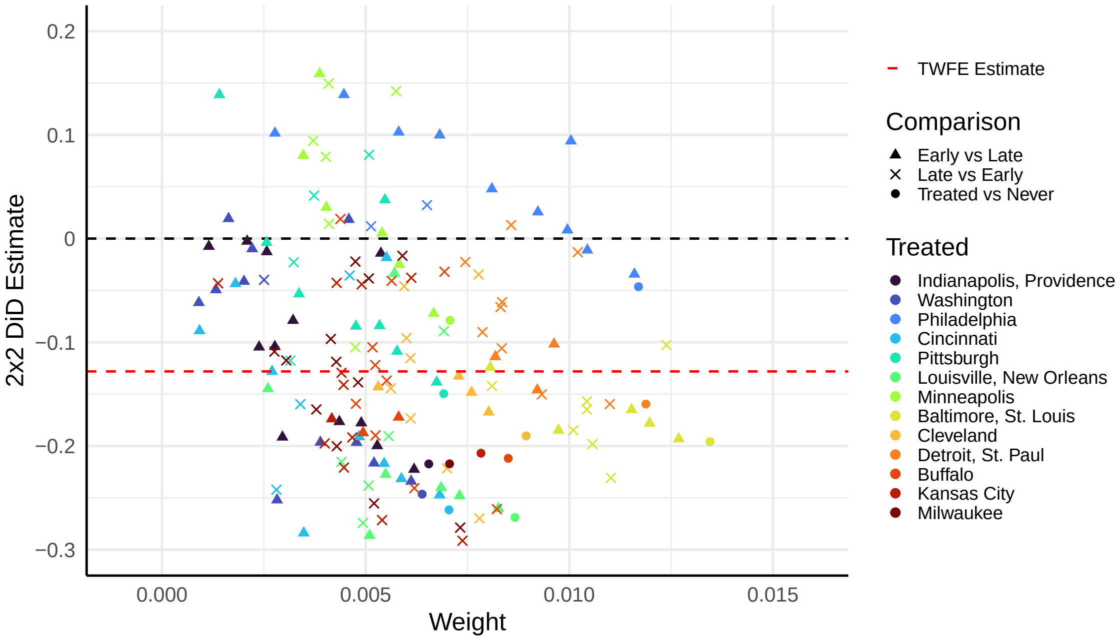

Figure 2: Goodman-Bacon Decomposition for Filtration and Infant Mortality

5 *Alternative Analysis*

To address concerns about treatment effect heterogeneity, we use the doubly-robust estimator developed by Callaway and Sant'Anna (2021b).[2] Their procedure eliminates the problematic comparison between newly treated and already treated units that is done by the two-way fixed effects estimator. Instead, it only makes comparisons between treated units and not yet treated or never treated units. This allows us to get unbiased estimates of group-time average treatment effects on the treated when treatment effects are heterogeneous across groups, time, or exposure. We apply this method to the data used by Anderson et al. to estimate the average treatment effect of filtration on infant mortality.

We aggregate our estimated group-time treatment effects by the length of exposure to treatment,

[2]While the papers proposing this and other modern difference-in-differences estimators were published before Anderson et al., long publishing timelines and the novelty of the methods likely explain their reliance on the standard estimator.

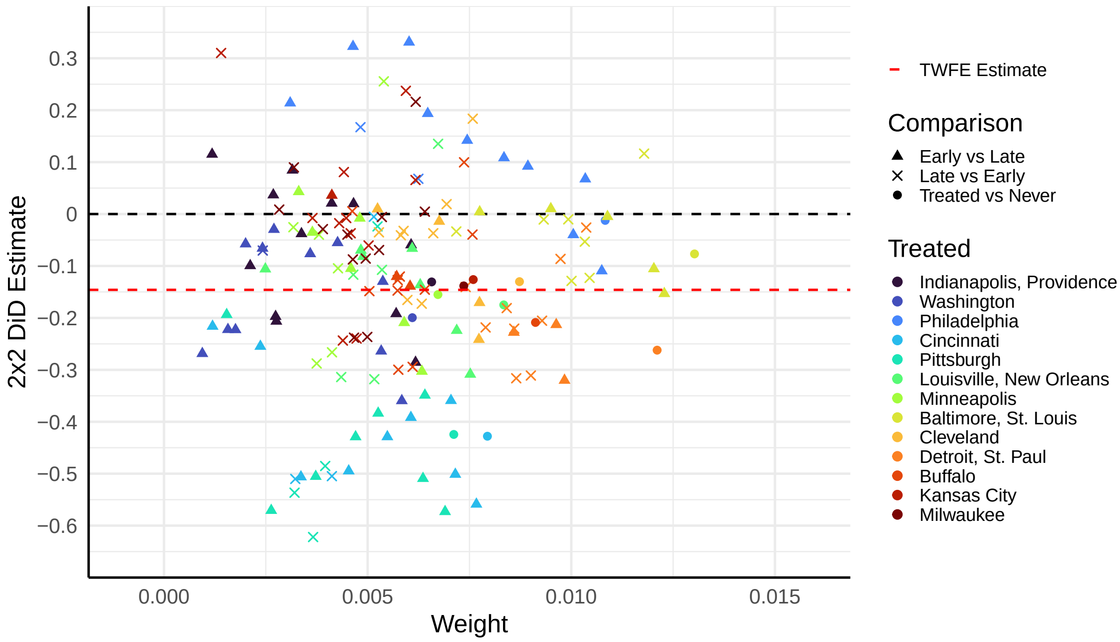

Figure 3: Goodman-Bacon Decomposition for Filtration and Typhoid Mortality

similarly to what Anderson et al. do in Figure 1. These estimates are presented in Figure 4 and Table 5. Consistent with Anderson et al. (2022), infant mortality outcomes are specified in natural logs of mortality rates.

Our results do not show pre-treatment trends. In contrast, the post-treatment effects are uniformly negative in the five years post-treatment, which provides some evidence in favor of the claim that filtration reduced infant mortality. However, and unlike Anderson et al., none of our treatment effects are statistically significant (the error bars show bootstrapped 90% confidence intervals).

To further investigate potential treatment effect heterogeneity, we aggregate our treatment effects by group. These aggregates are displayed in Figure 5 (with bootstrapped 90% confidence intervals, as in Figure 4) and Table 6. While two groups[3] show statistically significant reductions in infant mortality as a result of water filtration, the treatment effect is not significant for nearly half and is positive for five of them.[4] The wide variation in the estimated effects is a reminder that, as we noted in our discussion of concerns, the use of a single "filtration" dummy masks important differences in the nature of the intervention across cities. And this heterogeneity in what our treatment variable *signifies* means that aggregating treatment effects across groups will not give an easily interpretable result.

We also use the Callaway and Sant'Anna estimator to investigate the effect of water filtration

[3]Cleveland in 1918 and Milwaukee in 1939.

[4]Pittsburgh in 1908, Louisville and New Orleans in 1909, Minneapolis in 1913, Baltimore and St. Louis in 1915, and Kansas City in 1928.

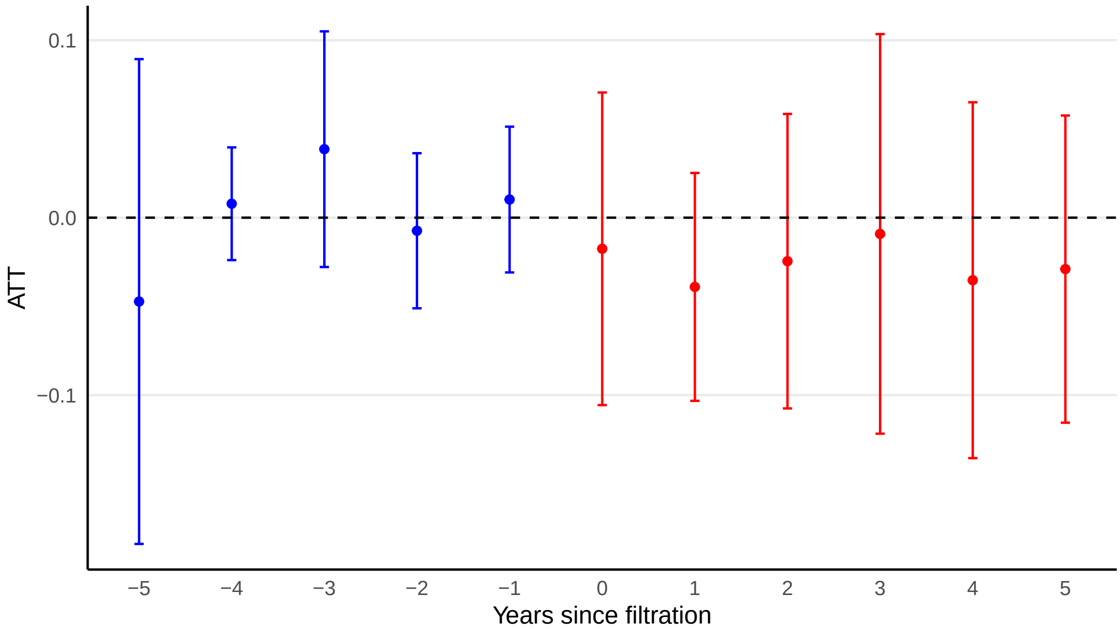

Figure 4: Average Treatment Effect on Infant Mortality by Length of Exposure

Table 5: Average Treatment Effect on Infant Mortality by Length of Exposure

	(1)	
	Est.	(std.error)
-5 years	-0.047	(0.060)
-4 years	0.008	(0.014)
-3 years	0.039	(0.029)
-2 years	-0.007	(0.019)
-1 year	0.010	(0.018)
0 years	-0.018	(0.039)
1 year	-0.039	(0.028)
2 years	-0.025	(0.037)
3 years	-0.009	(0.050)
4 years	-0.035	(0.044)
5 years	-0.029	(0.038)
Observations	25	

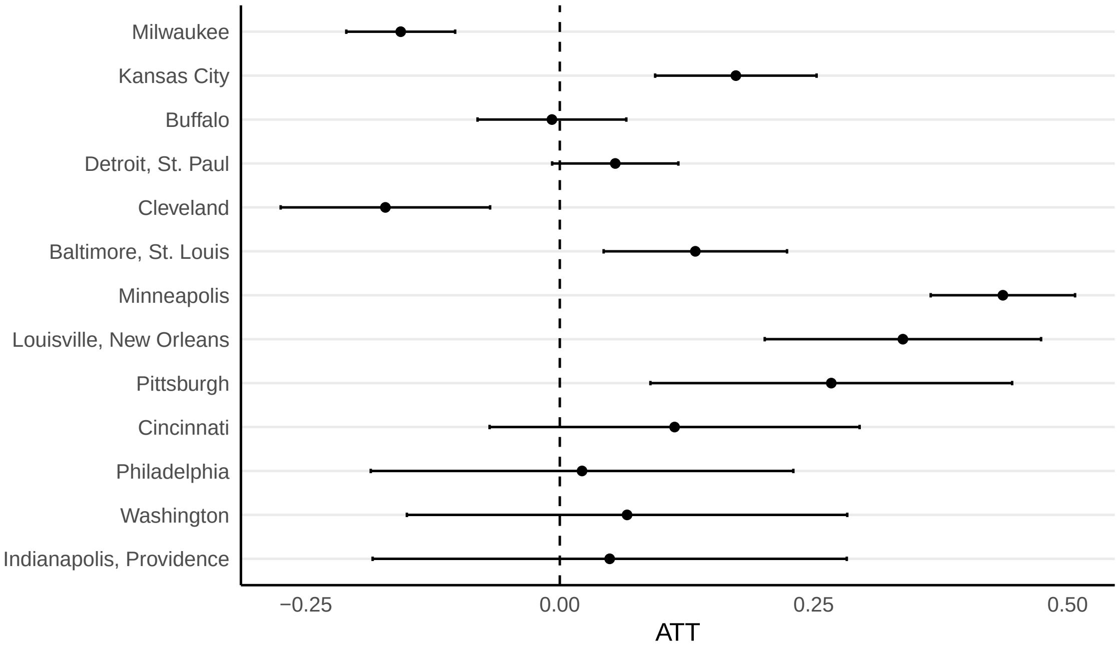

Figure 5: Average Treatment Effect on Infant Mortality by Treatment Cohort

Table 6: Average Treatment Effect on Infant Mortality by Treatment Cohort

	(1)	
	Est.	(std.error)
Combined	0.084	(0.062)
Indianapolis, Providence	0.049	(0.120)
Washington	0.066	(0.111)
Philadelphia	0.022	(0.107)
Cincinnati	0.113	(0.094)
Pittsburgh	0.267	(0.091)
Louisville, New Orleans	0.338	(0.070)
Minneapolis	0.436	(0.037)
Baltimore, St. Louis	0.133	(0.046)
Cleveland	-0.172	(0.053)
Detroit, St. Paul	0.055	(0.032)
Buffalo	-0.008	(0.038)
Kansas City	0.173	(0.041)
Milwaukee	-0.157	(0.028)
Observations	25	

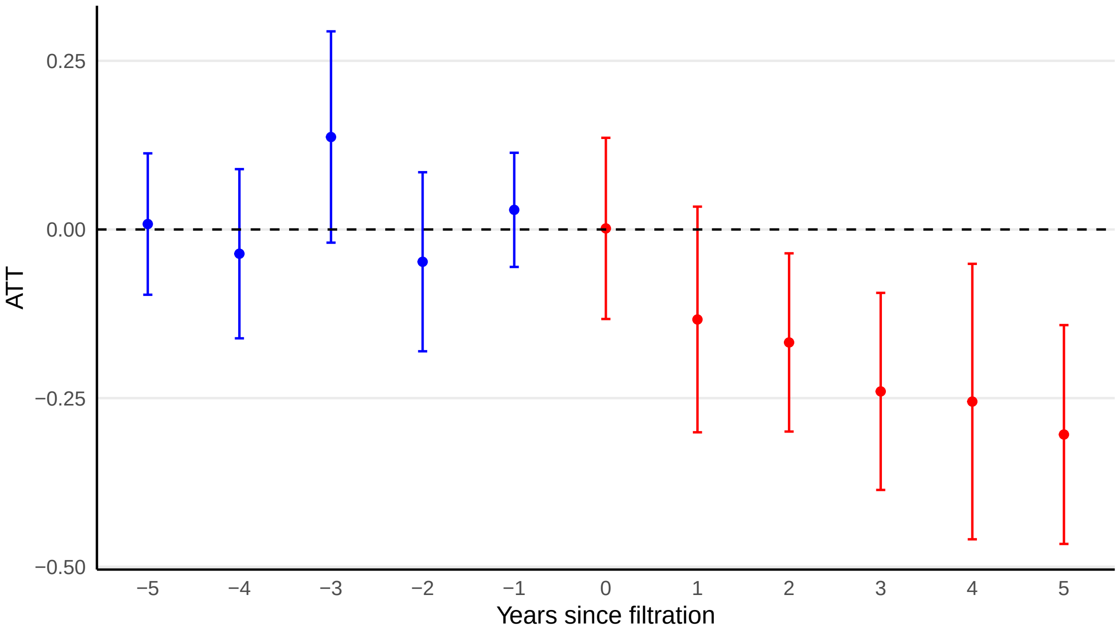

Figure 6: Average Treatment Effect on Typhoid Mortality by Length of Exposure

Table 7: Average Treatment Effect on Typhoid Mortality by Length of Exposure

	(1)	
	Est.	(std.error)
-5 years	0.008	(0.049)
-4 years	-0.036	(0.058)
-3 years	0.137	(0.073)
-2 years	-0.048	(0.061)
-1 year	0.029	(0.039)
0 years	0.002	(0.062)
1 year	-0.133	(0.077)
2 years	-0.167	(0.061)
3 years	-0.240	(0.068)
4 years	-0.255	(0.095)
5 years	-0.304	(0.075)
Observations	25	

on typhoid mortality.[5] Aggregating by exposure, we see a statistically significant reduction in typhoid mortality beginning two years after the construction of a water filtration plant. When we aggregate by treatment cohort, we see that the effect is negative for each group and statistically significant for all but four.[6] This suggests that the effect of water filtration on typhoid mortality is robust, in contrast to its effect on infant mortality.

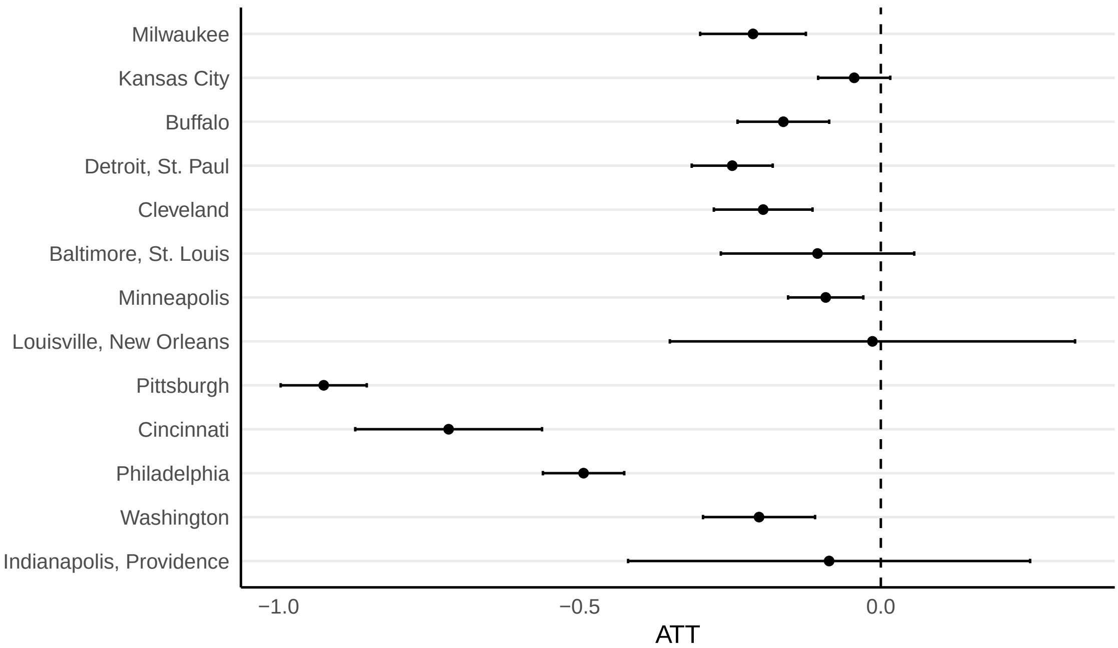

Figure 7: Average Treatment Effect on Typhoid Mortality by Treatment Cohort

6 *Discussion of Findings*

There are many possibilities as to why different cities display different patterns of treatment effects. As discussed by Anderson et al. (2022), treatment effects may evolve over time due to learning effects, where individuals gradually adopt and more effectively utilize improved water sources, or due to crowding-out mechanisms, where investments in filtration substitute for other public health interventions. Heterogeneity across cities may arise from differences in the implementation of water filtration across cities as well as variation in the populations served by the interventions.

The disparity between our reanalyses of infant and typhoid mortality likely results from differences in effect sizes and causal mechanisms. For typhoid mortality, the plausible mechanism is direct: typhoid is a waterborne disease caused by bacteria—water filtration lowers the bacterial load of drinking water and should therefore reduce transmission. This is reflected in both the original estimates and our reanalysis, which show large and statistically significant declines in

[5]Following Anderson et al., we use the quartic root of typhoid mortality rather than the natural logarithm because typhoid mortality is zero for some city-years.

[6]Kansas City in 1928, Baltimore and St. Louis in 1915, Louisville and New Orleans in 1909, and Indianapolis and Providence in 1904.

Table 8: Average Treatment Effect on Typhoid Mortality by Treatment Cohort

	(1)	
	Est.	(std.error)
Combined	-0.282	(0.068)
Indianapolis, Providence	-0.086	(0.168)
Washington	-0.202	(0.047)
Philadelphia	-0.494	(0.034)
Cincinnati	-0.718	(0.078)
Pittsburgh	-0.925	(0.036)
Louisville, New Orleans	-0.014	(0.170)
Minneapolis	-0.092	(0.032)
Baltimore, St. Louis	-0.105	(0.081)
Cleveland	-0.195	(0.041)
Detroit, St. Paul	-0.247	(0.034)
Buffalo	-0.162	(0.038)
Kansas City	-0.044	(0.030)
Milwaukee	-0.212	(0.044)
Observations	25	

typhoid mortality following filtration.

However, as noted by Anderson et al., typhoid mortality represented a relatively small share of overall mortality and an even lower share of infant mortality.[7] Even large reductions in typhoid mortality would translate into relatively small changes in aggregate or infant mortality. This helps explain why, despite strong causal effects on typhoid mortality, we do not observe similarly robust or statistically significant effects on infant mortality.

To add further context, we examined mortality dynamics at the city level by comparing each treated city's mortality rate trajectory to that of the population-weighted average of untreated and not-yet-treated cities (Figures 8 and 9). These comparisons reveal a clear pattern for typhoid mortality: in most cases, mortality in treated cities is greater than mortality in the control group at the moment of treatment and subsequently declines at least as rapidly, if not more so, following the adoption of filtration. Neither of these facts are true for infant mortality rates in treated cities, which noticeably diverge from the downward trend in untreated mortality in the final decade of the sample.

[7] Infants were far less exposed to the primary vector of typhoid spread—contaminated drinking water—due to the low adoption of baby formula during the sample period.

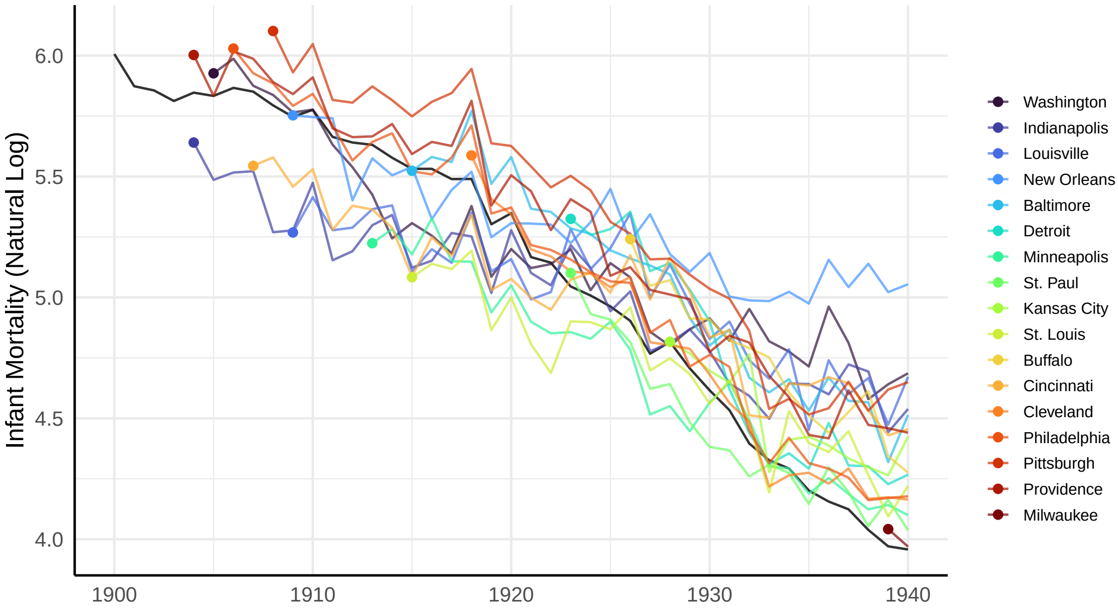

Figure 8: Post-Treatment Infant Mortality and the Untreated Average

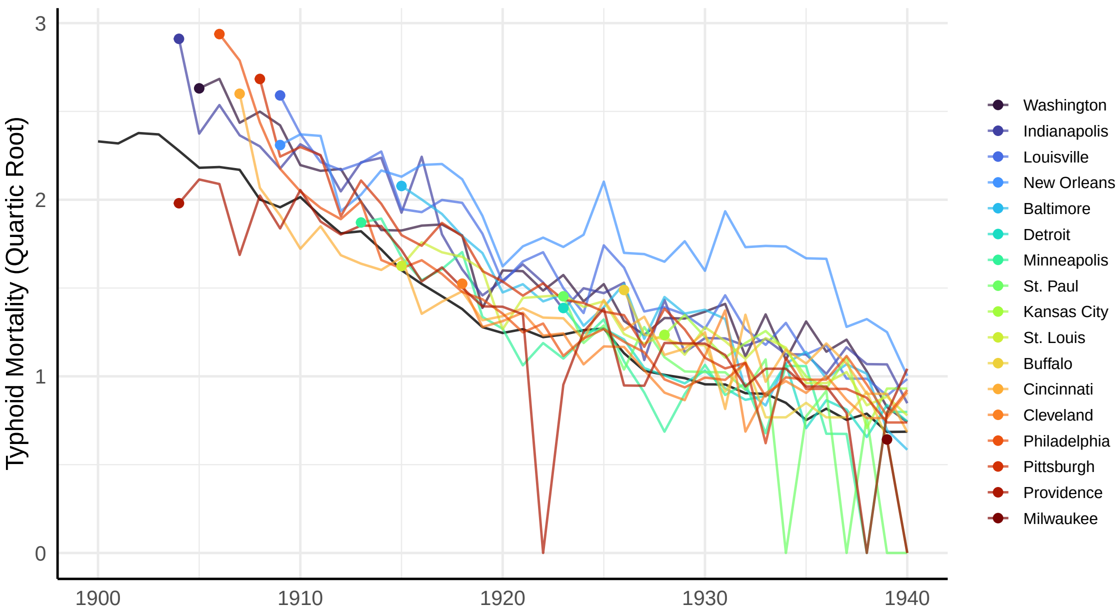

Figure 9: Post-Treatment Typhoid Mortality and the Untreated Average

7 Conclusion

Anderson et al. reexamine the conclusions of Cutler and Miller's study on the role of public health interventions in the American urban mortality transition during the turn of the 20th century. They find that most interventions had no significant effect on total or infant mortality, but that water filtration did have a significant effect on infant mortality, albeit a smaller effect than the one found by Cutler and Miller. They also document large and statistically significant reductions in typhoid mortality, consistent with the direct effect of improved water quality on waterborne disease. However, both studies rely on a two-way fixed effects estimator that is biased when confronted with heterogeneous treatment effects. To combat this, we apply a novel estimation procedure developed by Callaway and Sant'Anna. We find evidence of effect heterogeneity across groups, implying that the estimates of Anderson et al. are biased. When aggregating treatment effects by length of exposure, we do not find a significant effect of filtration on infant mortality. In contrast, when applying the same estimator to typhoid mortality, we continue to find statistically significant effects of similar magnitude, indicating that these results are robust to treatment effect heterogeneity.

Taken together, our results suggest that the role of public health interventions in the 20th century mortality decline may be even more limited than Anderson et al. claim, and that differences in interventions between cities makes it difficult to draw general conclusions about the impact of public health measures, like filtration, on mortality.

8 References

Anderson, D. M., Charles, K. K., and Rees, D. I. (2022). "Reexamining the Contribution of Public Health Efforts to the Decline in Urban Mortality." *American Economic Journal: Applied Economics*, vol. 14, no. 2, pp. 126–157. DOI: 10.1257/app.20190034.

Arel-Bundock, V. (2022). "`modelsummary`: Data and Model Summaries in R." *Journal of Statistical Software*, vol. 103, no. 1, pp. 1–23. DOI: 10.18637/jss.v103.i01.

Bergé, L. (2018). "Efficient Estimation of Maximum Likelihood Models with Multiple Fixed-Effects: The R Package FENmlm." *CREA Discussion Papers*, no. 13.

Butts, K. (2021). *`did2s`: Two-Stage Difference-in-Differences Following Gardner (2021)*. URL: https://github.com/kylebutts/did2s/.

Callaway, B., and Sant'Anna, P. H. C. (2021a). *`did`: Difference in Differences*. URL: https://bcallaway11.github.io/did/.

Callaway, B., and Sant'Anna, P. H. C. (2021b). "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, vol. 225, no. 2, pp. 200–230. DOI: 10.1016/j.jeconom.2020.12.001.

Costa, D. L. (2015). "Health and the Economy in the United States from 1750 to the Present."

Journal of Economic Literature, vol. 53, no. 3, pp. 503–570. DOI: 10.1257/jel.53.3.503.

Cutler, D. M., and Miller, G. (2005). "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States." *Demography*, vol. 42, no. 1, pp. 1–22. DOI: 10.1353/dem.2005.0002.

Gardner, J. (2022). *Two-Stage Differences in Differences*. URL: https://arxiv.org/abs/2207.05943.

Goodman-Bacon, A. (2021). "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, vol. 225, no. 2, pp. 254–277. DOI: 10.1016/j.jeconom.2021.03.014.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. URL: https://www.R-project.org/.

Sun, L., and Abraham, S. (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, vol. 225, no. 2, pp. 175–199. DOI: 10.1016/j.jeconom.2020.09.006.

Wickham, H. (2016). *`ggplot2`: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL: https://ggplot2.tidyverse.org.

Wickham, H. (2023). *`stringr`: Simple, Consistent Wrappers for Common String Operations*. URL: https://stringr.tidyverse.org.

Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). *`dplyr`: A Grammar of Data Manipulation*. URL: https://dplyr.tidyverse.org.

Wickham, H., Miller, E., and Smith, D. (2023). *`haven`: Import and Export "SPSS," "Stata" and "SAS" Files*. URL: https://haven.tidyverse.org.

Wickham, H., Vaughan, D., and Girlich, M. (2024). *`tidyr`: Tidy Messy Data*. URL: https://tidyr.tidyverse.org.

9 *Appendix*

In addition to the Callaway & Sant'Anna estimator, we considered the Gardner (2022) and Sun & Abraham (2021) estimators as well. The Gardner estimator uses binned years for the event study coefficients to enable bootstrapping ($n = 1000$) of the 90% confidence intervals. All three methods show directionally similar results (Figures 10 and 11).

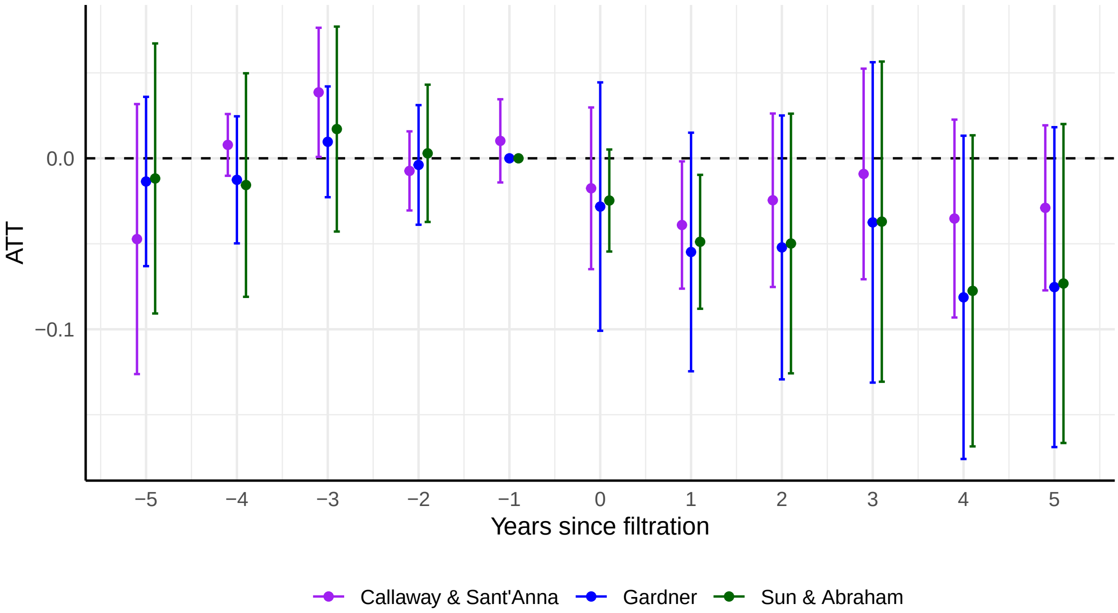

Figure 10: Comparison of Estimators (Infant Mortality)

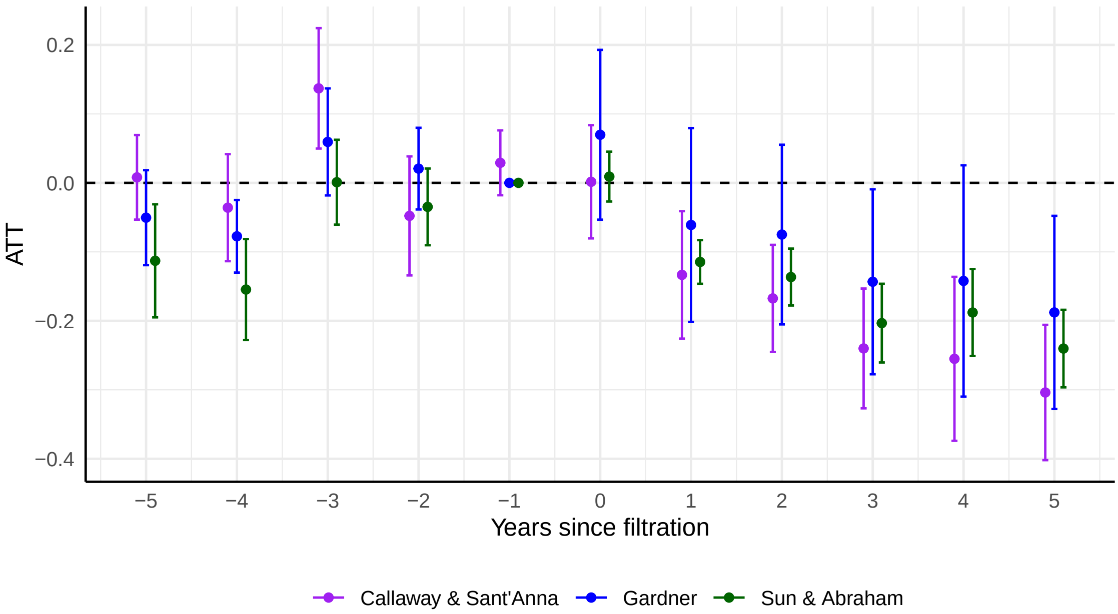

Figure 11: Comparison of Estimators (Typhoid Mortality)