Legislating Longevity: State Public Health Laws and Mortality in the Early Twentieth Century

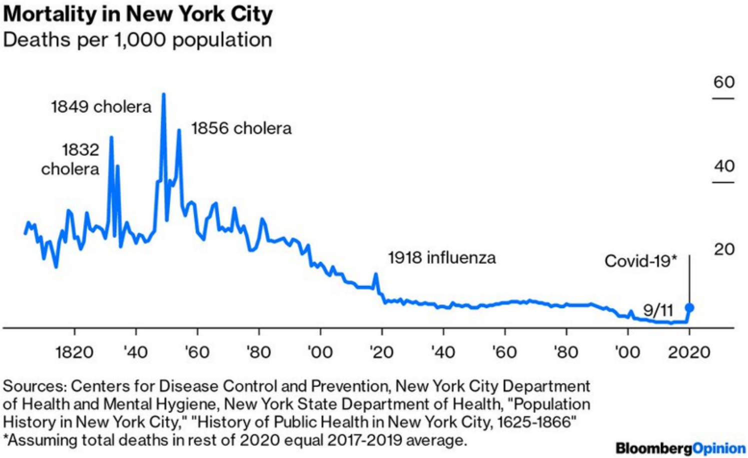
Martin Saavedra
Rutgers University

2026

Question

- Can the universe of state-level public health laws explain the mortality transition?
- Can we use text analysis to describe the most common types of laws?
- If so, which laws are most important?
- Strategy:
  - Digitize the universe of state-level public health laws (1911-1928)
  - Use topic modeling to categorize nearly 7,000 laws
  - Instrument for laws using the legislative schedule and legislator salary.

Motivation



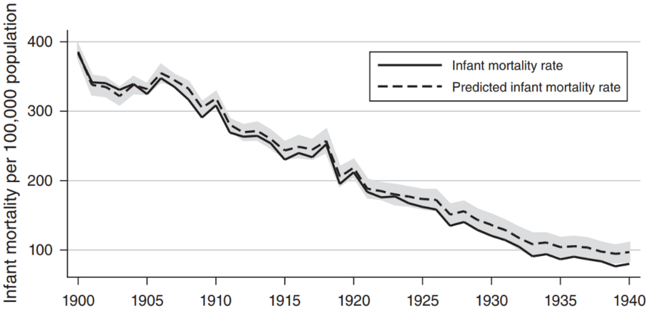
Mortality in New York City
Deaths per 1,000 population
1849 cholera
1856 cholera
1832 cholera
1918 influenza
Covid-19*
9/11
60
40
20
1820
'40
'60
'80
'00
'20
'40
'60
'80
'00
2020
Sources: Centers for Disease Control and Prevention, New York City Department of Health and Mental Hygiene, New York State Department of Health, "Population History in New York City," "History of Public Health in New York City, 1625-1866"
*Assuming total deaths in rest of 2020 equal 2017-2019 average.
BloombergOpinion

Motivation

- What caused the mortality transition?
- Some interventions can explain the declines in specific causes
  - Water filtration for typhoid fever
  - Vaccines for smallpox
- Most is still unexplained
- One theory: it was many small interventions

Motivation

Figure: Anderson, Charles, and Rees (2022, AEJ: Applied)



Preview of findings

- Baseline estimates: public health laws can explain 70% of the mortality decline during the sample.
- Text analysis identifies categories of laws that are well studied:
  - Sewers/water
  - Dairy
  - Health boards
- And less well studied:
  - Housing/tenements
  - Drug safety
  - Workplace
- Evidence on which types of laws mattered is mixed or noisy.

Previous research

- Health boards: Hoehn-Velasco and Wrigley-Field (2022)
- Waterworks/sanitation: Anderson et al (2022), Cutler and Miller (2005), Alsan and Goldin (2019), Troesken (2004), Ferrie and Troesken (2008)
- Vaccines: Brehm and Saavedra (2025), Troesken (2017)
- Milk/food inspections: Anderson et al (forthcoming)
- Physicians/hospitals: Hollingsworth et al (2024)

Background

Figure: NYC Tenement

Background

Figure: NYC Tenement Museum

Background

Figure: Bayer's Heroin

Background

- Maximum recommended dose of Aspirin in 24 hours is 4 grams.
- “These recommended doses (1000–1300 mg), with frequencies ranging from hourly to every 3 hours, resulting in daily doses of 8–31.2 grams, are above the maximum safe dose defined above and would lead to accumulation, as noted below.” -Starko (2009) in *Clinical Infectious Diseases*

Law Data

- "State Laws and Regulations Pertaining to Public Health" published annually by the USPHS
- Title is hand entered
- Full text is OCRed
- Merged with state-level vital statistics data:
    - Monthly all-cause mortality
    - Annual cause of death

Law Data

Figure: Example of a public health regulation

Smallpox—Quarantine. (Reg. Bd. of H., Jan. 13, 1922)

1301. All members of a household where smallpox exists shall be quarantined until released by the local health officer under the following provisions:

(*a*) Before release of a smallpox patient the skin must be free of scabs and the dark-colored plaques, often present under the outer layer of skin of the palms of the hands and the soles of the feet; the patient must take a full bath and shampoo the hair, and all clothing and other articles exposed to infection must be disinfected as directed by the local health officer.

(*b*) Persons, not protected by a recent successful vaccination or an attack of smallpox, residing on premises where smallpox exists or directly exposed by association with a case of smallpox, who refuse to be vaccinated shall be isolated and shall not be permitted to leave the premises until 21 days after last exposure.

(*c*) Persons who are protected by a recent successful vaccination, or an attack of smallpox, or who submit to vaccination within three days after first exposure to smallpox, may be given written authorization by the local health officer to go into and from the premises under quarantine for smallpox.

Shaving Brushes Containing Horse Hair—Manufacture, Possession, Sale, or Distribution of, Prohibited. (Reg. Bd. of H., May 2, 1922)

2000. No person shall manufacture, have, keep, offer for sale, sell, distribute, or give away, in the State of Minnesota, any shaving brush in which horsehair is used in whole or in part.

Law Data

Figure: Hand entered Table of Contents

Law Data

- OCR strategy to obtain the full text of law n with:
  - Title $title_n$
  - Page $page_n$
  - State $state_n$

Law Data

- OCR strategy to obtain the full text of law n with:
  - Title $title_n$
  - Page $page_n$
  - State $state_n$
- OCR every page of the annual report as a separate file
- Find the OCR to TOC page number correspondence.
- Locate the page range: $page_n$ to $page_{n+1}$.
- Clean the text: Keep alphabetic characters and spaces; treat consecutive spaces as 1.
- Continue to truncation step to remove text from other laws.

Law Data

- Truncation step:
  - Search for $title_n$ $\Rightarrow$ Remove text before that.
  - Search for $title_{n+1}$ $\Rightarrow$ Remove text after that.
  - If $state_n = state_{n+1}$, search for $state_{n+1}$ $\Rightarrow$ Remove text after that.
- If the left/right truncation of text is unsuccessful, there may be paragraphs of other laws appended (21% of laws)

Topic modeling: Intuition

- **Latent Dirichlet Allocation**: Bayesian method from NLP
- Data generating process:
  - Every law is a mixture over 10 topics.
  - Every topic a unique probability distribution over words
- Finds sets of vocabulary that tend to co-occur together.
- Text cleaning:
  - Remove rare words (fewer than 5 times), stop words.
  - Apply the Porter stemmer.
  - Count bigrams as words.

Topic modeling: Formal Setup

- For each document $d \in \{1, \ldots, D\}$:
  1. Draw topic proportions $\theta_d \sim \text{Dir}(\alpha)$
  2. For each word position $n \in \{1, \ldots, N_d\}$:
     - 2.1 Draw a topic assignment: $z_{d,n} \sim \text{Multinomial}(\theta_d)$
     - 2.2 Draw a word: $w_{d,n} \sim \text{Multinomial}(\beta_{z_{d,n}})$
- Estimated parameters:
  - $\hat{\beta}_k$: Probability distribution over words for topic k
  - $\hat{\theta}_d$: Topic proportion of document d

Topic modeling: Aggregating laws

- Laws passed in state s in time t be $dLaws_{s,t}$
- Cumulative laws passed is:

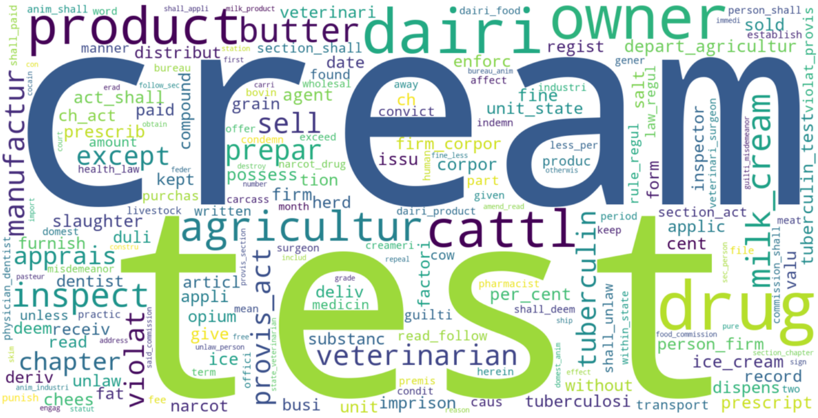
$$Laws_{s,t} = \sum_{\tau=t_0}^{t} dLaws_{s,\tau}$$

- Cumulative laws passed in topic k is

$$Topic_{s,t}^{k} = \sum_{\tau=t_0}^{t} \sum_{d\in s,\tau} \hat{\theta}_d^k$$

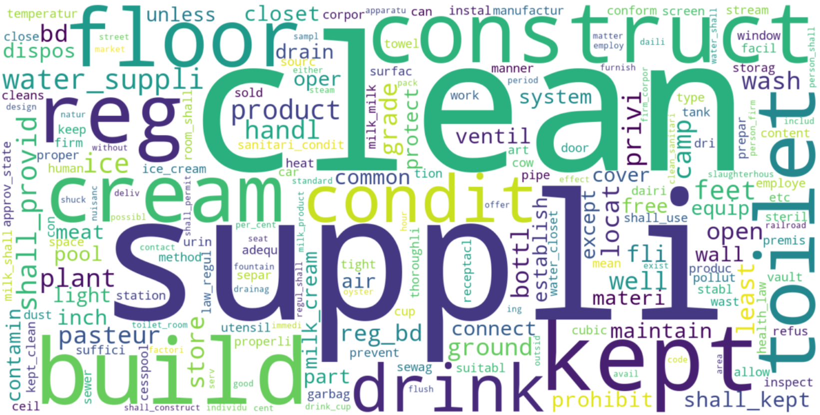
Topic modeling

Figure: Dairy and agriculture



Topic modeling

Figure: Dairy and sanitation

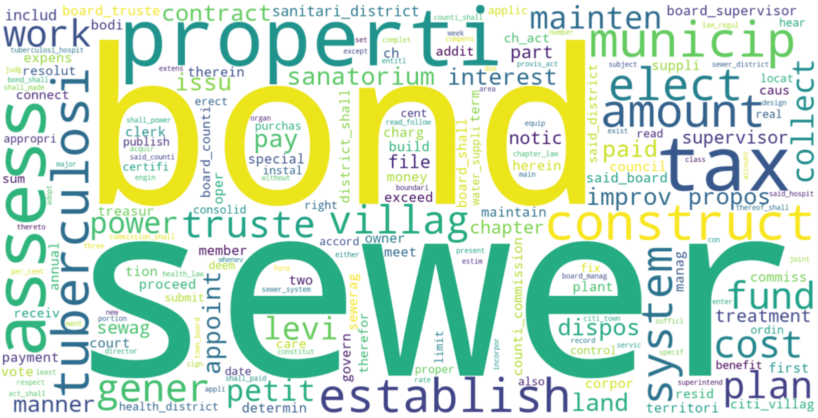


Topic modeling

Figure: Venereal disease

Topic modeling

Figure: Sewers and water

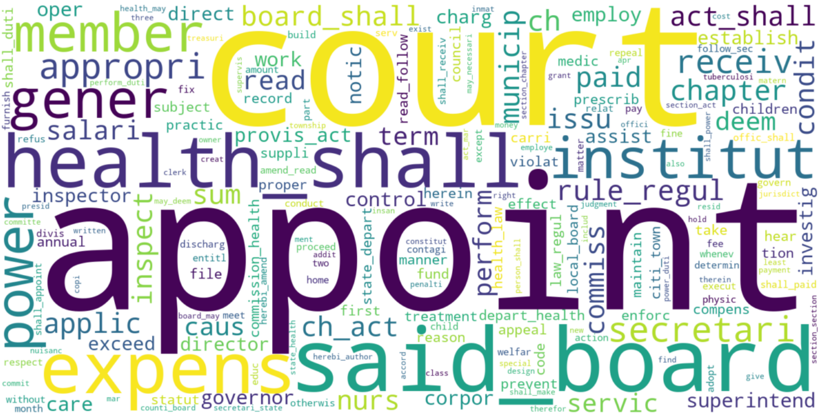


Topic modeling

Figure: Communicable disease

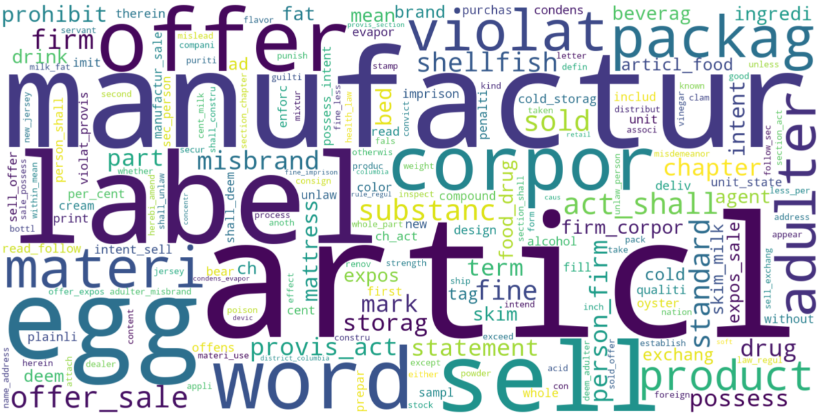
Topic modeling

Figure: Boards of health



Topic modeling

Figure: Food and drug safety



Topic modeling

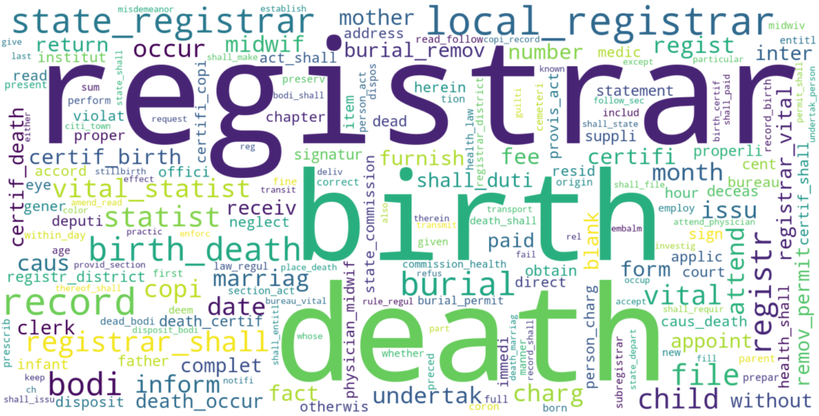
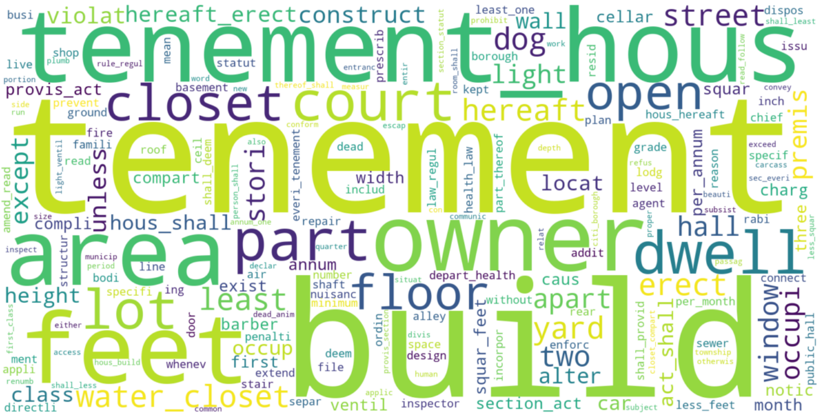


Figure: Vital statistics

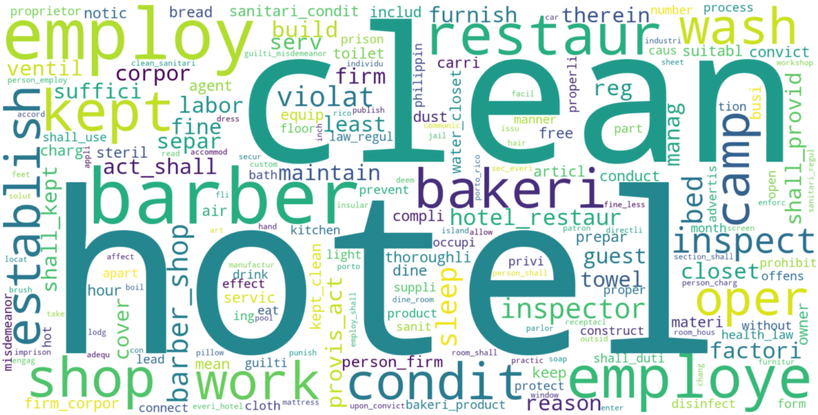
Topic modeling

Figure: Housing



Topic modeling

Figure: Work place



Data

Table: Summary statistics

	Mean	S.D.	Min	Max	N
Monthly mortality rate	105.5	30.2	46.9	636.5	6,850
Laws passed that month	0.7	2.3	0.0	41.0	6,850
Word count	965.0	3745.6	0.0	71919.0	6,850
Repeal laws	0.1	0.7	0.0	13.0	6,850
Laws with punishments	0.2	0.9	0.0	16.0	6,850
Cumulative laws	73.1	62.1	0.0	422.0	6,850
Cumulative word count	104.9	90.8	0.0	636.1	6,850
Cumulative repeal laws	14.6	13.9	0.0	70.0	6,850
Cumulative laws with punishments	24.6	19.5	0.0	113.0	6,850
Year	1920.8	4.9	1911.0	1928.0	6,850
Month	6.6	3.4	1.0	12.0	6,850
Salary	3.8	3.2	0.5	15.0	6,850
Session length	111.8	57.6	40.0	180.0	6,850
In session	0.2	0.4	0.0	1.0	6,850
Cumulative months in sessions	22.8	19.0	0.0	102.0	6,850

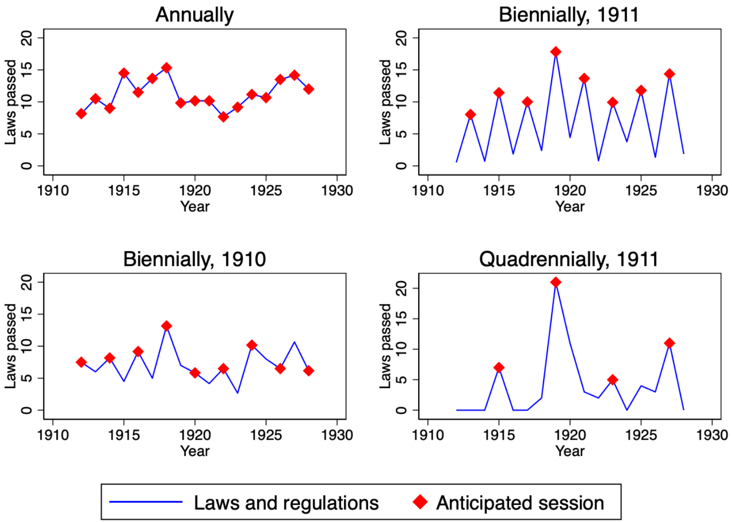
Data

Table: Summary statistics of cumulative laws by topic

Topic	Mean	S.D.	Min	Max	N
Dairy and agriculture	8.1	8.3	0.0	46.9	6,850
Dairy and sanitation	8.2	8.8	0.0	64.4	6,850
Venereal disease	4.1	3.2	0.0	16.2	6,850
Sewers and water	8.7	13.9	0.0	140.6	6,850
Communicable disease	7.2	7.8	0.0	57.9	6,850
Boards of health	18.8	18.1	0.0	111.3	6,850
Food and drug safety	5.9	7.4	0.0	60.3	6,850
Vital statistics	4.4	3.7	0.0	26.1	6,850
Housing	2.5	3.3	0.0	20.3	6,850
Work place	5.2	4.1	0.0	18.5	6,850

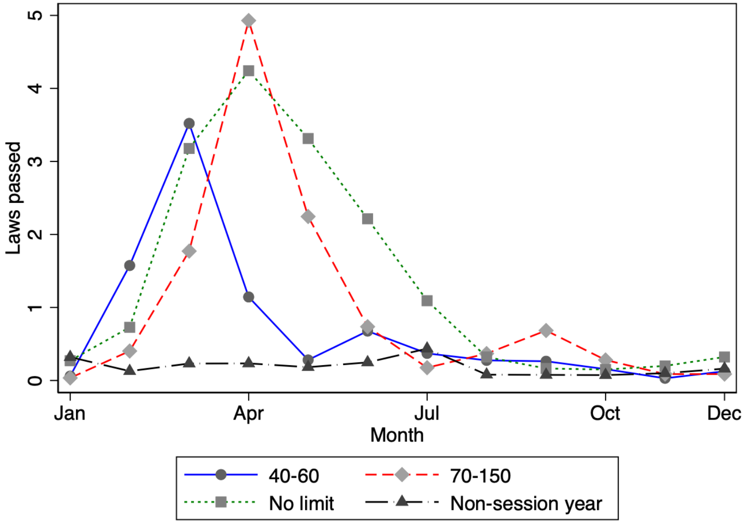
Identification strategy

Figure: Laws passed by session schedule

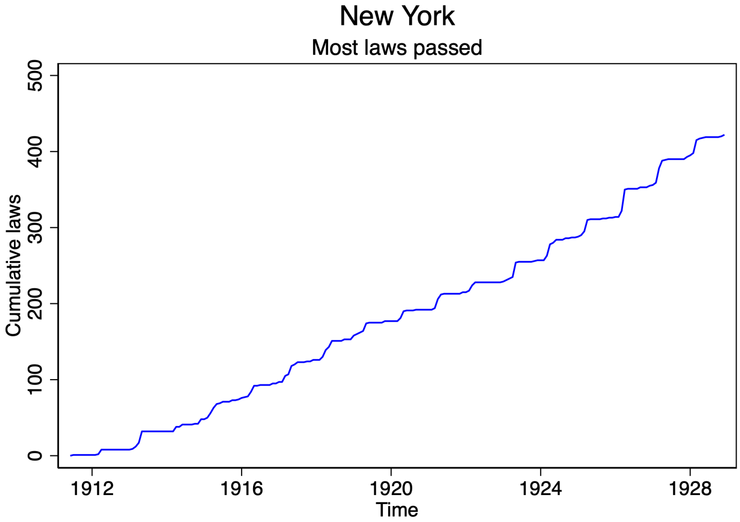


Identification strategy

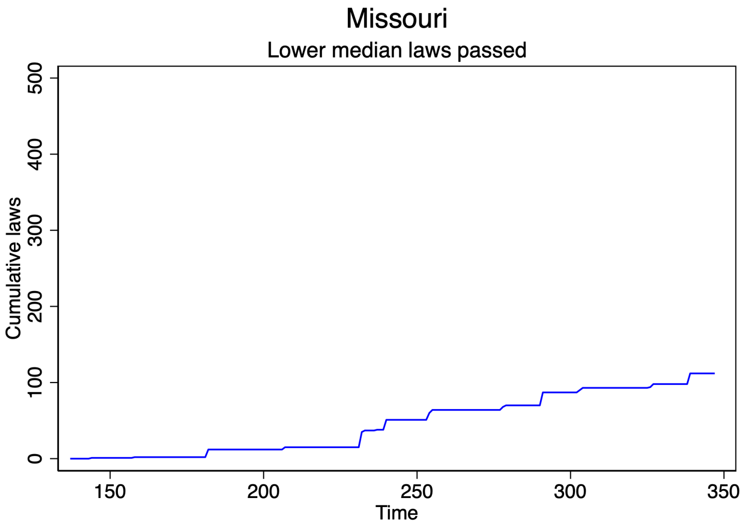
Figure: Laws passed by month and session limit



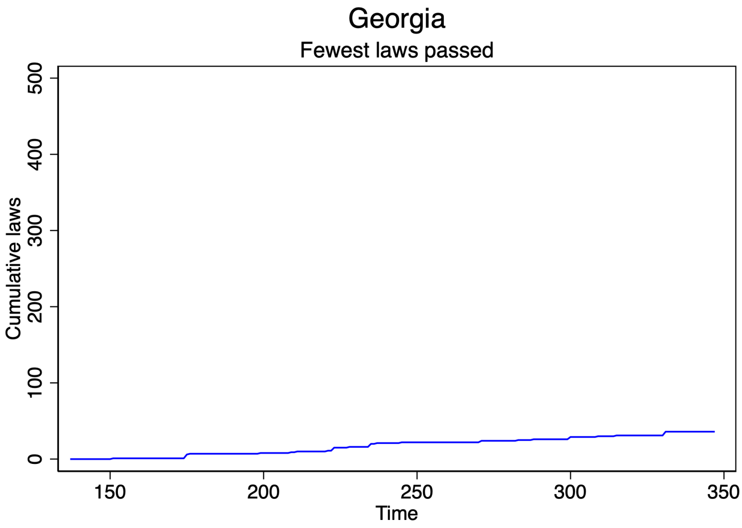
Identification strategy



Identification strategy



Identification strategy



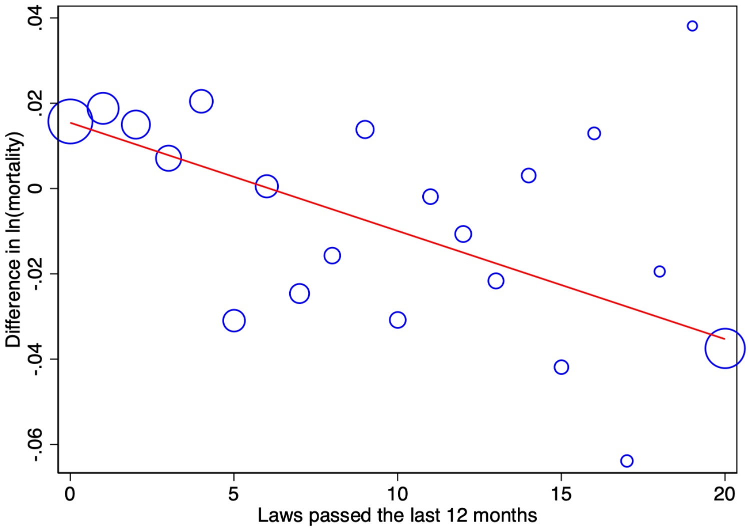
Identification strategy

- TWFE:

$$\ln(m_{st}) = \alpha_t + \beta_s + \delta Laws_{st} + \epsilon_{st}$$

- Panel IV:
  - Cumulative number of session months
  - Interacted with salary
- Event study version using Local Projection DiD

Results



Difference in ln(mortality)
.04
.02
0
-.02
-.04
-.06
0
5
10
15
20
Laws passed the last 12 months

Results

Table: First stage estimates

	(1)	(2)	(3)	(4)
Cumulative months in sessions	0.501	0.882**	0.493	0.026
	(0.488)	(0.422)	(0.491)	(0.564)
Cumulative sessions $\times$ salary	0.176***	0.154***	0.177***	0.222***
	(0.026)	(0.021)	(0.026)	(0.041)
F-stat of IVs	361.41	232.69	419.78	71.71
N	6,850	4,642	6,805	6,850
Balanced	N	Y	N	N
Lagged mortality	N	N	Y	N
Weighted	Y	Y	Y	N

Results

Table: OLS estimates of the effect of public health laws

	(1)	(2)	(3)	(4)
Cumulative laws	-0.000568***	-0.000535***	-0.000313***	-0.000535**
	(0.0000917)	(0.000107)	(0.0000495)	(0.000215)
Balanced	N	Y	N	N
Lagged mortality	N	N	Y	N
Weighted	Y	Y	Y	N

Results

Table: Second stage estimates

	(1)	(2)	(3)	(4)
Cumulative laws	-0.000713***	-0.000721***	-0.000746***	-0.000535**
	(0.000121)	(0.000149)	(0.000180)	(0.000215)
Balanced	N	Y	N	N
Lagged mortality	N	N	Y	N
Weighted	Y	Y	Y	N

Results

Table: Effects of laws classified by dictionary methods

	(1)	(2)	(3)	(4)
Cumulative non-repeal laws	-0.000668***		-0.000664***	
	(0.000101)		(0.000133)	
Cumulative repeal laws		-0.00183*	-0.0000271	
		(0.000912)	(0.000707)	
Cumulative laws with punishments				-0.000842
				(0.000922)
N	6,850	6,850	6,850	6,850

Results

Are more prolific legislatures more effective?

$$\text{Prolificness}_s = \frac{Laws_{s,\text{Dec 1928}} - \min_{s'}\left(Laws_{s',\text{Dec 1928}}\right)}{\max_{s'}\left(Laws_{s',\text{Dec 1928}}\right) - \min_{s'}\left(Laws_{s',\text{Dec 1928}}\right)}$$

Results

Table: Measuring laws by word count

	(1)	(2)
Cumulative word count	-0.000347***	0.0000374
	(0.0000873)	(0.000186)
Prolificness $\times$ word count		-0.000491**
		(0.000195)
N	6,850	6,850

Results: What about other laws?

Table: Controlling for other laws

	(1)	(2)	(3)	(4)
Panel A: Ordinary Least Squares				
Standardize public health laws	-0.0345***	-0.0279*	-0.0332***	-0.0251
	(0.00564)	(0.0160)	(0.00482)	(0.0165)
Standardize all law pages		-0.00771		-0.00929
		(0.0140)		(0.0148)
Women's suffrage			-0.0597***	-0.0590***
			(0.0169)	(0.0171)
Prohibition			-0.0264**	-0.0302**
			(0.0123)	(0.0124)

Results: What about other laws?

Table: Controlling for other laws

	(1)	(2)	(3)	(4)
	Panel B: Instrumental Variables			
Standardize public health laws	-0.0429***	-0.0641***	-0.0420***	-0.0656***
	(0.00749)	(0.0181)	(0.00687)	(0.0175)
Standardize all law pages		0.0196		0.0213*
		(0.0130)		(0.0120)
Women's suffrage			-0.0600***	-0.0618***
			(0.0161)	(0.0168)
Prohibition			-0.0222*	-0.0113
			(0.0134)	(0.0162)
N	556	556	556	556

Topic modeling results

Table: Two-way fixed effects estimates of topics

	Coef. (1)	p-value (2)
Dairy and agriculture	0.003	0.026
Dairy and sanitation	0.000	0.773
Venereal disease	0.006	0.118
Sewers and water	-0.000	0.574
Communicable disease	0.003	0.184
Boards of health	-0.001	0.466
Food and drug safety	-0.005	0.106
Vital statistics	-0.011	0.013
Housing	-0.007	0.211
Work place	0.004	0.111

Table: Two-way fixed effects estimates of grouped topics

	Coef. (1)	p-value (2)
Food, milk, and drug safety	0.001	0.578
Sanitation infrastructure	-0.001	0.045
Communicable/STI	0.001	0.026
Institutions/state capacity	-0.002	0.038

Table: Two-way fixed effects estimates for causes of death

	OLS		IV	
	Coef.	p-value	Coef.	p-value
	(1)	(2)	(3)	(4)
Typhoid	-0.001	0.161	-0.001	0.549
Malaria	-0.002	0.226	-0.002	0.127
Smallpox	-0.002	0.958	-0.003	0.846
Measles	-0.002	0.086	-0.002	0.242
Scarlet fever	-0.003	0.300	-0.003	0.065
Whooping cough	-0.000	0.221	-0.000	0.654
Diphtheria	-0.001	0.206	-0.001	0.242
Influenza	-0.003	0.072	-0.003	0.009
Meningitis	0.002	0.958	0.002	0.846
Diabetes	-0.001	0.161	-0.001	0.435
Circulatory	-0.000	0.943	-0.000	0.654
Pneumonia	-0.001	0.149	-0.001	0.005
Diarrhea (under 2)	-0.003	0.086	-0.003	0.077
Brights nephritis	-0.002	0.008	-0.002	0.001
Suicide	-0.000	0.958	-0.000	0.841
Tuberculosis	-0.001	0.161	-0.001	0.024
Cancer/tumor	-0.000	0.161	-0.000	0.236
Childbirth	-0.000	0.206	-0.000	0.435
Accidents/violence	-0.000	0.161	-0.000	0.435
Infant mortality	-0.001	0.039	-0.002	0.024

Table: Two-way fixed effects estimates for grouped causes of death

	OLS		IV	
	Coef.	p-value	Coef.	p-value
	(1)	(2)	(3)	(4)
Waterborne/drainage	-0.0024	0.0030	-0.0026	0.0230
Airborne	-0.0013	0.0010	-0.0015	0.0010
Noncommunicable	-0.0004	0.0030	-0.0005	0.0230

Table: Two-way fixed effects estimates of grouped topics

	Water/drainage (1)	Airborne (2)	Noncommunicable (3)
Food, milk, and drug safety	0.006	0.000	0.000
	(0.611)	(0.821)	(0.821)
Sanitation infrastructure	–0.002	–0.002	–0.001
	(0.793)	(0.120)	(0.742)
Communicable disease	–0.005	0.000	0.001
	(0.276)	(0.179)	(0.179)
Institutions/state capacity	–0.010	–0.003	–0.001
	(0.039)	(0.053)	(0.202)

Local Projection DiD

$$\begin{aligned}\ln(m_{s,t+h}) - \ln(m_{s,t}) &= \beta^h \Delta Laws_{s,t} + \delta_t^h + \\ &\quad \gamma_1^h \ln(m_{s,t-1}) + \gamma_2^h \ln(m_{s,t-2}) + e_{st}^h\end{aligned}$$

- Treatment group: $Laws_{s,t} = Laws_{s,t+h}$ and $Laws_{s,t-1} = Lawss, t-5$
- "Clean" control group: $Laws_{s,t+h} = Laws_{s,t-5}$
- Dube et. al (2022)
- Allows for treatment effects to be heterogeneous.
- Assumes dynamic effects stabilize after 5 months.

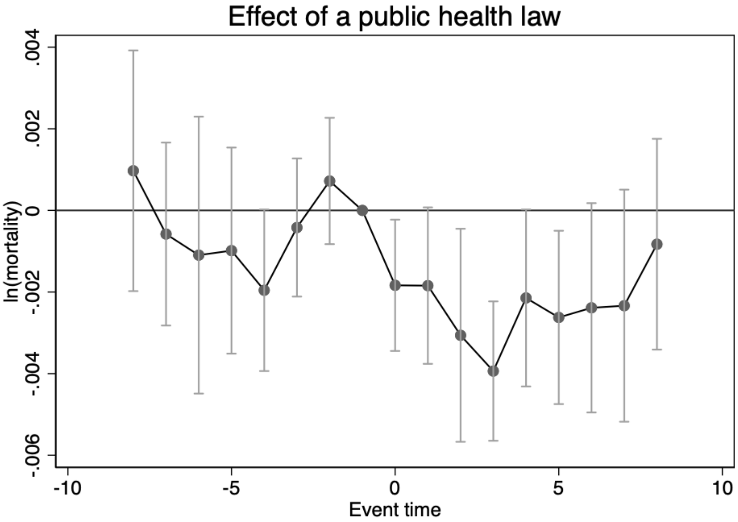
Local Projection DiD

$$\ln(m_{s,t+h}) - \ln(m_{s,t}) = \sum_{j=1}^{10} \beta_j^h \Delta Topic_{j,s,t} + \delta_t^h + \gamma_1^h \ln(m_{s,t-1}) + \gamma_2^h \ln(m_{s,t-2}) + e_{st}^h$$

- $Topic_{j,s,t}$ is the cumulative number of laws in Topic j.

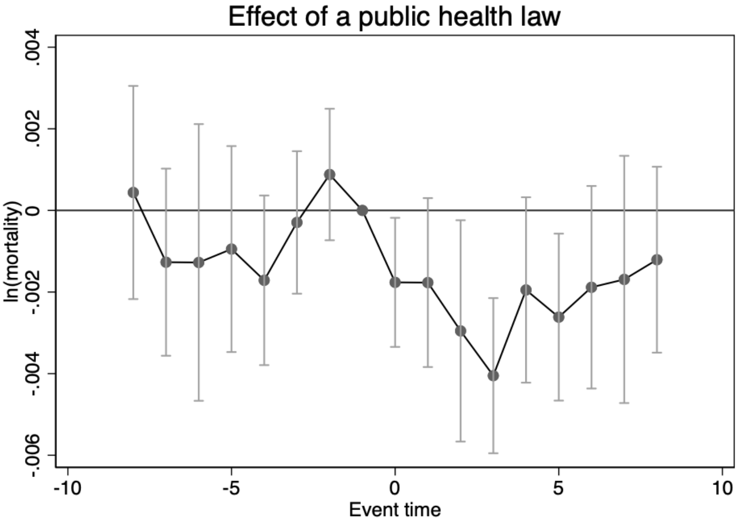
LP-DiD Results

Figure: Event study of Public Health Laws

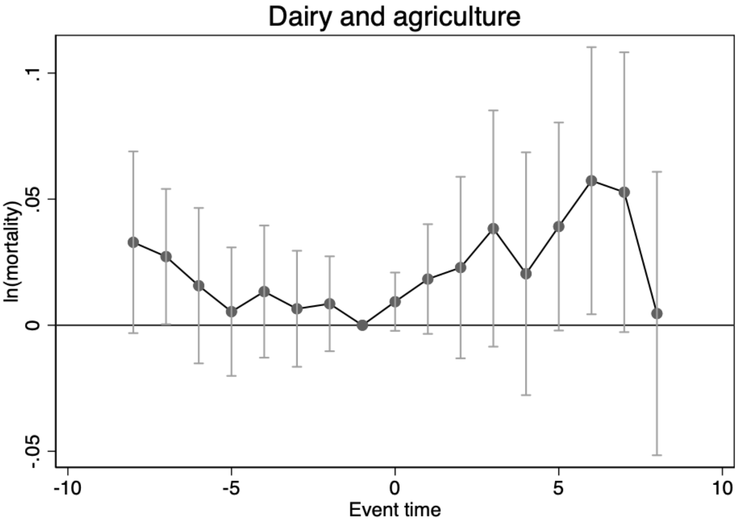


LP-DiD Results

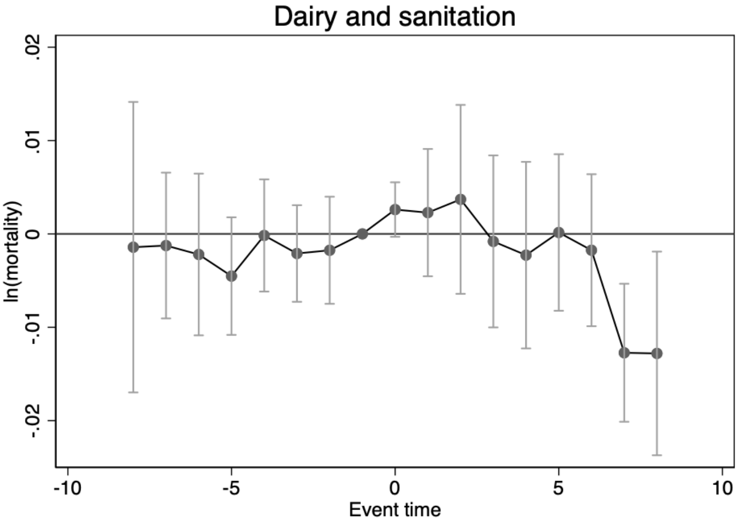
Figure: Laws effectively immediately



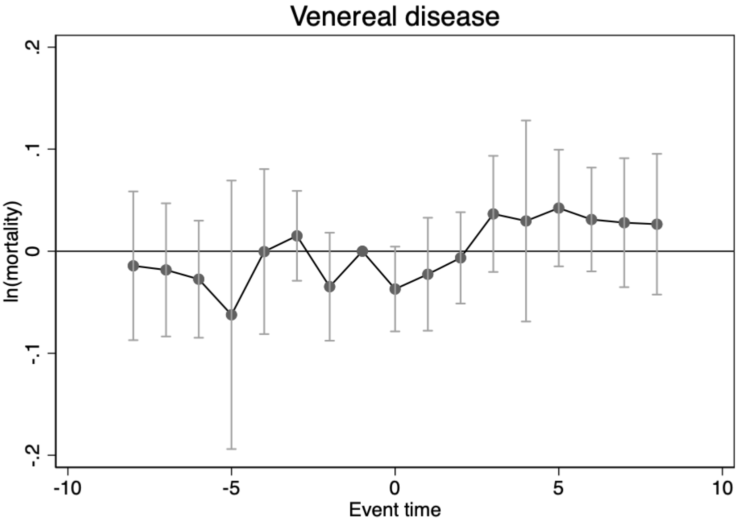
LP-DiD Results



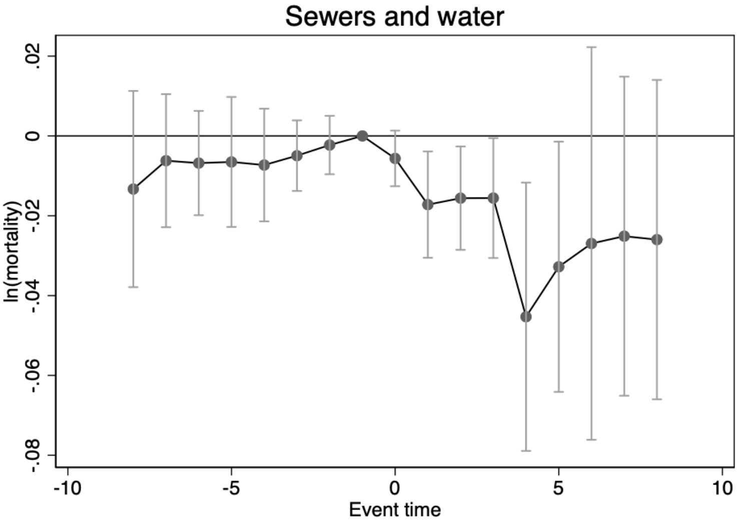
LP-DiD Results



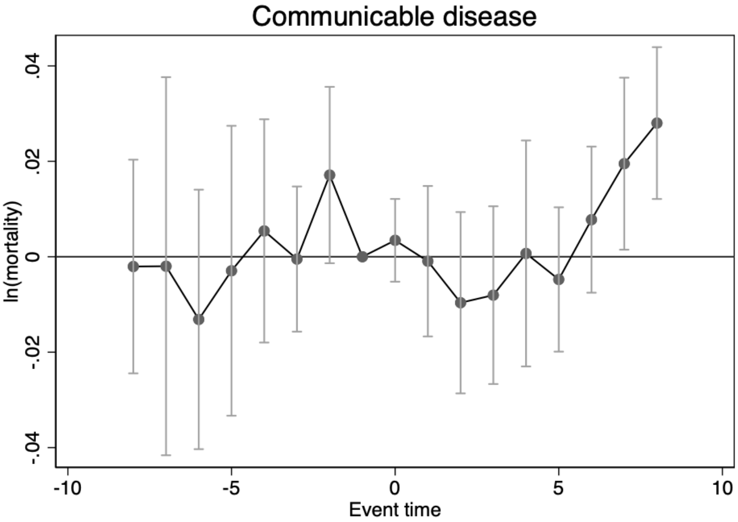
LP-DiD Results



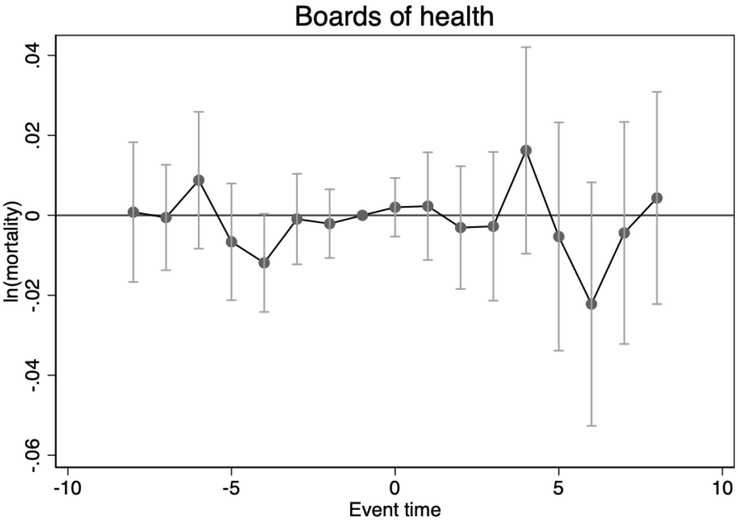
LP-DiD Results



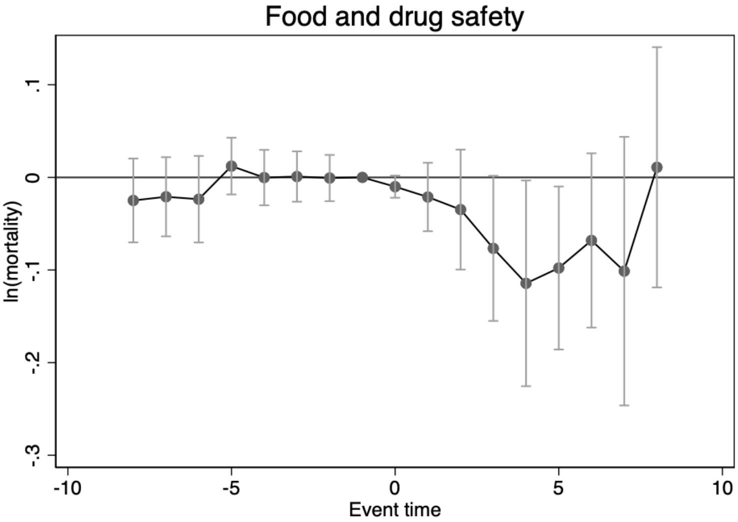
LP-DiD Results



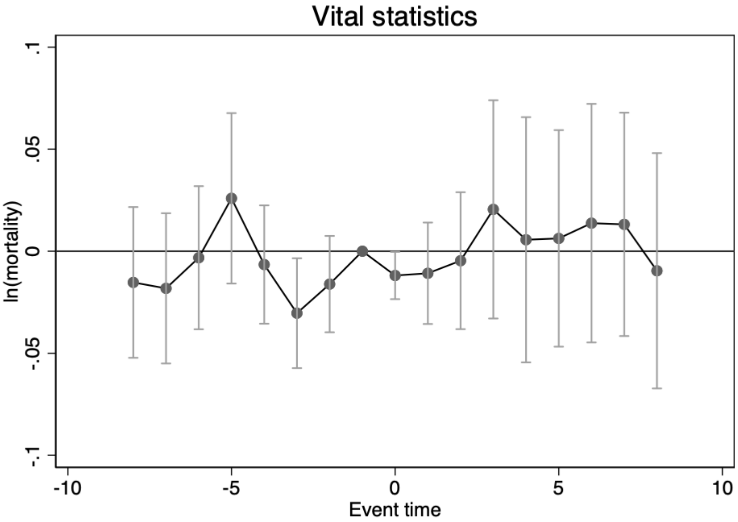
LP-DiD Results



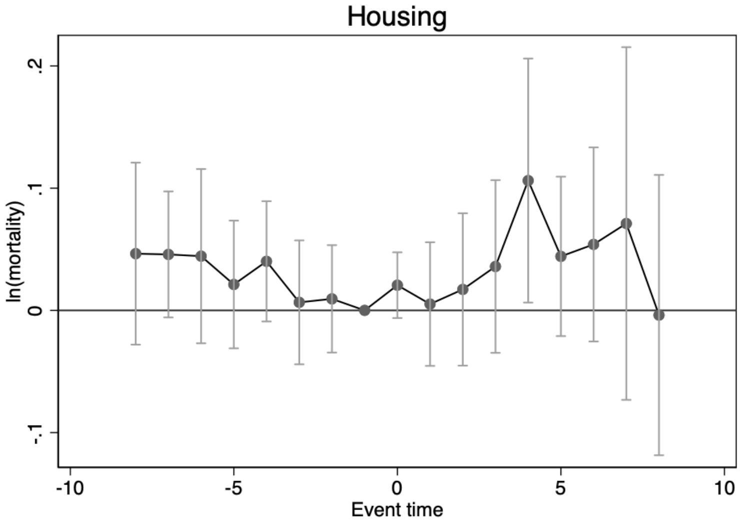
LP-DiD Results



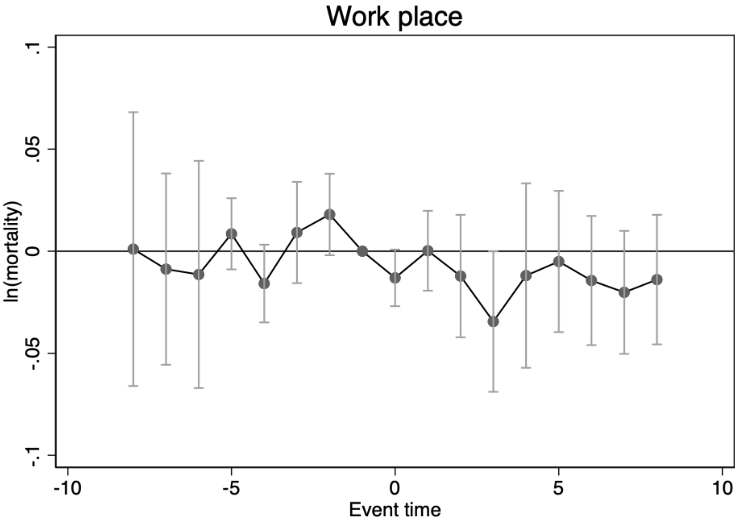
LP-DiD Results



LP-DiD Results



LP-DiD Results



Can laws explain the mortality transition?

- Collapse the balanced panel into a time series.
- Normalize time to a unit interval $t' \in [0, 1]$:

$$t' = \frac{t-1}{T-1}, \quad t \in \{1, 2, \ldots, T\}$$

- Estimate the regression:

$$\ln(m_t) = \alpha_{month} + \beta_1 1[t \in Wave1] + \beta_2 1[t \in Wave2] + \delta t' + \epsilon_t$$

  where $Wave1$ and $Wave2$ indicate waves of influenza.
- Estimate "no mortality transition" counterfactual:

$$\ln(m_t^{\text{NMT}}) = \ln(m_t) - \hat{\delta} t'$$

Can laws explain the mortality transition?

- Generate a no law counterfactual mortality:

$$\ln(m_t^{\mathrm{NL}}) = \ln(m_t) + 0.0005 \times Laws_t$$

- Estimate the regression:

$$\begin{aligned}\ln(m_t^{\mathrm{NL}}) &= \alpha_{month}^{\mathrm{NL}} + \beta_1^{\mathrm{NL}} 1[t \in Wave1] + \beta_2^{\mathrm{NL}} 1[t \in Wave2] + \\ &\quad \delta^{\mathrm{NL}} t' + \epsilon_t\end{aligned}$$

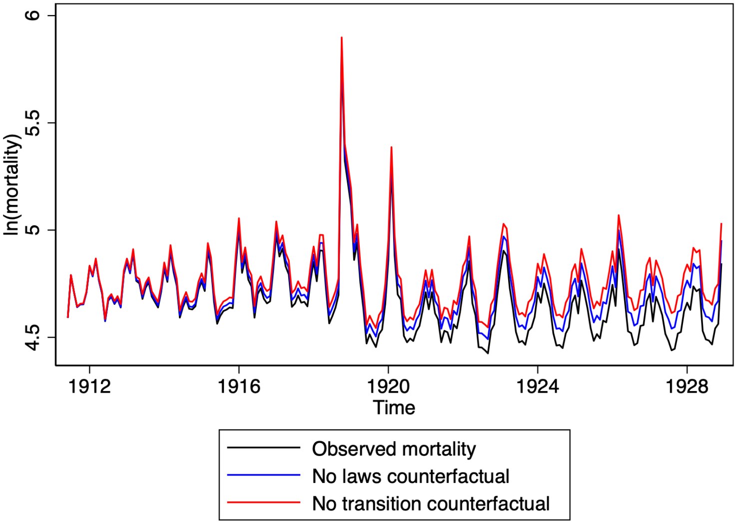
- $\hat{\delta}^{NL}$ is an estimate of how much mortality would have declined if no laws were passed.

Can laws explain the mortality transition?

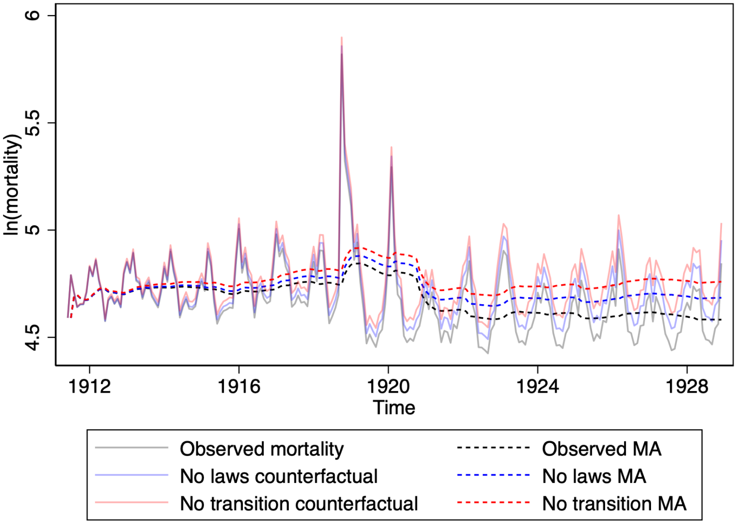
Table: The log mortality decline June 2011 to Dec 1928

	Observed	No law counterfactual
Unit time	-0.188***	-0.0737***
	(0.0147)	(0.0147)
R^2	0.708	0.676
N	211	211

Can laws explain the mortality transition?



Can laws explain the mortality transition?



Conclusion

- Preliminary estimates suggest that public health laws reduced mortality by approximately 10%.
- Laws worthy of more study:
  - Housing/tenements
  - Drug safety