THE PUBLIC HEALTH EFFECT OF MANDATORY AND VOLUNTARY FOOD SAFETY MEASURES: GENERALIZED SYNTHETIC CONTROL METHODS ON MILK PASTEURIZATION IN THE UNITED STATES

By

Huiqiang Wang
University of Rhode Island
huiqiang_wang@my.uri.edu

Thomas W. Sproul
University of Rhode Island

Corey Lang
University of Rhode Island

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting, Minneapolis, MN, July 27-29, 2014.

Abstract

This paper investigates the public health effects of mandatory and voluntary milk pasteurization adoptions in the United States. With generalized synthetic control methods (GSCM), we estimate the casual health impact of city pasteurization ordinances in 16 treated cities. Our results show the average treatment effect of city with pasteurization ordinances was obvious. Children diarrhea mortality rates on average in the treated group is 7.62 lower than their synthetic versions. And the number was only 1.04 in the control group. The treatment effect measured by GSCM is also different from the results measured by classical linear based Differences-in-Differences model. Next, we implement a FE analysis to examine the effect of voluntary pasteurization adoptions. We choose robust FE estimators to minimize the influence of observations of extreme values. Also, our robust panel estimations indicate the spread of pasteurization was significantly responsible for the improved health profile in US cities. In sum, we find pasteurization had its unique and significant contribution to public health in history.

Keywords: Public Health, Pasteurization, Generalized Synthetic Control; Robust Estimators

1. INTRODUCTION

Are mandatory measures necessary to provide food safety controls? Or market-based voluntary adoptions are more efficient than those "command-and-control" measures in providing consumer protections. This question deserves much attention in making food safety policies. And its theoretic framework has been well discussed in previous literatures (Antle 1995; Henson and Caswell 1999; Segerson 1999; Fares and Rouviere 2010). But, empirical studies that examine the health effects of those mandatory and voluntary measures are still rare. To understand the roles of those measures, we conduct a case study to show how proper policy evaluations can be made. With historical records of city milk pasteurization ordinances and changes in the extent of this new technology, this paper highlights the complexity in evaluating the health effect of food safety measures in the real world.

As a remarkable public health innovation, milk pasteurization played a critical role in the fight of milk diseases. Historically, biological contaminations made milk was a serious health threat, particularly to infants and children. As a solution, pasteurization was introduced from the late 1800s. In practice, it was attractive since it can effectively eliminated pathogens and with very small physical changes and nutritional losses (Kelly and Clement 1931). Also, compared with other methods, pasteurization was much cheaper and very preferable for large scale commercial liquid milk production. In the US, city milk depots that were supported by philanthropists began to provide pasteurized milk to the public from the early 1890s (Selitzer 1976). Remarkably, milk related epidemics were much controlled, especially the death of children diarrhea (Parker 1917). So pasteurization was recommended as a method that is "practically feasible to keep milk clean and pure" and the "essential safeguard" (Winslow 1952).

However, this innovation has been very controversial since its early commercial success. In fact, the pasteurization story also has many parallels in our modern food systems. One major concern was its health effect. As a food technology innovation, it is interesting to know whether the market was able to provide adequate protection to consumers. If mandatory policy interventions were introduced, how can we measure its effect? In history, pasteurization stands out other food safety innovations as it was an obvious mix-up of both mandatory and voluntary measures. It is a good example to show how the "co-regulation task" (Fares and Rouviere 2010) was shared by city health departments and private milk suppliers. From voluntary commercial application to mandatory city ordinances, the pasteurized milk switched from a technology to city health policy. From the early 1910s, cities like Chicago, New York and Philadelphia began to adopt their pasteurization ordinances and required all milk be pasteurized before sales[1]. In the US, the extent of pasteurization in city milk supply was generally improved in the 1920s. More cities followed this trend even without mandatory ordinances. Even some literature discussed the health effect of those mandatory or voluntary adoptions of pasteurization, less enough quantitative discussions have been made.

The purposes of this paper are four folds. First, we want to measure the health effect of city pasteurization ordinances. As many policies, pasteurization laws have been every controversial and are still a center in public debate. In the US, 30 states allow unpasteurized milk sales, while 20 states prohibit them. Second, we measure the role of voluntary pasteurization practices or the extent of this technology and its health implications. Historical facts remind us it may be a more

[1] More exactly, most ordinances required all milk but certified needs to be pasteurized before sale. But the share of certified milk was quite small in total milk supply. See Block (1999; 2009).

rigorous approach to illustrate the contribution of this technology than just focusing on policies. Third, we generalize the nonparametric synthetic control methods (SCM) to multiple treated units and compare its result with traditional linear regression based Difference-in-Difference method (DD). Last, we provide a case to examine the effect of robust estimator with panel data. It is a key point in our Fixed-Effects analysis to the extent of pasteurization and health outcomes.

Our first step is to capture the causal health effect of city pasteurization ordinances from the 1900s to 1930s. This identification is technically demanding for appropriate data and empirical strategies. We have two major problems, namely how to control unobserved heterogeneity and how to select proper comparison units. Using decennial census data, unobserved individual characteristics are a serious concern to the validity of DD methods and other ordinary least square (OLS) based methods. As a popular microeconometric method, DD relies on the strong assumption that such unobserved factors should be independent of both group status (treated or control) and time period (pre or post-intervention). Thus, such factors will be differenced out across periods and groups. However, this assumption could be less likely satisfied in this study. SCM, on the other hand, allows such factor to exist. It constructs a synthetic version to it using other control units. Moreover, DD is a "population average difference" (Imbens 2007). And SCM is the "population optimized difference". By matching the similarities in control variables, SCM put the optimized weights to control units instead of the average weight. Theoretically, SCM is a combination of both DD and matching method. With 16 treated cities (with ordinances) and 52 control cities (without), we study how children diarrhea mortality rates changed. SCM results show the real rate was 7.62 lower than its synthetic version in treated cities, while the number was only 1.04 in the control group.

Moreover, historical facts remind us the market could work better in adoption of new food safety innovations. That is to say mandatory regulation may be not the only key in improving the extent of pasteurization. Other incentives made milk suppliers tend to pasteurize milk voluntarily, for example to preserve milk and save their losses[2]. Considering such forces, more attention is given to the extent of pasteurization in this paper. With two cross-sectional surveys to the extent of pasteurization in 1921 and 1924, we made a FE estimation to the health effect of the spread of pasteurization. Here we met a new problem how to reduce the influence of outliers. Using historical data, we noticed that some observations were from different data generating processes (DGPs). These contaminated data could violate least-square based FE estimation (Verardi and Croux 2009; Vogel and Wanger 2011). For remedy, we chose robust estimators detect outliers and then make regressions. Our results show the results of median regression, M-estimator, S-estimator and MM-estimator are also consistent. Compared with regular FE models, they have much lower standard error. It reminds us we should care outliers in choosing empirical strategies. In sum, all our estimations suggest the beneficial health effect of pasteurization.

Our article proceeds as follows. It begins in Section 2 and reviews the adoption of pasteurization ordinances across cities. Also, it illustrates how this technology spread in early 1920s. Section 3 presents the data and summary statistics. The empirical model for SCM analysis is introduced in Section 4. It discusses how SCM could be used in the case of multiple treated units. Section 5 is the FE regressions with robust estimators. It focuses on health effect of voluntary pasteurization adoptions. Finally, our major findings and conclusions are summarized in Section 6.

[2] Ayers (1922, 1926, 1930) described the different purposes between milk dealers and public health experts in early uses of pasteurization. To milk dealers, pasteurization was for the "sole purpose of preserving the milk". But health experts were more concerning about its bacteria sterilizing function.

2. BACKGROUND

From the late 19th to early 20th century, clean and safe urban milk supply was regarded as one critical factor in the process of improved public health. Voluminous literature suggests the role of milk pasteurization can never be overestimated. From 1880s, it has been successfully applied to commercial liquid milk production. The obvious attractiveness is that this technology worked effectively to sterilize pathogens in milk and with minimum changes to milk sensory and physical attributes (Hall and Trout 1968). Also, compared with other options, it was cheaper and very preferable for large-scale production (MacNutt 1917; Kelly and Clement 1931). On contrary, another option, the certified milk, was much less population. Certification from medical milk commission made milk was far from a cheap source of nutrient (Block 1999). It was a lot more expensive and cost double as the regular milk (Lentzner 1987).

In the US, commercial pasteurization began in 1890s when city milk supply was seriously riddled with adulterations and biological contaminations. To the public, lower quality was a serious health threat, particularly children and infants. It usually caused extraordinarily high children mortality of diarrhea, tuberculosis, scarlet fever and other milk diseases. This problem exacerbated when more female labor were employed and artificial feeding became increasingly popular (Beaver 1973; Meckel 1990; Vögele and Woelk 2002). As a result, early city milk depots that sponsored by generous philanthropists began to provide pasteurized milk to the public (Selitzer 1976), while it was also operated secretively by milk dealers to preserve the milk (Ayers 1922). However, the share of pasteurization was still not quite noticeable around the early 1900s (Alvord and Pearson 1903). A 1902 survey of the United States Department of Agriculture (USDA) indicated there were only 18 cities had exact reported share of pasteurized milk. Only

two cities have 50% or more milk pasteurized. It was even much lower in other cities. However, things changed in the 1910s. In the study of Jordan (1913), the share of pasteurization rose from 5% to 40% in New York from 1902 to 1912. In Boston, the number increased from zero to 75%. Similar changes also occurred in smaller cities (Jordan 1913).

Compared with commercial applications, regulations about pasteurization came relatively later. The first city adopted pasteurization ordinance was Chicago in 1908. But it was subsequently banned by the state of Illinois. Its full pasteurization law was not activated until in 1916 (Wang, Sproul and Lang 2013). New York City adopted its pasteurization ordinance in 1912. In next years, other cities also introduced their pasteurization ordinances, for example Philadelphia, Cleveland, Detroit and Buffalo *et al.* (Table I). Actually, the relation between the extent of pasteurization and regulation did vary a lot across cities. For example, some cities did encourage it even without ordinance on the paper. So it was not rare that the extent of pasteurization was not that low in the pre-intervention period. Similarly, the enforcement of this regulation was also bended in some cases. As a result, the extent may be lower than expected[3]. However, the share of milk pasteurized was generally over 90% in cities with such interventions. It was much lower in those without city ordinances.

As an important public health policy intervention, it is meaningful to understand the real effect of those pasteurization ordinances in that time. Thus, one major objective of this paper is to capture the casual health effect of this intervention. Empirically, however, we have two concerns, i.e. unobserved heterogeneity and less appropriate counterfactual. They decide the validness of our

[3] Similar non-all-or-nothing option also exists in other policy interventions. Cutler and Miller (2005) explained why their urban water treatment was not in the all-or-nothing pattern.

estimations. First, unobserved heterogeneity increases the unexplained part in the variation (Cameron and Trivedi 2005). Ordinary least square (OLS) estimator has no direct control to it. For panel data, either FE or DD model has some control to it. Classical linear DD and other FE based methods rely on a strong assumption that unobserved factors should be independent of both group status (treated or control) and time period (pre or post-intervention). Thus, such factors will be differenced out across periods and groups. But this assumption is less likely satisfied with decennial data in our study, considering those dramatic changes occurred in the early 1900s. But it could be violated easily in our study period, e.g. 1900s to 1930s. This period was full of sharp and quick changes, including individual health behavior changes, clean water technology, medical innovations and other major city health campaigns (Cutler and Miller 2005). Compared with other factors, the role of city clean water treatment was even more influential. It was responsible for the drop in mortality rate of typhoid fever and other diseases, just like clean milk products did[4]. To explain the health effect of pasteurization ordinance, we need to squeeze out its influence. Of course, we can obtain some water treatment information in some cities, but not everyone. In most control cities, we don't know their status of water treatment. Second, a similarly important issue is how to choose proper comparison units for policy evaluations. Less appropriate control group will lead to inaccurate result (Abadie, Diamond and Hainemuller 2012). In fact, DD estimator measures the "population average difference" (Imbens 2007). And, most existing econometric methods are not helpful to this concern[5]. It is desirable if we can construct an appropriate control group and measure the "population optimized difference".

[4] Prior studies suggest two clean water technologies were generally used before 1930s. They were filtration and chlorination. The first was found significantly associated with the drop of infant and children mortality rates before 1940. But chlorination did contribute much to such drops (Cutler and Miller 2005).

[5] Propensity score matching (PSM) is an exception. In essence, our SCM approach is a combination of both DD and PSM, details in Abadie et al. (2010).

To stress these concerns, we choose nonparametric data driven synthetic control methods to estimate the effect of city ordinances. SCM builds a synthetic version for a real unit with matching on selected covariates and minimizing pre-intervention distance between the real and its synthetic version. And policy effect is the post-intervention difference between these two versions. With the matching process, the weights for control units are then optimized. That is why the SCM result is the "population optimized difference". Originally, SCM was designed for comparative case studies. But we extended it to multiple treated units (generalized synthetic control methods or GSCM). So our interest is the average treatment effect on the treated (ATT). More than mandatory regulations, we are also interested to understand how voluntary adoption of milk pasteurization changed the public health outcome. With holding other factors constant, we estimate the health outcomes that were caused by variations in the extent of pasteurization. In this process, a potential problem is the influence of outliers (observations with extreme values). It appears a wide concern in historical studies, especially when we are less certain about the reliability of data collection techniques. As a remedy, we choose robust panel data estimators to illustrate the casual connection between pasteurization and health outcomes.

In sum, we have three contributions to existing knowledge. First, we use GSCM to examine the health effect of city pasteurization ordinances. Also, we compared the predictive power of SCM with DD estimator. Second, we measured the effect of voluntary adoption of pasteurization with robust panel regressions. The results are different from regular OLS based FE models. Third, we discussed which measure is more appropriate to reflect the exact health effect of pasteurization. Regarding historical evidence, voluntary measure seems a better option in this study, at least in the period of early 1920s.

3. DATA

In the part of treatment evaluation, our study covers the period1900 to 1930 which witnessed the rapid expansion of pasteurization. Our dependent variable is the annual city children diarrhea and enteritis mortality rate under 2 years[6]. In this study, it is calculated as the number of death every 100,000 population per year. They are obtained from Census of Mortality Statistics 1900 to 1930. Population data is from decennial Census of Population 1900, 1910, 1920 and1930. The year 1900 is the first year we can obtain city mortality statistics in the US. And, we found a 1931 survey that have records of the extent of pasteurization and the *status quo* of city ordinances. It is a good ending point to discuss the policy effect.

Our treatment group includes 16 cities (Table I), including Baltimore, Buffalo, Chicago, Cincinnati, Cleveland, Dayton, Detroit, Indianapolis, Jersey City, Milwaukee, New York, Philadelphia, Richmond, St. Louis, San Francisco, and Toledo. They adopted their ordinances 1912 to 1918. For the control group, there are 52 cities. They had not adopted any pasteurization laws by 1931 (Frank and Moss, 1931). Also, their extents of pasteurization were all lower than 90%[7]. On average, the extent of pasteurization in the control group was 65.6% in 1931. And it was 99.1% in the treated group.

The SCM operation requires matching and the treated unit and its control counterparts with their similarities in selected covariates or "predictors" (Abadie, Diamond and Hainemuller 2010). Our SCM setup considers city average income, female share, white share, share of population under 5

[6] Our dependent variable is a good indicator to milk quality (Lee 2007). In addition, prior studies like Cutler and Miller (2005) have not discussed the effect of major public health campaign on this variable.

[7] The extent of pasteurization in cities with mandatory ordinances was definitely 90% and above. The level in most control cities was lower than that. But some control cities still have 90% or more. Thus, they were not included.

years, and population growth rate. In addition, we incorporate pre-intervention outcomes (lagged values) to reflect the dynamic correlation of our outcome of interest. Demographic data are obtained from decennial Census of Population 1900, 1910, 1920 and 1930. Income data is the average wage of worker in manufacture from the Census of Manufacture 1900, 1909, 1919 and 1929. Population growth rate is the percentage of population increase from 1900 to 1910.

Our first round SCM is done with intervention in 1910. It shows that those ordinances occurred in 1912 to 1918 (Table I). But our data are on decennial basis. And we don't have annual data. So 1900 and 1910 are the two points available for pre-intervention data. The synthetic version will be built with matching all covariates on the values in 1900 and 1910 (population growth rate and share under 5 years are only 1910 data). We also include two individual outcomes of 1900 and 1910. Post-intervention observations are 1920 and 1930. And the policy effect is measured as the real-synthetic gap (difference between the real unit and its synthetic version) in 1930. As our dependent variable is children mortality rate, the real value is supposed to be lower than its synthetic version (a negative gap). In addition, we noticed enforcement of city ordinance was somewhat bended in some cities. To make a more rigorous measure, we do a similar SCM again with intervention occurred in 1920. That is the pre-treatment period covers 1900, 1910 and 1920. And post-treatment outcome is only 1930[8].

To capture the effects of voluntary adoptions, we make a robust FE regression to the extent of pasteurization and health outcome, with holding all other variables constant in early 1920s. We are inspired by the historical fact that the application of milk pasteurization increased a lot in this

[8] There will be some changes in the value of predictors, if intervention occurred in 1920. For example, the income, white share and female share are the average of 1900, 1910 and 1920. Population growth rate and share under 5 years are the average of 1910 and 1920. We have 3 pre-treatment outcomes in 1900, 1910 and 1920 (not averaged).

period. It was even more noticeable in those cities without mandatory ordinances. This is the tricky thing because it may lead to some effects that cannot be reflected with treatment dummy in either DD or SCM approach. In other words, the parameter of treatment variable may not tell the whole story if our treated cities and their control counterparts did not follow the same contour in the extent of pasteurization after 1920 (Figure III). For a complete picture, more attention should be given to the role of voluntary measures and examine the extent of pasteurization with persuasive methods.

Some early city level surveys kept the records of the share of pasteurization. They were mainly conducted by the US Department of Agriculture (USDA) and the US Public Health Service (USPHS). USDA did the ones in 1915, 1921, 1924 and 1930. USPHS did 1927, 1931 and 1936. Data were collected by sending questionnaires to city health officers. Published data were only available from USDA 1921, 1924, 1930 and USPHS in 1931[9].

Overall, the extent of pasteurization went up in the 1920s. Firstly, both USDA (Table II) and USPHS (Figure II) data indicated the average share of milk pasteurized was higher in the late 1920s. The increases were particularly more in cities with population under 100,000[10]. The changes in large cities, however, were much smaller. But they still had the highest shares. Second, the increasing trend was high in the early 1920s and the leveled off. It was clear that a few cities had more than half milk pasteurized in 1915. More cities experienced fast growth. It continued until the middle 1920s (Figure III). For example, in the six subgroups of cities, the

[9] Only summary statistics were provided in other years. The status of city pasteurization ordinances was only recorded in USPHS 1931 and 1936.

[10] The case of cities with population 10,000 or less seems odd, with a decrease in the average share. It is possibly because of more small cities were included in the 1924 and 1930 surveys. It "diluted" the average share in 1921.

slope of the share of pasteurization was much higher before 1924. Also, it shows the slope of large cities (e.g. 500,000 above) was clearly lower than others. As we know, many of them had adopted their compulsory pasteurization ordinance by that time. Third, the wide spread of pasteurization was homogenous among regions. That is all regions had more milk pasteurized for city supply, from New England to Pacific (Figure IV). On absolute value, pasteurization was higher in cities of New England, Middle Atlantic and Northern Central. But more increases came from other regions. In 1931, regions had more than half milk pasteurized. East North Central was the highest with 95.6% pasteurization, and West South Central was the lowest at 61.1%.

4. SYNTHETIC CONTOL ANALYSIS

4.1 Specification

The specification of SCM first appeared in Abadie and Gardeazabal (2003), and became more formalized in the papers of Abadie, Diamond, and Hainemuller (2010, 2012)[11]. As mentioned, based on predictors and comparison units, SCM constructs a synthetic version of the unit exposed to policy interventions. This version is built with the same control units but at the optimized weights. For those weights, they are generated by matching covariates and minimizing real-synthetic gap in the pre-intervention period. The synthetic version is a convex combination of its controls. Thus, the effect of policy is transformed as the real-synthetic gap in post-treatment period, which is akin to DD. That is why some researchers treat SCM as a combination of SCM and DD. For inference, SCM users are recommended to use the "placebo tests", or the "cross unit" and "cross time". This approach randomly assigns the treatment to other units or other periods.

In practice, SCM was designed for comparative case studies, namely with only 1 treated unit. In our paper, we generalize SCM to multiple treated units (GSCM). To each treated unit, it follows the SCM process. With more than one treated unit, we can estimate the ATT with our sample. Our GSCM is conducted with a 4-period panel from decennial surveys. First, suppose we have *M* treated cities, each city *i* has its pasteurization ordinance. And there are *N* control cities *j* (*j* = 1, 2 … *N*). Intervention split the study period into pre-treatment period *T* and post-treatment *t*. Thus, Y_{iT} and Y_{jT} are pre-treatment outcomes of *T*. Similarly, Y_{it} and Y_{jt} are the post-treatment results in *t*. $\boldsymbol{X}$ is a vector of predictors (other covariates). For single treated unit *i*, the treatment effect Φ_{it}

[11] See Abadie and Gardeazabal, "The Economic Costs of Conflict"; Abadie, Diamond and Hainmueller, "Synthetic Control Method for Comparative Case Study"; and Abadie, Diamond and Hainmueller, "Comparative Politics and the Synthetic Control Method"

is measured as the difference between its real value and a convex combination of its control units at the optimized weights w_j^*, as Equation (1).

As mentioned, the optimized weight is driven by minimizing the distance D_m between $\mathbf{X}_i$ and $\mathbf{X}_i \cdot \mathbf{W}_j$ in the pre-intervention period T[12], as in Equation (2). With the choice of matrix V, we can minimize different distances.

$$\Phi_{it} = Y_{it} - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \quad \text{with s.t. } w_j^* \geq 0 \text{ and } \sum_{j=1}^{N} w_j^* = 1 \tag{1}$$

$$D_m = \min_{w_j} [(\mathbf{X}_i - \mathbf{X}_j \cdot \mathbf{W}_j)' \mathrm{V} (\mathbf{X}_i - \mathbf{X}_j \cdot \mathbf{W}_j)]^{1/2} \tag{2}$$

Graphically, it seems w_j^* can also be obtained by just minimizing real-synthetic gaps in pre-intervention period as Equation (3). That is to say, we can just use pre-intervention outcomes to estimate the parameter of policy dummy in post-intervention period as in Giannone, Lenza and Reichlin (2010).

$$D_m = \min_{w_j} (Y_{iT} - Y_{jT} \cdot w_j)' (Y_{iT} - Y_{jT} \cdot w_j) \tag{3}$$

In practice, the values of w_j^* may be different from Equation (2) or (3). But this topic is beyond the scope of our paper. Overall, the validness of SCM largely depends on pre-intervention fitting between the real unit and its synthetic version. The constraints w_j^* for is in Equation (4).

$$Y_{iT} - \sum_{j=1}^{N} w_j^* \cdot Y_{jT} \approx 0 \quad \text{with s.t. } w_j^* \geq 0 \text{ and } \sum_{j=1}^{N} w_j^* = 1 \tag{4}$$

For multiple treated units, the treatment framework of GSCM measures ATT. To formalize it, we define the treatment status as 1 or 0 in the parenthesis (1 treated unit, and 0 control unit). Thus, the population value of GSCM is in Equation (5), given good pre-intervention fitting.

[12] Abadie, Diamond and Hainmueller, "Synthetic Control Method for Comparative Case Study" p.496

$$\begin{aligned}\Phi^{SCM} &= E[Y_{it}(1) - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \mid \mathbf{X}_i, \mathbf{X}_j, D_i = 1] \\ &= E[Y_{it}(1) \mid \mathbf{X}_i, D_i = 1] - E[\sum_{j=1}^{N} w_j^* \cdot Y_{jt} \mid \mathbf{X}_j, D_j = 0] \\ \text{IF } & Y_{iT}(1) - \sum_{j=1}^{N} w_j^* \cdot Y_{jT} = 0\end{aligned} \tag{5}$$

As mentioned, SCM is a combination of PSM and DD. Equation (5) shows how SCM is akin to DD. Equations (2) to (4) have explained similarities between SCM and PSM. SCM allocates the optimized weights to comparison units and makes the combination of control units sufficiently close to the treated unit (without intervention). At this point, we can get the average treatment effect on the treated, Equation (6). We can see the unique benefit of SCM that is $Y_{it}(0)$becomes observable, given $D_i = 1$.

$$\begin{aligned}\Phi^{SCM} &= E[Y_{it}(1) - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \mid \mathbf{X}_i, \mathbf{X}_j, D_i = 1] \\ &\approx E[Y_{it}(1) - Y_{it}(0) \mid \mathbf{X}_i, D_i = 1\ ] \\ &= \Phi^{ATT} \\ \text{IF } & Y_{it}(0) - \sum_{j=1}^{N} w_j^* \cdot Y_{jt} \approx 0\end{aligned} \tag{6}$$

To summarize, DD relies on both pre-treatment and post-treatment difference from the treated and control group. SCM, on the other hand, minimizes pre-treatment difference and just focuses on the post-treatment difference. Moreover, weights on control units are also different.

$$\begin{aligned}\Phi^{DD} &= E[Y_{it}(1) - Y_{jt}(0) \mid D_i = 1] - E[Y_{iT}(0) - Y_{jT}(0) \mid D_i = 0] \\ &= E[Y_{it}(1) - \sum_{j=1}^{N} \frac{1}{N} \cdot Y_{jt}(0) \mid D_i = 1] - E[Y_{iT}(0) - \sum_{j=1}^{N} \frac{1}{N} \cdot Y_{jT}(0) \mid D_i = 0] \\ &= E[Y_{it}(1) - \sum_{j=1}^{N} w_j^* \cdot Y_{jt}(0) \mid D_i = 1] \\ &= \Phi^{SCM} \\ \text{IF } & w_j^* = \frac{1}{N}; \text{ and SCM makes } Y_{iT}(0) - \sum_{j=1}^{N} w_j^* \cdot Y_{jT} = 0\end{aligned} \tag{7}$$

With minimizing real-synthetic gap in the pre-intervention period, SCM generate the optimized w_j^*. But DD uses an equal weight w, and $w = 1/N$. Theoretically, DD and SCM result should be different. DD measures "population average difference", while GSCM measures "population optimized difference". Loosely speaking, DD is a special class of GSCM when $w_j^* = w = 1/N$

For the sample value, GSCM is defined as Equation (8). Here we have N control cities and M treated cities. GSCM measures the averaged real-synthetic gaps at period t.

$$\hat{\Phi}^{SCM} = \frac{1}{M} \cdot \sum_{i=1}^{M} [Y_{it}(1) - \sum_{j=1}^{N} w_j^* \cdot Y_{jt}(0) \mid D_i = 1] \tag{8}$$

A further point, GSCM of Equation (8) measures the averaged treatment effect of 16 treated units in the post-intervention period. And, prior literature suggested an alternative way to do SCM with multiple treated units. That is we can average of the treated group first and treat it as one unit. And we can do the single treated SCM (Abadie et al. 2012). Equation (9) and (10) show this subtle difference between these two approaches.

$$\begin{aligned} \hat{\Phi}^{SCM} &= \frac{1}{M} \cdot \sum_{i=1}^{M} [Y_{it}(1) - \sum_{j=1}^{N} w_j^* \cdot Y_{jt}(0) \mid \mathbf{X}_i, \mathbf{X}_j, D_i = 1] \\ &= \frac{1}{M} \cdot \sum_{i=1}^{M} Y_{it}(1) - \frac{1}{M} \cdot \sum_{i=1}^{M} \sum_{j=1}^{N} [w_j^* \cdot Y_{jt}(0) \mid \mathbf{X}_i, \mathbf{X}_j, D_i = 1] \\ &= \bar{Y}_{it}(1) - \frac{1}{M} \cdot \sum_{i=1}^{M} \sum_{j=1}^{N} [w_j^* \cdot Y_{jt}(0) \mid \mathbf{X}_i, \mathbf{X}_j, D_i = 1] \end{aligned} \tag{9}$$

Equation (10) is the outcome with using averaged treated as a comparative case study, namely take the average of the treated group first and then use SCM.

$$\begin{aligned} \hat{\bar{\Phi}}^{SCM} &= \bar{Y}_{it}(1) - \sum_{j=1}^{N} [w_j'^* \cdot Y_{jt}(0) \mid \bar{\mathbf{X}}_i, \mathbf{X}_j, D_i = 1] \\ &= \bar{Y}_{it}(1) - \sum_{j=1}^{N} [w_j'^* \cdot Y_{jt}(0) \mid \bar{\mathbf{X}}_i, \mathbf{X}_j, D_i = 1] \end{aligned} \tag{10}$$

The first term in Equation (9) and (10) is the same, the averaged treated value in post-treatment. Their difference largely comes from the second term. More exactly, GSCM in Equation (9) relies on w_j^* , which was optimized by $\boldsymbol{X_i}$. But the second method depends on w'^*_j, Equation (10). It was driven by $\overline{\boldsymbol{X}}_{\boldsymbol{i}}$ (the averaged predictors in the treatment group). Thus, the results of Equation (9) and (10) are different. We will explain this point later.

Our GSCM estimation was carried out as follows, namely building synthetic versions, make inference and do robustness tests.

(1) We use SCM to very treated unit. We suppose the treatment occurred in 1910[13]. We use 52 control cities to construct the synthetic versions for 16 treated cities. Our predictors include the averaged demographic variables (female share, white share, share of population under 5 years) and income, and pre-intervention mortality. If ordinances worked, the real mortality trajectory should be lower than its synthetic version. It should be a "negative" real-synthetic gap.

(2) For inference, we made both individual level and group level checking. For individual city, we did permutation based "placebo study", by randomly assigning treatment to control cities. If the treatment effect was not from random chance, the effect should be more noticeable in the treated cities. For the treatment group, we focus on the average of outcomes or ATT. Then we compare the percentage change of mortality rates in both groups.

(3) We also did robustness tests to check our specifications, including changing covariates, drop observations with less well real-synthetic fitting in pre-intervention period, and cross-time test. For example, we compare the ATT from GSCM using only pre-intervention outcomes as covariates and the GSCM including all covariates.

[13] The case of intervention in 1920 is also considered. This move makes sense with regard to historical facts. The high demand for milk in war time (World War I) and lack of funding made real enforcement was not rigorous until early 1920s. The stories of Milwaukee (Levitt 1996) and Chicago (Czaplicki 2007) are two good cases.

4.2 Results and Inference

Figure V illustrates average decennial children diarrhea mortality rates of the treated and control cities. In both groups, noticeable drop occurred after 1910. But the treated group was clearly at a higher rate. In 1900 to 1910, the average of treated mortality rate slightly increased from 109.93 to 115.73. Meanwhile, the control group dropped from 104.74 to 96.58. After that, the trend was reversed. In the 1910s, the treated dropped 61.44 and higher than the 50.26 of the averaged control units. This trend continued in the 1920s, with 39.48 lower in the treated and 28.34in the control group. As a consequence, the average mortality rate of the treated group (14.82) on absolute value was lower than its control counterpart (17.98).

The next step is to construct the synthetic versions. First, we suppose the unanimous city policy intervention occurred in 1910. So our study period is separated into two parts, the pre-treatment period 1900 to 1910 and post-treatment 1910 to 1930. Each synthetic trajectory is built with the chosen predictors in 1900 to 1910. We first checked whether our treated units are in the convex hull of their control counterparts. That is to say we need a substantial overlap between two groups. This assumption is required in general matching method, since ensures the probability of comparable control units (Cameron and Trivedi 2005). Also, estimators with low common support will be very model-depending (Gary and Zeng 2006)[14]. To each treated city, its synthetic version should be a weighted convex combination of 52 control units. Our 16 treated cities and their synthetic versions (intervention 1910) are in Figure VI. We find the optimized weights are different across cities. Some are very different from the averaged weight. Some are not and close

[14] Fertility rate should be added as one predictor, since it influences mortality rate. However, with data limit, we cannot get city level fertility or birth rate. Actually, age specific birth rate data was not available for the whole US until 1940s. There were only estimated city data. They were calculated by the ratio of the share of female and the share of children. See Haines (1989) “American Fertility in Transition”.

to the average. With 52 control units, the averaged weight should be 0.019 for each control. But, in many cases, the weight of each control unit is not the value. For example, the combination of Cleveland has three control cities. They are Meriden, Sacramento and San Antonio, with weight of 0.696, 0.165 and 0.138. Other control cities just have 0 weight. On the other hand, some control cities have non-zero but very small weight, e.g. below 0.001. It suggests they didn't contribute much. For example, the synthetic version of Jersey City has 27 control cities with weight of 0.01 or lower. In total, these cities counted 0.233. In sum, which one provides the better estimation, SCM or DD? The answer really depends on the treated unit.

For inference, we do "placebo study" or "cross units" by assigning a hypothetical intervention to every control unit in 1910. Then, standard SCM procedure is repeated as above. If the result of treated city is different from most controls, the treatment effect should be significant. The inference results are graphically represented in Figure VII. This graph depicts the real-synthetic gaps in the treated and control groups. The solid color lines are treated units. Gray dash lines represent control units. To each treated, if the treatment effect was substantial, its real-synthetic gaps should be larger than most control cities (negative in the sign but with larger absolute value). We see two cities are noticeably lower than their comparison units, i.e. Chicago and Cleveland. The gap of other treated, however, is not quite obvious. But the distribution of the entire treated group seems relatively lower than the control group.

The real-synthetic gaps of 1930 are summarized in Table III. The average of the treated group is -7.62. The number is -1.04. Considering the average mortality rate in the treated group was 14.82 in 1930. It means the real mortality rate was 34.0% lower than the scenario if there were no city

ordinances. On average, the mortality rate of the treated group was 26% lower than its synthetic versions[15] in 1930. Similarly, the control group was 18.0% higher than its synthetic version. The difference between the two samples is significant at 5% level. Also, we calculated the percentage change of mortality rate in 1920. The result indicates there is no significant difference between these two groups. It implies the health effect could not be very remarkable in a relatively shorter period. But the story was different in the long-run. To summarize, to the entire treated group, we find the policy did have positive health effect at some decent statistical significance level.

4.3 Robustness Tests

Our robustness tests are conducted in two ways. First, we examine GSCM results by changing the timing of intervention and using different covariates. Second, we compare our GSCM results with DD specifications and check whether SCM can provide better estimations.

Our first robustness test is to make the "cross time" test, by switching the intervention from 1910 to 1920 and see how outcomes would change in 1930. If our target policy worked in the 1910s, this change would alternate our outcome of interest (real-synthetic gap). If the policy did not work in that time, the result would be largely constant, regardless of intervention time. Then, the GSCM process is replicated. Results are in Table IV. Figure VIII and Figure IX describe the real-synthetic gaps and "cross units" tests. Graphically, they are similar to the intervention in 1910. But we find the percentage change of mortality rate is comparatively smaller than the previous 1910 scenario. It drops from 26.08 to 18.64 (in absolute value), and it is not significant at 5% level anymore. It suggests that the policy was effective in the 1910s. In addition, we also tried specifications with different "predictors". We find the SCM trajectories mainly depends on

[15] It is calculated as (real – synthetic)/synthetic*100%.

the values of two pre-intervention outcomes (lagged values). At least in this study, other non-lagged outcomes covariates do not change the synthetic trajectories. In addition, we also dropped some control cities which are supposed be less well fitted in the pre-intervention period. The results remain consistent. So our GSCM specifications are robust to the above tests.

Another concern is the difference between SCM and DD estimators. Theoretically, DD estimates "the population average difference". SCM, on the other hand, was designed to measure "the population optimized difference". Recall Equation (7), it shows DD could be treated as a special class GSCM, with $w_j^* = w = 1/N$. SCM and DD do share similarities, but they worked in different ways. DD measures the coefficient of treatment dummy holding values of other control variables constant. In other words, the *ceteris paribus* assumption is automatically satisfied. But SCM noticed it may not work some circumstances, for example a variable of a treated unit may not fall in the overlap of their control counterparts. We should be careful about that either DD or SCM are linear model based. If the control variables of SCM are highly nonlinear, the result can be very different (Abadie, Diamond and Hainelmuller 2010).

Our DD model is specified as follows, Equation (11). The panel includes 68 cities in 4 years, 1900, 1910, 1920 and 1930.

$$Y_{it} = \alpha_i + \delta_t + \mathbf{X}'_{it}\beta + D_{it} + Y_{it-1} + \varepsilon_{it} \tag{11}$$

In our model, Y_{it} is the mortality rate of city *i* in year *t*. Our regressors include city fixed-effect α_i, the year fixed-effect δ_t, a vector of covariates $\boldsymbol{X}$ and a lagged value Y_{it-1}. The covariates are the log value of average wage, female share, white share and share of population under 5 years. DD results are represented in Table V. The coefficient of the policy dummy was higher in

absolute value compared with GSCM. The real-synthetic gap 1930 is -7.62 in the case of 1910 intervention and -6.92 with 1920 intervention. In DD specifications the value ranges from -8.88 to -12.26. Here we can see the subtle divergence between DD and SCM.

To reiterate, the divergence between DD and SCM are in two aspects. First, as mentioned, SCM uses the optimized weights to choose control cities. Some get higher weight than the averaged, some lower. Second, DD estimations are based on the "constant trend" assumption. But this assumption is unlikely be satisfied, since many dramatic changes could occur in every 10 years. Water treatment and other uncontrolled factors could potentially alternate the health profile. DD results are generally lower than SCM. It suggests health effects from DD are higher[16]. But GSCM tells us the averaged health effect could be lower. In this study, compared with DD, SCM is preferable in providing a clear and direct estimate to ATT.

[16] In a comparative case study to Chicago's pasteurization ordinance (Wang, Sproul and Corey 2013), the result of SCM was also smaller in absolute value than DD.

5. FIXED-EFFECTS MODEL

Our next step is to investigate how voluntary adoption of pasteurization affected public health. A good proxy to voluntary adoptions is the extent of milk pasteurized (%) in the market. Without mandatory ordinances, pasteurized milk was also quite popular in many cities. Of course, with cross-sectional data, we can estimate variations between the extent of pasteurization and its health effect across cities in a particular year. It is not helpful to solve endogeneity or omitted variable bias. For solution, we resort to FE model assuming unobserved heterogeneity is time-invariant across units. Then, our estimation requires only intra-units variations over time instead of inter-unit variations.

Thus we use a two period panel for the extent of pasteurization in 1921 and 1924[17]. This panel is chosen for two reasons. First, the assumption of time-invariant unobserved characteristics is more likely to be violated if the study period is longer. Considering dramatic changes occurred in 1920 to 1930, a shorter panel is preferred. Second, it makes more sense to focus on the early 1920s, since the expansion of pasteurization was much faster in this period (Figure III). After dropping observation with missing values, the balanced panel includes 109 cities.

Our FE model is specified as below. $i = 1, \dots, 109$ and $t = 0$ for 1921 and 1 for 1924.

$$Y_{it} = \alpha_i + \delta_t + \mathbf{X}'_{it}\theta + \varepsilon_{it} \qquad (13)$$

In the model, Y_{it} is the mortality rate of city i in year t. City fixed-effect is α_i, and year fixed-effect is δ_t. Covariates vector $\boldsymbol{X}$ includes share of pasteurization and population.

[17] We actually have 3 cross-sectional pasteurization data. They are 1921 and 1924 USDA surveys and 1931 USPHS survey.

Empirically, outliers are a potential problem in this part. Here, observations with extreme values were possibly from other data generating processes (DGPs) and "contaminate" our data. This problem is not rare for historical studies. And it makes our least square (LS) based FE regression very sensitive to those observations. LS estimation relies on the key assumption that errors are normally distributed. But, "outliers have large effect on the mean, and drag it towards them" (Rousseeuw and Leroy 1987). In fact, LS estimator may be biased but efficient, as it is calculated by variance. Figure X plots within group variations of the extent of pasteurization and mortality rate. Those good leverage points, bad leverage points and vertical outliers are labeled[18].

To solve this problem, we used robust panel estimators as follows . LS estimator is defined as below[19]. It minimize the squared residual of r_i, Equation (14).

$$\hat{\hat{\theta}}_{LS} = \arg\min_{\theta} \sum_{i=1}^{n} r_i^2(\theta);\ r_i = y_i - \hat{\alpha} - x_i'\theta \quad \text{for } 1 \le i \le n \tag{14}$$

M-estimator, on the other hand, is based on the maximum likelihood estimation (MLE).

$$\hat{\theta}_M = \arg\min_{\theta} \sum_{i=1}^{n} \rho\left\{\frac{r_i(\theta)}{\sigma}\right\} \tag{15}$$

It minimizes residuals with ρ function rather than the square, Equation (15)[20]. In practice, M-estimator is calculated with iteratively reweighted LS process, Equation (16).

$$\hat{\theta}_M = \arg\min_{\theta} \sum_{i=1}^{n} \omega_i r_i^2(\theta);\ \text{where } \omega_i = \rho\left\{\frac{r_i(\theta)}{\sigma}\right\} \cdot \frac{1}{r_i^2} \tag{16}$$

M-estimator is a weighted LS-estimator. It is not robust to bad leverage points.

[18] Graphically, it follows the pattern of Verardi and Croux (2009). Definition of the types of outliers is in Rousseeuw and Leroy (1987).
[19] Notations follow Verardi and Croux (2009).
[20] Discussions about M-estimator, S- estimator and MM-estimator are in Verardi and Wanger (2013).

S-estimator was designed by Rousseeuw and Leroy (1987). It uses loss function ρ () and minimizes the robust dispersion of residuals. The robust dispersion is $\hat{\sigma}^S$. By doing so, it gives less weight to large residuals, Equation (17).

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left\{\frac{r_i(\theta)}{\hat{\sigma}^S}\right\}=b;\ \text{where } b=E[\rho(Z)] \text{ with } Z\sim N(0,1) \qquad (17)$$

Then, S-estimator minimizes $\hat{\sigma}^S$, as Equation (18).

$$\hat{\theta}_S=\arg\min_{\theta}\hat{\sigma}^S\{r_1(\theta),r_2(\theta),\ldots,r_n(\theta)\} \qquad (18)$$

S-estimator has high breakdown point but low Gaussian efficiency.

MM-estimator combines both S-estimators' high breakdown point and M-estimator's Gaussian efficiency. First, it uses S-estimator to obtain the scale parameter $\hat{\sigma}^S$ at a break down point of 50%. Next, it assumes the M-estimator and achieves some high Gaussian efficiency by choosing an appropriate ρ function. MM-estimator is defined by Yohai (1987), as Equation (19).

$$\hat{\theta}_{MM}=\arg\min_{\theta}\sum_{i=1}^{n}\rho\left\{\frac{r_i(\theta)}{\hat{\sigma}^S}\right\} \qquad (19)$$

In addition, we checked our model with regular Median or L-estimator, Equation (20).

$$\hat{\theta}_L=\arg\min_{\theta}\sum_{i=1}^{n}\left|r_i(\theta)\right| \qquad (20)$$

L-estimator minimizes the sum of absolute value of r_i. But the efficiency of L-estimator is not quite high (Verardi and Croux (2009).

Recently, a new category of outliers have been identified by Bramati and Croux (2007) in panel estimations. The authors noticed the case when extreme values largely concentrate on certain periods. As a remedy, they proposed the within group generalized M estimator (WGM) and the

within groups MS (WMS) estimator[21]. The principle of WGM and WMS are similar, outliers can be dropped by centering within time series data. WGM is defined as Equation (21). W_r is a diagonal matrix and awards lower weights to observations with extreme values. W_x is a weighting matrix. And $\tilde{x}_{it}$ and $\tilde{y}_{it}$ are values that first centered within every time-series.

$$\hat{\theta}_{WGM} = (\tilde{X}' W_x W_r \tilde{X})^{-1} \tilde{X}' W_x W_r \tilde{y} \tag{21}$$

WMS is formalized as Equation (22). It can be used for both continuous and categorical variables. WMS jointly minimizes α and θ.

$$\hat{\theta}_{WMS} = \arg\min_{\theta} S(r_1(\hat{\alpha}(\theta), \theta), \ldots, r_{NT}(\alpha(\theta), \theta)) \tag{22}$$

In practice, they are used as robust fixed-effect estimators (Verardi and Wagner 2012). Data are first centered and then be removed individual fixed effects. Thus the robust estimator is used to identify outliers, like S-estimator. Outliers are awarded zero weight. Other normal observations are fitted with standard FE model. For this study, regarding only 2 periods in our panel data, the above category of outliers is not our major concern. We do not make WGM or WSM estimations.

The results of our FE regressions are presented in Table VI. It suggests the number of mortality rate drop with 1% increase in the pasteurized milk. The values estimated by LS (regular FE model), L-estimator (median), M-estimator, S-estimator and MM-estimators are -0.248, -0.157, -0.164, -0.151 and -0.132. All of them are statistically significant. The coefficient estimated by LS estimator is relatively higher than other estimators. It means 1% increase in pasteurization can drop 0.13 to 0.25 lower in mortality rate. Even the coefficient is not of an extraordinarily high value, accumulated health outcome was still remarkable. In that time, the extent of pasteurization in many cities was still quite low. The average share of milk pasteurized in our

21 Details of WGM and WSM are in Bramati and Croux (2007)

109 cities was 68%. If the number rose 30% (to 98%), the health benefit of LS estimator is 7.2 lower in mortality rate[22]. Considering the population in 109 cities, there would be 1623 under 2-years children saved from diarrhea death in total. The effect could be even better, if we included more small cities, which generally had very low pasteurization.

In sum, the FE regressions also support the positive health effect of pasteurization. The result of voluntary adoption is consistent with mandatory measures. From the perspective of public health, the more pasteurization was the better, though only some cities adopted their mandatory ordinances. But mandatory regulations were still desirable, since almost 100% milk can be pasteurized. Table VII illustrates the average pasteurization in cities with ordinances was 99.1% and the number was only 65.6% in cities without ordinances. There was some potential to realize the health benefit of this innovation in those cities.

[22] Recall SCM results, mandatory ordinances led to only 6.94 lower in mortality rate from 1920 to 1930.

6. CONCLUSIONS

This paper investigates the results when researchers use different policy measures. With the case of milk pasteurization in the US, we make policy evaluations to both mandatory and voluntary adoptions of pasteurization in the period 1900 to 1930. As a remarkable food safety innovation, pasteurization played a critical role in the fight of milk diseases. But less enough attention has been given to how to estimate its health outcomes. As a mix-up of both mandatory and voluntary measures, milk pasteurization is a desirable case to investigate its health impacts.

We first estimated the casual health impact of city pasteurization ordinances with the synthetic control approach. SCM stresses the problem of unobserved heterogeneity over period. Also, it provides a better counterfactual with awarding optimized weights to control units. In addition, we generalize SCM from comparative case study to multiple treated units (GSCM). Our results show that the ATT (intervention in 1910) in 16 treated group is -7.62 lower than the average of their synthetic versions. The number was only -1.04 in the control group. Percentage dropped in mortality rate of the treated sample is statistically different from the control sample. In addition, we find our GSCM result is not the same as DD estimator.

Next, we made a FE analysis to examine the role of voluntary pasteurization adoption. Historical evidence suggests market may work better to adopt new food safety innovations. In fact, the extent of pasteurization was not quite low even in those cities without ordinances. To construct a clear health picture of pasteurization, the impact of voluntary adoptions should not be ignored. With the panel in 1921 and 1924, we make FE estimations to the health effect of the spread of pasteurization. We choose robust panel estimators to minimize the influence of outliers. More

than regular least-square FE models, we also use L-estimator, M-estimator, S-estimator and MM-estimators. All results suggested the spread of pasteurization was significantly responsible for the improvement in public health.

Overall, our results suggest the more pasteurization was the better in that period, although only some cities adopted city ordinances. Mandatory measures are more desirable as it can effectively increase the extent of pasteurization. Either mandatory or voluntary measures, the health effect was positive. There was much potential to implement this food safety innovation, particularly when the extent of pasteurization was not high around 1920s. Pasteurization had its unique and significant contribution to public health in history.

REFERENCES

Abadie A and Gardeazabal J . 2003. The Economic Costs of Conflict: A case study of the Basque Country. American Economic Review 93: 112-132

Abadie A Diamond A and Hainmueller J . 2010. Synthetic Control Method for Comparative Case Study: Estimating the effect of California's tobacco control program. Journal of American Statistical Association. 105: 493-505

Abadie A Diamond A and Hainmueller J . 2012. Comparative Politics and the Synthetic Control Method. http://www.hks.harvard.edu/fs/aabadie/comparative.pdf

Alvord H. E. and R. A. Pearson, The milk supply of two hundred cities and towns, U.S. Department of Agriculture, Bureau of Animal Industry, Bulletin, 46, Washington, DC: 1903

Atkins P J . 1992. White Poison? The Social Consequences of Milk Consumption, 1850-1930. Social History of Medicine 5: 207-227

Atkins P J . 2003. Mother's milk and infant death in Britain, circa 1900-1940. Anthropology of food . September. http://aof.revues.org/index310.html

Atkins P J . The pasteurization of England: the science, culture and health implications of milk processing. 1900-1950. pp. 37-51, in Smith D. F. and J. Phillips ed . 2000. Food, science, policy and regulation in the twentieth century: international and comparative perspectives. Routledge: London

Ayers S H . 1922. The present status of the pasteurization of milk, Bulletin No. 342. United States Department of Agriculture

Ayers S H . 1926. The present status of the pasteurization of milk, Bulletin No. 342. revised edition. United States Department of Agriculture

Bartlett R W . 1946. The Milk Industry: a comprehensive survey of production, distribution, and economic importance. New York: Ronald Press Company

Beaver M W . 1973. Population, infant mortality, and milk. Population Studies . 27: 243-254

Block D and DuPuis E M . 2001. Making the Country Work for the City Von Thünen's Ideas in Geography. The American Journal of Economics and Sociology . 60: 79-98

Block D R, Purity, economy, and social welfare in the Progressive Era milk movement. Journal for the Study of Food and Society. 1999. 3: 20-27

Block D R . 2009. Public health, cooperatives, local regulation, and the development of modern milk policy: the Chicago milkshed, 1900–1940, Journal of Historical Geography. 35: 128-153

Bramati M. C. and Croux C., Robust estimators for the fixed effects panel data model, The Econometrics Journal, 2007, 10: 521-540

Buzby J C et al. . 2013. Characteristics of Consumers of Unpasteurized Milk in the United States . Journal of Consumer Affairs 47: 153-166

Cameron A C . and Trivedi P K . 2005. Microeconometrics: Methods and Applications, New York: Cambridge University Press

Czaplicki A . 2007. Pure Milk Is Better Than Purified Milk Pasteurization and Milk Purity in Chicago, 1908-1916, Social Science History 31: 411-433

Eagan K . 2010. Raw Milk Sales. http://legis.wisconsin.gov/lrb/pubs/wb/10wb1.pdf [Wisconsin Briefs. April 2010]

Ernst M D . 2004. Permutation methods: A basis for exact inference. Statistical Science 19 (4): 676-685

Fares M. and E. Rouviere, The implementation mechanisms of voluntary food safety systems, Food Policy, 2010, (35) 5: 412-418

Fogel R W . 1994. Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy. American Economic Review 84: 369-395

Frank L C and Moss F J . 1931. The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931. US Public Health Service

Fuchs A W and Frank L C . 1938. Milk Supplies and Their Control in American Urban Communities of Over 1,000 Population in 1936, Public Health Bulleting No. 245. US Public Health Service

Giannone D., M. Lenza and L. Reichlin, 2010, Business cycles in the euro area. In Europe, and the Euro, Alesina A, Giavazzi F (eds). National Bureau of Economic Research: Cambridge, MA; 141–167.

Hall C W and Trout G M . 1968. Milk pasteurization. Westport, Conn., Co

Headrick M L . 1998. The epidemiology of raw milk-associated foodborne disease outbreaks reported in the United States, 1973 through 1992, American Journal of Public Health 88: 1219-1921

Hotchkiss J H . 2001. Lambasting Louis: lessons from pasteurization, in the US National Agricultural Biotechnology Council Report.
http://nabc.cals.cornell.edu/pubs/nabc_13/HOTCHKISS.pdf

Hsiao C, Ching S H and Wan K . 2012. A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. Journal of Applied Econometrics 27: 705-740

Kelly E and Clement C E . 1931. Market milk. J. Wiley & Sons, Inc: New York; Chapman & Hall Limited: London

Kiesel K and Villas-Boas S B . 2013. Can information costs affect consumer choice? Nutritional labels in a supermarket experiment, International journal of industrial organization 31: 153-163

King G and Zeng L . 2006. The Dangers of Extreme Counterfactuals. Political Analysis 14: 131-159

Langer A J et al. . 2012. Nonpasteurized Dairy Products, Disease Outbreaks, and State Laws-United States, 1993-2006 . Emerging Infectious Diseases 18: 385-391

Law M T . 2003. The Origins of State Pure Food Regulation, Journal of Economic History 63: 1103-1130

Lee K S . 2007. Infant Mortality Decline in the Late 19th and Early 20th Centuries: The Role of Market Milk. Perspectives in Biology and Medicine 50: 585-602

Lentzner H R., 1987, Seasonal Patterns of Infant Child Mortality in New York, Chicago, and New Orleans, 1870-1919, Ph.D. dissertation, University of Pennsylvania

Ludbrook J and Dudley L . 1998. Why Permutation Tests are Superior to t and F Tests in Biomedical Research. The American Statistician 52: 127-132

Meckel R A. 1990. Save the Babies: American public health reform and the prevention of infant mortality, 1850-1929. Johns Hopkins University Press

Meckel R. A., Save the Babies: American public health reform and the prevention of infant mortality, 1850-1929, Johns Hopkins University Press, 1990

Moore D, McCabe G and Craig B . 2007. Introduction to the Practice of Statistics, 6th ed., W. H. Freeman

North C E . Milk and its relation to public health, in Ravenel P M . 1921. ed. A Half Century of Public Health, American Public Health Association: New York

Painter J A et al. . 2013. Attribution of Foodborne Illnesses, Hospitalizations, and Deaths to Food Commodities by using Outbreak Data, United States, 1998-2008 . Emerging Infectious Diseases. 19: 407-415

Parker H. N., City milk supply, New York: McGraw-Hill, 1917

Preston S H and Haines M R . 1991. Fatal Years: child mortality in late nineteenth-century, Princeton, N.J.: Princeton University Press

Rousseeuw P. J., and A. M. Leroy, Robust Regression and Outlier Detection, New York: Wiley, 1987

Segerson K., Mandatory versus voluntary approaches to food safety, Agribusiness, 1999, 15: 53-70

Selitzer R . 1976. The dairy industry in America. Magazines for Industry: New York

Veradi V. and C. Croux, Robust regression in Stata, The Stata Journal, 2009, 3 (9): 439-453

Vogel A. and J. Wagner, Robust estimates of exporter productivity premia in German business services enterprises, Economic and Business Review, 2011, (13) 1-2: 7-26

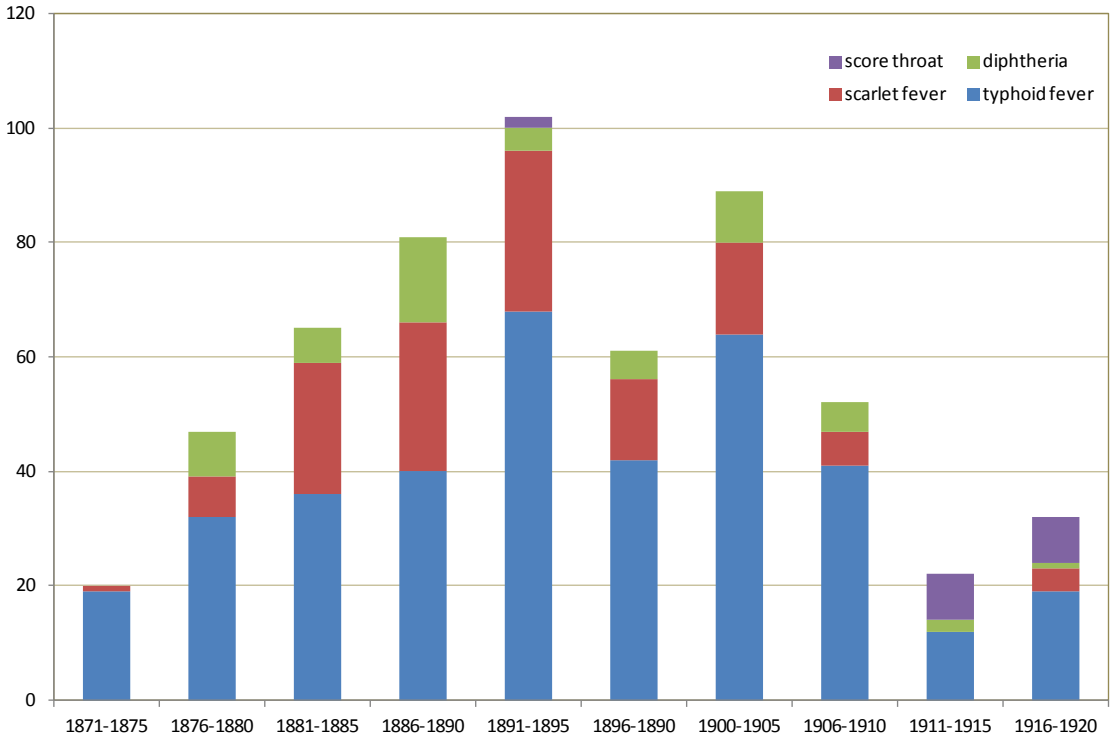
Wagstaff A . 2010. Estimating health insurance impacts under unobserved heterogeneity: the case of Vietnam's health care fund for the poor. Health Economics 19: 189-208

Winslow C-E A . 1952. Man and epidemics, Princeton University Press. Princeton: New Jersey

Wooldridge J . Difference-in-Difference Estimation. National Bureau of Economic Research. http://www.nber.org/WNE/lect_10_diffindiffs.pdf [2007 Summer Institute]

APPENDIX

Figure I. The number of reported milk diseases epidemics 1871 to 1920 (every 5 years)



Source: the original annual data are obtained from North (1921) and summed up every five years by the authors. The data did not include all epidemics in observed periods (more in North, 1921).

Table I. Dates of city milk pasteurization ordinances

State	City	Year	State	City	Year
Maryland	Baltimore	1917	New Jersey	Jersey City	1915
New York	Buffalo	1918	Wisconsin	Milwaukee	1916
Illinois	Chicago	1916	New York	New York	1912
Ohio	Cincinnati	1914	Pennsylvania	Philadelphia	1914
Ohio	Cleveland	1916	Virginia	Richmond	1916
Ohio	Dayton	1918	Missouri	St. Louis	1915
Michigan	Detroit	1915	California	San Francisco	1916
Indiana	Indianapolis	1916	Ohio	Toledo	1915

Source: (not in references)
Baltimore: Medical news, Journal of the American Medical Association, 1917, 68, p. 1418; Atlantic reporter, Volume 104, West Publishing Company, 1918, p. 181-182
Buffalo: Food Inspection, Annual report of Department of Health, 1919, p. 114; US Public Health Service, Public Health Reports 1917-1919, p. 318-319
Chicago: Illinois Health News, Illinois State Department of Health, 1922, p. 144-145
Cincinnati: US Public Health Service, Public Health Reports 1915, p. 2567
Cleveland: US Public Health Service, Municipal Ordinances and Regulations Pertaining to Public Health 1915, p. 217-224
Dayton: The Creamery and Milk Plant Monthly, 1918, August, p. 37
Detroit: Hedrick W. and Anderson A., Detroit Commission Plan of City Milk Administration, 1919, p. 12
Indianapolis: Journal of the Indiana State Medical Association, 1916, February, p. 71; US Public Health Service, Municipal Ordinances and Regulations Pertaining to Public Health 1916, p. 130-134
Jersey City: US Public Health Service, Municipal Ordinances and Regulations Pertaining to Public Health 1915, p. 291-296
Milwaukee: Hibbard B. and Erdmann H., Marketing Wisconsin milk, 1917, p. 49-50; Levitt J. W., The Healthiest City: Milwaukee and the Politics of Health Reform, 1996, p. 187
New York: Monthly Bulletin of the Department of Health in the City of New York, New York Department of Health, 1912, p. 4-9
Philadelphia: National Municipal Review, Vol. 2, National Municipal League, 1913, p. 716-717
Richmond: US Public Health Service, Municipal Ordinances and Regulations Pertaining to Public Health 1915, p. 364-365
St. Louis: Health Departments Reports and Notes, American Journal of Public Health, 1914, 7, p. 619
San Francisco: US Public Health Service, Reprint from the Public Health Reports, 1916, p. 160-173
Toledo: US Public Health Service, Municipal Ordinances and Regulations Pertaining to Public Health 1915, p. 386-389

Table II. Summary Extent of Pasteurization in Cities 1921 and 1924

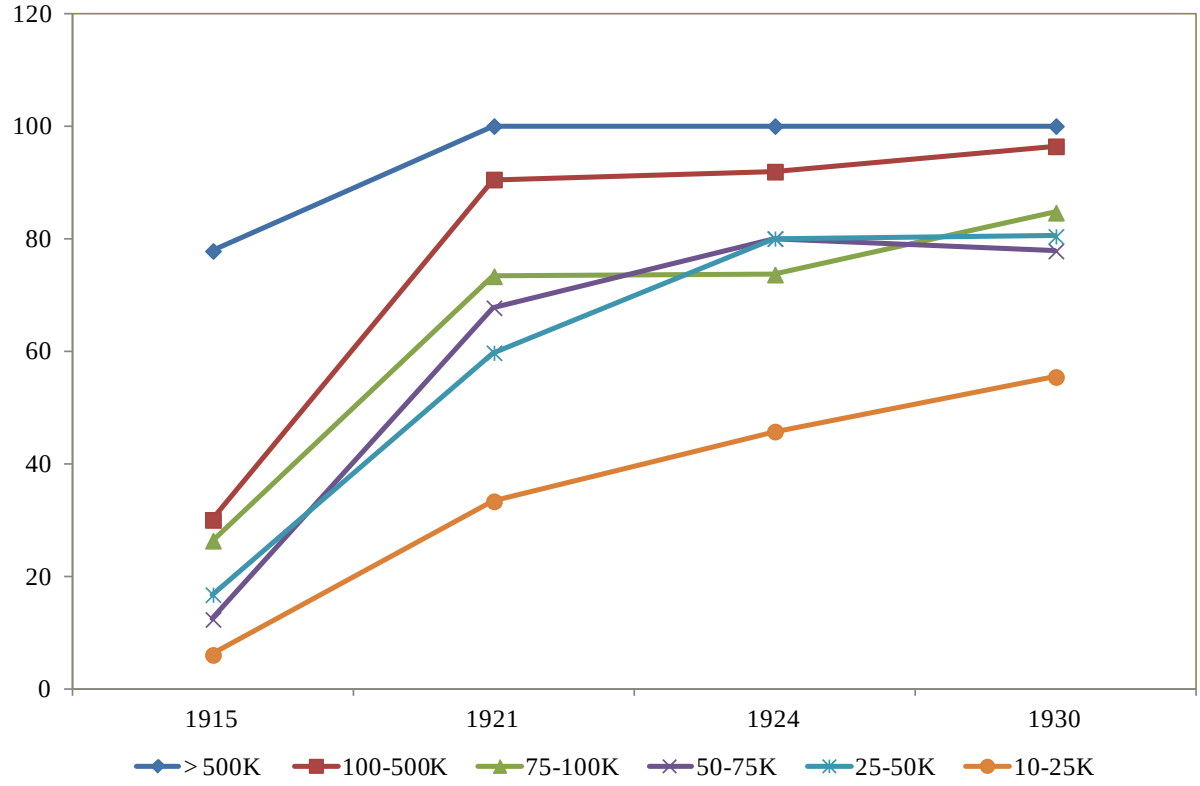
City Population	Number of cities reporting			Number of cities without milk pasteurized			Average % of milk pasteurized		
	1921	1924	1930	1921	1924	1930	1921	1924	1930
> 500K	12	9	11	0	0	0	95.0	98.1	97.1
100-500K	42	37	56	0	0	0	72.0	81.7	84.9
75-100K	15	19	13	0	0	0	68.0	66.6	81.5
50-75K	29	25	37	5	0	0	65.0	66.6	72.2
25-50K	55	60	56	7	2	0	58.0	67.0	73.1
10-25K	77	105	92	49	21	6	51.0	42.5	52.1
<10K	36	73	79	52	20	44	53.0	33.0	27.1
Total	266	328	344	113	43	50	----	----	----

Source: *Present status of the pasteurization of milk*, Bulletin No. 342. USDA (Ayers 1922, 1926, 1932)
Note: 1,000 people as 1 K

Figure II. The percentage of milk pasteurized in cities of population groups

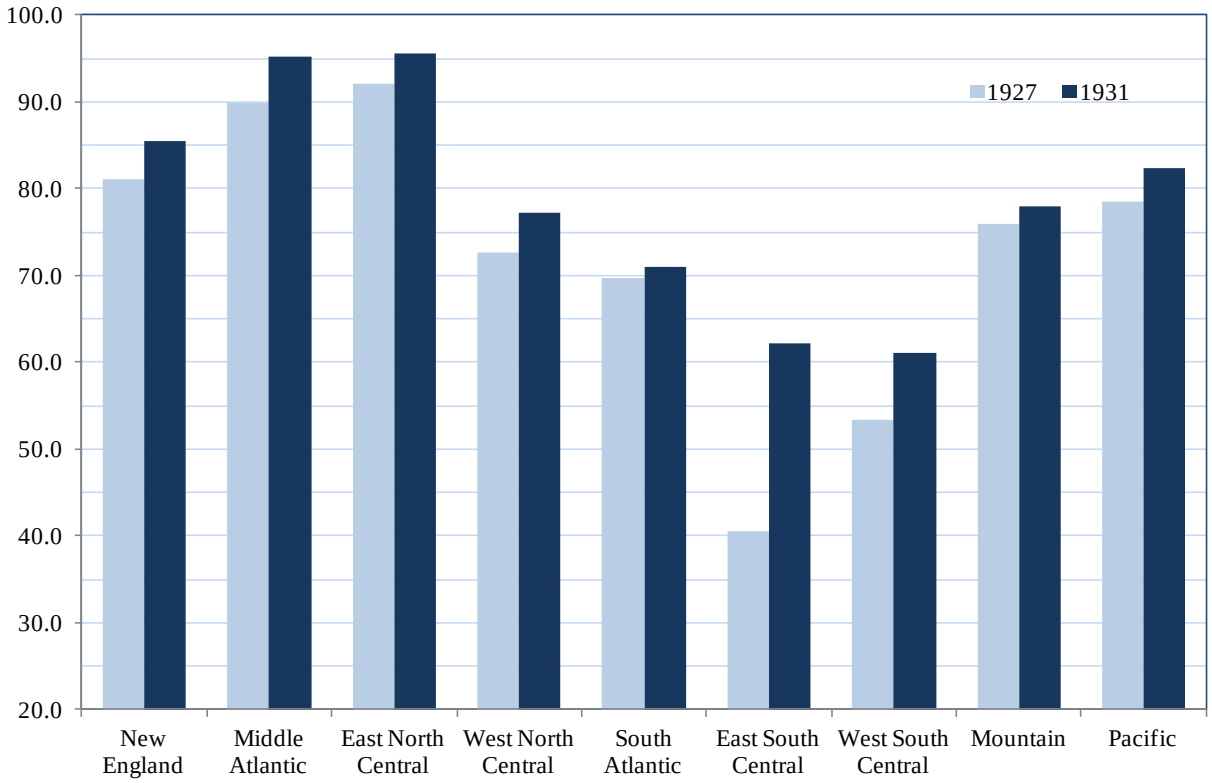
Source: Frank and Moss, *The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931*. US Public Health Service
Note: 1,000 people as 1 K

Figure III. Changes in the extent of pasteurization, 1920 to 1930



Source: *Present status of the pasteurization of milk*, Bulletin No. 342. USDA (Ayers 1922, 1926, 1932)
Note: 1,000 people as 1 K

Figure IV. The percentage of milk pasteurized in cities of regions



Source: Frank and Moss, *The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931*. US Public Health Service
Note: 1,000 people as 1 K

Figure V. Average decennial mortality rate 1900-1930, 16 treated and 52 control cities

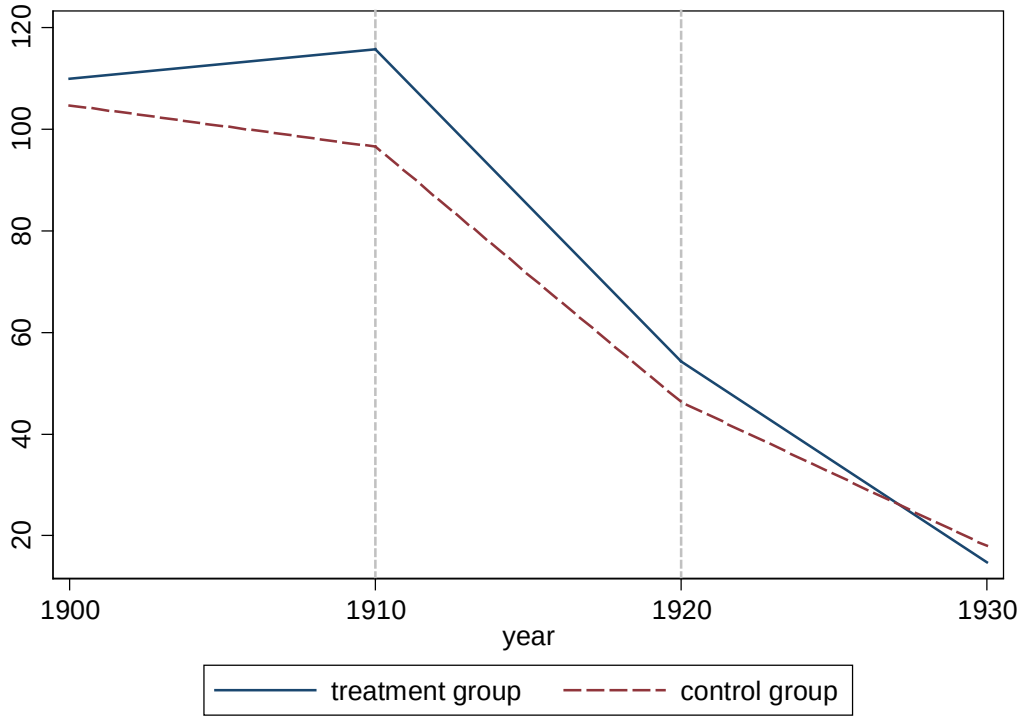


Figure VI. Mortality real treated cities and their synthetic versions, treatment 1910

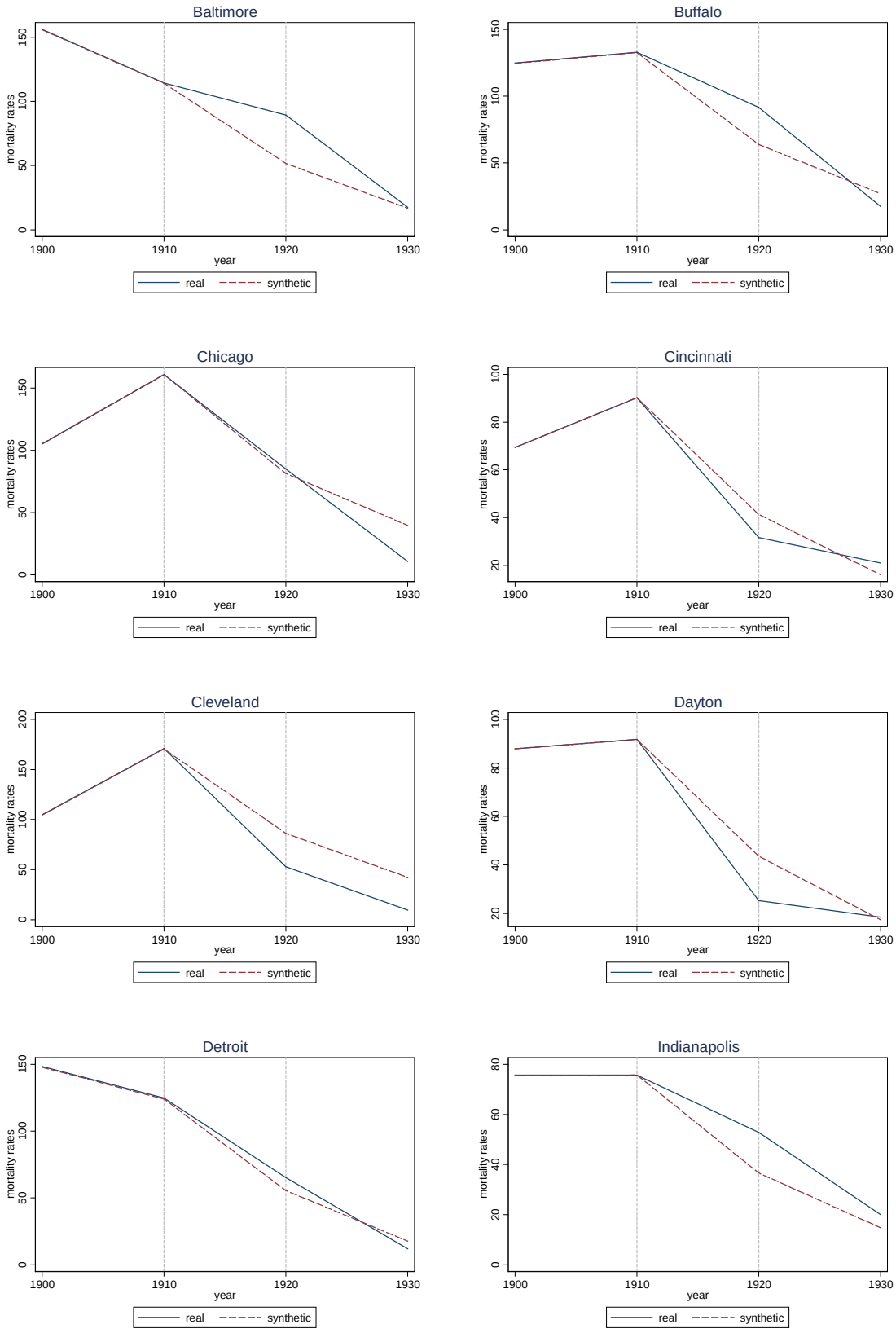


Figure VI. - continued

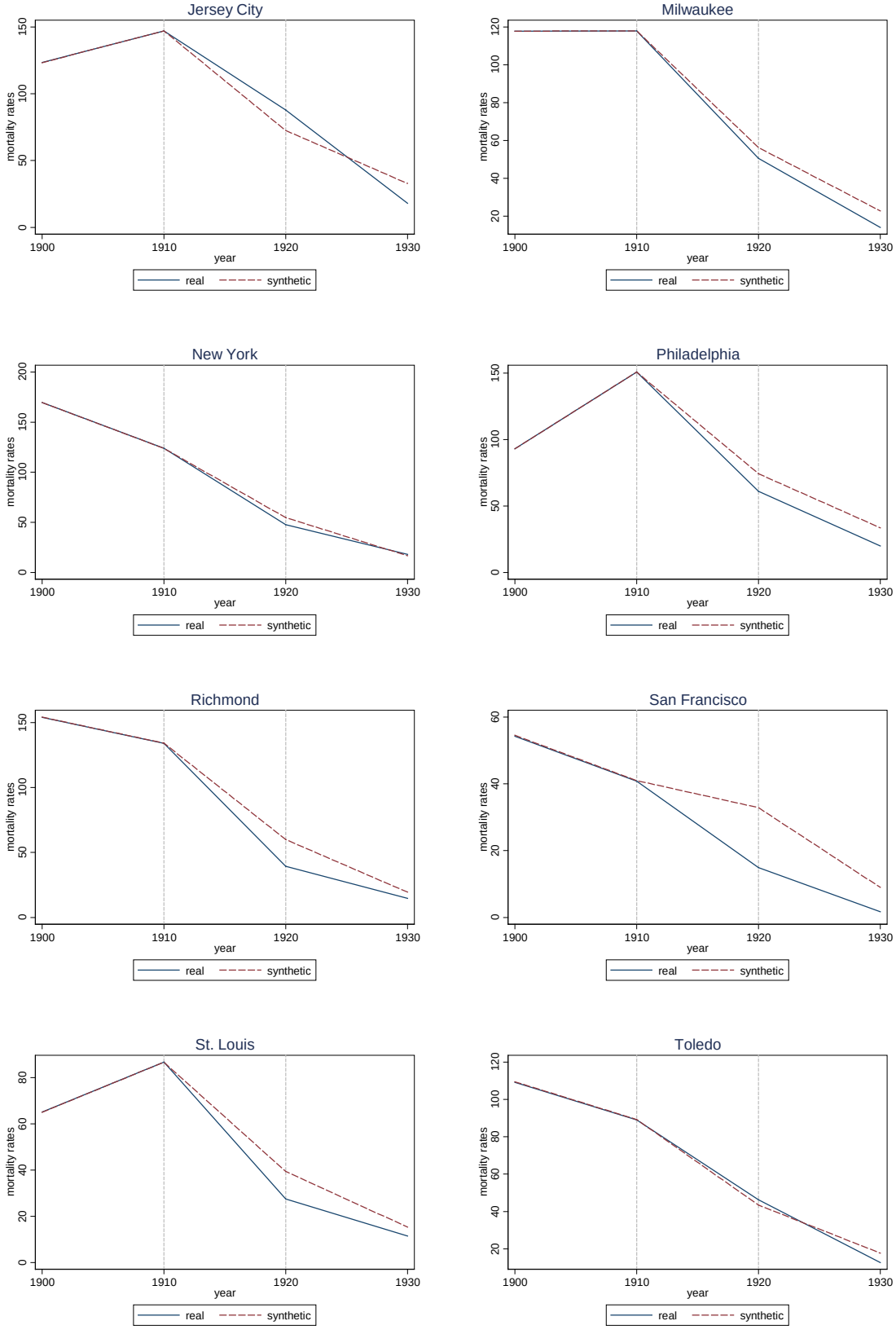


Table III. Real-synthetic mortality rate gaps across cities in 1930

Treated	Gap	Control	Gap	Control	Gap	Control	Gap
Baltimore	0.73	Brockton	-3.38	Memphis	1.79	Quincy	12.20
Buffalo	-9.62	Concord	-7.18	Meriden	-35.11	Raleigh	-24.07
Chicago	-29.01	Duluth	-13.03	Middletown	65.48	Sacramento	-64.61
Cincinnati	4.95	Evansville	3.83	Mobile	20.71	San Antonio	32.21
Cleveland	-32.61	Fitchburg	-15.90	Montclair	-11.55	San Diego	0.78
Dayton	1.00	Green Bay	-13.90	Muncie	8.96	Savannah	1.62
Detroit	-5.65	Harthord	7.38	Nashville	19.28	Seattle	-0.81
Indianapolis	5.23	Jackson City	0.69	New Britain	-9.95	Springfield	-9.01
Jersey City	-14.88	Jacksonville	4.36	New Haven	-14.33	St. Paul	-11.81
Milwaukee	-8.67	Jamestown	-5.37	New Orleans	-0.18	Superior	-11.60
New York	1.33	Kalamazoo	2.90	Omaha	-6.10	Tacoma	-1.29
Philadelphia	-13.72	Kansas city	2.38	Paducah	14.01	Troy	-8.63
Richmond	-4.77	La Fayette	-1.08	Petersburg	-1.83	Utica	-17.99
San Francisco	-7.30	Lancaster	33.55	Pittsfield	-15.66	Wheeling	24.70
St. Louis	-3.94	Lincoln	-5.37	Plainfiled	-8.83	Wichita	-0.52
Toledo	-5.00	Los Angeles	5.20	Portland	-12.29	Wilmington	1.98
		Lynchburg	23.31	Portland	0.36		
Avg. treated	*-7.62*	Manchester	2.30	Providence	-12.68	*Avg. control*	*-1.04*

Figure VII. Inference with “placebo study” to treated group, intervention 1910
(Treated: color solid line; Control: gray dash line)

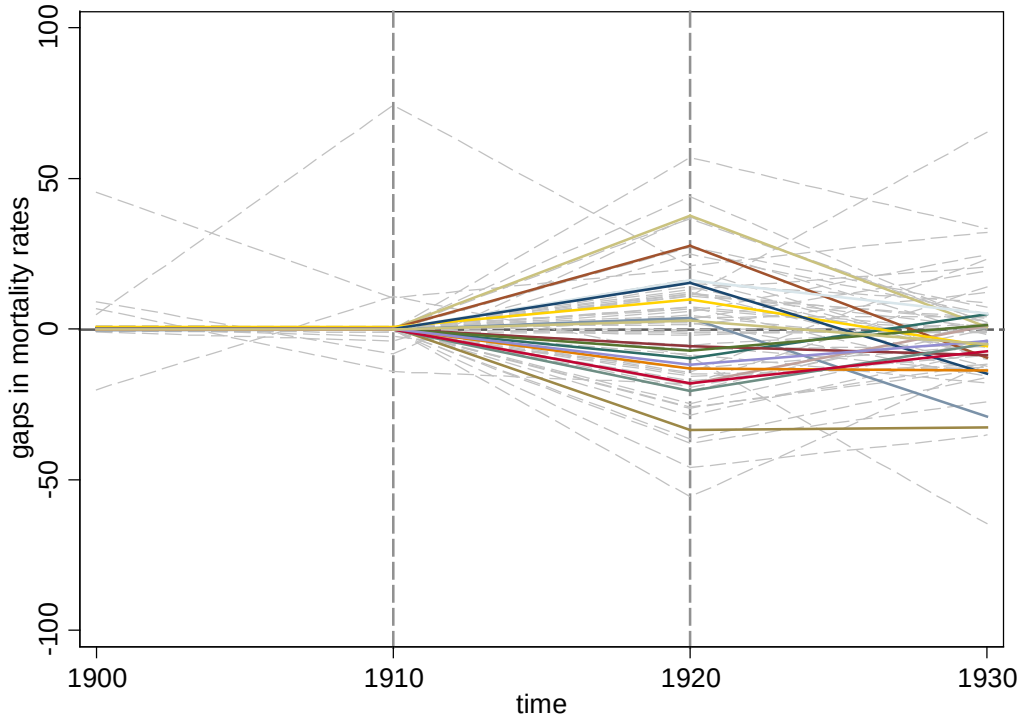


Figure VIII. Mortality real treated cities and their synthetic versions, treatment 1920

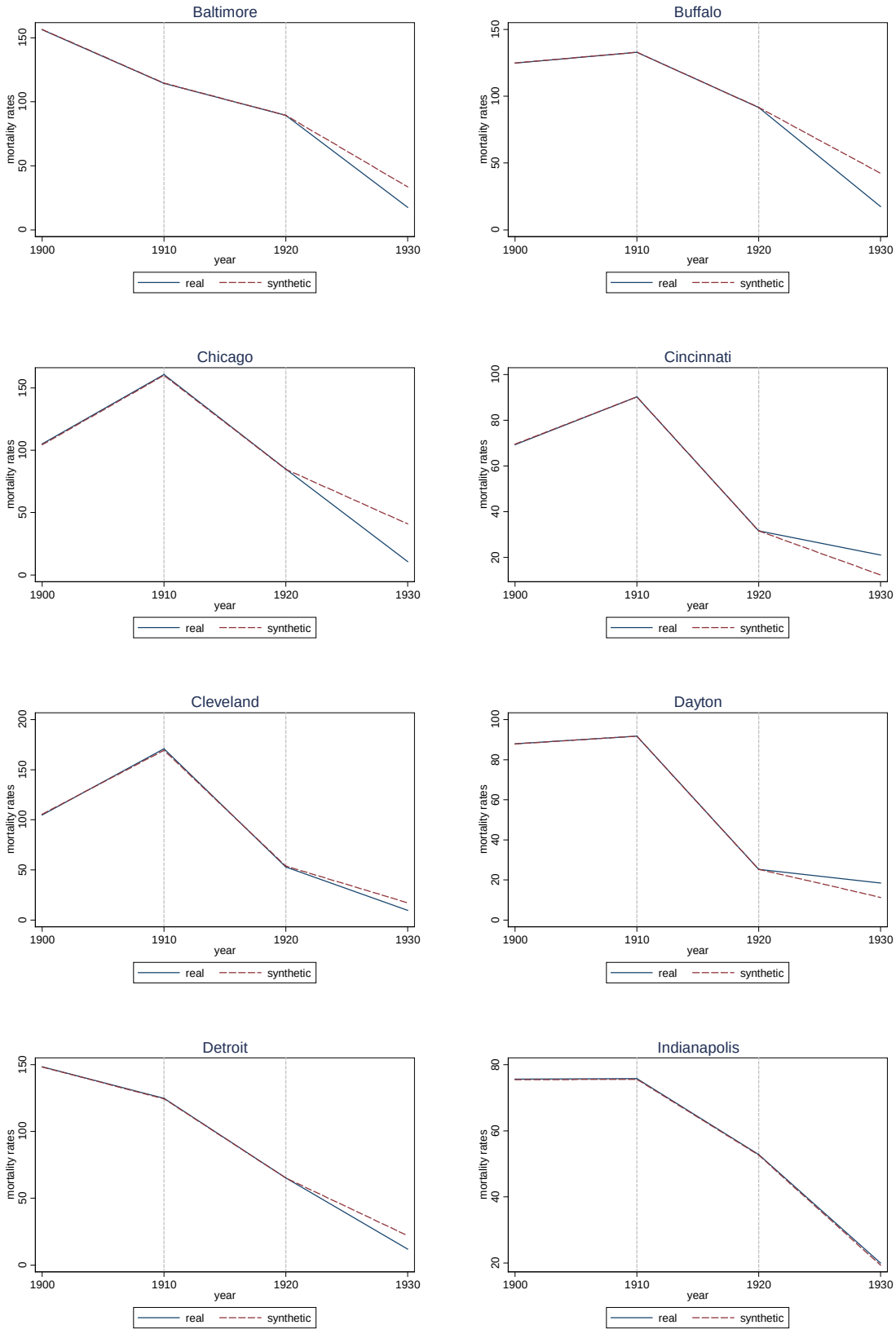


Figure VIII. - continued

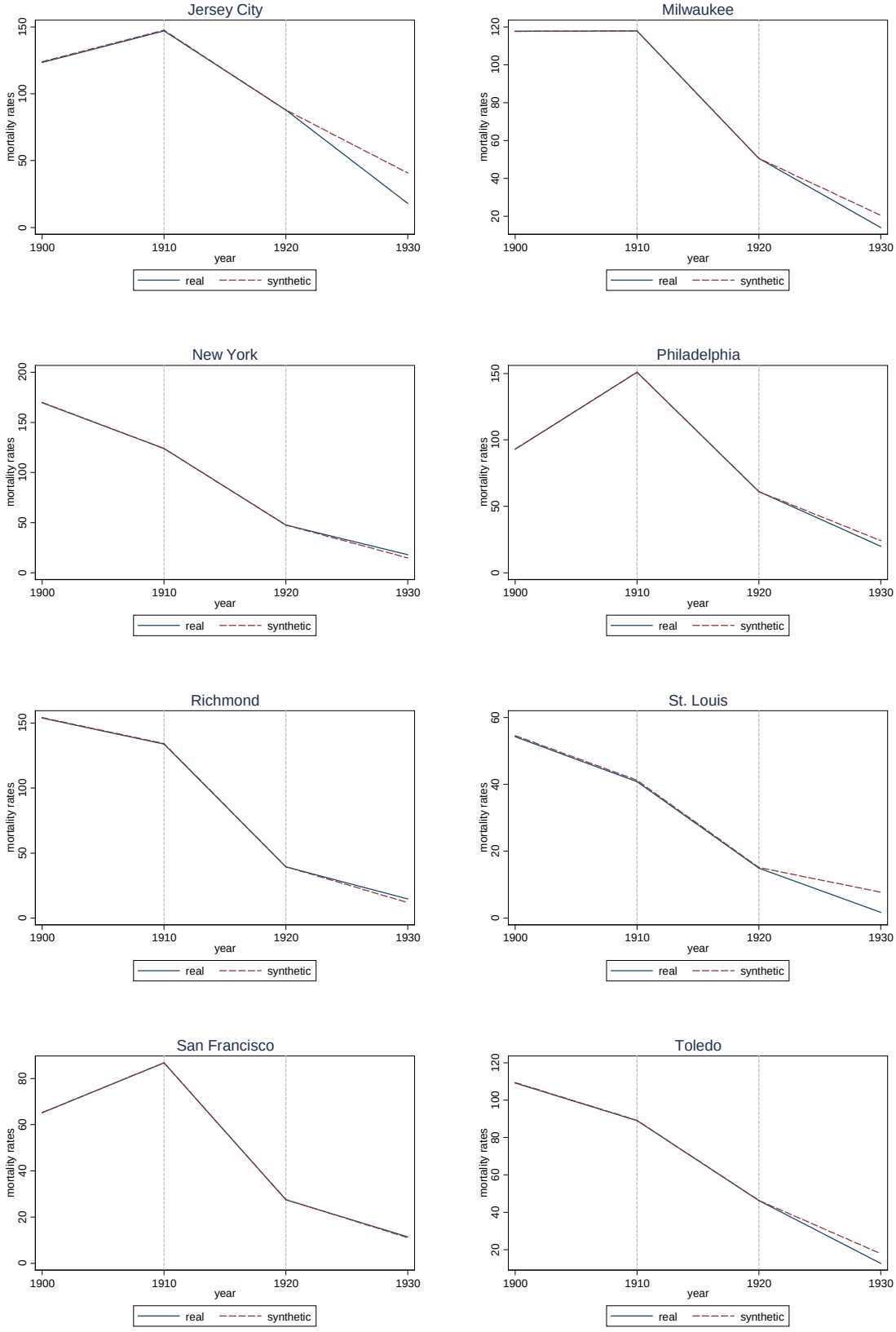


Figure IX. Inference with “placebo study” to treated group, intervention 1920
(Treated: color solid line; Control: gray dash line)

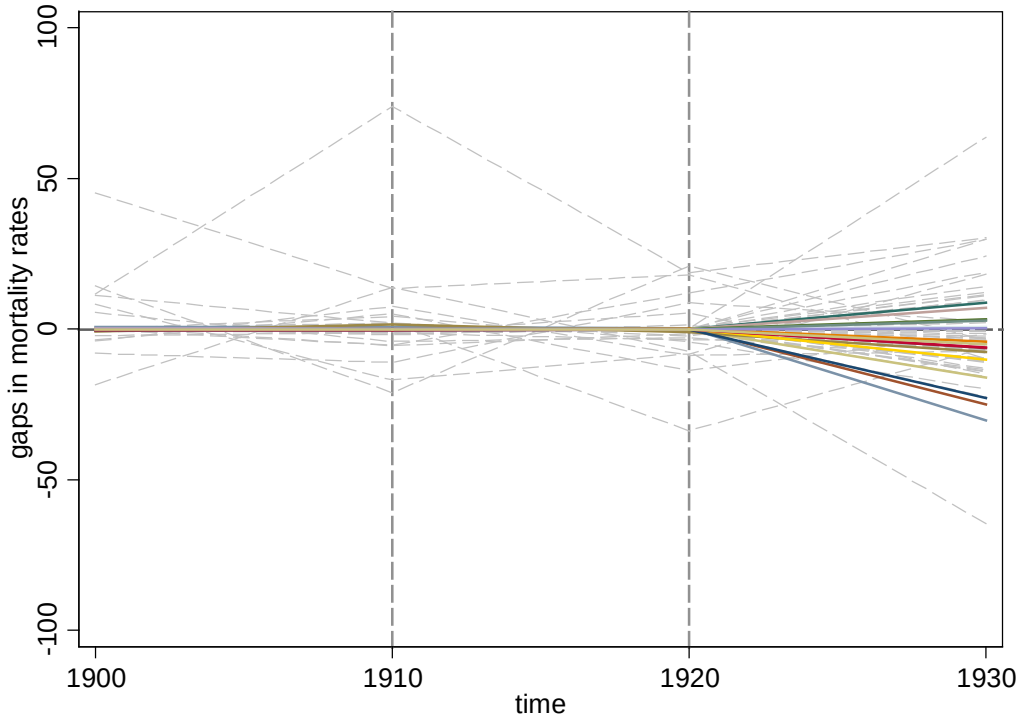


Table IV. Inference of percentage change (%) in mortality rate with intervention

	Intervention 1910		Intervention 1920
	% 1930	% 1920	% 1930
Treated Group	-26.08**	-3.41	-18.64*
	[8.86]	[8.96]	[11.35]
Control Group	17.97	5.31	18.62
	[20.15]	[7.03]	[18.07]

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table V. DID Estimation of City Pasteurization Ordinances 1900 to 1930

	Specifications			
	DD(1)	DD(2)	DD(3)	DD(4)
Ordinance	-12.261	-9.989	-10.926	-8.881
	(9.451)	(8.987)	(8.543)	(9.032)
Population growth	-0.020	-0.060	-0.041	-0.069
	(0.119)	(0.134)	(0.132)	(0.133)
Income (log)	-44.684	-42.987	-68.278***	-46.615
		(31.969)	(7.675)	(33.135)
Female share	----	0.654	-0.059	0.530
		(2.512)	(2.110)	(2.512)
White share	----	-1.095	-1.389*	-1.201
		(0.771)	(0.791)	(0.816)
Share under 5 years	----	14.119***	13.633***	15.105***
		(2.686)	(3.167)	(2.382)
Mortality-1	----	0.089	0.079	----
		(0.093)	(0.084)	
City fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	No	Yes
Observations	204	204	204	204
R^2	0.718	0.758	0.756	0.755

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; no value ----

Figure X. Outliers detection of FE model

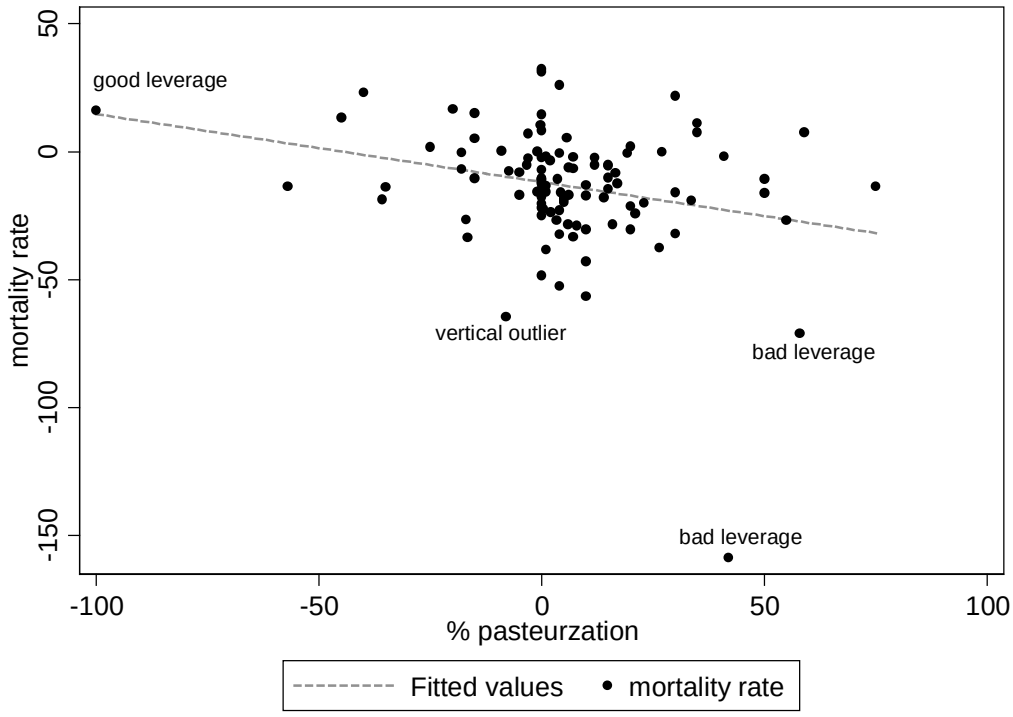


Table VI. FE Regressions of Extent (%) of Pasteurization, 1921-1924

Variables	Specifications				
	LS	L-est.	M-est.	S-est.	MM-est.
% pasteurized milk	-0.248**	-0.157***	-0.164***	-0.151***	-0.132***
	(0.123)	(0.048)	(0.050)	(0.033)	(0.040)
Population (log)	38.520	4.237***	3.895***	6.487***	4.669***
	(24.363)	(1.027)	(1.082)	(1.123)	(1.116)
Year dummy	-14.509***	-4.330***	-3.787***	-17.732***	-12.440***
	(2.373)	(0.766)	(0.798)	(2.236)	(2.330)
Observations	218	218	218	218	218
Groups	109	----	----	----	----
R^2 (Pseudo)	0.315	0.1033	0.165	----	----
Δ% SCM (-6.94)	27.98	44.20	42.32	45.96	52.58

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; no value ----

Table VII. Share (%) of pasteurized milk in cities 1931

Treated	Gap	Control	Gap	Control	Gap	Control	Gap
Baltimore	98.5	Brockton	84.0	Memphis	74.0	Quincy	70.0
Buffalo	100.0	Concord	28.0	Meriden	75.0	Raleigh	33.3
Chicago	99.5	Duluth	58.0	Middletown	88.0	Sacramento	89.5
Cincinnati	100.0	Evansville	85.0	Mobile	15.0	San Antonio	69.0
Cleveland	99.0	Fitchburg	66.2	Montclair	82.3	San Diego	75.5
Dayton	99.9	Green Bay	80.0	Muncie	75.0	Savannah	33.0
Detroit	100.0	Harthord	89.0	Nashville	60.0	Seattle	87.9
Indianapolis	97.5	Jackson City	58.0	New Britain	68.0	Springfield	85.0
Jersey City	98.0	Jacksonville	40.0	New Haven	80.0	St. Paul	79.7
Milwaukee	99.5	Jamestown	25.0	New Orleans	----	Superior	41.0
New York	98.3	Kalamazoo	84.0	Omaha	70.0	Tacoma	54.0
Philadelphia	99.7	Kansas city	50.0	Paducah	60.0	Troy	39.5
Richmond	100.0	La Fayette	35.0	Petersburg	64.0	Utica	79.9
San Francisco	99.5	Lancaster	70.0	Pittsfield	64.0	Wheeling	76.0
St. Louis	96.2	Lincoln	80.0	Plainfiled	71.0	Wichita	66.0
Toledo	99.5	Los Angeles	82.3	Portland	86.7	Wilmington	40.0
		Lynchburg	33.3	Portland	75.0		
Avg. treated	99.1	Manchester	85.0	Providence	86.9	Avg. control	65.6

Source: Frank and Moss, *The extent of pasteurization and tuberculin testing in American cities of 10,000 population and over in 1927 and 1931*. US Public Health Service