Essays in Applied Microeconomics

Yuchen Jiang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2024

Abstract

Essays in Applied Microeconomics

Yuchen Jiang

This dissertation consists of three essays in applied microeconomics.

The first chapter investigates the effect of coroner partisanship on COVID-19 death reporting. The politicization of the COVID-19 pandemic in the United States has raised questions about the integrity and accuracy of death reporting, particularly in jurisdictions with elected, partisan coroners. Using mortality data from the CDC and manually collected data on county-level death certification systems and coroner party affiliation where applicable, I examine the parallel systems of appointed medical examiners and elected coroners and analyze the effect of partisanship on reported COVID-19 deaths. Cross-sectional comparisons do not seem to suggest counties with coroners report fewer deaths than those with medical examiners, and difference-in-differences specifications reveal limited evidence of a statistically significant but not economically meaningful effect of partisanship on reported COVID death counts.

The second chapter examines the effect of new information on lead water pipes on housing prices. In 2016, the Water and Sewer Authority of Washington, DC released an online map that contains information on lead service lines (LSLs) for all properties in the district. Using the release as a natural experiment, I estimate the effect of the new information on prices of properties with and without LSLs. Recent literature has found that housing lead reduction policies such as remediation mandates have significant price

effects. In DC, while the map's release was followed by a marked increase in requests for water lead tests, neither a difference-in-differences model nor a repeat sales model captures a significant divergence between housing prices of the two types of properties after the release, implying the housing market response to the information was limited.

The second chapter considers the effect of the marriage tax subsidy on the marriage decision of same-sex couples. The U.S. Supreme Court's ruling on United States v. Windsor in June 2013 compelled the federal government to recognize state-sanctioned same-sex marriages, including for tax purposes. The switch in the income tax filing status for same-sex couples meant that the marriage penalty or subsidy as a result of joint filing became a relevant factor that may enter couples' marriage decisions. I construct a sample of married and cohabiting same-sex couples in 2012 and 2014 from public-use data of the American Community Survey. Using a difference-in-differences methodology, I do not find evidence that same-sex couples who would earn a higher marriage subsidy became more likely to marry after the Supreme Court ruling.

Table of Contents

List of Figures

List of Tables

Acknowledgments

None of this dissertation would have materialized without the warm support and valuable advice of Douglas Almond and Michael Best during my Ph.D. years. I shall always remain deeply grateful to them.

I would also like to thank Miguel Urquiola for his advising, as well as Suresh Naidu and Adam Sacarny for serving on my defense committee.

My studies—like those of countless Ph.D. students before me—benefited immensely from Amy Devine's tireless work, which I truly appreciate.

Krzysztof Zaremba, Matthew Isbell and Miao Wang offered me tremendous help during the development of Chapters 1 and 2. Many thanks to you all.

To Josh, Ryan, Krzys (again), Xinru and Ernest: thank you for being good friends and helping me through the darkest times. Your faith and encouragement were priceless along this journey.

And Lucas, because of you, I am now a better version of myself.

Dedication

For my parents.

獻給我的父母。

Chapter 1: Death Investigation Systems, Coroner Partisanship and Reported COVID-19 Mortality

1.1 Introduction

Since the initial outbreak of COVID-19 (Coronavirus Disease 2019) in the United States in 2020, measuring deaths that result directly or indirectly from the pandemic has become an essential part of accurately assessing its impact (Zylke & Bauchner, 2020). Challenges abound, and official death numbers are widely believed to be a massive undercount of the true toll (Kiang et al., 2020; Stokes, Lundberg, Bor & Bibbins-Domingo, 2021; Woolf et al., 2020; Woolf et al., 2021). One of those challenges involves the proper attribution of deaths to COVID-19. Public health experts have underscored the importance of death certification in informing the public and guiding the policy response to the pandemic, and identified drawbacks in the current system, such as the lack of COVID testing in many cases and the inadequate training received by death certifiers (Gill & DeJoseph, 2020; Stokes, Lundberg, Bor & Bibbins-Domingo, 2021).

Concerns have been raised, and anecdotal evidence documented, about the fragmented death investigation systems in the United States. The American population is served by a mix of coroner and medical examiner (ME) systems at both state and local levels. Medical examiners are appointed officials who are medical professionals that received specialized training in death certification and forensic autopsy. In contrast, coroners are usually elected and politically partisan, and often laypeople who are neither trained in death certification nor medicine at large (Hanzlick & Combs, 1998; Institute of Medicine (US) Committee for the Workshop on the Medicolegal Death certification System [IOM Committee], 2003). This lack of professional knowledge and training is sometimes assumed to lead to a larger number of unattributed COVID-19 deaths in jurisdictions with coroners compared to those with MEs (Stokes, Lundberg, Bor & Bibbins-Domingo, 2021). Possibly exacerbating the problem is the politicization of the pandemic in the United States, with views clearly divided along party lines among both elected officials and the public. Some media reports and analysis have pointed to possible underreporting of COVID deaths by elected Republican Party coroners out of political motivations (Bergin et al., 2021; "Politics of Death", 2022). A Republican coroner of a Missouri county reportedly said his office "[didn't] do COVID deaths", and attributed no deaths to COVID-19 in 2021 (Bergin et al., 2021). Such reports suggest the possibility that partisan politics may incentivize partisan coroners to manipulate death numbers in their jurisdictions.

Despite the theorizing and anecdotes, to my knowledge, no empirical analysis has been conducted on the effect of coroner partisanship on COVID-19 death reporting. A major obstacle may be data availability on partisanship: states and counties differ vastly in whether and how information about coroner elections and party affiliations is made available, and a comprehensive account of the nationwide picture would require extensive manual data collection and verification. In this paper, I employ an original, manually-collected data set containing such information, combine it with mortality data from the United States Centers for Disease Control and Prevention (CDC) and investigate the effect of death certification systems and partisanship. Taking into consideration the interval-censored nature of the mortality data due to privacy concerns, I estimate both regular, left-censored tobit models and interval-censored models to account for suppressed death counts, and also use the probability of having a low and suppressed death count as a dependent variable in alternative specifications. A cross-sectional comparison of coroner and ME counties does not reveal different levels of underreporting, and my models produce mixed evidence on systemic underreporting in Republican-coroner counties compared to Democratic-coroner ones.

This paper contributes to a nascent and growing literature that seeks to understand the toll of COVID-19 in America. A major focus of the literature is to estimate excess deaths, i.e. the difference between actual and expected numbers of deaths (Zylke & Bauchner, 2020). When performed at subnational levels, these exercises consistently reveal large geographic variations in the percentage of excess deaths not directly attributed to COVID-19, as pointed out by Woolf et al., (2020) and Woolf et al., (2021) at the state level, and by Stokes, Lundberg, Elo, et al. (2021) and Ackley et al. (2022) at the county level, to name a few examples. My work partially aims to test one possible explanation for such disparity. It is also closely related to research that examines the politicization of the COVID-19 pandemic in the United States and its public health consequences, ranging from individual behaviors (e.g. Allcott et al., 2020; Grossman et al., 2020) to local government policy (e.g. Holman et al., 2020) to possible fraudulent death reporting practices (Eutsler et al., 2023). This paper is the first to focus on the death investigation system and seeks to test the theory of political influence on death certifiers.

The rest of the paper is structured as follows. Section 1.2 reviews the facts and findings on the politicization of COVID-19 in the United States and outlines the country's two main types of death investigation system. Section 1.3 provides a summary of the data and sample. Section 1.4 describes the empirical methodology. Section 1.5 presents the results and offers possible interpretations. Section 1.6 discusses the findings and Section 1.7 concludes.

1.2 Background

1.2.1 Politicization of COVID-19 in the United States

The emergence of COVID-19 in the United States coincided with a period of intense political polarization in the country. Unsurprisingly, the public discourse on the pandemic also largely evolved into a polarized political debate during its course. Ever since the virus's initial outbreak, top-ranking and high-profile Republican Party figures repeatedly downplayed the threat posed by the virus, endorsed conspiracy theories and pseudoscientific treatments, used racist language to refer to the disease, contradicted public health recommendations issued by the CDC, and dismissed Democratic officials' concerns and policy responses as political stunts (Bolsen & Palm, 2022; Halpern, 2020). Both polarization and politicization were also amplified by the news media (Hart et al., 2020); conservative media in particular spread and promoted COVID-related misinformation (Motta et al., 2020).

As a consequence, perceptions, attitudes and behavior among the American public all displayed sharp partisan divisions. Two longitudinal and cross-national studies conducted by Stroebe et al. (2021) show that the extent of such politicization increased over time, and was greater in the U.S. than in a comparison group of countries. Compared to self-identified liberals or Democrats, conservatives or Republicans were less concerned about the health risk posed by COVID-19 and less trusting in mainstream media's reporting on the pandemic and public health recommendations from medical experts (Allcott et al., 2020; Kerr et al., 2021; Rothgerber et al., 2020). They also reported less adherence to health-protection protocols such as hand-washing, quarantining, mask-wearing and social distancing (Allcott et al., 2020; Rothgerber et al., 2020; Kerr et al., 2021; Stroebe et al., 2021), were more skeptical of and less likely to receive COVID vaccines (Bolsen & Palm, 2022), and less supportive of aggressive government policies both in pandemic control and on related public issues (Gadarian et al., 2021). Moreover, Gadarian et al. (2021) find that the partisan divide cannot be fully explained by other correlating variables, such as consumption of conservative media or the local COVID-19 death toll.

The self-reported differences are corroborated by empirical data. Google search data indicates Democrats showed greater interest in social distancing (Grossman et al., 2020). Using mobile phone location data, Allcott et al. (2020), Grossman et al. (2020) and Gollwitzer et al. (2020) all find that residents in U.S. counties with higher Democratic vote shares in the 2016 presidential election were more likely to practice social distancing and comply with stay-at-home orders. These disparities between counties are also subsequently linked to higher COVID-19 infection and death rates in Republican-leaning counties (Gollwitzer et al., 2020).

Furthermore, local policymaking seems to reflect the partisan differences, too. Democratic governors were generally more prompt in adopting a variety of social-distancing policies than Republican ones (Adolph et al., 2021; Grossman et al., 2020). Holman et al. (2020) find the ideological leaning of local populations to be one of the factors that affected how early municipal governments issued stay-at-home orders. There is also evidence on political influences on COVID death reporting: Eutsler et al. (2023) employ Benford's law, a phenomenon observed in naturally occurring numerical data sets, and find evidence of underreporting of COVID-19 deaths; in addition, the extent of such underreporting in a county is related to the county's partisan leaning in the 2016 presidential election vote as well as the party affiliation of the state governor. All of this points to the possibility of a partisan line that divides death investigation systems in America as well.

1.2.2 Death certification during COVID-19 in the United States

Death certificates are a crucial source of information about public health. During the COVID-19 pandemic, data from death certificates formed the basis of the mortality statistics published by the CDC's National Center for Health Statistics, and informed national and subnational monitoring of the pandemic's progression and severity (Gill & DeJoseph, 2020). Normally, natural-cause deaths (such as those resulting from viral infections) that occur in the hospitals or long-term care/hospice facilities are certified by a facility physician (Bhullar et al., 2022; Gill & DeJoseph, 2020). These deaths thus do not require reporting to medical examiner or coroner (ME/C) offices, which together form the medicolegal death investigation system in the United States and are legally mandated to investigate unnatural or unexpected deaths (IOM Committee, 2003). Nevertheless, ME/Cs played an important role in certifying COVID-related deaths during the pandemic for a number of reasons. Firstly, all deaths that take place outside hospitals or care facilities are reportable to ME/C offices (Bhullar et al., 2022), and such deaths accounted for a meaningful portion of COVID deaths (Pathak et al., 2021). More importantly, while statutes that specify the types of death reportable to a

ME/C differ by jurisdiction, most jurisdictions require reporting of deaths that involve diseases that may constitute a threat to public health, shifting confirmed and suspected COVID deaths into ME/Cs' purview, and allowing ME/Cs to revise death certificates when necessary (Gill & DeJoseph, 2020; Kiang et al., 2020; National Vital Statistics System, 2023). In practice, ME/Cs are extensively involved in certifying and counting COVID-related deaths (Zavattaro, 2023).

ME and coroner systems coexist in the United States today, and the type of office overseeing medicolegal death investigation varies by state and county. According to the CDC (2023b), 22 states and the District of Columbia exclusively use ME systems; among them, 16 states and D.C. have a centralized system at the state level, and 6 have a county- or district-based system. The remaining 28 states use coroner systems in at least some parts of the state, with 14 of them using a county- or district-based coroner system, and the other 14 using a county- or district-based system with a combination of MEs and coroners. The coroner system originated in 9th- or 10th-century England, and its current use in the United States is a vestige of the British colonist era (Hanzlick & Combs, 1998). The modern incarnation of the ME systems first emerged in 1877, and it is the consensus among today's public health experts that ME systems are clearly preferable to coroner systems (IOM Committee, 2003), but the latter persist. From the 1960s to the 1980s, a period of rapid transition from coroner systems to ME systems took place nationwide, followed by a "lull in the action" starting in the 1990s (Hanzlick, 2007); in recent years, the pace of conversion seems to be picking up again (Denham et al., 2022).

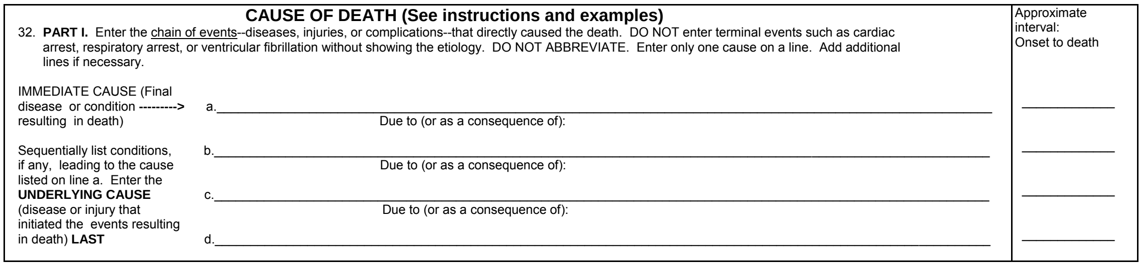
Two main differences distinguish ME and coroner systems from each other, and both point to reasons for the former's advantage over the latter. The first difference concerns qualification and professionalism. ME offices are held by medical professionals—usually physicians, often pathologists or forensic pathologists—who additionally receive special training and certification in death investigation (Hanzlick & Combs, 2007; IOM Committee, 2003). This type of training is seldom provided in medical schools or healthcare facilities, making MEs more competent death certifiers than other healthcare professionals. On the other hand, neither such qualification nor training is required of coroners, and they are almost always laypeople who need as little as a high school diploma to qualify for the job (Choi & Gulati, 2017; IOM Committee, 2003). The deficiency in knowledge and training makes coroners less capable of the task of investigating deaths: coroner systems have been found to be less efficient and more error-prone than ME systems (Choi & Gulati, 2017; Denham et al., 2022; Flynn, 1955).

The second difference that sets the two systems apart pertains to the method of selection for each type of office. MEs are appointed, whereas coroners are usually elected (IOM Committee, 2003). The perceived electoral mandate of coroners is one of the main arguments made against the conversion to ME systems (Flynn, 1955; IOM Committee, 2003). But the flip side of representing the will of the electorate is that the electoral system provides incentives for coroners to respond to political pressure, and voters' demands may not always align with what is good for society (Choi & Gulati, 2017; "Politics of Death", 2022). In the case of COVID-19, where politics and public health have become so inextricably intertwined, it is conceivable that the conflict of interests can lead elected coroners to make questionable decisions. Given their authority in death certification, coroners may, for instance, choose to omit COVID-19 as a cause of death in the absence of a positive laboratory test, even though CDC guidelines say it should be listed as long as certain clinical criteria are met; they may also remove COVID-19 from a death certificate bowing to pressure from the family of the deceased (Bergin et al., 2021; Bordelon, 2021; "Politics of Death", 2022). In their Benford's law analysis, Eutsler et al. (2023) find descriptive evidence that counties with MEs were less likely to see politically-motivated underreporting than those with coroners. Some descriptive works that estimate excess deaths from COVID-19, such as Paglino et al. (2023), point out that regions with higher discrepancies between reported COVID deaths and estimated excess deaths were more likely to have coroners rather than MEs.

Laws governing ME/C offices vary greatly by state, and the offices differ not only in the manner of selection of ME/Cs, but also in their structure, operation and procedures (CDC, 2023b; Hanzlick et al., 1993). The United States Department of Justice conducts five-yearly censuses of ME/C countries nationwide. The most recent one, conducted in 2018 (Brooks, 2021), shows that ME/C offices vary in size of jurisdiction, manpower, budget and caseload, all of which are positively correlated. For example, on average, coroner offices serving counties with a population larger than 250,000 have a staff of 15 people, whereas those in counties with a population smaller than 25,000 only employ two people. Employees of the office act on behalf and under the authority of the ME/C, who serves as the primary death investigator and is responsible for completing the death certificate (Hanzlick, 1996). It is customary in the literature examining the effect of death investigation systems to treat the ME/C office as a unitary entity without distinguishing between the roles of the overseeing ME/C and other staff members of the office (e.g. Denham et al., 2022; Klugman et al., 2013), and I adopt the same approach in this paper.

1.3 Data

The data used for my analysis mainly consists of two parts. Mortality data comes from the Wide-ranging ONline Data for Epidemiologic Research (WONDER) system maintained by the CDC (2023a), and includes monthly all-cause (i.e. total) death and COVID death counts at the county level in 2020 and 2021, along with each county's urban–rural classification according to the National Center for Health Statistics' 2013 scheme (the most recent update to the scheme). While CDC WONDER contains arguably the highest-quality mortality data for the United States, its privacy rules introduce one complication to my analysis. In order to avoid revealing individual identities, CDC wonder suppresses death counts when they are below 10; instead of the actual value, the death count is replaced by a dedicated code. Counts of zero are not suppressed. The suppression hence results in interval-censored data, with death counts between 1 and 9 concealed. I will address this issue when I introduce my empirical models in Section 1.4.



CAUSE OF DEATH (See instructions and examples)

32. **PART I.** Enter the chain of events--diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.

Approximate interval: Onset to death

IMMEDIATE CAUSE (Final disease or condition ---------> resulting in death) a.______

Due to (or as a consequence of):

Sequentially list conditions, if any, leading to the cause listed on line a. Enter the **UNDERLYING CAUSE** (disease or injury that initiated the events resulting in death) **LAST**

b.______

Due to (or as a consequence of):

c.______

Due to (or as a consequence of):

d.______

Figure 1.1: Part of the U.S. standard certificate of death

I include three types of death counts in my data: all-cause deaths, deaths where COVID-19 is listed as a cause of death, and deaths where COVID-19 is listed as the *underlying* cause of death. On the U.S. standard death certificate, up to 20 causes of death can be listed to form a "chain of events that directly cause the death" (see Figure 1.1). The first listed cause is the immediate one, and the last the underlying one, which initiated the chain. The main analysis focuses on deaths where COVID-19 is mentioned anywhere in the list of causes, i.e. the most broadly-defined COVID deaths based on death certificate data. Death counts with COVID-19 listed as the underlying cause are used for supplemental analysis.

Additionally, I obtained data on types of death investigation office and party affiliations of elected coroners in both 2017 and 2021. This data is generously provided by by Matthew Isbell, a political data analyst. He manually compiled the data from state and local government websites and directories depending on where states store and publish such information, and crosschecked it with multiple sources, including the CDC (2023b). To the best of my knowledge, this is the first time such data has been gathered and used in empirical research: while earlier research on the U.S. death investigation systems at the local level, such as Denham et al. (2022), has utilized data on office type, mine is the first to focus on the effect of coroner partisanship.

Figures 1.2 and 1.3 provide a visualization of the data Isbell collected. In these maps, I only highlight states with at least one coroner office; the rest have either statewide or county/district-based ME systems.[1] As of 2021, 1,276 counties across America elected coro-

[1] The highlighted states in the maps exclude three states despite their being listed in CDC (2023b) as ones

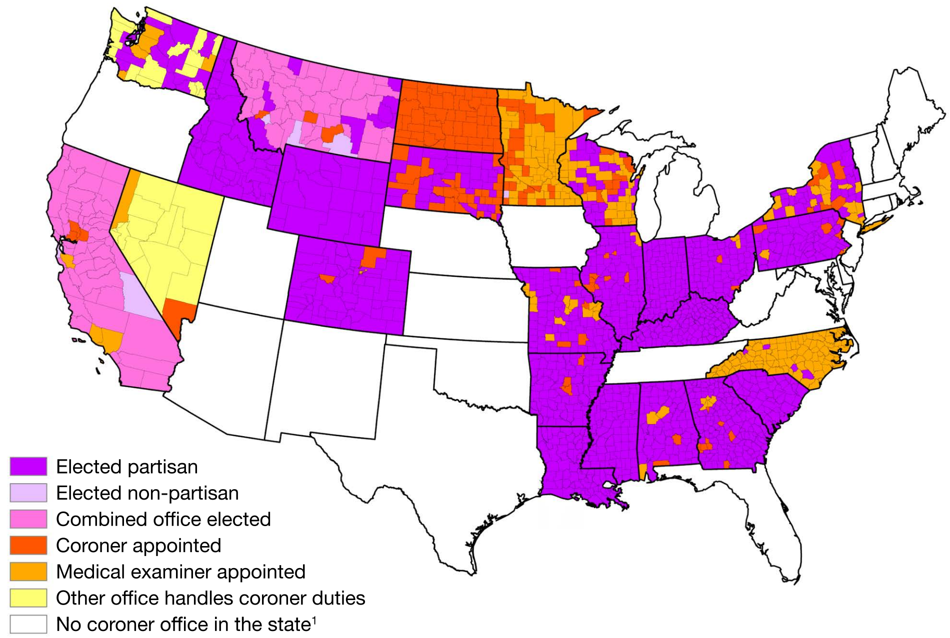


Figure 1.2: Type of death investigation office and method of selection in 2021

ners, with all but six conducting the elections in a partisan manner. A small number (176) of counties appoint rather than elect coroners, including the entirety of North Dakota. Republicans held more coroner offices than Democrats in both years, and made moderate gains in the 2020 election. Among coroner offices held by either of the two major parties in both 2017 and 2021, Democrats and Republicans held on to 352 and 725 counties, respectively, after the election; 119 flipped from Democrats to Republicans, and 20 in the opposite direction.

I augment the mortality and ME/C office data with two additional sources. The mid-2020 county resident population estimates from the United States Census Bureau (2022) are used to derive demographic characteristics of counties, including the shares of gender, race and age groups in the population. The 2020 county-level presidential election results are obtained from the MIT Election Data and Science Lab (2018).[2]

with coroner or mixed ME/C systems. Texas has a mixed system, with medical examiners in some counties and the office of justice of the peace handling coroner duties in others. In Nebraska, county attorneys perform coroner duties. Kansas has a district-based system. These states are excluded from the empirical analysis, as none have strictly-defined, county-based coroners.

[2]In addition, I manually collected 2020 county coroner election results, where possible, from various state and county sources. However, because a vast majority of coroner races were uncontested, the results are not very informative for my purposes and therefore not included in the data.

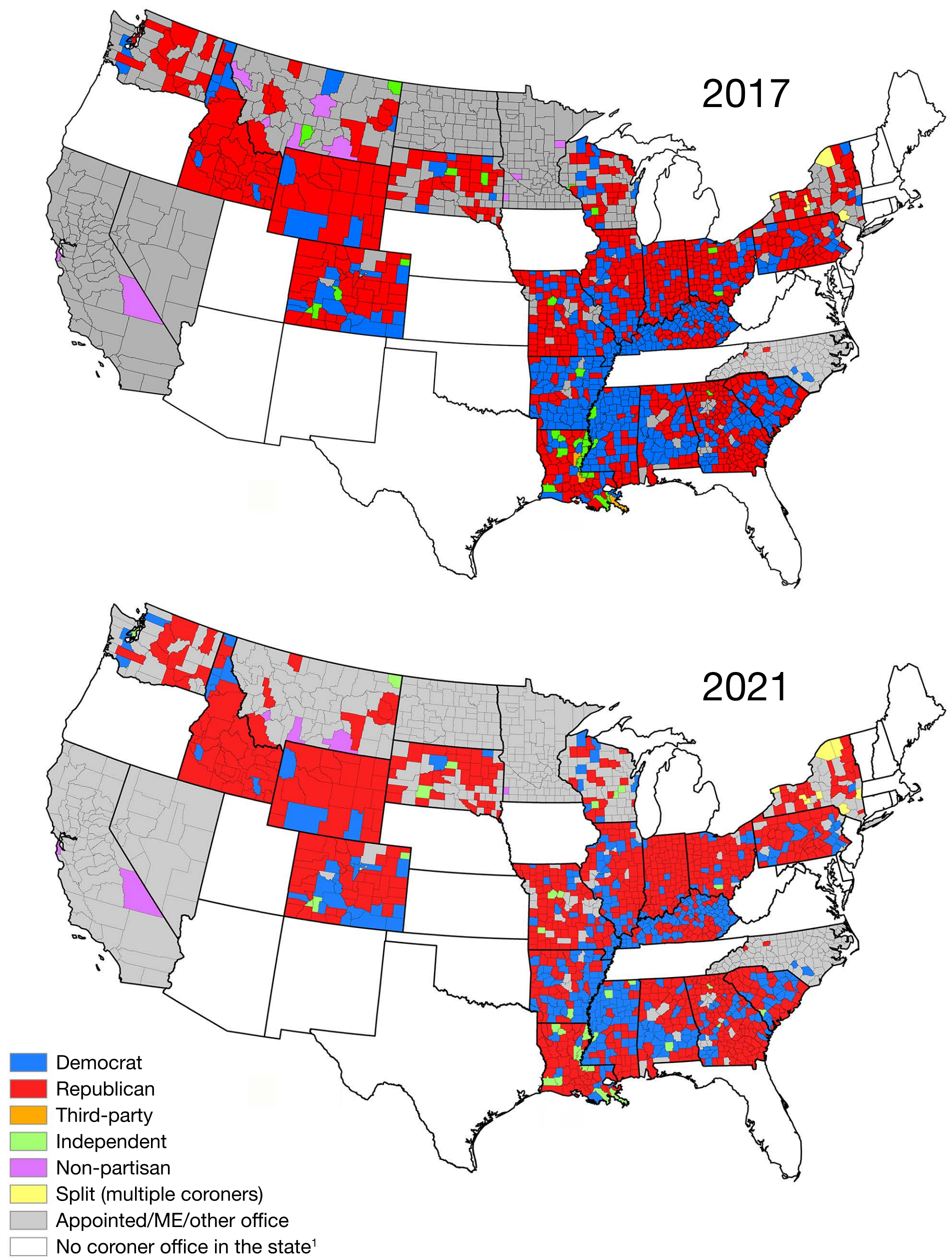


Figure 1.3: Partisan makeup of death investigation offices in 2017 and 2021

My main sample consists of 2,367 counties over 22 months, from March 2020 to December 2021, for a total of 52,074 observations. The counties are from 46 states and the District of Columbia. Three states are excluded as explained in footnote 1. Among the non-ME counties, I only include those with elected partisan coroners, and drop those with appointed coroners, elected non-partisan coroners and combined offices to allow for straightforward comparisons. As can be seen in Figure 1.2, this selection criterion excludes the entire state of North Dakota (which appoints coroners), a majority of counties in California, Montana (both of which elect combined offices) and Nevada (where coroners' offices are combined with sheriffs' offices [CDC, 2023b]), about half of Washington and South Dakota, a third of Minnesota, and a few counties in eight other states. Some counties in Alaska and Hawaii are also excluded because of non-perfect matching between jurisdiction definitions from the three different sources in my data set. Five counties in the sample reported zero COVID-19 deaths in 2020 and 2021. Among the 2,367 counties, 1,270 (53.7%) have elected partisan coroners, and the remaining 1,097 (46.3%) have MEs. Table 1.1 shows the partisan breakdown within the coroner counties before and after the 2020 election. As pointed out previously, coroners are more likely to be Republican than Democratic.

Table 1.2 summarizes and compares characteristics both between ME and coroner counties, and between Democratic-coroner and Republican-coroner counties in 2021. Both pairs of groups are largely similar in terms of age structure and sex ratio, and any significant differences are small in magnitude. The racial composition is different within each pair: ME counties tend to have fewer black residents and more Hispanic ones than coroner counties; Republican-coroner counties have much more white residents and much fewer black ones than Democratic-coroner counties. Politically, as expected, Republican-coroner counties were more in favor of Donald Trump, the incumbent president and Republican presidential nominee, than Democratic-coroner ones in the 2020 election; ME counties were more supportive of Joe Biden, the Democratic nominee, and saw a greater leftwards shift compared to 2016 than coroner counties. Geographically, ME counties are more likely located in the

		Democrat	Republican	Other/No party	Split
Year	2017	478 (37.6%)	755 (59.4%)	32 (2.5%)	5 (0.4%)
	2021	374 (29.4%)	853 (67.2%)	35 (2.8%)	8 (0.6%)

Table 1.1: Partisan makeup of elected partisan coroners

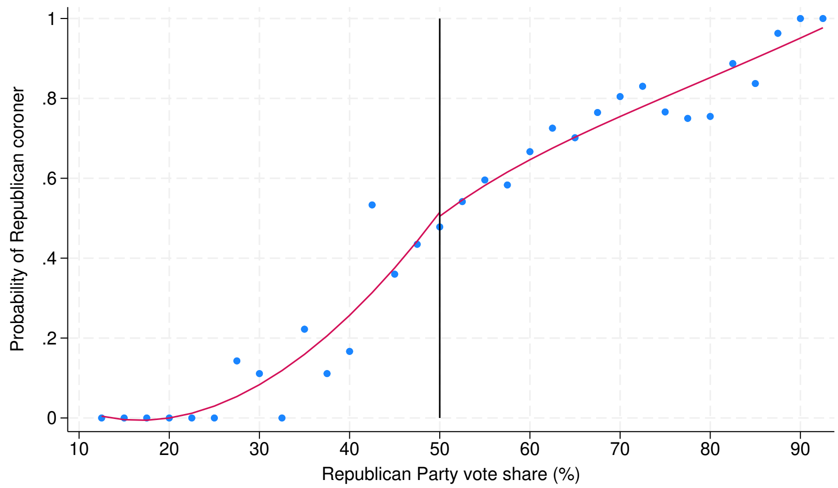


Figure 1.4: 2020 county Republican vote share and coroner partisanship

Northwest and less likely in the Midwest; Republican-coroner counties are much more likely to be in the Midwest and much less likely to be in the South. Finally, ME counties are more likely to be metropolitan, defined as being in a metropolitan statistical area. It is also worth pointing out that while the average population is marginally different between Democratic- and Republican-coroner counties, a closer look at their distributions reveals that both include counties of all sizes in roughly similar proportions. This alleviates concerns about different practices between offices in large and small jurisdictions described at the end of Section 1.2, as any effect of such differences should be balanced out between the two parties.

Figure 1.4 plots the probability of a county coroner being Republican against the county's Republican Party vote share (between the two major parties) in the 2020 presidential election. Coroner party affiliation is highly correlated with presidential vote share, but a regression-discontinuity (RD) style quadratic fit shows no discrete jump at a cutoff of 50% vote share. This indicates ticket splitting between the presidential candidate and the coroner candidate was common, which was in turn likely due to the large proportion of uncontested coroner

	Office type			Coroner party (2021)		
	Coroner	ME	Difference (ME – C)	Democrat	Republican	Difference (R – D)
Population, Jul 2020 est.	61,322	165,426	104,104***	70,858	57,926	$-12,933$*
	(109,287)	(468,183)	(13,545)	(150,140)	(87,539)	(6,848)
Share of age group (%):						
0 to 19	24.658	23.708	−0.950***	24.346	24.819	0.474**
	(3.147)	(3.478)	(0.136)	(2.947)	(3.182)	(0.193)
20 to 29	12.034	11.973	−0.061	12.491	11.813	−0.677***
	(2.822)	(3.024)	(0.120)	(3.175)	(2.629)	(0.174)
30 to 49	23.702	23.627	−0.075	23.669	23.718	0.049
	(2.325)	(2.793)	(0.105)	(2.288)	(2.315)	(0.143)
50 to 64	20.430	20.627	0.196**	20.412	20.445	0.033
	(2.083)	(2.390)	(0.092)	(2.050)	(2.105)	(0.130)
65 to 74	11.447	11.894	0.447***	11.429	11.445	0.015
	(2.230)	(2.874)	(0.105)	(2.075)	(2.292)	(0.138)
75 to 84	5.692	6.002	0.310***	5.646	5.713	0.067
	(1.258)	(1.774)	(0.063)	(1.254)	(1.265)	(0.078)
85 and above	2.037	2.170	0.133***	2.007	2.047	0.040
	(0.621)	(0.704)	(0.027)	(0.588)	(0.632)	(0.038)
Share of male (%)	50.230	50.207	−0.023	49.917	50.327	0.410***
	(2.325)	(2.056)	(0.091)	(2.409)	(2.221)	(0.141)
Share of racial group (%):						
White	82.712	83.794	1.083	73.836	86.988	13.153***
	(18.049)	(15.081)	(0.690)	(23.773)	(12.587)	(1.042)
Black	13.010	8.796	−4.214***	22.145	8.689	−13.456***
	(17.896)	(12.289)	(0.641)	(24.080)	(11.901)	(1.029)
Hispanic	5.754	8.989	3.235***	5.499	5.901	0.402
	(6.867)	(11.337)	(0.380)	(7.302)	(6.763)	(0.430)
2020 pres. vote share (%):						
Democratic (Biden)	28.313	33.315	5.003***	34.963	24.959	−10.004***
	(14.188)	(16.337)	(0.628)	(17.049)	(11.368)	(0.828)
Republican (Trump)	61.256	51.122	−10.134***	53.237	64.959	11.721***
	(17.626)	(18.493)	(0.743)	(19.938)	(15.323)	(1.046)
Shift to GOP vs 2016	−3.728	−7.146	−3.418***	−3.458	−3.999	−0.541
	(10.988)	(11.900)	(0.471)	(9.266)	(11.822)	(0.689)
Region (1 = yes):						
Northeast	0.072	0.105	0.033***	0.048	0.076	0.028*
	(0.258)	(0.306)	(0.012)	(0.214)	(0.265)	(0.016)
Midwest	0.339	0.268	−0.071***	0.217	0.400	0.183***
	(0.474)	(0.443)	(0.019)	(0.412)	(0.490)	(0.029)
South	0.469	0.509	0.039*	0.650	0.390	−0.259***
	(0.499)	(0.500)	(0.021)	(0.478)	(0.488)	(0.030)
West	0.120	0.119	−0.001	0.086	0.134	0.048**
	(0.325)	(0.323)	(0.013)	(0.280)	(0.340)	(0.020)
Metro area (1 = yes)	0.354	0.470	0.117***	0.318	0.374	0.056*
	(0.478)	(0.499)	(0.020)	(0.466)	(0.484)	(0.030)
n	1,270	1,097	—	374	853	—

Levels of significance: *** = .01, ** = .05, * = .10.

Table 1.2: Descriptive statistics for sub-samples

races. The close relationship between the two variables, shown here and in Table 1.2, calls for controlling for the presidential vote in my empirical specification. On the other hand, the lack of a jump at the 50% level precludes the adoption of a fuzzy-RD design.

Before moving on to econometric analysis, I take inspiration from works like Eutsler et al. (2023) and Campolieti (2022) and check for signs of underreported COVID-19 deaths using a mathematical tool. Benford (1938) described the following phenomenon. In many sets of naturally-occurring numbers, the leading digits follow a probability distribution where the smaller numbers occur more often than larger numbers: the probability of the leading digit being $d \in [1, 9]$ equals $P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$, with $P(1) \approx 0.301$, $P(2) \approx 0.176$, ..., and $P(9) \approx 0.046$. Similar distributions also exist for non-leading digits: for example, 0 is the most likely second digit (probability 0.120), and 9 the least likely (probability 0.085). Deviations from the Benford distributions in supposedly naturally-occurring data may indicate data manipulation, and Benford's law has been successfully used to detect fraud in a wide range of contexts (Mebane, 2006; Nigrini, 2012). Eutsler et al. (2023) argue that daily COVID death counts are likely to meet the necessary conditions for the data to satisfy the law, and find higher frequencies for small leading digits in the reported death counts than the theoretical distribution predicts, pointing to possible underreporting. In the case of monthly CDC WONDER data, despite the censoring of death counts below 10, the law should still hold for the remaining data with higher orders of magnitude (Benford, 1938) and, in any case, for the second digits of the death counts as they are not affected by the censoring.

Figure 1.5 illustrates the comparison of observed frequencies for the leading and second digits with the theoretical Benford's law distributions across subsamples grouped by office type, or by the party affiliation of coroner in each year (for example, the subsample "Democratic 2020" only includes death counts in the year of 2020 in counties with a Democratic coroner in that year). Two patterns stand out from the charts. First of all, for both the first and second digits, low (high) digits appear more (less) frequently in COVID death counts

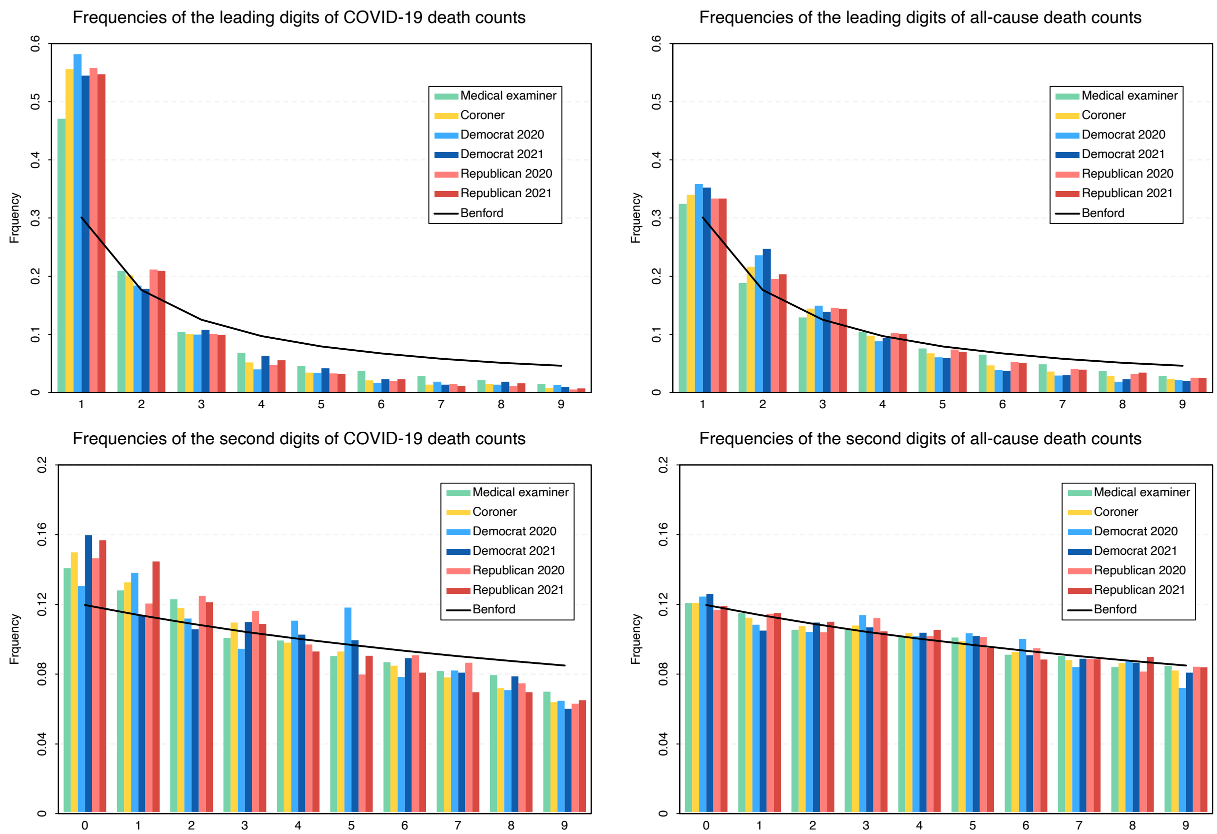


Figure 1.5: Observed frequencies of leading and second digits vs. Benford's law distributions

than predicted by the corresponding Benford distribution, but the observed and theoretical distributions seem to match much better for all-cause death counts, confirming the applicability of Benford's law to my data set while indicating possible underreporting of COVID deaths. Secondly, the extent of deviation of COVID death counts from Benford distributions is similar across all subsamples, and there does not appear to be clear visual evidence suggestive of systemic underreporting by either office type or either party in either year.

Nigrini (2012) proposes the use of the mean absolute deviation (MAD) as a testing statistic for conformity to Benford's law. It is defined as

$$MAD = \frac{1}{D}\sum_{d=1}^{D} |OP_d - TP_d| \, ,$$

where D is the total number of possible digits in the position (9 for the leading digit and

10 for all thereafter), and OP_d and TP_d are the observed and theoretical proportions of digit d. The smaller the statistic, the closer the conformity. In Table 1.3, I calculate the MAD for both the leading and the second digits across all subsamples and compare them to Nigrini's (2012, p.160) critical values for different levels of conformity. The statistical tests reported in panels (a) and (b) confirm my visual observation: the MAD is vastly higher for COVID-19 deaths than for all-cause deaths in every case. Tests for both the leading and second digits also indicate that ME counties are subject to less underreporting than coroner counties. In terms of coroner party affiliation, however, neither party seems to reliably and significantly outperform the other in both years. Panel (c) focuses on the leading digits of COVID death counts in major-party-coroner counties, and calculates the MAD in each year for counties that saw a party switch and those that didn't. The tests again suggest non-conformity across the board, although there now appears to be indicative evidence of a pre-existing gap between switch and no-switch counties: Democratic-coroner counties that flipped Republican already had a higher MAD in 2020 than those that didn't flip, and Republican-coroner ones that flipped had a lower MAD in 2020 than their no-flip counterparts. However, the gaps seem to have persisted in 2021 without growing in size (a smaller MAD in Republican-to-Democratic counties in 2021 is likely due to a small sample size of 110), suggesting the actual party switch did not exacerbate the pre-existing differences. Instead, these differences are probably attributable to factors other than the death investigation system, such as public attitudes towards COVID death certification and prevalent practice in the healthcare facilities.

1.4 Empirical Methodology

1.4.1 OLS difference-in-differences framework

I employ a difference-in-differences strategy in my empirical analysis. In order to examine the effect of coroner partisanship on reported deaths, I leverage party flips in coroners'

(a)

Leading digits:	ME	Coroner	Dem 2020	Dem 2021	Rep 2020	Rep 2021
COVID deaths	$0.0452^{\times}$	$0.0623^{\times}$	$0.0642^{\times}$	$0.0547^{\times}$	$0.0650^{\times}$	$0.0621^{\times}$
All-cause deaths	0.0104**	$0.0220^{\times}$	$0.0315^{\times}$	$0.0303^{\times}$	$0.0172^{\times}$	$0.0182^{\times}$

(b)

Second digits:	ME	Coroner	Dem 2020	Dem 2021	Rep 2020	Rep 2021
COVID deaths	0.0098**	$0.0127^{\times}$	$0.0140^{\times}$	0.0101*	$0.0123^{\times}$	$0.0170^{\times}$
All-cause deaths	0.0019**	0.0020***	0.0059***	0.0036***	0.0033***	0.0020***

(c)

Leading digits:	Dem hold	Dem to Rep	Rep hold	Rep to Dem
2020	$0.0605^{\times}$	$0.0807^{\times}$	$0.0654^{\times}$	$0.0581^{\times}$
2021	$0.0603^{\times}$	$0.0800^{\times}$	$0.0601^{\times}$	$0.0339^{\times}$

Levels of conformity:
*** = close conformity, ** = acceptable conformity, * = marginally acceptable conformity, $^{\times}$ = nonconformity

Table 1.3: Mean absolute deviation (MAD) test for Benford's law conformity

offices in the 2020 election. Because of the minimal number of counties that had third-party/independent coroners or offices with split party control, it becomes impractical to estimate the effect of party flips either to and from these affiliations. Hence I confine the analysis to counties where the coroner belonged to one of the two major parties in both years. This leaves me with 1,216 counties (out of 1,270 coroner counties) for a total of 26,752 observations. The basic diff-in-diff specification is given by the regression model

$$\begin{aligned} y_{it} = \alpha_0 + \alpha_t + \eta \times Post_t + \beta_1 \times DtoR_i + \beta_2 \times DtoR_i \times Post_t \\ + \gamma_0 \times R_i + \gamma_1 \times RtoD_i + \gamma_2 \times RtoD_i \times Post_t + Z_{it}'\theta + \epsilon_{it}, \end{aligned}$$

where y_{it} is the death count in county i in month t; α_t is the month fixed effect; $DtoR_i$ and $RtoD_i$ are indicator variables that equal 1 if the coroner's office in county i flipped from Democratic to Republican, or from Republican to Democratic, in the 2020 election, respectively; $R_i = 1$ if county i had a Republican coroner in 2020; $Post_t = 1$ if month t is January 2021 or later, when newly elected coroners were in office; X_{it} is a vector of covariates, including 2020 presidential election vote shares, the shift towards the Republican Party between 2016 and 2020, population and its polynomials (up to third-order) and logarithm, fixed effects for state, urbanization status, and demographic (age, gender, race) groups, as well as region and month fixed effects and their interaction (to roughly account for the different timing of COVID-19 waves in different U.S. regions). The coefficients of interest are β_2 and γ_2, which capture the effects on reported deaths of a Democratic-to-Republican or a Republican-to-Democratic flip, under the identifying assumption of parallel pre-trends between counties that had coroners from the same party in 2020 but different ones in 2021. Because the coroner's office only has an effect on COVID death counts through its capacity for medicolegal death investigation and does not affect local public health policy, and assuming the latter has been sufficiently controlled for using the presidential election vote shares, these coefficients of interest represent the effects of coroner party changes alone.

Given the structure of my data set, I estimate the equivalent (and more symmetric)

$$y_{it} = \alpha_0 + \alpha_t + \eta Post_t + \sum_{k=1}^{3} (\phi_k Change_{ki} + \delta_k Change_{ki} \times Post_t) + Z'_{it}\theta + \epsilon_{it}, \tag{1.1}$$

where $Change_{1i}$, $Change_{2i}$ and $Change_{3i}$ are a set of indicator variables for the coroner party change in the 2020 election ($k = 1$ means "Republican hold", 2 means "Democratic to Republican flip", 3 means "Republican to Democratic flip", and "Democratic hold" is the omitted group). In this specification, the coefficients of interest are represented by δ_2 and $\delta_3 - \delta_1$, which are equal to β_2 and γ_2 from Equation (1.1), respectively. I also allow for dynamic treatment effects by estimating the alternative specification

$$y_{it} = \alpha_0 + \alpha_t + \sum_{k=1}^{3} (\phi_k Change_{ki} + \tau_{kt} Change_{ki}) + Z'_{it}\theta + \epsilon_{it}, \tag{1.2}$$

where the effect of a Democratic-to-Republican (or Republican-to-Democratic) shift in month t compared to the reference month is captured by τ_{2t} (or $\tau_{3t} - \tau_{1t}$).

1.4.2 Considerations about data censoring

As mentioned in Section 1.3, CDC WONDER censors low death counts due to privacy concerns, i.e. to avoid the identification of individuals. One straightforward solution is to estimate Equation (1.2) using a standard tobit model (Tobin, 1958) and treating the dependent variable as left-censored below 10.[3] The model, which is included in common statistical packages, is estimated using maximum likelihood estimation (MLE). Without loss of generality and for the sake of simplicity, I can designate all suppressed death counts y_{it} as equal to 9, a value not observed for y_{it} elsewhere in the data. Then, assuming a conditional normal distribution for the true (latent) death count y_{it}^*, the likelihood function for the sample is

$$\mathcal{L}(\beta, \sigma) = \prod_{y_{it}>9} \left[\frac{1}{\sigma} \varphi \left(\frac{y_{it} - X_{it}'\beta}{\sigma} \right) \right] \prod_{y_{it} \leq 9} \Phi \left(\frac{9 - X_{it}'\beta}{\sigma} \right),$$

and the log-likelihood function is

$$\log \mathcal{L}(\beta, \sigma) = \sum_{y_{it}>9} \log \left[\frac{1}{\sigma} \varphi \left(\frac{y_{it} - X_{it}'\beta}{\sigma} \right) \right] + \sum_{y_{it} \leq 9} \log \Phi \left(\frac{9 - X_{it}'\beta}{\sigma} \right), \tag{1.3}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and cumulative density function of a standard normal distribution, respectively.

There are two main drawbacks to this standard approach. Firstly, since death counts of zero are actually not censored, treating them as such means discarding a large amount of information in the data (out of the 52,074 county-month observations, 12,895, or close to a quarter, had a death count of zero). Secondly, the tobit model assumes a conditional normal distribution for the dependent variable, which may not be the most suitable assumption for death counts. I therefore propose two alternative regression models to address the censoring issue.

[3] Censored regression models are only suitable for data where the dependent variable is censored based on a fixed threshold, which is the main reason why I use raw death count rather than death rate as the dependent variable.

The first model is a slight modification of the standard tobit model that allows me to make use of the zero counts. The data is treated (correctly) as interval-censored on $[1, 9]$. Additionally, under the assumption of normal distribution, any $y_{it} = 0$ can be considered a "censored" value for a true $y_{it}^* < 0$. The log-likelihood function then becomes

$$\log \mathcal{L}(\beta, \sigma) = \sum_{y_{it}>9} \log \left[\frac{1}{\sigma} \varphi \left(\frac{y_{it} - X_{it}'\beta}{\sigma} \right) \right] + \sum_{y_{it}=9} \log \left[\Phi \left(\frac{9 - X_{it}'\beta}{\sigma} \right) - \Phi \left(\frac{-X_{it}'\beta}{\sigma} \right) \right] + \sum_{y_{it}=0} \log \Phi \left(\frac{-X_{it}'\beta}{\sigma} \right), \quad (1.4)$$

and I can estimate the coefficients β and σ using MLE.

My second model is a modification of the standard Poisson regression model. I assume that the dependent variable y_{it}^* follows a Poisson instead of normal distribution, a more accurate assumption for death counts (Scott, 1981), and that the logarithm of its conditional expectation is a linear combination of the covariates, i.e. $\log E(y_{it} | X_{it}) = X_{it}'\beta$, or $E(y_{it} | X_{it}) = \exp(X_{it}'\beta)$. Because the probability mass function of a Poisson distribution with expectation λ is

$$f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!},$$

in the standard Poisson regression model, which is also estimated using MLE, the likelihood function would be given by

$$\mathcal{L}(\beta) = \prod_{i,t} \frac{\exp(y_{it} X_{it}'\beta) \exp\left(-e^{X_{it}'\beta}\right)}{y_{it}!},$$

and the log-likelihood function

$$\log \mathcal{L}(\beta) = \sum_{i,t} \left[y_{it} X_{it}'\beta - \exp(X_{it}'\beta) - \log(y_{it}!) \right].$$

Now I modify the Poisson regression model to allow for interval censoring. When y_{it} is interval-censored on $[1,9]$ and designated a value of 9, the likelihood function becomes

$$\mathcal{L}(\beta) = \prod_{y_{it} \neq 9} \frac{\exp\left(y_{it} X_{it}' \beta\right) \exp\left(-e^{X_{it}'\beta}\right)}{y_{it}!} \prod_{y_{it}=9} \sum_{k=1}^{9} \frac{\exp\left(k X_{it}' \beta\right) \exp\left(-e^{X_{it}'\beta}\right)}{k!},$$

which gives the log-likelihood function

$$\log \mathcal{L}(\beta) = \sum_{y_{it} \neq 9} \left[y_{it} X_{it}' \beta - \exp\left(X_{it}'\beta\right) - \log\left(y_{it}!\right) \right] + \sum_{y_{it}=9} \log \left(\sum_{k=1}^{9} \frac{\exp\left(k X_{it}' \beta\right) \exp\left(-e^{X_{it}'\beta}\right)}{k!} \right). \tag{1.5}$$

I can then estimate the parameter β using MLE. I prefer this interval-censored Poisson model to the interval-censored tobit model because of the more realistic distribution assumption.

1.5 Estimation Results

1.5.1 ME counties vs. coroner counties

I begin by presenting comparisons between counties with different types of death investigation offices. It is important to stress from the outset that the following results are only descriptive, since my identification strategy does not extend to ME–coroner comparisons. Instead of diff-in-diff, these are simple-difference regressions with the same set of controls as in my main specifications and with the ME indicator interacted with the month indicators. Figure 1.6 displays the expected deaths by month based on a standard tobit (T) regression, an interval-censored tobit (T-IC) regression, and an interval-censored Poisson (P) regression, counterparts of Equations (1.3), (1.4) and (1.5), respectively. Each regression is run twice, with COVID-19 deaths and all-cause deaths as the dependent variable. Standard errors are clustered at the state level. The plots display point estimates and 95% confidence intervals.

I make a few observations from these estimates. Firstly, the shapes of the curves, even after controlling for covariates, closely follow the monthly COVID death tolls in the U.S., which saw its first three peaks of mortality in April 2020, January 2021 and September 2021. Secondly, the three sets of results have clear qualitative similarities, but the standard tobit model produces very large standard errors whereas estimates from the Poisson model are the most precise, with the exception of March and April, 2022, when the U.S. outbreak was concentrated in a small region. Thirdly, although results from the COVID-19 death regressions seem to suggest ME and coroner counties sometimes see statistically significant gaps in either direction in the number of reported COVID deaths in certain months, the same is true of all-cause deaths, with the gap often similar in sign and often at least as large in magnitude, and neither measure shows consistent underreporting by one type of county compared to the other. This indicates that both sets of differences are almost certainly driven by other county characteristics which are not sufficiently accounted for. Therefore, the results underscore the fact that these comparisons are correlative and descriptive and should not be used to derive causal conclusions about the effect of death investigation system types.

1.5.2 Democratic-coroner counties vs. Republican-coroner counties

I now move to the main analysis of the paper and examine the effect on reported deaths of party switches in the 2020 election. In the first four columns of Table 1.4, I report coefficient estimates from counterparts of Equation (1.1), the single-period diff-in-diff specification. I use both the interval-censored tobit and Poisson models, with either COVID or all-cause deaths as the dependent variable. Figures 1.7 and 1.8 illustrate the dynamic effects from counterparts of Equation (1.2), showing point estimates and 95% confidence intervals. December 2020 is the reference month in Figure 1.8. Standard errors are clustered at the state level. In order to more accurately capture any change in death counts immediately before and after the party change in January 2021, and to strip out the volatile initial outbreak in Spring 2020, the models reported in Table 1.4 restrict the sample to the 10-month period between August 2020 and May 2021, leaving me with slightly less than half of my coroner partisanship subsample.

I first examine the results from the simple diff-in-diff models, with one set of $Change_{ki} \times$

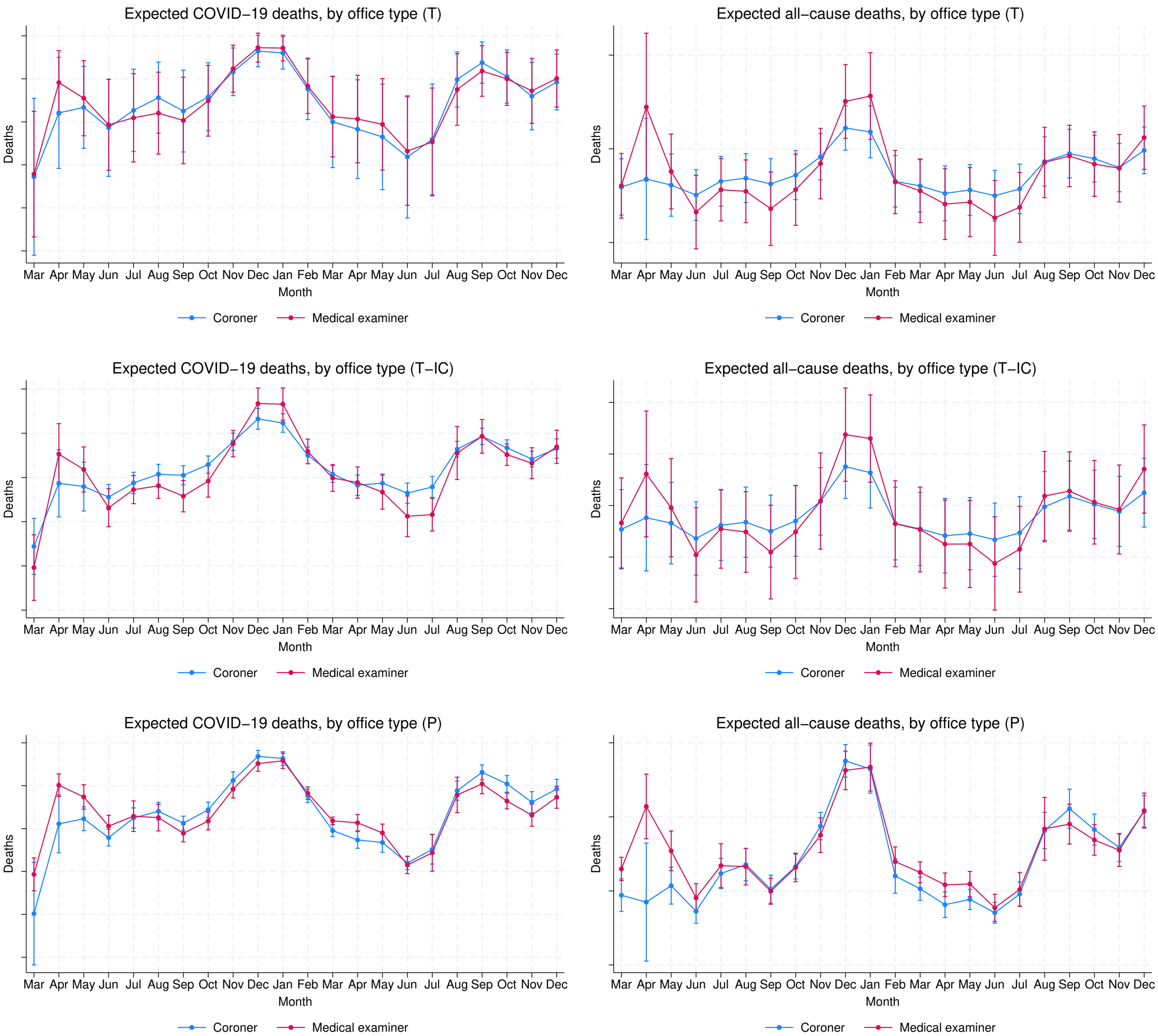


Figure 1.6: Comparison of reported COVID-19 and total death counts between ME and coroner counties

	COVID-19 deaths (T-IC)	COVID-19 deaths (P)	All-cause deaths (T-IC)	All-cause deaths (P)	COVID deaths ≤ 10	All-cause deaths ≤ 10
$Post = 1$	5.212	0.256	−0.115	0.000553	−0.0455	−0.0439**
	(4.676)	(0.588)	(1.624)	(0.0301)	(0.0319)	(0.0165)
Party flip (*Change*):						
Rep hold (ϕ_1)	−1.729*	−0.111***	−5.648**	−0.0785***	−0.0116	0.0114
	(0.889)	(0.0370)	(2.420)	(0.0118)	(0.0180)	(0.0168)
D–R flip (ϕ_2)	−0.0743	0.0141	−1.220	−0.00627	0.000862	−0.00220
	(0.763)	(0.0589)	(1.776)	(0.0188)	(0.0207)	(0.0176)
R–D flip (ϕ_3)	−3.077	−0.0934	−9.149	−0.0608**	0.00950	0.0123
	(2.945)	(0.0741)	(8.280)	(0.0296)	(0.0359)	(0.0386)
Interactions between *Change* & *Post* = 1:						
R hold (δ_1)	0.844	-0.0339	1.124	-0.00885	0.0421*	−0.00469
	(0.643)	(0.0508)	(0.764)	(0.0120)	(0.0207)	(0.00886)
D–R flip (δ_2)	−0.612	−0.163***	0.629	−0.0334***	0.0130	0.0113
	(0.498)	(0.0534)	(0.792)	(0.0117)	(0.0267)	(0.0172)
R–D flip (δ_3)	1.389	0.0920	1.141	0.0342***	−0.0461	−0.0449*
	(2.861)	(0.0657)	(2.785)	(0.0118)	(0.0275)	(0.0219)
$\delta_3 - \delta_1$	0.545	0.126	0.0162	0.0431***	−0.0882**	−0.0402*
	(3.006)	(0.0786)	(2.832)	(0.0116)	(0.0385)	(0.0223)
Population ('00,000)	6.872***	−0.152**	89.03***	−0.0592*	−0.325***	0.648***
	(1.969)	(0.0660)	(7.363)	(0.0331)	(0.0399)	(0.0527)
Population^2	0.402	0.0191*	−1.989	0.00788	0.0489***	−0.100***
	(0.442)	(0.0103)	(2.049)	(0.00502)	(0.00811)	(0.0114)
Population^3	−0.0318	−0.000783	0.00926	−0.000331	−0.00219***	0.00464***
	(0.0227)	(0.000482)	(0.122)	(0.000237)	(0.000470)	(0.000725)
log (population)	5.949***	1.056***	8.568***	1.020***	−0.0937***	−0.406***
	(0.641)	(0.0372)	(1.775)	(0.0222)	(0.0192)	(0.0244)
2020 pres. vote share (%):						
Republican	0.227***	0.0189***	0.598***	0.0106***	−0.00385***	−0.00331**
	(0.0395)	(0.00277)	(0.113)	(0.00124)	(0.000516)	(0.00128)
Shift to GOP	−0.135**	−0.0101*	−0.278	−0.00534	0.00209*	0.00306*
	(0.0682)	(0.00565)	(0.301)	(0.00380)	(0.00116)	(0.00150)
Share of age group (%):						
0 to 19	−1.473**	−0.288***	−6.698***	−0.161***	0.0335***	−0.00291
	(0.607)	(0.0380)	(1.960)	(0.0282)	(0.00678)	(0.0205)
20 to 29	−1.570**	−0.297***	−6.851***	−0.152***	0.0357***	−0.00694
	(0.634)	(0.0339)	(1.996)	(0.0271)	(0.00665)	(0.0214)
30 to 49	−1.697***	−0.301***	−7.372***	−0.155***	0.0340***	−0.00632
	(0.621)	(0.0351)	(1.939)	(0.0278)	(0.00834)	(0.0223)
50 to 64	−1.526**	−0.295***	−6.334***	−0.144***	0.0321***	−0.00550
	(0.678)	(0.0329)	(2.043)	(0.0279)	(0.00523)	(0.0224)
65 to 74	−1.475***	−0.320***	−4.618***	−0.148***	0.0298***	−0.000189
	(0.569)	(0.0369)	(1.660)	(0.0267)	(0.00954)	(0.0212)
75 to 84	−0.998	−0.227***	−7.432***	−0.108**	0.0389***	−0.0337
	(0.749)	(0.0630)	(2.815)	(0.0494)	(0.0108)	(0.0302)
Share of male (%)	−0.0210	−0.00641	0.132	−0.0174**	0.000744	0.0112***
	(0.106)	(0.0120)	(0.366)	(0.00695)	(0.00240)	(0.00319)
Share of racial group (%):						
White	−0.0731	0.00878	−0.266	0.00176	0.00120	0.000598
	(0.0560)	(0.00849)	(0.267)	(0.00436)	(0.000870)	(0.000966)
Black	0.123*	0.0228***	0.220	0.0111***	−0.00151	−0.00264***
	(0.0655)	(0.00828)	(0.308)	(0.00422)	(0.00124)	(0.000611)
Hispanic	0.123**	0.00579	0.164	0.00207	−0.00127	−0.00136
	(0.0612)	(0.00362)	(0.263)	(0.00212)	(0.000781)	(0.00127)
Intercept	161.3***	30.08***	770.9***	19.47***	−2.412***	−0.190
	(61.64)	(3.442)	(176.5)	(2.797)	(0.648)	(2.186)
Fixed effects	State, region, month, region × month, urbanization					
n	12,160					

Levels of significance: *** = .01, ** = .05, * = .10.

Table 1.4: Diff-in-diff regression estimates

$Post_t$ interaction terms to capture the effect of party changes. In Table 1.4, the coefficients of interest are δ_2, which represents the effect of a switch from a Democratic to a Republican coroner, and $\delta_3 - \delta_1$, which measures the effect of the opposite switch. In both the interval-censored tobit model and my preferred Poisson model, the point estimates for these coefficients in the COVID death regression have signs that are consistent with the theory of political influences on Republican coroners ($\delta_2 < 0$ and $\delta_3 - \delta_1 > 0$). The tobit model produces no statistically significant estimate out of all four (two coefficients each for COVID and all-cause deaths), but three from the Poisson model are significant at the 1% level.

The fact that estimates for both δ_2 and $\delta_3 - \delta_1$ are highly significant with all-cause deaths as the dependent variable should raise concerns, as it suggests that a change in coroner partisanship has an effect on the overall number of deaths in a county, which seems unlikely and points to issues with model specification. On the other hand, it should be noted that these point estimates are much larger in magnitude in the COVID death regressions than in those for all-cause deaths. In a Poisson regression, a coefficient on an independent variable is interpreted as the marginal expected effect of the variable on the *log*, not actual, dependent variable; as all-cause deaths are, by definition, larger in value than corresponding COVID deaths, the difference in magnitude is to be expected even if the true marginal effects are in fact the same on the pre-log death count. Nevertheless, it turns out the larger estimates in the COVID death regressions do translate to a larger effect on death counts, which can be seen from the bottom four panels in Figure 1.8.[4] Here, I give one possible interpretation of these results. Although I am unable to perfectly control for correlating factors that affect the difference in all-cause death counts between counties before and after the party change, such differences in COVID deaths move in the same direction but to a much larger extent, implying an increase in COVID deaths as a share of all-cause deaths.[5] Given that the estimated δ_2

[4]Normally, I can use the **margins** command in STATA to explicitly calculate the marginal effect of independent variables on the pre-log dependent variable. But the command fails to work properly in this case, most likely due to empty cells.

[5]Ideally, I would test this using the share of COVID deaths as the dependent variable in regressions, but data censoring makes this impossible.

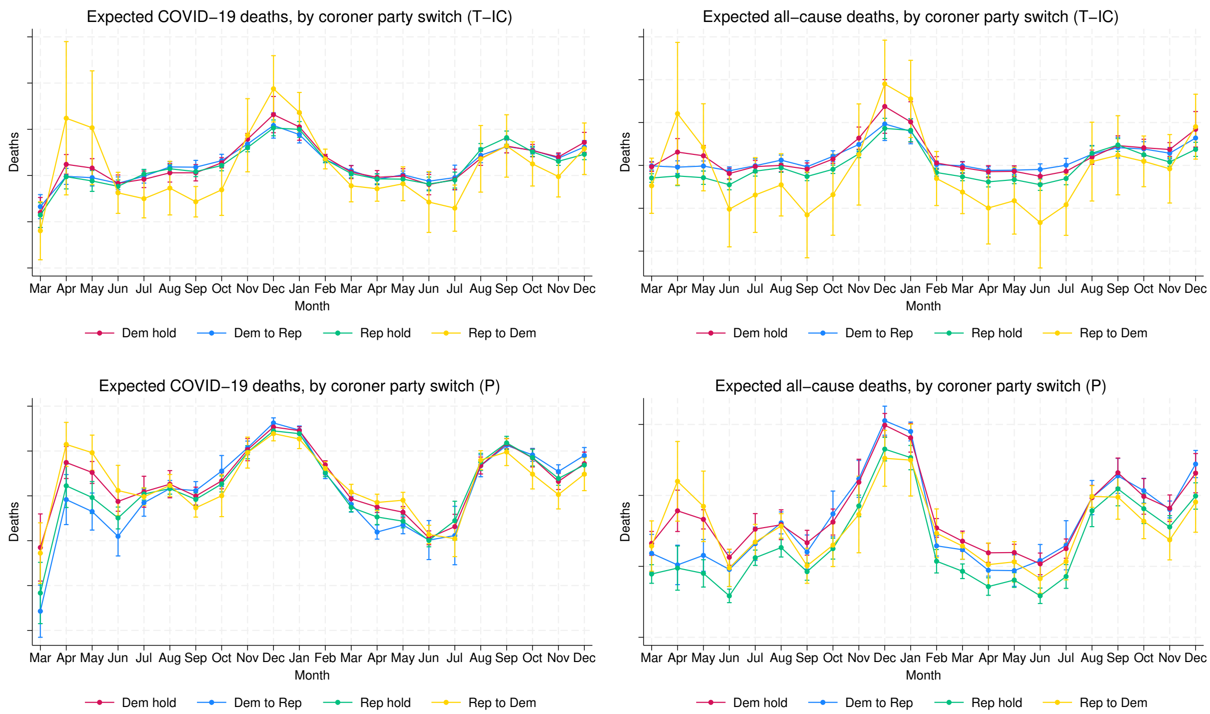


Figure 1.7: Comparison of reported COVID-19 and total death counts, by coroner party switch

in the COVID death regression is significant and $\delta_3 - \delta_1$ is not, this interpretation would mean that a county's switch from a Democratic to a Republican coroner led to a decrease in reported COVID deaths compared to counties held by Democratic coroners, but the reverse party flip had no such effect in the opposite direction.

In the last two columns of Table 1.4, I estimate OLS regressions where the dependent variable is an indicator for death counts smaller than 10, i.e. either counts of zero or suppressed counts. The estimates for $\delta_3 - \delta_1$ are significant and negative in both the COVID and all-cause regressions, and those for δ_2 are not significant in either. The results seem to suggest that a switch from a Republican to a Democratic coroner makes a county less likely to have such small COVID death counts (a larger effect than on all-cause deaths) than it otherwise would have. But because the threshold of 10 is completely arbitrary, I refrain from drawing stronger conclusions based on these results.

I now consider the dynamic effects illustrated in Figures 1.7 and 1.8. Similar to the preceding subsection on ME/C comparisons, estimates from the interval-censored tobit model are much less precise relative to their magnitude, as Figure 1.8 most clearly demonstrates. Both figures also show that estimates involving Republican-to-Democrat switches are less precise in general, due to the small number of such flips in the 2020 election.

The Poisson regression plots in Figure 1.8 (bottom four panels) reveal some interesting patterns. Focusing on the immediate vicinity of the party switch (August 2020–May 2021), estimates from the COVID death regressions indicate statistically significant effects of party changes in both directions. Compared with December 2020, COVID death counts were not statistically different between switch and no-switch counties between August and November, indicating parallel pre-trends. Death counts between the two groups of counties started to significantly diverge in February or March 2021. Democratic-coroner counties that switched Republican began to see a decrease in reported COVID deaths compared to their counterparts that stayed Democratic, and Republican-coroner counties that switched saw higher death counts. The effect lasted for several months before mostly tapering off through the rest of the year. Whereas qualitatively similar trends can be observed in the all-cause death plots, the estimates are less precise and have smaller magnitude just like in the single-period diff-in-diff regression in Table 1.4. This is the strongest evidence yet from this analysis that supports the theory of politically incentivized death reporting.

Finally, I repeat the analysis in this subsection after replacing my death measure with deaths where COVID-19 was the *underlying* cause, instead of those with COVID-19 listed as a cause of death. The results are almost identical, only with slightly larger standard errors of the estimated coefficients of interest. This is to be expected: the two death count measures turn out to be highly correlated, but the underlying death count is smaller than or equal to the more widely defined measure, leading to slightly more suppressed counts. As they do not affect my analysis, I omit those results.

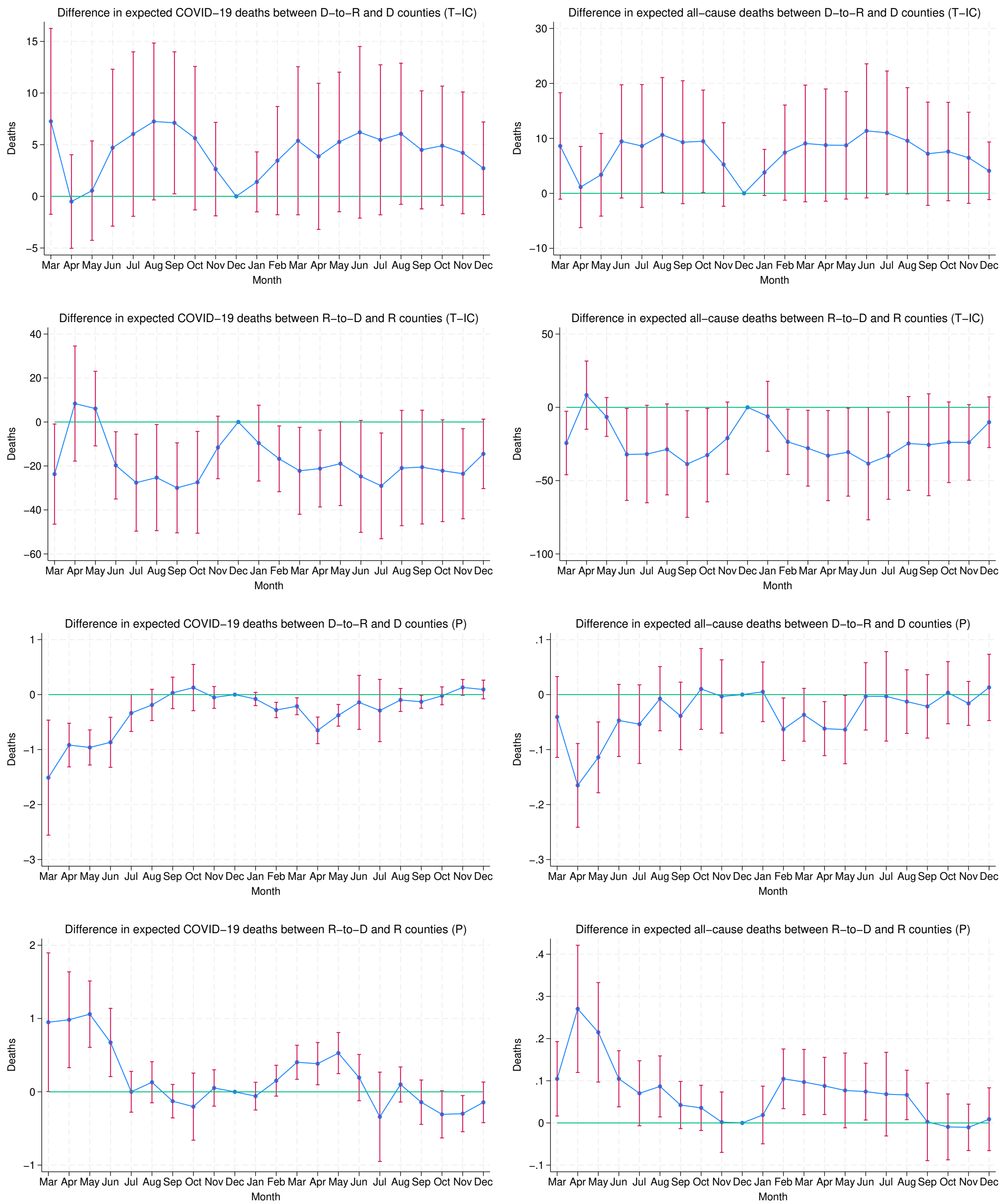


Figure 1.8: Dynamic diff-in-diff effect estimates

1.6 Discussion

Do partisan coroners manipulate reported COVID-19 death counts for political expediency? The analysis of mortality and coroner party affiliation data yields evidence that is decidedly mixed. On one hand, both the single-period and dynamic diff-in-diff regressions produce estimates that seem consistent with the concerns raised about Republican coroners. On the other hand, a few countervailing factors cast doubt on the validity of such theories. The main issue is that I find non-zero effects in my intended placebo tests that use all-cause deaths as the dependent variable, although these are much smaller in magnitude and often less precise. In addition, the dynamic effects I find were relatively short-lived compared to the duration of the COVID pandemic, suggesting any effects from party switches dissipated soon into the new coroner's term. Perhaps more importantly, my most precise estimates (from the Poisson model) are too small to be economically meaningful, and do not point to large-scale, systemic underreporting of COVID-19 deaths by Republican coroners.

It may be helpful to consider the matter of politically-motivated underreporting in the bigger context of COVID death reporting in general. As mentioned in the introduction, there were immense challenges in properly certifying COVID-19 deaths. During the first two years of the pandemic, estimates put excess deaths in the United States at over 1.1 million (Paglino et al., 2023; Rossen et al., 2022), while the officially reported COVID-19 death toll was around 950,000, meaning over 15% of excess deaths were not attributed to the disease. Some of these deaths will have been due to undocumented or unrecognized COVID infections (Woolf et al., 2020). As my Benford's law test in Figure 1.5 shows, there is strong indication of underreporting of COVID deaths across death investigation office types, coroner party affiliations and years. It is conceivable that any politically-motivated manipulation paled beside such inevitable underreporting and was masked by it. Tactics that would be used to manipulate death certificates, such as skipping autopsies or requiring positive COVID-19 laboratory tests (Bergin et al., 2021; "Politics of Death", 2022), may resemble constrained testing and investigation capacity faced by ME/C offices across the country.

The significant yet small results may also reflect the small number of coroners engaged in malpractice, the limited scope coroners have for manipulating death certificates, and/or the relative small share of COVID-19 deaths passing through ME/C offices as opposed to being handled in the healthcare system. While I am not equipped to determine the role of these factors, my findings do suggest that coroners of all affiliations seem to have performed their duties during the pandemic better than many have feared.

1.7 Conclusion

Using public-use mortality data from the CDC and original data on medicolegal death investigation office types and elected coroners' partisanship, I use a number of tools to test the effect of death investigation on COVID-19 reporting. I find suggestive evidence that there is widespread underreporting in COVID-19 deaths, but a cross-sectional comparison between medical examiner counties and coroner counties does not reveal differing extents of such underreporting. Employing econometric specifications in a difference-in-differences framework, I find some evidence that counties reported fewer deaths when they switched from a Democratic coroner to a Republican one after the 2020 election, and vice versa. However, the magnitude of these differences is too small to have had any meaningful impact in relation to the scale of the underreporting problem during the pandemic.

Considerable accommodation has had to be made due to the data censoring adopted by the CDC, which results in more than half of my observations containing a suppressed death count. If manipulation did happen to a marginal degree as my findings seem to indicate, having low-death-count observations in the data would probably help unveil a much clearer picture of such practice. Therefore, it will be worthwhile to carry out follow-up analysis if and when such data can be obtained. It might also be possible to obtain measures of coroner partisanship that can augment party affiliations per se: Farris and Holman (2023), for example, use a survey to gauge county sheriffs' agreement with right-wing extremist ideology and examine its correlation with their strictness in enforcing mask mandates.

Some empirical research has been done to compare the performance of ME and coroner systems. The political aspect of coroner systems, however, remains overlooked and under-researched. As the United States continues to become more politicized and polarized, the subject may merit further research, in areas including infectious disease and beyond, such as crimes, mental health issues and the opioid crisis. Unfortunately for America, the COVID-19 pandemic is unlikely to be the last time that the integrity of the death investigation system is put under the microscope. Future research endeavors should aim to strengthen the system's resilience against potential biases and maintain public trust in the work it produces.

Chapter 2: Lead Pipe Information and Housing Prices: An Analysis in Washington, DC

2.1 Introduction

Ever since Rosen's (1974) seminal work, the hedonic model has been widely used in a number of fields in economics. Applied in the housing literature, this approach considers housing as a differentiated good composed of multiple attributes, and the equilibrium price schedule in the competitive market can be used to derive the marginal value of a given attribute. Numerous studies have been devoted to measure the value of a wide range of attributes including non-market, "environmental amenities", such as school quality (Black, 1999), air quality (Chay & Greenstone, 2005), and local crime risk (Linden & Rockoff, 2008). This paper adopts the framework and focuses on the presence (or lack thereof) of lead hazards in housing, in particular lead service lines (LSLs), more commonly known as lead pipes.

LSLs may be of particular interest to homebuyers because of the demonstrated adverse health effects of exposure to the metal lead, with children being the most susceptible group. Links have been established between children's lead exposure to negative impacts on various health and other outcomes, including cognitive function (Needleman & Gatsonis, 1990), development delays (Selevan et al., 2003), violent and criminal behavior (Reyes, 2007), ADHD (Goodlad, 2013), and academic achievement (Aizer et al., 2018). Importantly, many such effects seem to be long-term and persist well into adulthood (Reyes, 2007; Grönqvist et al., 2020). During the past several decades, as the scientific consensus evolved, the acceptable upper limit (later renamed the "reference level") of blood lead concentration in children has been lowered six times, from 60 micrograms per 100 milliliters of blood (μg/dl) in the early 1960s to 3.5 μg/dl since 2021 (American Academy of Pediatrics Subcommittee on Accidental

Poisoning, 1961; Pueschel et al., 1996; Centers for Disease Control and Prevention, 2021). Furthermore, while the thresholds are those that "should prompt public health actions", it is commonly acknowledged that there is no known safe level for blood lead in children (Binns et al., 2007).

Few attempts have been made to actually estimate parents' willingness to pay for reducing their children's exposure to lead hazard. In one of these, under fairly restrictive assumptions, Agee and Crocker (1996) use data on enrollment in chelation therapy and estimate parents were willing to pay between $39.01 and $364.23 in 2022 dollars for a one-percent reduction in their children's lead burden, which was significantly higher than the U.S. Environmental Protection Agency's estimate of savings from lead reduction efforts. However, given the overwhelming consensus of the health risk posed to children by even low levels of lead exposure, it is conceivable that in the hedonic framework, homebuyers, especially parents, will place a positive valuation on the absence of potential lead hazards in housing, including lead paint and LSLs. In this paper, I use data on housing prices and presence of LSLs in Washington, DC to examine the extent to which lead hazards affect housing prices.

In the early 2000s, the District of Columbia experienced a public health crisis due to lead contamination of drinking water, an unintended consequence of a change in the District's water disinfection procedure and resulting water pipe corrosion. The episode saw a notable rise in blood lead levels of affected children (Edwards et al., 2009; Jane Brown et al., 2010). The crisis attracted nationwide attention, led to a congressional investigation, and highlighted the potential health hazards from the prevalent use of LSLs in the district. LSLs were widely used in the Unites States before they were banned in new constructions after 1986, and may are still in service in older buildings, including in Washington, DC. Since the lead contamination crisis, the District's government has been working on tackling the issue through replacement and awareness campaigns. In June 2016, as part of the effort, the District of Columbia Water and Sewer Authority published a lead pipe map on their website that allows anyone to identify the types of water pipes used in the service lines for any address in the District. With the explicit goals of increasing transparency, informing residents about LSLs and eliciting them to help the water authority identify and solve related problems, this easy-to-use visual tool represented a massive upgrade from the previously undigitized records of pipe material, and made such data readily accessible to the public for the first time. The release of the map also coincided with a significant and sustained rise in the American public's interest in the issue of lead contamination of drinking water following the initial national media coverage in early 2016 of the water crisis in Flint, Michigan, and the availability of the lead map has made it possible for any DC resident to obtain information related to LSLs if they wish, enabling them to better inform their housing decisions.

Treating the release of the lead map as a natural experiment, I try to estimate the resulting information effect on housing prices due to such negative valuation associated with LSLs. I use two different model specifications for identification: a straightforward difference-in-differences model, and a repeat sales model. Based on the recorded service line material, all housing units are divided into three groups: those with LSLs, those without, and those for which the Water and Sewer Authority does not have information on service line material. The grouping is done multiple times, using information on both the public and private sides of the service line. I set the date of the lead map's release, June 6, 2016, as the cutoff for the pre- and post-information-shock periods, and focus on a two-year window centered on that date for the diff-in-diff specification, and a seven-year window for the repeat sales model. Neither method points to a discernible information effect on housing prices.

This paper joins a number of other studies that adopt the hedonic model to estimate the information effect on housing prices related to disclosure of negative environmental amenities. Negative price effects have been found from disclosure of such unpleasantries as airport noise (Pope, 2008), proximity to polluting firms (Mastromonaco, 2015), health hazards from nearby waste sites (Gayer et al., 2002), and flood risk (Votsis & Perrels, 2016). Each of those studies also utilizes new or increased disclosure (usually originating from policy changes) of information that was previously known only to an limited extent, similar to the release of the lead map. In addition, this paper also contributes to the existing literature on the effectiveness of lead reduction policies as well as their effects on the housing market. Gazze (2021) makes use of data from multiple states and finds that lead hazard mitigation mandates lead to a decrease in prices of old houses, where such hazards are more likely to be present. Theising (2019) finds that a mandate to replace private LSLs in Madison, Wisconsin had a large, positive price effect on post-replacement units that exceeds the cost of replacement, implying homebuyers place high value on the absence of LSLs in addition to the explicit cost saving under the mandate. Both papers point out the potentially important information effect of disclosure. Billings & Schnepel (2017) find similar returns to lead paint remediation from a voluntary program in Charlotte, North Carolina. My study is most similar to Theising (2019) in that I use building-level service line data instead of resorting to a proxy (such as year of construction) for likely presence of LSLs, enabling me to exploit variation between otherwise similar properties that would have been buried in less granular data; but these papers differ in an important manner because I am able to examine the information effect in isolation by studying a city without a mitigation mandate. As far as I am aware of, this paper is the first to examine information effects on housing prices in the context of LSLs.

The rest of the paper is organized as follows. Section 2.2 provides a brief introduction of the use of LSLs in the United States and the attendant health hazards, along with the background of DC Water's release of the lead map. Section 2.3 describes the two empirical specifications I use. Section 2.4 summarizes the data. Section 2.5 presents the results, and Section 2.6 discusses possible interpretations. Section 2.7 concludes.

2.2 Background

2.2.1 LSLs and water-borne lead

The metal lead has a long history of being used as a piping material for water supply and distribution, dating back to Roman times (Hodge, 1981). Its wide adoption in United States started in mid to late 1800s and was most notable in large cities, appearing in 85 percent of the biggest American cities by 1897 (Troesken, 2006). Compared to alternative piping materials such as steel, iron and cement, lead is both more malleable and more durable, making it ideal for municipal water systems from an engineering perspective (Clay et al., 2006). Around the turn of the twentieth century, however, there were growing public concerns about the potential danger of lead poisoning posed by LSLs, and many cities started placing restrictions on their use by the 1920s (Rabin, 2008). Nonetheless, the relentless and carefully orchestrated lobbying from the lead industry successfully slowed down the abandonment of LSLs and ensured their common and continual use for supplying water to homes and buildings in the United States (Rabin, 2008). It was not until 1986 that their installation was finally banned by a set of amendments to the Safe Drinking Water Act. The Act now prohibits the use of LSLs in new constructions, but existing LSLs remain in use in many buildings built before 1986 in large cities. A 2016 survey estimated that 15 to 22 million people (out of a total of 297 million) served by community water systems in the United States had a LSL serving their home (Cornwell et al., 2016).

There is usually little or no lead in water from raw sources or water treatment plants; however, lead in service lines can leach into tap water through the corrosive chemical process that occurs between water and the pipe material, and corrosion control is essential in preventing such contamination (Triantafyllidou & Edwards, 2012). In 1991, the United States Environmental Protection Agency issued the Lead and Copper Rule (LCR), a regulation intended to oversee effective corrosion control measures. The LCR sets an upper limit (of 15 parts per billion) for the concentration of lead in tap water; in the event that the limit is breached, it requires public water utilities to take actions to control plumbing corrosion, inform the public and, if necessary, replace the related LSLs. Despite the majority of public water utilities being in compliance with the LCR,[1] drinking water is still an important source of environmental lead exposure (Triantafyllidou & Edwards, 2012; Brown & Margolis, 2012),

[1] It has been pointed out that the current sampling protocol under the LCR can fail to detect lead levels in breach of the upper limit and lead to a false conclusion of compliance (Del Toral et al., 2013).

and LSLs remain the biggest source of lead of drinking water, accounting for 50%–75% of the metal by mass (Sandvig et al., 2008).

The 21st century has seen two major public health crises in the United States related to lead contamination of drinking water, both of which were results of the release of lead in service lines into tap water. The Washington DC water crisis from 2001 to 2004 was triggered by the water authority's switch of the type of disinfectant, from free chlorine to chloramine. While the former had been used by the water industry for more than a century, not until after the crisis did researchers discover that it had the side benefit of reducing lead solubility in water, and the change in disinfectant effectively increased water corrosivity (Edwards et al., 2009). The Flint, MI water crisis from 2014 to 2016 was the joint work of two factors: the switch to a temporary new water source with a different chemical makeup, and the interruption of corrosion control treatment (Roy & Edwards, 2019). While the scale and severity of both incidents were direct results of government misconduct and oversight, they also highlighted the public health risk in the use of LSLs, especially as the science on related safety standards and acceptable practice seems to be continuously evolving.

2.2.2 The DC Water lead map

On June 6, 2016, the District of Columbia Water and Sewer Authority (branded as DC Water) launched an online interactive map tool that allows users to identify the materials of the water service lines for any of the over 120,000 properties in the district and, in particular, find out if those service lines contain lead. The map includes a circular marker for every address, divided in two halves, with colors denoting the materials of the public and private service lines, respectively: green means the line does not contain lead; gray means it does; white means there is no information for the line. Figure 2.1 is a sample screenshot of part of the map. Users can either view the overall map or enter an address to search for a particular property, much like they can when using a smartphone map application.

The release of the map meant that DC residents were able to get information on the

Figure 2.1: A sample screenshot of DC Water's lead map

material of their water service lines for the first time, as the determination of pipe material usually requires excavation and cannot be easily done by individuals. It was also a big leap from the district's previous record-keeping of service line materials, when such information only existed on a variety of physical records in a haphazard fashion. DC Water publicized the map in multiple ways, including via email to users and through social media campaigns (on Facebook and Twitter). The announcement was also covered in a handful of national and local online news outlets including as Vox, DCist (the website of NPR's local station), and Fox 5 DC, as well as popular DC local news blogs including *Petworth News* and the *Georgetown Metropolitan*.

Evidence seems to point to DC residents' rising awareness of lead hazards after the release of the map. Via a FOIA request from DC Water, I obtained a data set containing the daily number of water lead test kits requested from DC Water. Similar to other cities like Chicago and New York City, the DC government sends residents such kits to conduct lead tests on request, free of charge. A kit contains two bottles for collecting water samples, along with instructions and a simple questionnaire. Upon receiving the kit in the mail, a resident can fill the bottles with tap water as instructed, and send them back to DC Water using a prepaid shipping label for the water samples to be tested for lead. Results are then delivered electronically after four to six weeks. The simplicity and low cost of requesting a kit and submitting samples mean that such a test is the one of the easiest and most accessible ways for residents to act on their concerns about lead in drinking water, and the number of tests may respond strongly when there are elevated concerns among the public. Figure 2.2 shows the weekly number of requests from 2015 to 2017 with the twenty-third week of 2016 (the week of the map's release) highlighted; the fitted curve is a visual aid derived from a simple regression-discontinuity estimation, using the map's release as a cutoff. Requests exhibited a distinct pattern. Having remained flat for the entirety of 2015, they started increasing shortly after the Flint crisis gained national attention, reaching a height in late April 2016 about ten times the average count in 2015 before dropping again through May. Following the map's release, the number of requests shot up in the following two weeks, more than doubling the post-Flint peak, before settling back down to a level not as high as in the immediate aftermath of Flint, but higher than in 2015. These numbers demonstrate that lead map did significantly raise DC residents' interest in getting informed about their tap water quality, even more so than a major national news story like Flint; as a matter of fact, it may have precisely been the local nature of the map that prompted more people to get their water tested. I now investigate whether such interest translated into changes in housing market outcomes.

2.3 Empirical Methodology

2.3.1 The hedonic price function

My analysis adopts two parallel empirical strategies, both of which builds upon the hedonic model pioneered by Rosen (1974). The model describes housing as a vector of various utility-

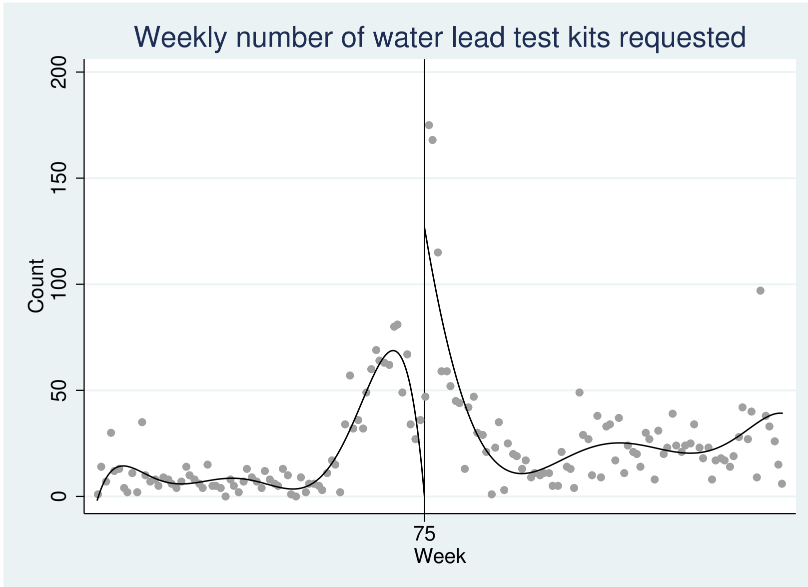


Figure 2.2: Weekly number of water lead test kit requests in DC, 2015–2017

generating attributes, which can be structural, environmental, neighborhood, etc. Market equilibrium prices of housing are determined by the interactions of consumers and producers, and are a function of those attributes. It can be shown that the partial derivative of price with respect to a given attribute reveals the consumer's marginal willingness to pay for the attribute. In the context of this analysis, assuming a log-linear form, the hedonic price function can be written as

$$\ln P_{it} = \alpha_0 + \beta_0 \times lead_i + X'_{it}\theta_0 + \epsilon_{it}, \tag{2.1}$$

where P_{it} is the price of housing unit i traded at time t, $lead_i$ is a dummy variable that equals 1 if unit i has a LSL, and X_{it} is a vector that contains other housing attributes; the coefficient β_0 represents consumers' valuation of the presence of LSLs in housing.

Assuming buyers have perfect information about lead pipes, a direct estimation of the price coefficient of LSLs, β_0, using (2.1) would not be reliable: because it is impossible to include all relevant housing attributes in X_{it} in the empirical analysis, coefficient estimates would be biased due to omitted variables. Hence I explore two alternative identification strategies that address the issue of bias, using the lead map's release as a natural experiment and estimating the interactive price effect of LSLs and the revelation of new information.

2.3.2 Specification #1: difference-in-differences

In the first strategy, I employ a simple difference-in-differences approach to identify the information shock's effect on house prices. I estimate the following regression:

$$\ln P_{it} = \alpha + \beta \times lead_i + \gamma \times map_t + \eta \times lead_i \times map_t + Z'_{it}\theta + \epsilon_{it}. \tag{2.2}$$

Here map_t is a dummy that equals 1 if the time of the transaction, t, was after the release of the lead map, and the control vector Z_{it} includes neighborhood, month of the year and use code fixed effects. The coefficient of interest is η, which captures the change in the effect of presence of LSLs on house prices before and after the map's release, i.e. the information effect. The diff-in-diff design assumes that the trends in sale prices of LSL and non-LSL housing units were parallel before the release of the lead map. Although this might sound restrictive given the high correlation of building age and presence of LSLs, I will demonstrate graphically in Section 2.5 that the parallel-trend assumption indeed seem to be the case.

2.3.3 Specification #2: the repeat sales model

The second empirical strategy uses the repeat sales model first proposed by Palmquist (1982). With this approach, the price of housing unit i at time t is assumed to depend on a time-varying housing price index (B_t) that is true but unknown, the age of the house (A_{it}), a composite measure of time-invariant housing attributes (Z_i) and time-varying environmental variables, in this case the interaction of $lead_i$ and map_t, in the form of

$$\ln P_{it} = \ln B_t + \ln Z_i - \delta A_{it} + \eta \times lead_i \times map_t + \epsilon_{it}, \tag{2.3}$$

where δ is the coefficient of depreciation. If the same housing unit is traded twice at t and t' respectively, taking the difference between equation (2.3) at the both points in time will lead to

$$\ln \frac{P_{it}}{P_{it'}} = (\ln B_t - \delta A_{it}) - (\ln B_{t'} - \delta A_{it'}) + \eta \times lead_i \times (map_t - map_{t'}) + (\epsilon_{it} - \epsilon_{it'}) \,. \quad (2.4)$$

Here, the coefficient of interest, η, can be estimated by regressing the log price ratio on the interaction term and a set of year dummies that equal 1 for year t, -1 for year t' and 0 for other years. This estimation does not yield reliable estimates for the log price indices B_t and $B_{t'}$ because the change in the house's age $(A_{it'} - A_{it})$ is collinear with the two year dummies, but that does not affect the unbiasedness of the estimate for η; adjustments would have been necessary only if we were interested in the indices themselves. The main identifying assumption here is that other housing characteristics remain constant between sales. This can be a strong assumption, and is more likely to be relatively realistic when a shorter time span is studied. Still, I maintain this assumption due to limitations of my data sets, which, among other things, do not include information on renovations.

The repeat sales method focuses on housing units that are traded at least twice. When the number of transactions is n, we are able to derive $(n-1)$ independent equations in the form of (2.4). One complication here is that when $n \geq 3$ for any housing unit, the error covariance matrix is no longer diagonal because of error correlation between sales of the same unit: given three sales at time t, t' and t'', the error terms from the two resulting equations are $\epsilon_{it} - \epsilon_{it'}$ and $\epsilon_{it} - \epsilon_{it''}$, and $\text{cov}\,(\epsilon_{it} - \epsilon_{it'}, \epsilon_{it} - \epsilon_{it''}) = \text{var}\,(\epsilon_{it}) = \sigma^2$. This calls for a generalized least squares estimation, and I will return to the topic when presenting my results.

As the discussion above makes clear, the two specifications used here are similar in spirit but use different identifying assumptions. By presenting results from both models, I hope they will complement each other and help paint a more robust picture about the map's effect on housing prices.

2.4 Data

The data I use comprises public records from three sources. Firstly, records of real property transactions are obtained from a public database from the Real Property Tax Administration of the DC Office of Tax and Revenue, and contain such information as premise address, property use code (which details the purpose of use and belongs to broad categories like residential, commercial and office), sale price, sale date, current-year (i.e. 2022) assessment, etc. Secondly, information about materials of water service lines is obtained by scraping DC Water's lead map website, and describes the types of pipes used for both public and private service lines for each property, along with succinct descriptions on the method of determination. Additionally, I obtain data on housing characteristics from the Office of Tax and Revenue's Computer-Assisted Mass Appraisal system, which includes year of construction along with structural attributes, e.g. numbers of rooms and stories. For reasons unclear to me, the lead map contains a small number of duplicate entries with conflicting service line material records; I dealt with these manually by keeping only the apparently correct, up-to-date information. The transactions data and characteristics data are first merged using the property identifier (SSL); the combined data is then merged with the pipe material data based on the building address, joining transaction records to pipe information. In the transactions data set, "premise address" contains both the building address and a unit number for housing units in multi-unit premises such as apartment buildings; in order to perform the merge by address, the premise address field is parsed to drop the unit number, so that both data sets contain comparable address fields. As a result, transactions of different units in the same building are matched to the same, unique service line record and remain distinct observations in the merged data set.

The sample selection processes of the two empirical strategies are different in some aspects to account for the different priorities and requirements of each approach. What is common between the two samples is that they both includes only residential property transactions[2] that were carried out at with a non-zero price and categorized as a market sale (as opposed to a transfer of ownership for other reasons such as gifting and divorce).

For the diff-in-diff sample, I include all transactions in a two-year window centered on the release date of the lead map, June 6, 2016. The final sample contains 12,719 observations. Column (2) of Table 2.1 presents summary statistics for this sample; those for all housing housing units in the intersection of my three data sets are shown in column (1) for comparison. With the exception of land area, almost all other housing characteristics are largely comparable between the sample and the population; in particular, the shares of housing units with LSLs and without are very similar between them. Columns (4) and (5) contain summary statistics for the subsample where lead is used on either the public or the private side of the service line, and the subsample where it is not used on either side. As expected, properties with LSLs are on average significantly older; they also tend to be bigger, and more likely to be single-unit.[3]

For the repeat-sales sample, I include all properties that were traded more than once in a seven-year window centered on the map's release date. In total, there were 9,713 such transactions of 4,743 properties: 4,521 were traded twice, 217 thrice, and 5 four times. After taking differences across transactions for the same unit, this yields 4,970 equations, which is the number of observations for the regression analysis. As shown in column (3) of Table 2.1, these properties are also mostly similar to the overall housing stock in DC, apart from covering smaller land areas and containing fewer single-unit properties. In addition to these statistics, among the twice-traded units, the median gap between the two sales is 1,197 days

[2] The use codes present in the sample are 001, 011, 012, 013, 016, 017, and 021 through 029. Details about the designation of use code are available at https://otr.cfo.dc.gov/sites/default/files/dc/sites/otr/publication/attachments/Use%20codes.pdf.

[3] Because of the presence of housing units for which DC Water does not have service line material information (and which I code as a separate category), the percentages for LSL and non-LSL units do not add up to 100%, and the numbers of observations in columns (4) and (5) do not add up to the total in column (2).

	All DC units (1)	Diff-in-diff sample (2)	Repeat-sales sample (3)	Within diff-in-diff sample: With LSLs (4)	Within diff-in-diff sample: Without LSLs (5)
Log sale price	13.096	13.272	13.375	13.483	13.241
	(.849)	(.589)	(.567)	(.458)	(.599)
Log 2022 assessment	13.301	13.388	13.412	13.615	13.353
	(.642)	(.580)	(.569)	(.445)	(.589)
Without LSLs (public)	.821	.843	.837	.422	1
	(.383)	(.363)	(.370)	(.494)	(0)
With LSLs (public)	.065	.052	.055	.403	0
	(.246)	(.221)	(.228)	(.491)	(0)
Without LSLs (private)	.739	.757	.744	.005	1
	(.439)	(.429)	(.437)	(.070)	(0)
With LSLs (private)	.143	.127	.134	.991	0
	(.350)	(.333)	(.341)	(.095)	(0)
Land area (sq. ft.)	2233.467	1910.209	1556.981	2142.161	1876.043
	(2792.916)	(2470.581)	(1909.968)	(2281.402)	(2389.915)
Number of stories	2.090	2.107	2.116	2.106	2.099
	(1.103)	(.436)	(.426)	(.394)	(.447)
Number of rooms	6.083	5.838	5.584	6.779	5.695
	(2.680)	(2.660)	(2.576)	(2.195)	(2.663)
Number of bedrooms	2.727	2.607	2.474	3.268	2.519
	(1.400)	(1.429)	(1.397)	(1.166)	(1.442)
Number of bathrooms	1.899	1.998	1.990	2.309	1.958
	(1.001)	(1.013)	(1.008)	(.961)	(1.009)
Year built (actual)	1943.984	1946.924	1943.091	1919.268	1952.440
	(36.272)	(39.803)	(39.470)	(22.152)	(40.197)
Year built (effective)	1967.562	1971.182	1967.997	1966.173	1973.354
	(26.516)	(29.859)	(31.369)	(22.320)	(29.959)
Is single-unit	.645	.576	.513	.827	.545
	(.479)	(.494)	(.500)	(.378)	(.498)
n	119,606	12,719	4,743	1,632	9,607

Table 2.1: Summary statistics of the diff-in-diff and repeat-sales samples

(3.28 years), and the mean gap is 1,205 days.

It should be pointed out that one issue with the lead map is that the data at any point only contains the most up-to-date records of the service lines, and DC Water does not release information on changes of service line material. As a result, the scraped data I use, obtained in March 2021, does not reflect the most accurate picture during the relevant period when housing transactions were made. I will present possible remedies for the issue in my later analysis.

2.5 Results and Discussion

2.5.1 Difference-in-differences

My analysis does not seem to suggest there exists an information effect of the lead map's release on the real property market in Washington DC. I first present some visualized results in the diff-in-diff spirit. Figure 2.3 shows the weekly average log transaction price for each week starting from 10 weeks before the release till 10 weeks after for different groups of housing units, controlling for neighborhood, property use code and housing attributes including land area, (effective) year built and number of rooms. Although the figure focuses on the 21-week window, the estimates and confidence intervals are derived from the full sample for statistical power. Panel (a) groups observations based on the pipe material of the public service line, creating three groups of housing units: those with LSLs, those without, and those for which information is missing. Panel (b) groups the observations based on the private service line material instead. Panels (c) through (f) are created by replicating the first two panels but bunching the missing-information group with either the LSL or the non-LSL group, creating two pairs of comparison: the former way of grouping compares units that *may* have LSLs (red) with those that definitely don't (blue); the latter compares those that definitely have LSLs (red) with those that *may not* (blue). Such regrouping is done in an attempt to narrow the confidence intervals. Finally, panels (g) and (h) group observations taking into account both the public and private sides: (g) is the counterpart of (c) and (d), where units in one group *may* have LSLs on at least one side, and those in the other group do not have LSLs on either; (h) is the counterpart of (e) and (f), where units in one group definitely have LSLs on at least one side, and those in the other are not known to have LSLs for sure on either side.

First of all, the figures show that the parallel trend assumption needed for a diff-in-diff analysis is satisfied, as there was no significant difference between the price trends of different groups before the release of the lead map. But they also suggest that there was no divergence in the trends after the release. These figures, while controlling for neighborhood and use code, do not look qualitatively different from those plotted with simple averages by group. One takeaway from the figures is that the relatively small number of transactions conducted each week (about 100 on average, of which about 10 or fewer were for LSL units depending on the definition) probably resulted in the high fluctuations, lack of clear trends as well as wide confidence intervals for the average sale price estimates. Using cruder time measures such as fortnights, months and quarters does not appear to alleviate such problems to an extent enough for different conclusions to be drawn.

Now I present my estimates for η, the coefficient of interest in the diff-in-diff regression (2.2), in columns (a) through (d) of Table 2.2, along with robust standard errors. I similarly estimate the regression multiple times, changing between the public and private lines and different grouping methods. Columns (a) and (b) correspond to panels (a) and (b) of Figure 2.3, respectively, grouping housing units based on either public or private service line material and including an extra dummy variable to indicate units with missing information; column (c) correspond to panel (h), combining information on the public and private sides and grouping missing-information units with non-LSL units. Columns (a') through (c') repeats the estimation in (a) through (c), but includes housing attributes as controls (the sample size is slightly smaller because attributes are not available for some units). While all the estimates of γ all have the expected negative sign, none of the specifications produce a

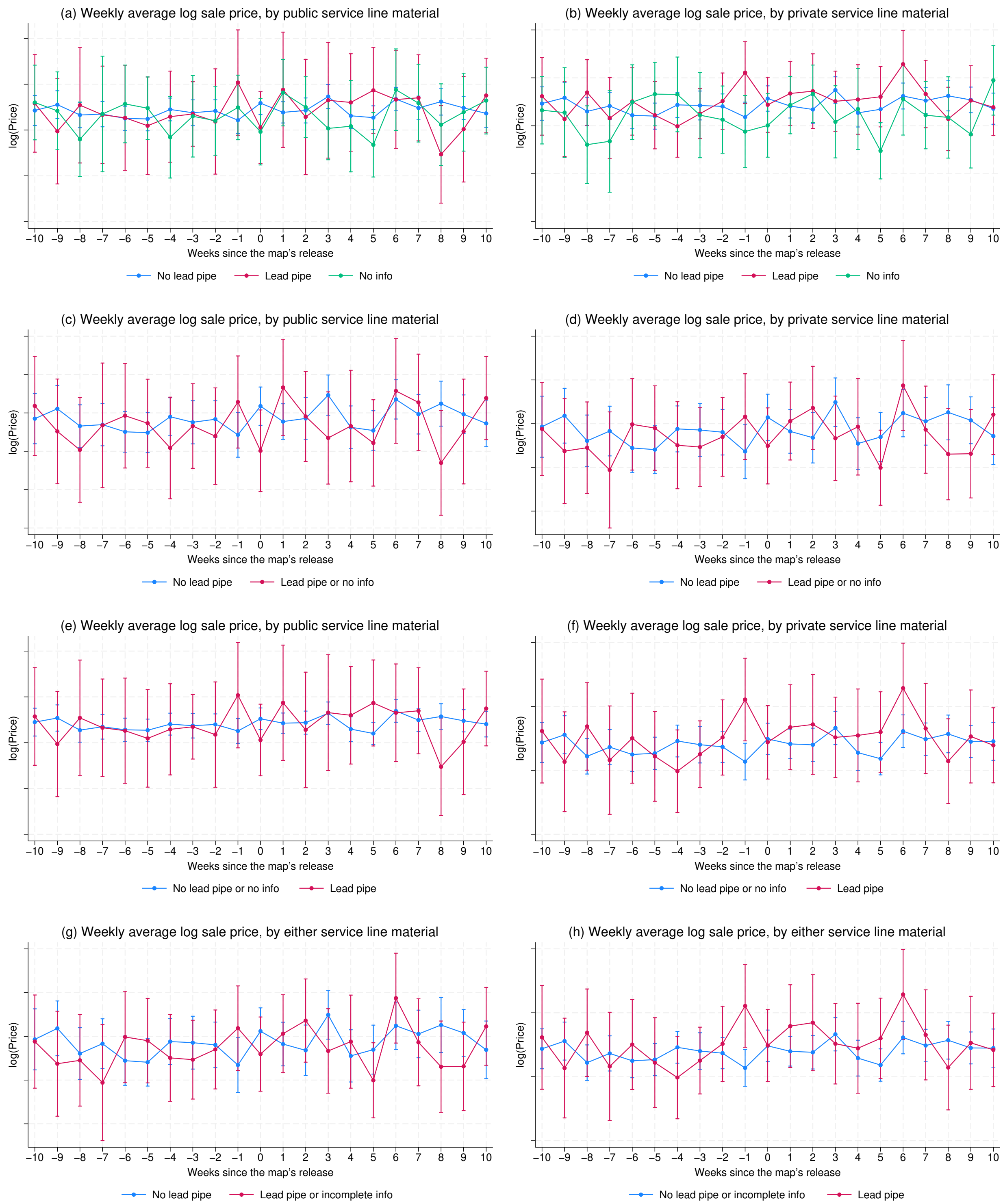


Figure 2.3: Weekly average log sale price by service line material

Variable	Specification					
	(a)	(a')	(b)	(b')	(c)	(c')
map_t	.069***	.068***	.072***	.071***	.072***	.063***
	(.007)	(.006)	(.008)	(.006)	(.007)	(.006)
$lead_i$	.020	.026*	.006	.031*	.013	.032***
	(.018)	(.015)	(.013)	(.010)	(.013)	(.010)
$lead_i \times map_t$	−.026	−.012	−.020	−.014	−.020	−.006
	(.025)	(.020)	(.017)	(.014)	(.017)	(.014)
Year built	—	.0036***	—	.0036***	—	.0036***
		(.0001)		(.0001)		(.0001)
Number of rooms	—	.102***	—	.102***	—	.102***
		(.002)		(.002)		(.002)
Land area ('000 sq. ft.)	—	.0386***	—	.0387***	—	.0385***
		(.0040)		(.0039)		(.0040)
Service line included	Public		Private		Public & private	
Missing-information units	Separate group		Separate group		Together with non-LSL units	
n	12,719	12,601	12,719	12,601	12,719	12,601

Levels of significance: *** = .01, ** = .05, * = .10.

Table 2.2: Estimation results for the main diff-in-diff specification

statistically significant estimate, which remains true with or without the housing attribute controls.

In order to check if the insignificant estimate conceals any heterogeneity across different types of homebuyers and housing units, I perform three additional exercises. First, I consider a triple-difference model, and expand the sample to include commercial- and office-use property. I estimate

$$\ln P_{it} = \alpha + \beta \times lead_i + \gamma \times map_t + \tau \times res_i + \eta \times lead_i \times map_t + \xi \times lead_i \times map_t \times res_i + Z_{it}'\theta + \epsilon_{it},$$

where the new indicator res_i is equal to one if unit i is residential, and zero otherwise (the resulting new two-way interaction terms are included in the regression but omitted from the equation above). The coefficient of interest is ξ, and the aim is to see if there was any divergence in the effect captured by η in the previous model between the two types of units, which could be the case if LSLs started to become more of a concern for homeowners than

Variable	Specification					
	(a)	(b)	(c)	(d)	(e)	(f)
map_t	.286**	.236**	.072***	.057***	.020	.022
	(.103)	(.107)	(.011)	(.012)	(.021)	(.022)
$lead_i$	−.042	−.376*	−.000	−.002	.100	.005
	(.248)	(.200)	(.036)	(.032)	(.073)	(.045)
$lead_i \times map_t$	.031	.166	−.077	−.021	−.034	−.026
	(.305)	(.264)	(.047)	(.043)	(.096)	(.061)
$lead_i \times map_t \times res_i$	−.057	−.185	—	—	—	—
	(.305)	(.265)				
$lead_i \times map_t \times avglevel_n$	—	—	.046	−.047	—	—
			(.369)	(.174)		
$lead_i \times map_t \times pctkids_a$	—	—	—	—	−.010	−.003
					(1.138)	(.735)
Service line included	Public	Private	Public	Private	Public	Private
n	12,913	12,913	12,719	12,719	12,719	12,719

Levels of significance: *** = .01, ** = .05, * = .10.

Table 2.3: Heterogeneity checks

commercial or office users following the release of the lead map. The results are reported in columns (a) and (b) of Table 2.3 (they correspond to columns (a) and (b) Table 2.2 in terms of grouping no-information units). Similarly, estimates of ξ have the expected sign but are not statistically significant enough to suggest the existence of diverging effects. It is worth noting that the majority of real property transactions that took place during the time period were for residential properties, and the expanded sample did not increase in size by a lot.

In the remaining two exercises, I revert to my original sample. The first check is for treatment effect heterogeneity across neighborhoods with different percentages of units with LSLs. The effects may differ because, say, higher availability of non-LSL units may make switching away from LSL units easier, and hence lead to a bigger price effect; the map in Figure 2.4 illustrates the considerable geographic variation in the prevalence of LSL units. The specification for this heterogeneity check includes an additional continuous variable, $avglevel_n$, which measures the prevalence of LSLs in neighborhood n, and interacts it with the original interaction term:

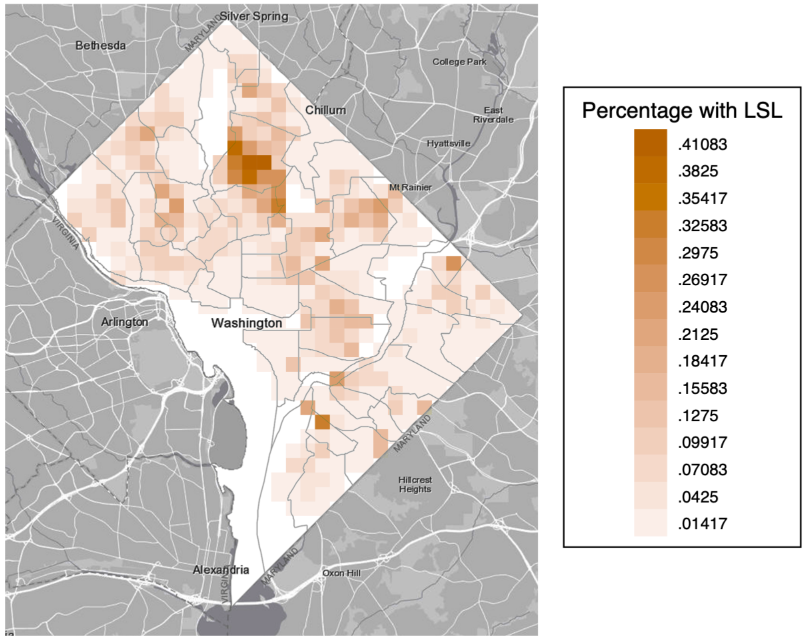


Figure 2.4: Heat map of housing units with LSLs

$$\ln P_{int} = \alpha + \beta \times lead_i + \gamma \times map_t + \nu \times avglevel_n + \eta \times lead_i \times map_t + \rho \times lead_i \times map_t \times avglevel_n + Z_{it}'\theta + \epsilon_{it}.$$

The coefficient of interest is ρ. The estimation results are reported in columns (c) and (d) of Table 2.3; again, no evidence of diverging trends is found.

The third exercise checks for heterogeneity across areas with different proportions of households with young kids. The perceptibility of children to damage from lead exposure may lead to more pronounced price effects in neighborhoods with more families with children. I use the American Community Survey Public Use Microdata Sample (ACS PUMS) to calculate the share of such households in each of the five Public Use Microdata Areas (PUMAs) in DC. The PUMA is the finest geographical level that the ACS data allows me to perform the analysis on, and the share of households with one or more related young children (defined as those under 5 years of age) range from 3.3% to 10.1% between these PUMAs. The new specification includes this share, $pctkids_a$, and interacts it with the original interaction term:

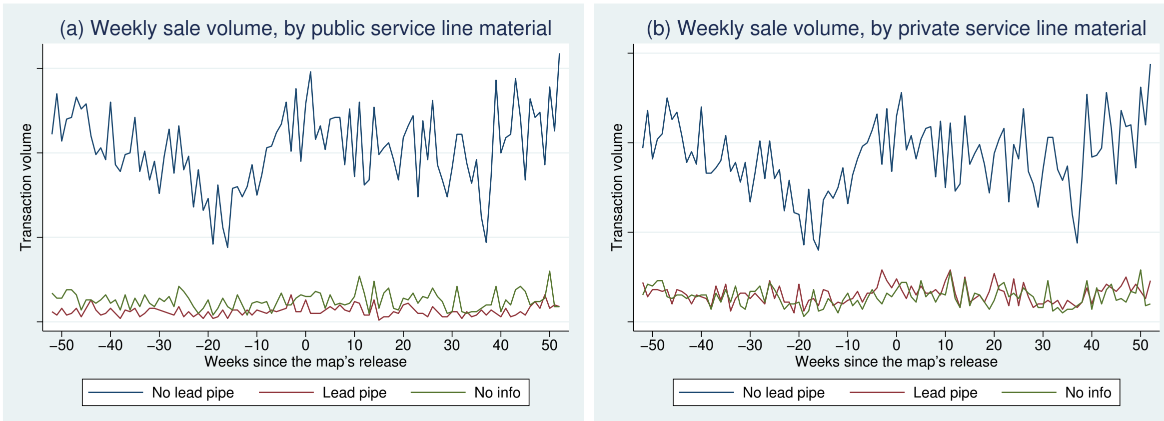


Figure 2.5: Weekly transaction volumes by service line material

$$\ln P_{iat} = \alpha + \beta \times lead_i + \gamma \times map_t + \psi \times pctkids_a + \eta \times lead_i \times map_t + \omega \times lead_i \times map_t \times pctkids_a + Z'_{it}\theta + \epsilon_{it}.$$

The coefficient of interest is ω, and estimates are reported in columns (e) and (f) of Table 2.3, which once again do not point to heterogeneous effects. Changing the definition of kids and including all minors (under 18 years of age) yields the same conclusion.

In addition to heterogeneity, I also look at transaction volumes, instead of prices, in the year before and the year after the map's release. Because of a general upward time trend in the total volume of transactions, the second one-year period saw about 20% more transactions than the first. On the other hand, the shares of units with and without LSLs remained consistent between both periods, suggesting the absence of any selection response from property buyers after the map's release. The two panels of Figure 2.5 show the number of transactions per week in each group of housing units, based on the public and private service lines, respectively. No noticeable divergence of trends seem to exist in either panel. The same conclusion can be reached from a diff-in-diff analysis using specifications similar to the one described earlier in the section, which I do not report here.

At the end of Section 2.4, I pointed out that the data from the lead map may suffer from accuracy issues. In particular, data used in my analysis reflects the up-to-date records as of

March 2021. Because of constant updating of the records as well as replacement of LSLs, the service line materials displayed on the map for any address can be different from those displayed when the map was released or when a potential homebuyer would check the map to look up a property of interest. With historical records unavailable, I have to resort to other less-than-perfect remedies. One feature of the lead map is that it specifies the method DC Water used to determine the service line material for each property. Determination was based on a variety of records including permits, meter records, maintenance records, replacement records, etc. When the service line material is determined because of replacement work, the map lists the date when the replacement was carried out. Therefore, I am able to omit from my main sample observations where replacement work was done on either the public or the private service line after the date of transaction (so that what I observe may not be what the buyer observed at the time), thus eliminating the main source of inaccuracy due to record updates.[4] After removing 476 such observations (with 12,236 remaining in my main sample), I repeat all the preceding exercises in this section, and all of them give virtually the same results as before. Therefore, it seems unlikely that data issues related to record updating are the main reason for my findings.

2.5.2 Repeat sales

Next, I present estimation results from the repeat sales model. OLS point estimates and robust standard errors from the main specification are reported in columns (a) and (b) of Table 2.4 (coefficients on the year dummies are not reported because they don't have the simple interpretation as housing price indices). Neither estimate of η is significant, implying the map's release did not have differential effects on LSL and non-LSL units. As a partial remedy to the aforementioned pipe replacement problem, I repeat the estimation after removing from the sample observations where replacement work was done after the

[4] All replacement work recorded in the data is relatively recent, so the replaced pipes are always non-lead. I am removing these observations instead of assuming the properties had LSLs pre-replacement because a significant number of non-LSLs are also replaced during water main work and emergency repairs, according to DC Water. Such an assumption would introduce a new source of inaccuracy in the data.

Variable	Specification							
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
$lead_i \times map_t$	.001	$-.003$	.003	$-.000$	.001	$-.003$	.016	.029
	(.004)	(.004)	(.005)	(.004)	(.005)	(.005)	(.026)	(.018)
Service line included	Public	Private	Public	Private	Public	Private	Public	Private
Missing-information units			Separate group				Separate group	
n	4,970	4,970	4,657	4,657	4,970	4,970	9,713	9,713

Levels of significance: *** = .01, ** = .05, * = .10.

Table 2.4: Estimation results for the repeated-sales specification

earlier transaction, so that it is less likely that the LSL status changed either between transactions or between the latest transaction and map data collection. This procedure removes 313 observations from the sample (with 4,657 remaining), and estimation results are reported in columns (c) and (d). Furthermore, in columns (g) and (h), I use all 9,713 repeat sales transactions to run a basic fixed-effects regression with the purpose of increasing statistical power. Sale date and property fixed effects are included alongside the *lead* and *map* indicators and their interaction. No clear evidence emerges that indicates the map had a price effect.

In addition, as mentioned in section 2.3, the existence of units that were traded more than twice causes the error covariance matrix to be non-diagonal and calls for a GLS estimation approach. I performed this using the main sample, and report the new standard errors in columns (e) and (f). Unsurprisingly, these results do not change the conclusion derived from the OLS estimates.

2.6 Interpretation of the Results

The lack of a discernible information effect seems to contradict findings by such studies as Theising (2019) and Billings & Schnepel (2017), where homebuyers are found to place a positive value on the absence of lead exposure hazards. But my results are in line with Bae's (2016) findings on the effect of disclosure alone. The fact that the release of the lead map did not appear to have any price effects in the real property market points to several possibilities, which I will not be able to fully investigate. For example, it could be the case that residents responded to the newly available information only in ways not reflected in the housing market, such as increased water testing, remediation efforts and blood testing for children. Alternatively, it could be that many homebuyers were simply unaware of the availability of the map at the time of their transaction, due to limited media coverage and government publicity given to it.

In order to further examine the price response to concerns about lead hazards in the housing market, I perform the following two exercises. Firstly, I divide housing units into those built before 1986 ("old" units), when the federal lead-pipe ban came into effect, and those built after 1986 ("new" units), and examine their price trends before and after the start of 2016, when the Flint water crisis came to national attention. At the time, because the lead map was not yet available, there was no easy way for residents to obtain information about service line materials, and the best proxy of the likelihood of LSL presence in a building is its year of construction. If health concerns led people to shun units with LSLs, there would be a divergence in the average price trends of old and new units. Figure 2.6 presents these trends in 2015 and 2016 in a similar way to Figure 2.3; neighborhood and use code are controlled for. Prices for both categories moved broadly in tandem with each other, with new units selling for higher prices than old ones, as expected. While both the monthly and weekly plots seem to show some movements in the price trend around the beginning of 2016, those are in line with seasonal movements observed in other years, and unlikely to be related to the Flint crisis. Thus it seems even renewed concerns about lead exposure in response to a major incident like Flint still had little effect on the housing market. In addition, this exercise also partially address the issue about updated service line material records raised in Section 2.4 and further discussed in Section 2.5, since the results only rely on construction year as a proxy for service line material, instead of accurate records.

A second, related exercise similarly chooses the Flint crisis as a cutoff point, but divide housing units according to their service line material just like before. The rationale is to

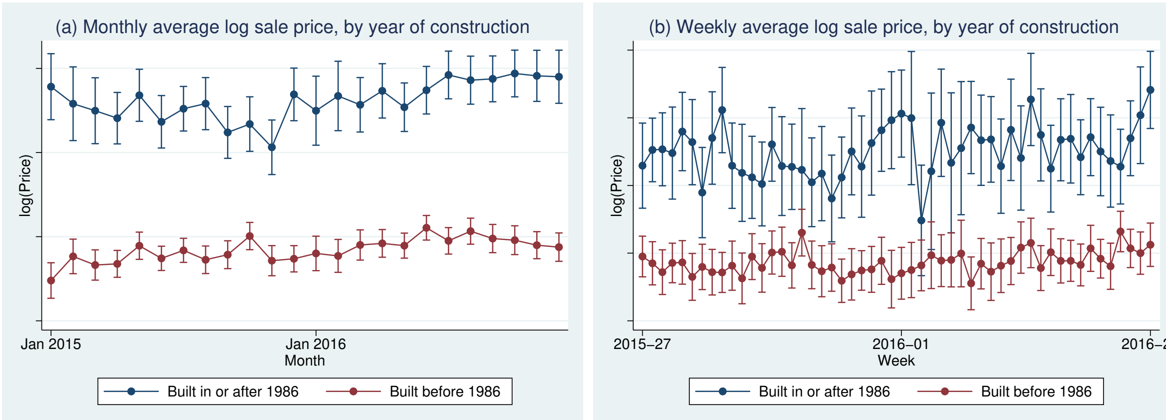


Figure 2.6: Monthly and weekly average log sale price by year of construction

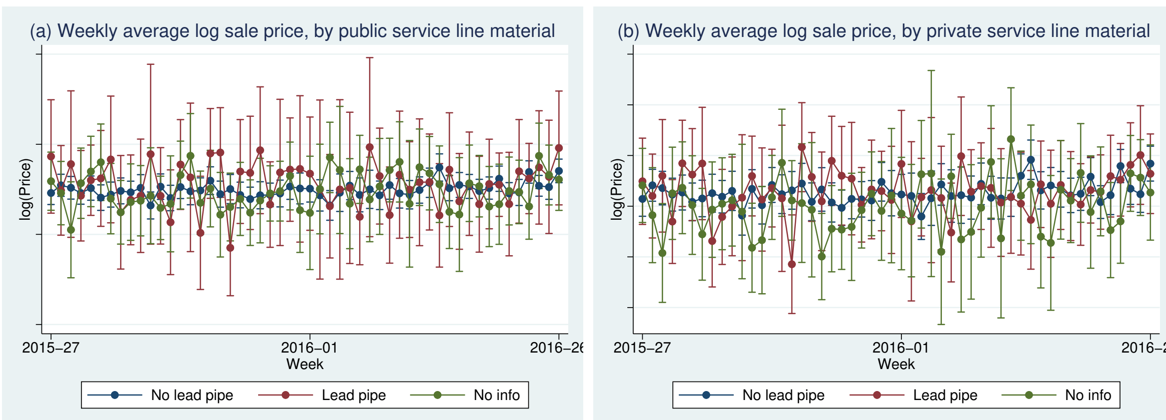


Figure 2.7: Weekly average log sale price by service line material

further check the effect of lead risk salience on the housing market. This exercise would reveal no new information compared to the previous one if people truly had no way of learning about their service lines before the map was released; however, if price trends did diverge post-Flint, it would most probably imply people had alternative ways of finding out about LSLs before the map (despite possibly high barriers), and new concerns led to behavioral responses in the market. However, as shown in Figure 2.7, this does not appear to be the case.

To gain more insights into the possible reasons for the lack of significant positive effects, I present findings from a few other data sources. First, I examine Google search data for terms related to both DC's lead map and lead in drinking water in general. Figure 2.8 displays the

Google Trends for the term "DC lead pipe map".[5] The term started to appear in searches the week following the announcement, although the number of queries has been so small for this search term that the plot exhibits a pattern of occasional spikes along a horizontal line at zero, instead of a smooth trend line. On the other hand, the next time the search term showed up in the data was in March and April 2017 as another two separate spikes. It was not until June 2019 that there started to be more consistent interest in the term, up to today; the reasons for such newfound interest are unclear to me. It is worth noting that Google search volume probably does not directly reflect actual usage of the service. The second group of search terms I examine come from an analysis for Pew Research Center by Matsa et al. (2017), which identified dozens of terms people used when looking up information about the Flint water crisis. Interest in such terms reflects people's concerns about lead exposure from water, and can be used to gauge the effect of the map's release on the salience of the topic. Looking at the search frequency of the leading terms identified (including "lead pipes", "lead water", "lead in water", "lead poisoning", etc.) in DC on Google Trends, I find their popularity among DC Google users to broadly mirror the nationwide trends, with elevated interest starting in January 2016, when the Flint crisis gained national media coverage. However, just like the national trends, there did not seem to be such a spike following the map's release. In addition, the search trends for these terms do not seem to mirror that for "DC lead pipe map" discussed earlier, which experienced a new and sustained rise starting in mid-2019, further suggesting that interest in the map does not fully arise from or lead to concerns about lead in drinking water.

A second set of data I analyze contains the number of blood lead tests carried out in DC, obtained from the DC Department of Health through a Freedom of Information Act (FOIA) request. It is required by law that every child in DC get tested twice for blood lead by the age of two, but parents also can take initiative and ask their doctor for tests whenever they suspect their child is at risk of lead exposure. Therefore, higher levels of such concern

[5] Other similar terms directly related to the map, such as "DC lead pipes", "DC lead map" and "DC Water lead map", did not produce search data above the threshold for Google Trends to display them.

Figure 2.8: Google Trends results for “DC lead pipe map”, 2016–2021

among parents would plausibly lead to more tests conducted. I obtained daily test count data during the three-year period from 2015 to 2017. Figure 2.9 shows the weekly number of tests, along with the four-week moving average, with the first and twenty-third weeks of 2016 highlighted: these were the weeks of the first national media coverage of the Flint crisis and the release of the lead map, respectively. The figure clearly shows seasonal patterns, but there does not seem to be increases in the number of tests following either event; the hike during late April and early May of 2016 was due to an isolated incident, where hundreds of students from three elementary schools were tested because of findings of elevated lead levels in their drinking water. The fact that there was no noticeable response to the Flint crisis suggests that parents may not react to general concerns about lead in drinking water by having their children tested more, possibly because of the time and efforts required. This stands in contrast with the pattern of water lead test kit requests presented at the end of Section 2.2, where a more notable response can be observed when the cost of responding is much lower.

To summarize, based on existing evidence, it appears that the reason for the absence of an information effect is multi-faceted. On one hand, usage of the map service seems low, as suggested by the Google Trends data. On the other hand, it also seems that residents only exhibited a rather limited range of behavioral responses to concerns about lead exposure in general, and not through the housing market, which implies the housing price effect of the map's release would still be limited even if take-up had been higher.

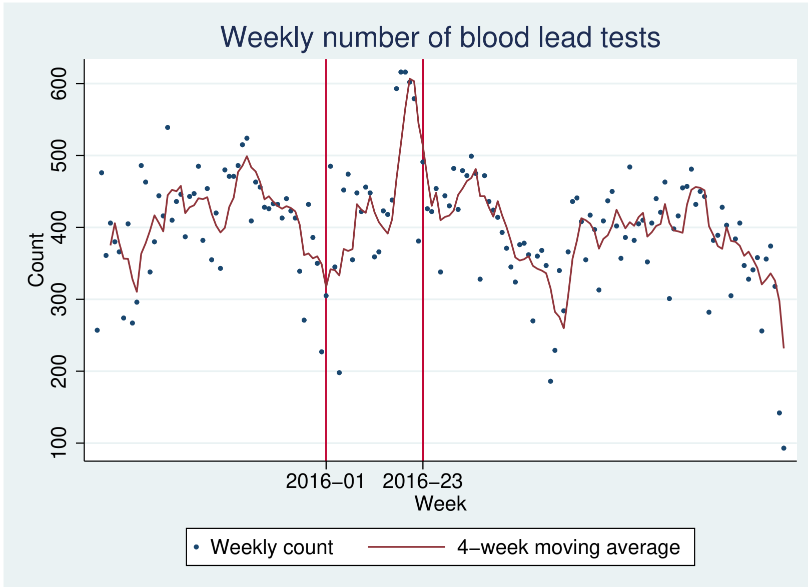


Figure 2.9: Weekly blood lead test count in DC, 2015–2017

2.7 Conclusion

Using real property transaction data, I examine the prices of LSL and non-LSL housing units around the time of the lead map's release, and find no discernible price effect of the newly available information regarding lead service lines. This remains true after I take into account the possibility of heterogeneous price effects between residential and commercial-use property or between neighborhoods with different base-level lead service line prevalence; nor did any differential effect seem to have taken place through transaction volumes.

It is difficult to determine why the map's release seems to have had no price effect. Google search data appear to suggest that few people sought out the map, and hence property buyers and sellers probably did not use it to inform their decisions. It is possible that responses to the new information mostly manifested themselves in subtler ways that were not reflected in the housing market, such as replacement of lead pipes and at-home lead testing of tap water. That points to other potential avenues to further investigate the effect of the new information on the housing market. For example, if records of lead pipe replacements become available, it would be possible and interesting to look at how the number of replacements differed before and after the release, as well as any price effects associated with replacements. Such a study would also give a clearer picture on the extent to which the map has achieved its intended goal as a public health policy. Examining price data on rental properties, which have higher turnovers and whose prices may respond to market forces more quickly, could also be shed new light on the topic.

Chapter 3: Marriage Tax Subsidy and the Decision to Marry: Same-Sex Couples After *United States v. Windsor*

3.1 Introduction

It has long been recognized that the income tax system of the United States creates a wide array of incentives and disincentives for marriage by treating taxpayers differently based on their marital status. Becker's (1973, 1974) seminal work models the marriage decision as an economic one, which depends on whether entering a marriage will generate positive gains for the individual. The tax implications of marriage have been studied as one of the economic factors that inform such decision making, and tax reforms and other legislative changes sometimes provide ideal settings for empirical work that looks to examine the importance of tax incentives in individuals' marriage decision.

In this paper, I focus on one particular type of tax incentive: the marriage "penalty" or "subsidy" that arises as a result of joint income tax filing in the United States tax system. When married couples file their taxes jointly, they effectively split their income and one or both spouses are often pushed into a different marginal tax bracket compared to if they were single, thereby altering their joint tax liability. To quantify the effect of this marriage non-neutrality, I exploit the 2013 United States Supreme Court ruling on *United States v. Windsor*. The ruling struck down a federal law that denied recognition of same-sex marriages, obliging the federal government to recognize such marriages sanctioned at the state level, including for federal income tax purposes. The resulting shift from individually filing to joint filing for same-sex married couples creates a natural experiment for studying how a change in tax incentives can influence decisions to enter a marriage.

Using the Public Use Microdata Sample (PUMS) of the American Community Survey

(ACS) from 2012 to 2017 and employing a difference-in-differences empirical strategy, I am able to examine whether same-sex couples with different levels of marriage subsidy or penalty responded differently to the Supreme Court decision. The results do not point to any statistically significant or economically meaningful response from couples: it does not appear that those who would enjoy a larger marriage subsidy became more likely to get married after the ruling. The absence of an effect remains the case across couples of both sexes. The findings add to the decidedly mixed evidence from the existing literature on the subject. For example, Alm and Whittington (1999) find that marriage penalties have a modest but significant impact on marriage decisions of women in opposite-sex couples, and both Baker (2004) and Fisher (2013) find larger effects using Canadian and U.S. data, respectively. Meanwhile, works like Sjoquist & Walker (1995) and Dickert-Conlin & Houser (2002) fail to find any such effect.

Early studies attempting to use microdata to quantify the effect of the transfer system, including the marriage subsidy, typically adopt unsophisticated empirical strategies on cross-sectional data; Alm et al. (1999) and Moffet (1992) provide extensive reviews. More recent papers (Rosenbaum, 2000; Alm and Whittington, 1999) utilize panel data but have to predict potential spouse earnings for single individuals. This paper contributes to the literature on the marriage non-neutrality of the U.S. tax system by exploiting a new natural experiment setting as well as focusing on the population of same-sex couples. It also adds to the fast-growing body of literature that examine the economic implications of same-sex marriage legalization, shedding new light on marriage decisions of same-sex couples, which did not become a feasible topic for empirical research because of data availability issues until fairly recently, with the ACS being one of the rare microdata sets that identify same-sex couples and their marital status.

The remainder of the paper is structured as follows. Section 3.2 provides background on the marriage non-neutrality of the tax system and the recent history of same-sex marriage legalization in the United States. Section 3.3 introduces the empirical strategy and Section

3.4 describes the data set and the sample construction procedure. Section 3.5 presents estimation results and discusses, giving special care to the comparison with a forthcoming study similar to this in focus and scope. Section 3.6 concludes.

3.2 Background

3.2.1 Same-Sex marriage legalization in the United States

The legalization of same-sex marriage has been a prolonged and complicated process in the Unites States. Until the landmark Supreme Court ruling on *Obergefell v. Hodges* in June 2015, marriage between two people of the same sex was not legalized at the federal level. State-level legalization efforts, however, preceded federal recognition. The first state to legalize same-sex marriage was Massachusetts, whose state supreme court ruled in 2003 that the state constitution requires the state to recognize same-sex marriage. In the dozen years after that, 35 more states and the District of Columbia followed suit with legalization at the state level, before the *Obergefell* decision granted legal marriage rights to same-sex couples nationwide.

The expansion of legalization coincided with drastic increases in both the number of cohabiting same-sex couples and the percentage of married ones among them. According to estimates by the United States Census Bureau (2022), between 2008 and 2018, the share of cohabiting same-sex couples in total U.S. households rose by 0.34 percentage point, or 71.8%, and the marriage rate among them rose from 26.4% to 59.5%, with the biggest change occurring between 2012 and 2016, when each year saw an increase of more than five percentage points. The surge was probably partially driven by both the *Obergefell* ruling and, prior to that, legalization in 29 states during this period.

For the purpose of this paper, the most relevant legal milestone is a different landmark Supreme Court ruling related to same-sex marriage: the ruling on *United States v. Windsor* in June 2013. Unlike the *Obergefell* decision, which required the federal government to issue

Marginal tax rate	Income brackets ($)	
	Single	Married filing jointly
10%	0—9,075	0—18,150
15%	9,076—36,900	18,151—73,800
25%	36,901—89,350	73,801—148,850
28%	89,351—186,350	148,851—226,850
33%	186,351—405,100	226,851—405,100
35%	405,101—406,750	405,101—457,600
39.6%	406,751+	457,601+

Table 3.1: 2014 Federal Income Tax Brackets

marriage licenses to same-sex couples, the Court decided that Section 3 of the Defense of Marriage Act (which defines marriage for federal purposes as the union of one man and one woman) was unconstitutional, and thereby ruled that the federal government must recognize, for all purposes, such licenses issued in states where same-sex marriage was legal. In August 2013, the Internal Revenue Service issued Revenue Ruling 2013-17 to comply with the decision, and started recognizing same-sex marriage for federal income tax purposes. For same-sex couples already married at the time, this implied that they would have to file as a married couple starting in tax year 2013, rather than filing separately as they would have otherwise; as for couples not yet married, the change meant that any tax incentives that pertained to joint taxation became relevant to their marriage decisions. Whereas the ruling did not have any direct effect on any couple's ability to enter a legal marriage, the tax implications may have changed their willingness to marry, which is the focus of my analysis.

3.2.2 Joint tax filing, tax subsidy and same-sex couples

Joint income tax filing for married couples was first introduced in the U.S. tax system by the Revenue Act of 1948. Initially, the joint-filing income tax schedule simply used marginal rate bracket thresholds that were exactly double those in the single-filing schedule, which meant that joint filing effectively allowed many couples with dissimilar earnings to evenly split their income and land in lower brackets, reducing their joint tax burden and leading to the so-called marriage tax subsidy. The Tax Reform Act of 1969 served as a corrective measure against such “discrimination” and established a new tax schedule for single taxpayers, reducing their tax liability relative to married couples and turning the marriage subsidy into a marriage penalty for some couples (Alm and Whittington, 1999). The penalty arises mainly from the fact that certain tax bracket threshold for couples filing jointly are now less than double those for single filers, which means single taxpayers stay in the lower-rate brackets for longer (compared to, say, if they were filing jointly with a spouse who earned the same income); the Earned Income Tax Credit is also a source of marriage subsidies or penalties for lower-income couples (Alm et al., 1999). Generally speaking, a couple is more likely to receive a marriage subsidy if their earnings are very different, and be subject to a marriage penalty if they have similar and relatively high earnings. For example, Table 3.1 lists the federal income tax schedule for tax year 2014, and Figure 3.1 uses simulated earnings profiles to illustrate the size of the marriage subsidy or penalty a hypothetical couple faces based on their respective individual earnings, assuming no other income and no other person in the household. As a result of the Tax Cuts and Jobs Act of 2017, in federal income tax schedules of tax year 2018 onwards, thresholds for all but the top income bracket for joint-filing payers again became exactly twice those for single-filing payers, and the marriage penalty was eliminated for the a large number of Americans.

The tax penalty from joint filing has become a topic of interest for economists and legal scholars in the context of same-sex marriage legalization. During the policy debate surrounding the issue, academic studies and government reports recognized and provided descriptive evidence that same-sex couples are less likely than their opposite-sex counterparts to engage in traditional household specialization, and more often form double-earner households (see e.g. Alm et al., 2000; Müller, 2002; Black et al., 2007). This suggests that same-sex couples are more vulnerable to marriage penalties and less likely to receive marriage subsidies because of the tax schedule. Kahng (2016, p.326) argues that “the tax law, through the fictitious construction of the married couple as an irreducible economic unit, continues to reward this anachronistic model of marriage and to penalize other, more egalitarian models

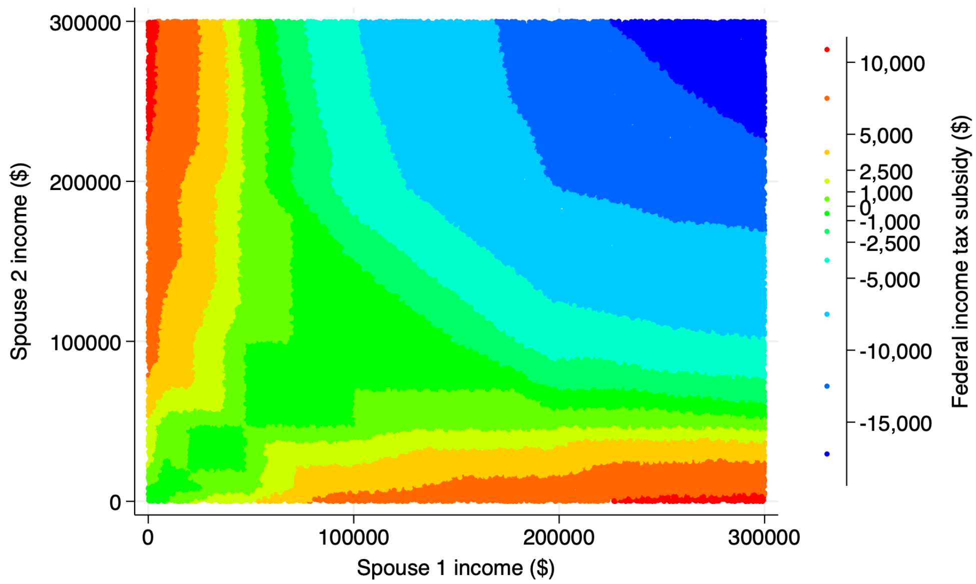


Figure 3.1: Relationship between marriage subsidy/penalty and spousal incomes

of marriage", and that the *Windsor* ruling has failed to achieve the purported tax equality through marriage equality.

The changes in marriage incentives for same-sex married couples that resulted from the *Windsor* ruling or other major milestones in the legalization process have been the focus of more recent research. A number of studies also employ ACS data and take advantage of its inclusion of same-sex couple indicators. For instance, Isaac (2018) estimates labor supply elasticities by using the shocks to tax liability and the marginal tax rate as a natural experiment, and finds moderate to large effects on different margins depending on the earning status of an individual in the household. Hamermesh & Delhommer (2020) consider a longer time span and more variation in state-level legalization status and find that the right to marry has led same-sex couples to make more investment in their relationship. Moreover, independently of this paper, Friedberg & Issac (2024) similarly examine the effect of the marriage penalty on the marriage decision of same-sex couples. Despite the common research question, their approach differs from mine both in terms of sample selection and identification strategy, and possibly contributes to the differences in our results and conclusions. I will compare the two papers in further detail in Section 3.5.

3.3 Empirical Methodology

The marriage penalty or subsidy can be incorporated into the Becker model as one of the factors that affect potential gains from marriage. A higher marriage subsidy or lower marriage penalty increases such gains and should therefore lead to a higher probability that an individual decides to enter marriage. By changing the filing status of same-sex married couples, the *Windsor* ruling made these tax incentives relevant for all same-sex couples. To the extent that these incentives play a role in individuals' decision making process, the marriage decision should respond differently to the size of the tax subsidy before and after the ruling. Since the legal consequence of the ruling is an exogenous shock to any individual couple, the identification strategy is straightforward. To capture this difference, I estimate the following logit model in a generalized difference-in-differences spirit:

$$\log \frac{P\left(married_{it}=1\right)}{1-P\left(married_{it}=1\right)} = \alpha + \beta \times post_t + \theta \times subsidy_{it} + \gamma \times post_t \times subsidy_{it} + X_{it}'\eta, \quad (3.1)$$

where $married_{it}$ is an indicator variable for marital status of couple i in year t, $subsidy_{it}$ is the marriage subsidy (penalty if negative) from joint filing for couple i in year t, $post_t$ is a binary variable that equals 1 if the decision to marry or not to was made after the *Windsor* ruling, and X_{it} is a vector of control variables that include the couple's average earnings and its square, age and its square, differences in spousal ages and incomes, state and year fixed effects and their interactions. The coefficient of interest is γ, which captures how a couple with a higher marriage subsidy responds differently to the *Windsor* ruling compared to a couple with a lower subsidy and therefore reflects the importance of the tax incentives in influencing the marriage decision. Here I am treating the coefficient on the interaction term as the "treatment effect" in this non-linear generalized difference-in-differences model, as shown in Puhani (2012).

The model suffers from endogeneity issues because earnings, tax liability and hence marriage tax subsidies can all be affected by the marriage decision, and such effects may differ from those known for opposite-sex couples (Hansen et al., 2020; Martell & Nash, 2020). I address the issue by instrumenting for the subsidy and earnings variables with predicted versions of them. I regress wage earnings on individual characteristics, and use the fitted values as a predicted wage earnings measure to calculate tax liability and marriage subsidies. The characteristics used in the earnings regression are sex (and its interaction with the other variables), race, age and its square, education, field of degree, state and year fixed effects, and their interactions. In order to calculate marriage subsidies, I make use of NBER's TAXSIM 35 tool to predict individual tax liability under different filing statuses using predicted earnings and take the difference between a couple's actual and counterfactual total tax liabilities to obtain the marriage subsidy measure. In generating the predicted earnings, tax liability and subsidies, I also make some adjustments to account for the fact that the year earnings were reported may be different from the year a marriage decision was made, which I will discuss in more detail in the following section.

3.4 Data

The ACS arguably provides the most ideal microdata currently available for empirical work related to same-sex marriage legalization because starting in the 2012 wave, data users are able to identify the marital status of cohabiting same-sex couples. In addition, the year of marriage is also observable for married couples. I construct a two-part sample using the PUMS from the 2012–2017 ACS waves. The first part of the sample consists of same-sex married couples in all waves who got married in either 2012 or 2014. The other part consists of cohabiting same-sex couples from the 2012 and 2014 waves who were not married at the time when they were surveyed. By construction, the sample includes same-sex cohabiting couples who made a decision on whether to marry or not in either year, which allows me to examine how their decisions responded differently to the size of the marriage subsidy before and after the ruling. The year 2013 is excluded because the ACS collects responses throughout the year and there is no indicator for the month in which a response was collected (or information about the month of marriage), making it impossible to determine whether a 2013 observation or a 2013 marriage was before or after the Supreme Court ruling in June. In addition, the sample only includes two-person households to abstract from the complication that arises from childbearing decisions, and is also restricted to households in the seven states that had legalized same-sex marriage by the end of 2011, namely Massachusetts (legalization effective in 2004), Connecticut (2008), Iowa, Vermont (both 2009), New Hampshire, Washington D.C. (both 2010) and New York (2011). For residents of these states, legal access to same-sex marriage has been available since before 2012, and the 2013 ruling serves as a clear natural experiment because the only change for these couples was the fact that same-sex marriage was now recognized at the federal level. Only couples where both individuals are working-age adults, i.e. between the ages 18 and 64, are included in the sample. The full sample consists of 2,290 individuals, or 1,145 couples. Because of the relatively small sample size, when calculating predicted earnings, I run the earnings regression on an expanded sample, which spans one more year (to include 2016 in addition to 2012 and 2014; no earlier years are included due to lack of same-sex married couple identifier) and includes all states and the full age range, in order to increase statistical power. The R^2 of the regression is 0.247. The two panels in Figure 3.2 plot the predicted earnings and subsidy against the reported ones for individuals in the main sample; the red lines are the 45-degree line. The predicted subsidies align remarkably with the ones calculated from reported earnings.[1]

Observations in the sample span all six ACS waves, and the survey year is generally different from the imputed year of marriage decision. As mentioned at the end of Section 3.3, the earnings prediction needs to take into consideration of the discrepancy, which affects predictions in two ways: through inflation, and through the effect of both the year and

[1]The F-statistic of this first-stage regression cannot be computed due to insufficient rank of the variance–covariance matrix, likely because some of the fixed-effects indicator variables equal 1 for only one observation in the sample.

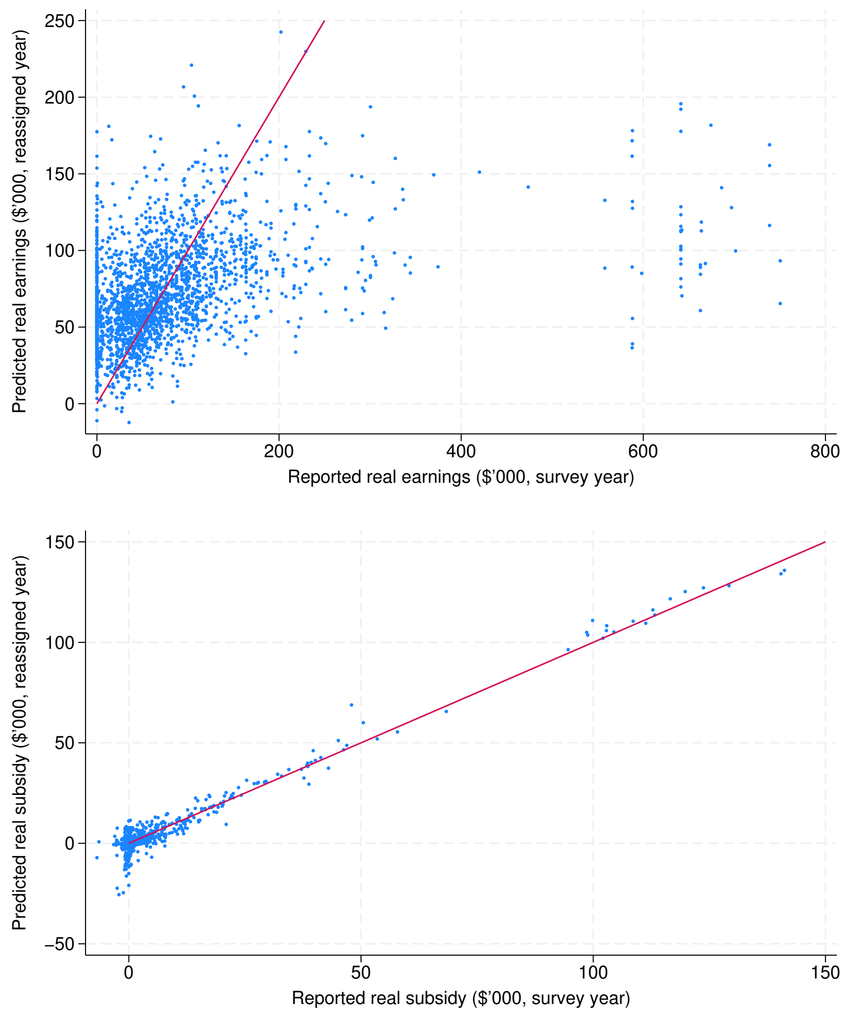


Figure 3.2: Scatterplots of predicted vs. reported earnings and subsidy

a person's age on real earnings. To adjust for both, the earnings and subsidy prediction is performed in the following steps: (a) earnings are converted to constant 2017 dollars, and used as the dependent variable in the earnings regression; (b) predicted real earnings are calculated from coefficient estimates, but adjusted for the time difference between survey and marriage decision, using estimates of coefficients on age, year and any related interaction terms; (c) based on year reassignment, predicted earnings are converted into nominal 2012 or 2014 dollars for the TAXSIM calculation, whose input and output are both in nominal dollars in the tax year in question; and (d) nominal predicted earnings, tax liabilities and marriage subsidies are all converted back to constant 2017 dollars. Throughout the paper,

	Opposite-sex	Same-sex	Gay	Lesbian	Year 2012	Year 2014
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.493	0.360	0.000	1.000	0.355	0.382
	(0.500)	(0.480)	(0.000)	(0.000)	(0.479)	(0.486)
Age	35.93	41.84	42.48	40.70	41.99	41.30
	(11.89)	(11.18)	(10.82)	(11.69)	(11.14)	(11.28)
Married	0.275	0.150	0.134	0.179	0.105	0.317
	(0.446)	(0.357)	(0.341)	(0.384)	(0.307)	(0.466)
Non-hispanic white	0.791	0.793	0.793	0.793	0.793	0.793
	(0.407)	(0.405)	(0.405)	(0.405)	(0.405)	(0.406)
Hispanic	0.088	0.102	0.104	0.098	0.103	0.097
	(0.283)	(0.303)	(0.306)	(0.297)	(0.305)	(0.295)
Black	0.058	0.059	0.054	0.069	0.061	0.052
	(0.234)	(0.236)	(0.226)	(0.254)	(0.240)	(0.222)
Asian	0.048	0.035	0.040	0.027	0.033	0.042
	(0.214)	(0.184)	(0.197)	(0.158)	(0.179)	(0.200)
College deg. or above	0.475	0.645	0.663	0.614	0.646	0.642
	(0.499)	(0.479)	(0.473)	(0.487)	(0.478)	(0.480)
Reported real earnings ($)	51278.38	78642.47	90369.10	57870.77	78130.94	80529.92
	(64461.49)	(94688.35)	(106375.90)	(64317.42)	(90659.61)	(108326.20)
Predicted real earnings ($)	46921.60	70698.25	77508.50	58635.06	69220.07	76152.47
	(22026.11)	(32136.24)	(33559.86)	(25304.31)	(32180.57)	(31407.38)
n (unweighted)	26,806	2,290	1,348	942	1,178	1,112
Predicted real subsidy ($)	1714.33	4341.93	5233.83	2763.94	4318.83	4426.29
	(7197.94)	(15377.09)	(17633.35)	(10061.96)	(14979.64)	(16776.38)
n (unweighted)	13,403	1,145	674	471	589	556

Table 3.2: Descriptive statistics of the sample

all related measures are reported in real dollars, and the main analysis aims to capture the effect of real, not nominal, differences in marriage subsidies.

Table 3.2 presents the summary statistics for selected demographic and income variables, both for the main sample and for an expanded sample that includes opposite-sex couples for comparison.[2] Columns (1) and (2) show that cohabiting (regardless of marital status)

[2]For predicted earnings and marriage subsidies for opposite-sex couples, a separate earnings regression is run on the expanded sample; those measures for the main sample and its subsamples in the subsequent columns are from the original regression. As a result, the statistics for these measures should not in principle be compared between same-sex and opposite-sex couples, and mean predicted earnings for same-sex couples using the expanded regression instead are indeed lower and more in line with those for opposite-sex couples. Nevertheless, mean predicted subsidies for same-sex couples remain largely similar quantitatively, regardless of the regression estimates used.

	Minimum	Maximum	Percentage that have: Penalty	Subsidy	Neither
All same-sex ($n = 1{,}145$)	−6865	142607	17.2%	52.4%	30.5%
Gay ($n = 674$)	−6865	142607	22.4%	51.9%	25.7%
Lesbian ($n = 471$)	−2628	121866	7.9%	53.3%	38.9%

	Percentiles 1st	5th	10th	25th	50th	75th	90th	95th	99th
All same-sex ($n = 1{,}145$)	−1683	−711	−390	0	24	1584	10890	20164	102902
Gay ($n = 674$)	−2114	−756	−540	0	26	1607	12172	29267	111318
Lesbian ($n = 471$)	−1034	−383	0	0	20	1307	5712	12522	53489

Table 3.3: Summary statistics for predicted marriage subsidy for same-sex couples

same-sex couples tend to be older and better educated, less likely to be married, and earn more than opposite-sex ones. Columns (3) and (4) decompose the main sample by sex. Gay couples outnumber lesbian ones in the sample, but most demographic characteristics are very similar between the two groups. On average, gay men are slightly more likely to be Asian, slightly less likely to be Black, and more likely to have a college degree or above. More importantly, gay couples report notably higher real earnings than lesbian couples and are predicted to incur a much bigger marriage penalty. Columns (5) and (6) compare same-sex couples who made a decision on whether to marry or not in 2012 and those who did so in 2014. The two groups are roughly equal-sized. Most demographic characteristics are similar across both years, but there was a significant increase in the marriage rate despite the fact all seven states had already legalized same-sex marriage by 2012, suggesting either a steady time trend or a response to other changes between the two years. Predicted real earnings also increased between the two years, as expected.

One notable feature of the summary characteristics is that contrary to the theoretical predictions discussed in Section 3.2, same-sex couples on average are predicted to earn a higher marriage subsidy than opposite-sex ones. A closer inspection reveals that the higher mean for same-sex couples is driven by a larger number of outliers on the right tail: while the 95th and 99th percentiles of predicted subsidy for opposite-sex couples are \$7,673 and \$27,022, respectively, they are \$20,164 and \$102,902 for same-sex couples. The medians, on the other hand, are similar between both groups ($83 and $24). Almost all of the outlying, high-marriage subsidy couples comprise a very high-earning spouse and one with little to no earnings, a phenomenon that appears to be more frequent among same-sex couples. Data in this sample, at least, does not corroborate previously mentioned theories and descriptive evidence suggesting higher incidence of marriage penalties among same-sex couples.

Table 3.3 provides further summary statistics for the distribution of the predicted marriage subsidy. Both gay and lesbian couples are on average predicted to enjoy a marriage subsidy, but large proportions of both groups neither enjoys a subsidy nor incurs a penalty. (These proportions are similar across same-sex and opposite-sex couples when the latter are included.) For the majority of households, the magnitude of the subsidy or penalty is also rather small. Furthermore, the distribution of predicted subsidy is more dispersed for gay couples than for lesbian couples, with more extreme outlier values in both tails.

3.5 Results and Discussion

Table 3.4 presents the result from estimating the logit model (3.1) separately using the full sample as well as subsamples by sex, with the household as the unit of observation. Instead of clustered standard errors at the state level, standard errors are obtained by manually bootstrapping the earnings prediction and main regression estimation using the 80 sets of household replicate weights provided in the ACS.[3] Column (1) of the table reports estimation results for the full sample. The coefficient estimate on the post-*Windsor* indicator is big and significant, consistent with the sharp increase in marriage rate in 2014 reported in Table 3.2. Being a lesbian couple is also associated with a higher probability of marriage. However, all the other coefficient estimates are insignificant. In particular, the estimate on the $post \times subsidy$ interaction term suggests no evidence showing that the magnitude of the marriage subsidy or penalty entered couples' marriage decisions. Columns (2) and (3) report

[3]This can only be done manually due to the use of TAXSIM, which does not run as part of major statistical packages.

	Full sample	Gay	Lesbian	Including 2015 & 2016	Including "late states"
	(1)	(2)	(3)	(4)	(5)
$post_t$	2.424***	2.527**	2.537**	1.488***	3.099**
	(0.630)	(1.004)	(0.985)	(0.503)	(1.149)
$subsidy_i$ ($'000)	0.0013	0.0018	0.0123	0.0011	0.0010
	(0.0060)	(0.0077)	(0.0119)	(0.0062)	(0.0058)
$post_t \times subsidy_i$	0.0030	0.0047	−0.0364	−0.0011	−0.0009
	(0.0091)	(0.0120)	(0.0251)	(0.0073)	(0.0083)
Female	0.3926**	—	—	0.3375***	0.4357***
	(0.1726)			(0.1379)	(0.1150)
Average predicted earnings ($'000)	0.0086	0.0426*	−0.0534*	0.0127	0.0236**
	(0.0181)	(0.0222)	(0.0319)	(0.0121)	(0.0105)
Average predicted earnings squared	−0.00006	−0.00024*	0.00038*	−0.00007	−0.00012*
	(0.00011)	(0.00013)	(0.00023)	(0.00008)	(0.00006)
Difference in predicted earnings ($'000)	0.0027	0.0028	−0.0003	0.0026	0.0012
	(0.0046)	(0.0059)	(0.0082)	(0.0039)	(0.0028)
Average age	0.0651	−0.0301	0.0217**	0.0175	−0.0138
	(0.0749)	(0.1045)	(0.0108)	(0.0653)	(0.0489)
Average age squared	−0.00086	0.00008	−0.00243*	−0.00035	0.00004
	(0.00089)	(0.00125)	(0.00128)	(0.00077)	(0.00056)
Age difference	−0.0331	−0.0158	−0.0670**	−0.0047	−0.0154
	(0.0167)	(0.0222)	(0.2995)	(0.0147)	(0.0112)
State fixed effects	Yes (7)	Yes (7)	Yes (7)	Yes (7)	Yes (22)
n (unweighted)	1,145	674	471	2,092	3,172

Levels of significance: *** = .01, ** = .05, * = .10.

Table 3.4: Estimation results

coefficient estimates when only gay or only lesbian couples in the sample are included in the regression. While a few estimates now become significant at the 5% or the 10% level, the coefficient of interest on the interaction term remains insignificant, in addition to being too small in magnitude to be economically meaningful. To take advantage of the full sample size, I additionally run the regression on the whole sample while allowing the coefficient on $post \times subsidy$ to vary by sex; the results remain qualitatively similar, which I do not report here.

A major drawback about the analysis concerns the sample size. As a trade-off for leveraging the *Windsor* ruling as a natural experiment, the sample selection is based on criteria that can limit the statistical power of the analysis. A combination of factors leads to a limited sample size, including the preference for a narrow bandwidth around the time of the ruling, the lack of same-sex couple identifiers in pre-2012 ACS waves, the inability to identify the month of survey or marriage in 2013, the desire to avoid confounding policy changes such as state-level legalization, and the small group of early-legalization states. To include more households, either temporally or geographically, would put identification at risk. To cover more years after 2014 would weaken the case for the event study. The problem with expanding the sample beyond the seven states currently included, on the other hand, is that numerous signifiant legislative changes regarding same-sex marriage took place across the United States in the two years between *Windsor* and *Obergefell*, and to include states that legalized same-sex marriage during the period would surely risk capturing the effect of such changes. Meanwhile, although the ACS data includes couples who lived in states that did not legalize same-sex marriage until *Obergefell* (so their legal standing remained the same between 2012 and 2014) but still got married before 2015, the fact that they would have needed to travel to another state to do so implies that legalization progress made in other, possibly neighboring states between 2012 and 2014 may have had an impact on their marriage decision. Nevertheless, in columns (4) and (5) of Table 3.4, I present regression estimates using either a temporally expanded sample to include years 2015 and 2016 in the post-*Windsor* period, or a geographically expanded one to include the 15 "late-legalization" states, which increase the sample size by 83% and 177%, respectively. In neither case does the estimated coefficient of interest see a meaningful change, which appears to lend more credibility to the findings in the original analysis.

3.5.1 Comparison with Friedberg & Issac (2024)

As acknowledged in Section 3.2, Friedberg & Issac (2024) independently carry out a study similar in focus and scope to mine. Their sample is also constructed primarily using the 2012 to 2017 waves of the ACS, and they similarly employ an IV approach to address the endogeneity problem for income and tax subsidy. However, they use an otherwise different

	This paper	Friedberg & Issac (2024)
Data	· ACS PUMS 2012–2017	· ACS PUMS 2012–2017, augmented with Survey of Income and Program Participation 2014
Sample	· Constructed to include same-sex couples in early-legalization states who made a marriage decision in 2012 or 2014	· All same-sex couples surveyed in all six ACS waves, from all states
	· 18–64 years old	· 18–60 years old
	· Only includes two-person households	· Includes couples with children
	· $n = 1145$	· $n = 37234$
Methodology	· IV and diff-in-diff	· IV without explicit diff-in-diff component
	· Treats marriage as choice variable	· Treats marriage as state variable
	· Standard regression for first-stage earnings & subsidy prediction	· Machine-learning LASSO for first-stage earnings & subsidy prediction
	· Examines federal tax subsidy alone (state subsidy remains unchanged for couples in sample)	· Examines total federal and state tax subsidies
	· Diff-in-diff identification strategy; uses 2013's Windsor decision as natural experiment	· Exploits variations created by state-level legalization, 2015's federal legalization from Obergefell, as well as 2013's tax status change from Windsor; embeds variation in instrument by setting federal/state subsidy to zero before same-sex marriage was recognized for tax purposes at corresponding level
	· Entails predicting counterfactual subsidy in different year from survey year	· No year adjustment needed
	· Only (federal) tax filing status changes for couples in sample	· Legalization status also changes for couples in sample; controls with dummy variable

Table 3.5: Comparison of my and Friedberg & Issac's (2024) approaches

empirical strategy and arrive at different results, which suggest the marriage tax subsidy has a small but significant positive effect on the marriage decision. Table 3.5 highlights the main differences in approaches between my paper and theirs. Chiefly among these are the sample selection method and the identification strategy, which are directly related to each other. Instead of using *Windsor* as a natural experiment and constructing "before" and "after" groups in states where legalization status remained unchanged in the meantime, they include all surveyed same-sex couples from all states, lending considerable statistical power to the analysis. This is made possible because they exploit broader variations in tax

	Full sample (1)	Gay (2)	Lesbian (3)
$legal_{st}$	0.3899***	0.2725***	0.5203***
	(0.0616)	(0.0953)	(0.0898)
$subsidy_{ist}$ ($'000)	−0.0009	−0.0006	−0.0020
	(0.0012)	(0.0015)	(0.0030)
Female	0.2117***	—	—
	(0.0455)		
Average predicted earnings ($'000)	−0.0004	−0.0023	−0.0033
	(0.0049)	(0.0065)	(0.0083)
Average predicted earnings squared	0.00002	0.00004	0.00005
	(0.00003)	(0.00004)	(0.00007)
Difference in predicted earnings ($'000)	0.0012	0.0003	0.0034
	(0.0013)	(0.0014)	(0.0025)
Average age	0.0730***	0.0648***	0.0932***
	(0.0189)	(0.0240)	(0.0238)
Average age squared	−0.00028	−0.00014	−0.00055**
	(0.00022)	(0.00028)	(0.00027)
Age difference	−0.0191***	−0.0137***	−0.0287***
	(0.0035)	(0.0041)	(0.0066)
State and year fixed effects	Yes	Yes	Yes
n (unweighted)	28,658	15,794	12,864

Levels of significance: *** = .01, ** = .05, * = .10.

Table 3.6: Estimation results from alternative methodology

subsidies with a different method: taking into consideration changes in both federal and state income tax filing statuses following *Windsor*, *Obergefell*, or state-level legalization, they set the portion of the subsidy from federal and/or state tax to zero in years prior to the relevant legislation or judicial ruling, so as to capture the fact that the amount of the subsidy was irrelevant for same-sex couples before legalization or recognition by federal/state tax law. Additionally, since marriage is considered a state—not choice—variable and households are treated as observed in the same year they were surveyed, their approach obviates the need to make adjustments to predicted earnings and subsidies due to year reassignment as in this paper.

Given that most of the differences between our papers are related to or dependent on each other, it is not feasible (in most of the cases) for me to isolate each single difference and investigate its contribution to our contradicting results. Hence I repeat my analysis after borrowing most of Friedberg & Issac's (2024) approach, namely abandoning the diff-in-diff identification strategy and adopting their way of constructing the predicted (state plus federal) subsidy variable, and examine if it yields different results.[4] I estimate the second-stage regression

$$\log \frac{P\left(married_{ist}=1\right)}{1-P\left(married_{ist}=1\right)} = \alpha + \beta \times subsidy_{ist} + \theta \times legal_{st} + X_{it}'\eta,$$

where $legal_{st}$ is an indicator variable that equals 1 if same-sex marriage was legalized in state s on or before year t, and X_{it} includes year fixed effects in addition to my original controls. The coefficient of interest is β. Estimation results are reported in Table 3.6; no estimate of β is significant.[5]

The failure to bridge the gap between the two sets of results suggests that these main differences in our approaches are not the primary contributor to the discrepancy in our results. Considering that they also include controls for the presence and number of children in the household, the remaining factor that is most likely to underly the differences is the first-stage prediction of earnings and subsidy. Friedberg & Issac (2024) state that they choose machine-learning LASSO for the prediction because they see the first stage as "effectively a prediction exercise" (Mullainathan & Spiess, 2017, p. 100). Compared to my standard regression approach and given the year adjustments I have to make, such an approach can conceivably lead to less noise in the predicted instrument, which may be the reason for the discrepancy in our results. At the moment, investigating the effect of the method of prediction is beyond the scope of this paper.

[4] I do not, however, include couples with children in the sample or change the age range.

[5] I additionally change the definition of $legal_{st}$ so that the year of legalization itself is coded as 0 instead of 1. This does not significantly impact the results.

3.6 Conclusion

Using ACS data on same-sex married and unmarried cohabiting couples in 2012 and 2014, I am able to estimate how the Supreme Court's 2013 ruling on *United States v. Windsor* affected these couples' behavioral response to the marriage subsidy. A predicted earnings measure is constructed from individual characteristics to tackle endogeneity issues, and the measure is used to derive tax liability and marriage penalties. Estimation results do not point to evidence of a significant effect of the ruling for the full sample or either subsample of gay or lesbian couples. The relatively small subsidy or penalty faced by the majority of couples may partially explain its lack of significance in marriage decisions. Perhaps more importantly, given the wide variety of factors that may go into such decisions, in particular the symbolic importance of marriage for same-sex couples during a period of rapid change in the cultural and legal environment, it would not be surprising if the tax incentives only have a subdued effect on their decisions.

The behavioral response of same-sex couples to marriage tax incentives is a worthy research topic both as part of the broader endeavor to understand how households respond to distortions created by the transfer system, and in its own right because it facilitates understanding of the differences between same- and opposite-sex couples in their economic decision-making. Although the nature of the *Windsor* ruling makes it an ideal natural experiment for revisiting the topic of marriage tax incentive in a new context, data availability on same-sex couples has imposed limits on the possibility of empirical research. As shown by Friedberg & Issac (2024), it is possible to adopt different methods to analyze the ACS data and continue to test the robustness of the findings in this paper, and incorporating state-level variations to complement my current approach may be helpful in relaxing some of the constraints this study has faced.

References

Ackley, C. A., Lundberg, D. J., Ma, L., Elo, I. T., Preston, S. H., & Stokes, A. C. (2022). County-level estimates of excess mortality associated with COVID-19 in the United States. *SSM - Population Health*, *17*, 101021.

Adolph, C., Amano, K., Bang-Jensen, B., Fullman, N., & Wilkerson, J. (2021, 04). Pandemic Politics: Timing State-Level Social Distancing Responses to COVID-19. *Journal of Health Politics, Policy and Law*, *46*(2), 211-233.

Agee, M. D., & Crocker, T. D. (1996). Parental Altruism and Child Lead Exposure: Inferences from the Demand for Chelation Therapy. *The Journal of Human Resources*, *31*(3), 677–691.

Aizer, A., Currie, J., Simon, P., & Vivier, P. (2018). Do Low Levels of Blood Lead Reduce Children's Future Test Scores? *American Economic Journal: Applied Economics*, *10*(1), 307–341.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*, 104254.

Alm, J., Badgett, M. L., & Whittington, L. A. (2000). Wedding Bell Blues: The Income Tax Consequences of Legalizing Same-Sex Marriage. *National Tax Journal*, *53*(2), 201–214.

Alm, J., Dickert-Conlin, S., & Whittington, L. A. (1999). Policy Watch: The Marriage Penalty. *The Journal of Economic Perspectives*, *13*(3), 193–204.

Alm, J., & Whittington, L. A. (1999). For Love or Money? The Impact of Income Taxes on Marriage. *Economica*, *66*(263), 297–316.

American Academy of Pediatrics Subcommittee on Accidental Poisoning. (1961). Report Of Subcommittee On Accidental Poisoning : Statement on Diagnosis and Treatment of Lead Poisoning in Childhood. *Pediatrics*, *27*(4), 676–680.

Bae, H. (2016). The impact of the residential lead paint disclosure rule on house prices: findings in the American Housing Survey. *Journal of Housing and the Built Environment*, *31*(1), 19–30.

Baker, M., Kantarevic, J., & Hanna, E. (2004). The Married Widow: Marriage Penalties Matter! *Journal of the European Economic Association*, *2*(4), 634–664.

Becker, G. (1973). A Theory of Marriage: Part I. *Journal of Political Economy*, *81*(4), 813–846.

Becker, G. (1974). A Theory of Marriage: Part II. In *Marriage, family, human capital, and fertility* (pp. 11–26). National Bureau of Economic Research, Inc.

Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, *78*(4), 551–572.

Bergin, D., Ladyzhets, B., Kincaid, J., Kravitz, D., Haselhorst, S., White, A., ... Hassanein, N. (2021, December 22). Uncounted: Inaccurate death certificates across the country hide the true toll of COVID-19. *USA Today*. Retrieved from https://www.usatoday.com/in-depth/news/nation/2021/12/22/covid-deaths-obscured-inaccurate-death-certificates/8899157002/

Bhullar, M. K., Gilson, T. P., & Lee, J. (2022). The Public Health Role of Medical Examiner Offices During COVID-19 and Other Mass Fatality Events. *The American journal of forensic medicine and pathology*, *43*(2), 101–104.

Billings, S. B., & Schnepel, K. T. (2017). The value of a healthy home: Lead paint remediation and housing values. *Journal of Public Economics*, *153*, 69–81.

Binns, H. J., Campbell, C., & Brown, M. J. (2007). Interpreting and Managing Blood Lead Levels of Less Than 10 μg/dL in Children and Reducing Childhood Exposure to Lead: Recommendations of the Centers for Disease Control and Prevention Advisory Committee on Childhood Lead Poisoning Prevention. *Pediatrics*, *120*(5), e1285–e1298.

Black, D. A., Sanders, S. G., & Taylor, L. J. (2007). The Economics of Lesbian and Gay Families. *Journal of Economic Perspectives*, *21*(2), 53–70.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, *114*(2), 577–599.

Bolsen, T., & Palm, R. (2022). Chapter Five - Politicization and COVID-19 vaccine resistance in the U.S. In T. Bolsen & R. Palm (Eds.), *Molecular Biology and Clinical Medicine in the Age of Politicization* (Vol. 188, pp. 81–100). Academic Press.

Bordelon, C. (2021, January 12). Conspiracies attack coroner: Families demand COVID-19 diagnoses be removed from loved one's death certificates. *KOAA News5*. Retrieved from https://www.koaa.com/news/coronavirus/conspiracies-attack-coroner-families-demand-covid-19-diagnoses-be-removed-from-loved-ones-death-certificates

Brooks, C. (2021, November). *Medical Examiner and Coroner Offices, 2018*. U.S. Department of Justice, Bureau of Justice Statistics. Retrieved from https://bjs.ojp.gov/content/pub/pdf/meco18.pdf

Brown, M. J., & Margolis, S. (2012). *Lead in Drinking Water and Human Blood Lead Levels in the United States*. Centers for Disease Control and Prevention.

Brown, M. J., Raymond, J., Homa, D., Kennedy, C., & Sinks, T. (2011). Association between children's blood lead levels, lead service lines, and water disinfection, Washington, DC, 1998–2006. *Environmental Research*, *111*(1), 67–74.

Campolieti, M. (2022). COVID-19 deaths in the USA: Benford's law and under-reporting. *Journal of Public Health*, *44*(2), e268–e271.

Centers for Disease Control and Prevention. (2021, October). *CDC Updates Blood Lead Reference Value for Children.* (Press release)

Centers for Disease Control and Prevention. (2023a). *CDC WONDER, Multiple Cause of Death, 2018-2021, Single Race.* Retrieved from https://wonder.cdc.gov/mcd-icd10-expanded.html

Centers for Disease Control and Prevention. (2023b, February 8). *Death Investigation Systems.* Retrieved from https://www.cdc.gov/phlp/publications/coroner/death.html

Chay, K. Y., & Greenstone, M. (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy*, *113*(2), 376–424.

Choi, S. . J., & Gulati, M. (2017). Adjudicating Death: Professionals or Politicians? *Vanderbilt Law Review*, *70*(6), 1709–1727.

Clay, K., Troesken, W., & Haines, M. (2006). Lead Pipes and Child Mortality. National Bureau of Economic Research Working Paper No. 12603.

Cornwell, D. A., Brown, R. A., & Via, S. H. (2016). National Survey of Lead Service Line Occurrence. *Journal (American Water Works Association)*, *108*(4), E182–E191.

Del Toral, M. A., Porter, A., & Schock, M. R. (2013). Detection and Evaluation of Elevated Lead Release from Service Lines: A Field Study. *Environmental Science & Technology*, *47*, 9300–9307.

Denham, A., Vasu, T., Avendano, P., Boslett, A., Mendoza, M., & Hill, E. L. (2022). Coroner county systems are associated with a higher likelihood of unclassified drug overdoses compared to medical examiner county systems. *The American Journal of Drug and Alcohol Abuse*, *48*(5), 606–617.

Dickert-Conlin, S., & Houser, S. (2002). EITC and Marriage. *National Tax Journal*, *55*(1), 25–40.

Edwards, M., Triantafyllidou, S., & Best, D. (2009). Elevated Blood Lead in Young Children

Due to Lead-Contaminated Drinking Water: Washington, DC, 2001–2004. *Environmental Science & Technology*, *43*(5), 1618–1623.

Eutsler, J., Kathleen Harris, M., Tyler Williams, L., & Cornejo, O. E. (2023). Accounting for Partisanship and Politicization: Employing Benford's Law to Examine Misreporting of COVID-19 Infection Cases and Deaths in the United States. *Accounting, Organizations and Society*, 101455.

Farris, E. M., & Holman, M. R. (2023). Sheriffs, right-wing extremism, and the limits of U.S. federalism during a crisis. *Social Science Quarterly*, *104*(2), 59–68.

Fisher, H. (2013). The Effect of Marriage Tax Penalties and Subsidies on Marital Status. *Fiscal Studies*, *34*(4), 437–465.

Flynn, J. L. (1955). The Office of Coroner vs. the Medical Examiner System. *The Journal of Criminal Law, Criminology, and Police Science*, *46*(2), 232–238.

Friedberg, L., & Isaac, E. (2024). Same-Sex Marriage Recognition and Taxes: New Evidence about the Impact of Household Taxation. *The Review of Economics and Statistics*, *106*(1), 1–17. (Advance online publication)

Gadarian, S. K., Goodman, S. W., & Pepinsky, T. B. (2021). Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. *PLOS ONE*, *16*(4), 1–13.

Gayer, T., Hamilton, J. T., & Viscusi, W. K. (2002). The Market Value of Reducing Cancer Risk: Hedonic Housing Prices with Changing Information. *Southern Economic Journal*, *69*(2), 266–289.

Gazze, L. (2021). The price and allocation effects of targeted mandates: Evidence from lead hazards. *Journal of Urban Economics*, *123*, 103345.

Gill, J. R., & DeJoseph, M. E. (2020). The Importance of Proper Death Certification During the COVID-19 Pandemic. *JAMA*, *324*(1), 27–28.

Gollwitzer, A., Martel, C., Brady, W. J., Pärnamets, P., Freedman, I. G., Knowles, E. D., & Van Bavel, J. J. (2020). Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nature Human Behaviour*, *4*(11), 1186–1197.

Goodlad, J. K., Marcus, D. K., & Fulton, J. J. (2013). Lead and Attention-Deficit/Hyperactivity Disorder (ADHD) symptoms: A meta-analysis. *Clinical Psychology Review*, *33*(3), 417–425.

Grönqvist, H., Nilsson, J. P., & Robling, P.-O. (2020). Understanding How Low Levels of

Early Lead Exposure Affect Children's Life Trajectories. *Journal of Political Economy*, *128*(9), 3376–3433.

Grossman, G., Kim, S., Rexer, J. M., & Thirumurthy, H. (2020). Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proceedings of the National Academy of Sciences*, *117*(39), 24144–24153.

Halpern, L. W. (2020). The Politicization of COVID-19. *American Journal of Nursing*, *120*(11), 19–20.

Hamermesh, D. S., & Delhommer, S. (2020). *Same-Sex Couples and the Marital Surplus: The Importance of the Legal Environment* (NBER Working Papers No. 26875). National Bureau of Economic Research, Inc.

Hansen, M. E., Martell, M. E., & Roncolato, L. (2020). A labor of love: The impact of same-sex marriage on labor supply. *Review of Economics of the Household*, *18*, 256–283.

Hanzlick, R. (2007). The Conversion of Coroner Systems to Medical Examiner Systems in the United States: A Lull in the Action. *The American journal of forensic medicine and pathology*, *28*(4), 279–283.

Hanzlick, R., & Combs, D. (1998). Medical Examiner and Coroner Systems: History and Trends. *JAMA*, *279*(11), 870–874.

Hanzlick, R., Combs, D., Parrish, R. G., & Ing, R. T. (1993). Death Investigation in the United States, 1990: A Survey of Statutes, Systems, and Educational Requirements. *Journal of Forensic Sciences*, *38*(3), 628–632.

Hanzlick, R., & Parrish, R. G. (1996). The role of medical examiners and coroners in public health surveillance and epidemiologic research. *Annual Review of Public Health*, *17*, 383–409.

Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and Polarization in COVID-19 News Coverage. *Science Communication*, *42*(5), 679–697.

Hodge, A. T. (1981). Vitruvius, lead pipes and lead poisoning. *American Journal of Archaeology*, *85*(4), 486–491.

Holman, M. R., Farris, E. M., & Sumner, J. L. (2020). Local Political Institutions and First-Mover Policy Responses to COVID-19. *Journal of Political Institutions and Political Economy*, *1*(4), 523–541.

Institute of Medicine (US) Committee for the Workshop on the Medicolegal Death Investigation System. (2003). Medicolegal Death Investigation System: Workshop Summary. In (pp. 23–26). National Academies Press.

Isaac, E. (2018). *Suddenly Married: Joint Taxation And The Labor Supply Of Same-Sex Married Couples After U.S. v. Windsor* (Working Papers No. 1809). Tulane University, Department of Economics.

Kahng, L. (2016). The Not-So-Merry Wives of Windsor: The Taxation of Women in Same-Sex Marriages. *SSRN Electronic Journal*, *101*, 325–384.

Kerr, J., Panagopoulos, C., & van der Linden, S. (2021). Political polarization on COVID-19 pandemic response in the United States. *Personality and Individual Differences*, *179*, 110892.

Kiang, M., Irizarry, R., Buckee, C., & Balsari, S. (2020, 15). Every Body Counts: Measuring Mortality From the COVID-19 Pandemic. *Annals of Internal Medicine*, *173*(12), 1004–1007.

Klugman, J., Condran, G., & Wray, M. (2013). The Role of Medicolegal Systems in Producing Geographic Variation in Suicide Rates. *Social Science Quarterly*, *94*(2), 462–489.

Linden, L., & Rockoff, J. E. (2008). Estimates of the Impact of Crime Risk on Property Values from Megan's Laws. *The American Economic Review*, *98*(3), 1103–1127.

Martell, M. E., & Nash, P. (2020). For Love and Money? Earnings and Marriage Among Same-Sex Couples. *Journal of Labor Research*, *41*, 260–294.

Mastromonaco, R. (2015). Do environmental right-to-know laws affect markets? Capitalization of information in the toxic release inventory. *Journal of Environmental Economics and Management*, *71*(C), 54–70.

Matsa, K. E., Mitchell, A., & Stocking, G. (2017). *Searching for News: The Flint water crisis.* Pew Research Center.

McGivern, L., Shulman, L., Carney, J. K., Shapiro, S., & Bundock, E. (2017). Death Certification Errors and the Effect on Mortality Statistics. *Public Health Reports*, *132*(6), 669–675.

Mebane, W. (2006). *Election Forensics: Vote Counts and Benford's Law.* Summer Meeting of the Political Methodology Society, University of California, Davis.

MIT Election Data and Science Lab. (2018). *County Presidential Election Returns 2000-2020.* Harvard Dataverse. Retrieved from https://doi.org/10.7910/DVN/VOQCHQ

Moffitt, R. (1992). Incentive Effects of the U.S. Welfare System: A Review. *Journal of*

Economic Literature, *30*(1), 1–61.

Motta, M., Stecula, D., & Farhart, C. (2020). How Right-Leaning Media Coverage of COVID-19 Facilitated the Spread of Misinformation in the Early Stages of the Pandemic in the U.S. *Canadian Journal of Political Science/Revue canadienne de science politique*, *53*(2), 335–342.

Mullainathan, S., & Spiess, J. (2017, May). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Müller, C. (2002). *An Economic Analysis of Same-Sex Marriage* (German Working Papers in Law and Economics No. 2002-1-1045). Berkeley Electronic Press.

National Vital Statistics System. (2023, February). *Guidance for Certifying Deaths Due to Coronavirus Disease 2019 (COVID-19)*. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/nchs/data/nvss/vsrg/vsrg03-508.pdf

Needleman, H. L., & Gatsonis, C. A. (1990). Low-Level Lead Exposure and the IQ of Children: A Meta-analysis of Modern Studies. *Journal of the American Medical Association*, *263*(5), 673–678.

Nigrini, M. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Wiley.

Paglino, E., Lundberg, D. J., Zhou, Z., Wasserman, J. A., Raquib, R., Hempstead, K., ... Stokes, A. C. (2023). Differences Between Reported COVID-19 Deaths and Estimated Excess Deaths in Counties Across the United States, March 2020 to February 2022. *medRxiv*.

Palmquist, R. B. (1982). Measuring Environmental Effects on Property Values Without Hedonic Regressions. *Journal of Urban Economics*, *11*(3), 333–347.

Pathak, E. B., Garcia, R. B., Menard, J. M., & Salemi, J. L. (2021). Out-of-Hospital COVID-19 Deaths: Consequences for Quality of Medical Care and Accuracy of Cause of Death Coding. *American Journal of Public Health*, *111*(S2), S101–S106.

The politics of death. (2022, January 29). *The Economist*, *442*(9281), 22.

Pope, J. (2008). Buyer information and the hedonic: The impact of a seller disclosure on the implicit price for airport noise. *Journal of Urban Economics*, *63*(2), 498–516.

Pueschel, S. M., Linakis, J. G., & Anderson, A. C. (1996). Lead Poisoning in Childhood. In S. M. Pueschel, J. G. Linakis, & A. C. Anderson (Eds.), (pp. 1–13). P. Brookes Publishing Company.

Puhani, P. (2012). The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-in-Differences" Models. *Economics Letters*, *115*(1), 85–87.

Rabin, R. (2008). The Lead Industry and Lead Water Pipes: "A Modest Campaign". *American Journal of Public Health*, *98*(9), 1584–1592.

Reyes, J. W. (2007). Environmental Policy as Social Policy? The Impact of Childhood Lead Exposure on Crime. *The B.E. Journal of Economic Analysis & Policy*, *7*(1), 1–43.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55.

Rosenbaum, D. T. (2000). *Taxes, the Earned Income Tax Credit, and Marital Status* (JCPR Working Papers No. 177). Northwestern University/University of Chicago Joint Center for Poverty Research.

Rossen, L. M., Nørgaard, S. K., Sutton, P. D., Krause, T. G., Ahmad, F. B., Vestergaard, L. S., . . . Nielsen, J. (2022). Excess all-cause mortality in the USA and Europe during the COVID-19 pandemic, 2020 and 2021. *Scientific reports*, *12*(1), 18559.

Rothgerber, H., Wilson, T., Whaley, D., Rosenfeld, D. L., Humphrey, M., Moore, A., & Bihl, A. (2020). Politicizing the COVID-19 Pandemic: Ideological Differences in Adherence to Social Distancing. *PsyArXiv*.

Roy, S., & Edwards, M. A. (2019). Preventing another lead (Pb) in drinking water crisis: Lessons from the Washington D.C. and Flint MI contamination events. *Current Opinion in Environmental Science & Health*, *7*, 34–44.

Sandvig, A., Kwan, P., Kirmeyer, G., Maynard, B., Mast, D., Trussell, R. R., . . . Prescott, A. (2008). *Contribution of Service Line and Plumbing Fixtures to Lead and Copper Rule Compliance Issues. Research Report 91229*. American Water Works Association Research Foundation.

Scott, W. F. (1981). Some Applications of the Poisson Distribution in Mortality Studies. *Transactions of the Faculty of Actuaries*, *38*, 255–263.

Selevan, S., Rice, D., Hogan, K., Euling, S., Pfahles-Hutchens, A., & Bethel, J. (2003). Blood Lead Concentration and Delayed Puberty in Girls. *The New England journal of medicine*, *348*, 1527–36.

Sjoquist, D. L., & Walker, M. B. (1995). The Marriage Tax and the Rate and Timing of Marriage. *National Tax Journal*, *48*(4), 547–558.

Stokes, A. C., Lundberg, D. J., Bor, J., & Bibbins-Domingo, K. (2021). Excess Deaths During the COVID-19 Pandemic: Implications for US Death Investigation Systems.

American Journal of Public Health, *111*(S2), S53–S54.

Stokes, A. C., Lundberg, D. J., Elo, I. T., Hempstead, K., Bor, J., & Preston, S. H. (2021). Assessing the Impact of the Covid-19 Pandemic on US Mortality: A County-Level Analysis. *medRxiv*.

Stroebe, W., VanDellen, M. R., Abakoumkin, G., Lemay Jr, E. P., Schiavone, W. M., Agostini, M., ... Leander, N. P. (2021). Politicization of COVID-19 health-protective behaviors in the United States: Longitudinal and cross-national evidence. *PloS one*, *16*(10), e0256740.

Theising, A. (2019). Lead Pipes, Prescriptive Policy and Property Values. *Environmental and Resource Economics*, *74*(3), 1355–1382.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*(1), 24–36.

Triantafyllidou, S., & Edwards, M. (2012). Lead (Pb) in Tap Water and in Blood: Implications for Lead Exposure in the United States. *Critical Reviews in Environmental Science and Technology*, *42*(13), 1297–1352.

Troesken, W. (2006). *The Great Lead Water Pipe Disaster*. MIT Press.

United States Census Bureau. (2022). *Characteristics of Same-Sex Couple Households: 2005 to Present.* Retrieved from https://www.census.gov/data/tables/time-series/demo/same-sex-couples/ssc-house-characteristics.html

United States Census Bureau. (2022, June). *Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2021*. Retrieved from https://www2.census.gov/programs-surveys/popest/datasets/2020-2021/counties/asrh/cc-est2021-all.csv

Votsis, A., & Perrels, A. (2015). Housing Prices and the Public Disclosure of Flood Risk: A Difference-in-Differences Analysis in Finland. *The Journal of Real Estate Finance and Economics*, *53*, 450–471.

Woolf, S. H., Chapman, D. A., Sabo, R. T., Weinberger, D. M., Hill, L., & Taylor, D. D. H. (2020). Excess Deaths From COVID-19 and Other Causes, March-July 2020. *JAMA*, *324*(15), 1562–1564.

Woolf, S. H., Chapman, D. A., Sabo, R. T., & Zimmerman, E. B. (2021). Excess Deaths From COVID-19 and Other Causes in the US, March 1, 2020, to January 2, 2021. *JAMA*, *325*(17), 1786–1789.

Zavattaro, S. M. (2023). Death managers, public health, and COVID-19: An exploratory study. *Public Administration Review*, *83*(5), 1339–1350.

Zylke, J. W., & Bauchner, H. (2020). Mortality and Morbidity: The Measure of a Pandemic. *JAMA*, *324*(5), 458–459.