Applications of the difference-in-differences approach

Volha Lazuka

vola@sam.sdu.dk

LUPOP seminar, 07/10/2021

Common sense seems to dictate that it is almost impossible to estimate causal effects of antibiotics, maternity wards, medical procedures, etc.

Yet, the difference-in-differences method can help to obtain those.

Outline of the talk

1. The canonical difference-in-differences (DD) approach
2. DD with a continuous treatment: Lazuka, 2020
3. DD with variation in treatment timing: Lazuka, forthcoming
4. Triple-differences: Lazuka, WP

"Credibility revolution" in economics

Currie et al. (2020) have documented the boom of experimental and quasi-experimental methods:

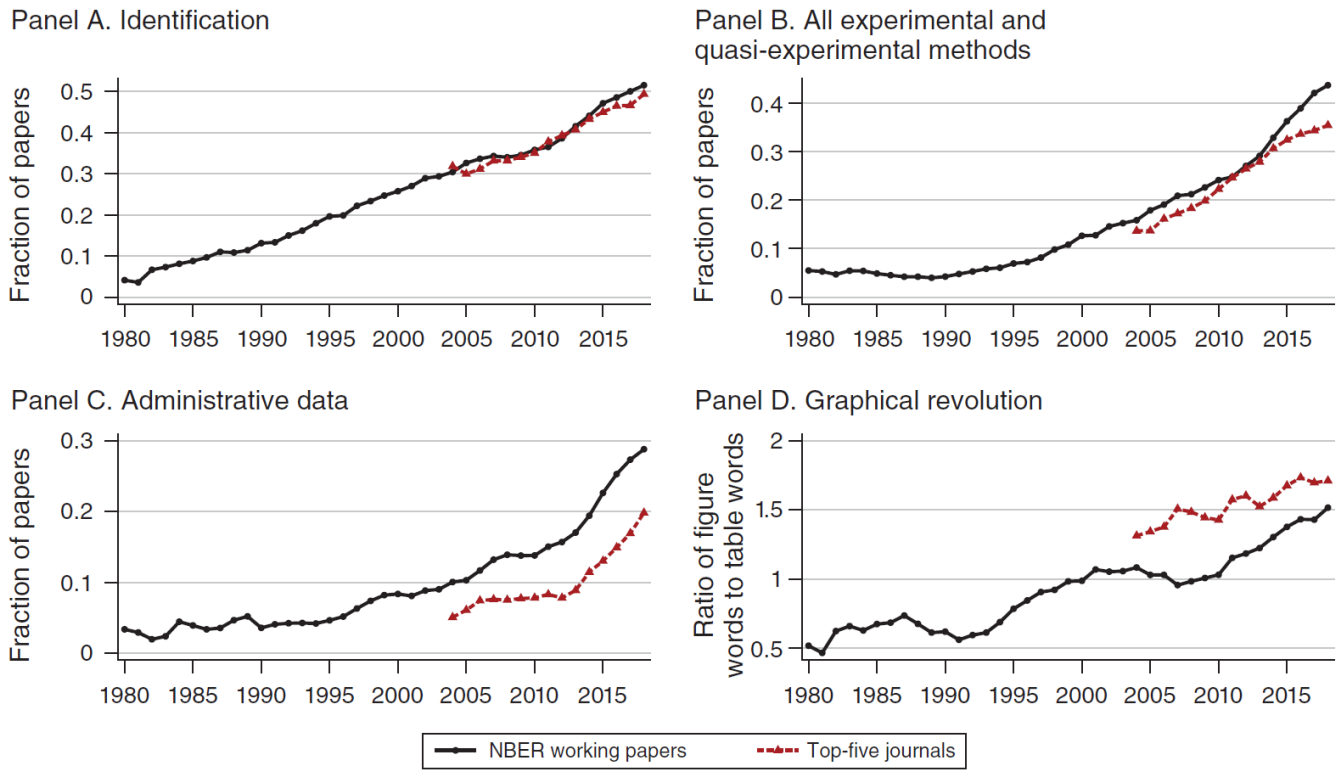


FIGURE 2. THE CREDIBILITY REVOLUTION

Notes: This figure shows different dimensions of the "credibility revolution" in economics: identification (panel A), all experimental and quasi-experimental methods (panel B), administrative data (panel C), and the graphical revolution (panel D). Panel D shows the ratio of the number of "figure" terms to the number of "table" terms mentioned. See Table A.I for a list of terms. The series show five-year moving averages.

"Credibility revolution" in economics

Main "research designs" by Anqrist and Pischke (2009):

- Randomized control trial
- Selection on observables
- Instrumental variables
- **DD**
- Regression discontinuity

"Credibility revolution" in economics

Currie et al. (2020):

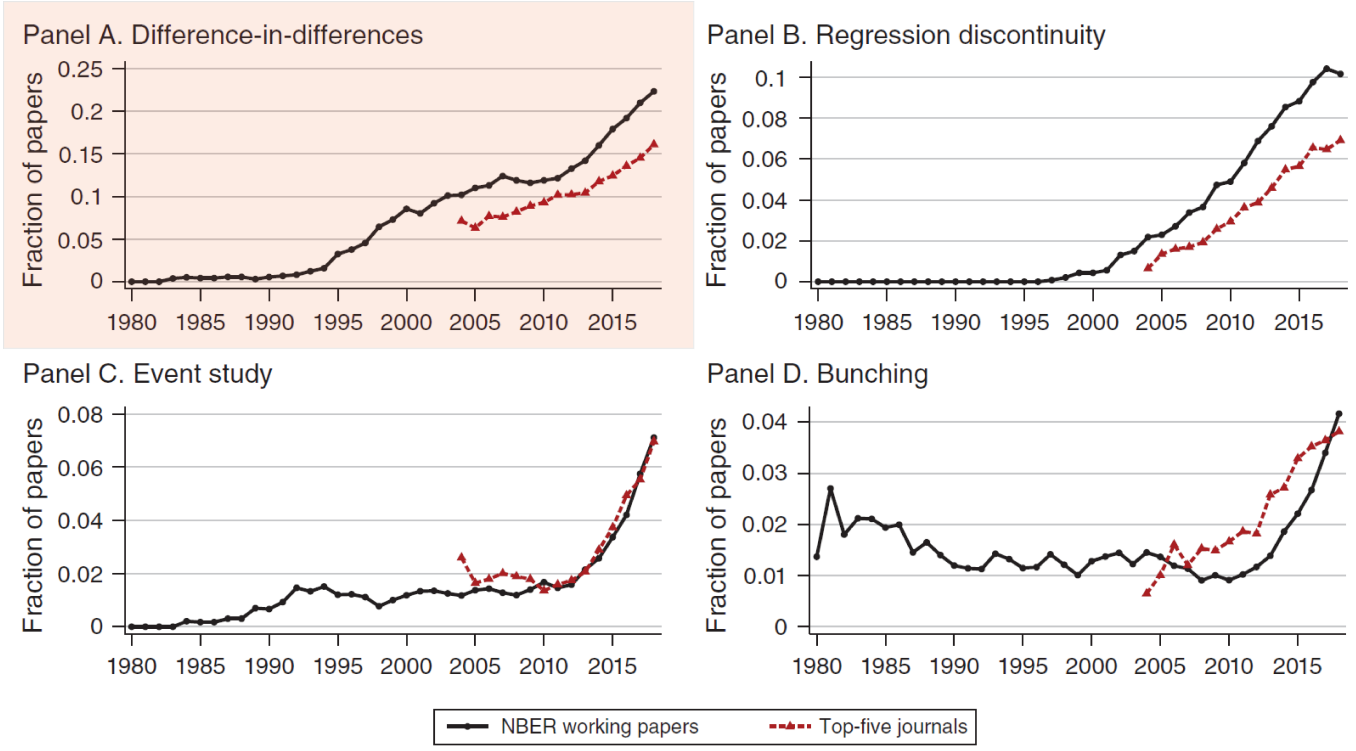


FIGURE 4. QUASI-EXPERIMENTAL METHODS

Notes: This figure shows the fraction of papers referring to each type of quasi-experimental approach. See Table A.I for a list of terms. The series show five-year moving averages.

Canonical DD

- DD methods exploit variation in time (Pre vs Post) and across groups (Treated vs Control) to recover causal effects of interest.
- Pre vs Post comparisons:
  Compares the same groups of units before and after reform.
  Limitation: do not account for potential trends in outcomes.
- Treated vs Control comparisons:
  Compares units to those who have not experienced treatment.
  Limitation: do not account for selection into treatment group.
- DD allows the researcher to avoid both limitations.

Canonical DD

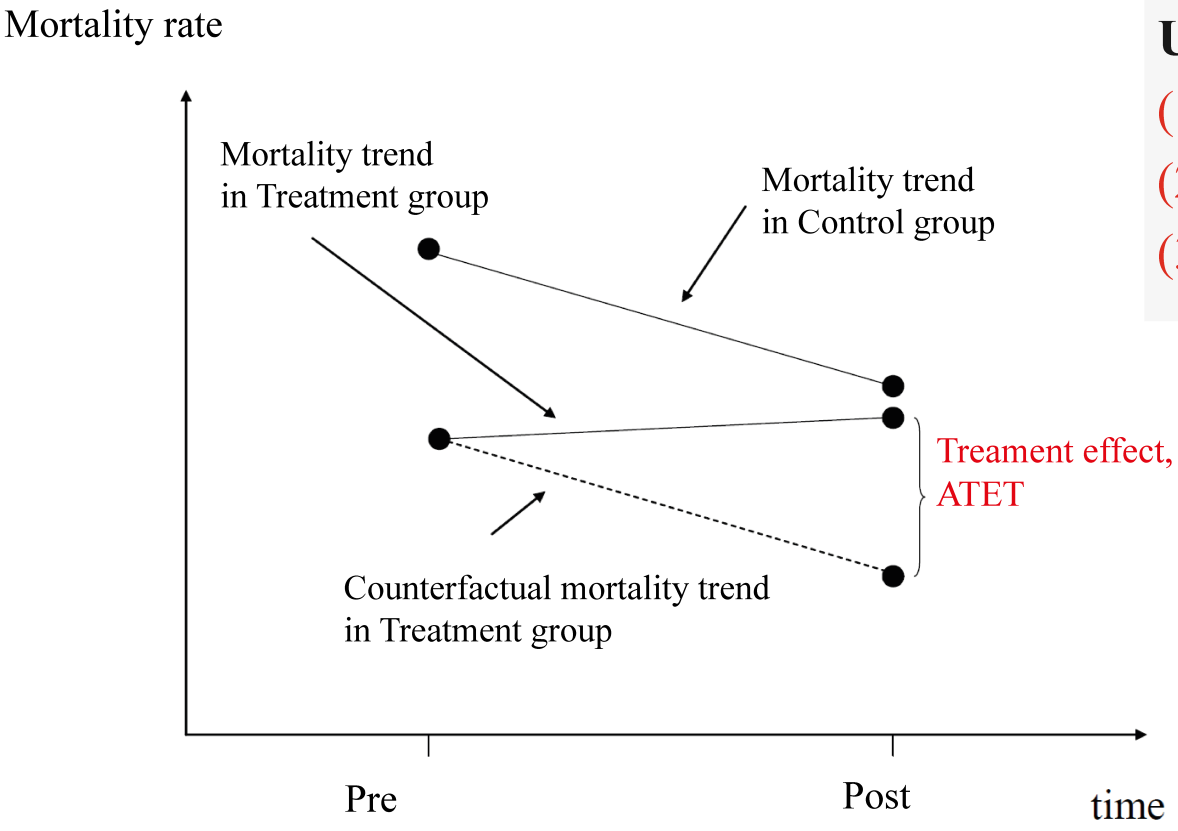
The canonical 2 x 2 DD estimator:

$$\texttt{DD} = (\bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{PRE}}) - (\bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{PRE}})$$

or obtain from the regression:

$$\texttt{y}_{\texttt{it}} = \alpha_{\texttt{i}} + \alpha_{\texttt{t}} + \beta^{\texttt{DD}}\texttt{POST}_{\texttt{t}} \times \texttt{TREAT}_{\texttt{i}} + \varepsilon_{\texttt{it}}$$

Causal effects in DD



Under the assumptions:

(1) Stable Unit Treatment Value;

(2) No anticipation;

(3) Parallel trends.

The rest of the talk:

Applications of DD in research of the impacts of health reforms in Sweden:

1. Lazuka (2020) exploits continuous variation in $\texttt{TREAT}_i$ based on pre-treatment characteristics: **the impact of sulpha antibiotics.**

2. Lazuka, forthcoming exploits variation in $\texttt{POST}_t$ based on staggered rollout of the reform: **the impact of maternity wards' reform.**

3. Lazuka WP exploits variation in $\texttt{POST}_t \times \texttt{TREAT}_i$ based on groups affected with different intensity: **the impact of medical innovations.**

(1) The impact of sulpha antibiotics

Infant Health and Later-Life Labour Market Outcomes: Evidence from the Introduction of Sulpha Antibiotics in Sweden

Journal of Human Resources, 2020, 55(2), pp.660-98

Volha Lazuka

Methods: DD approach: it compared outcomes of individuals born in regions with high baseline pneumonia mortality rate versus those born in regions with low rates, before and after the arrival of sulpha antibiotics.

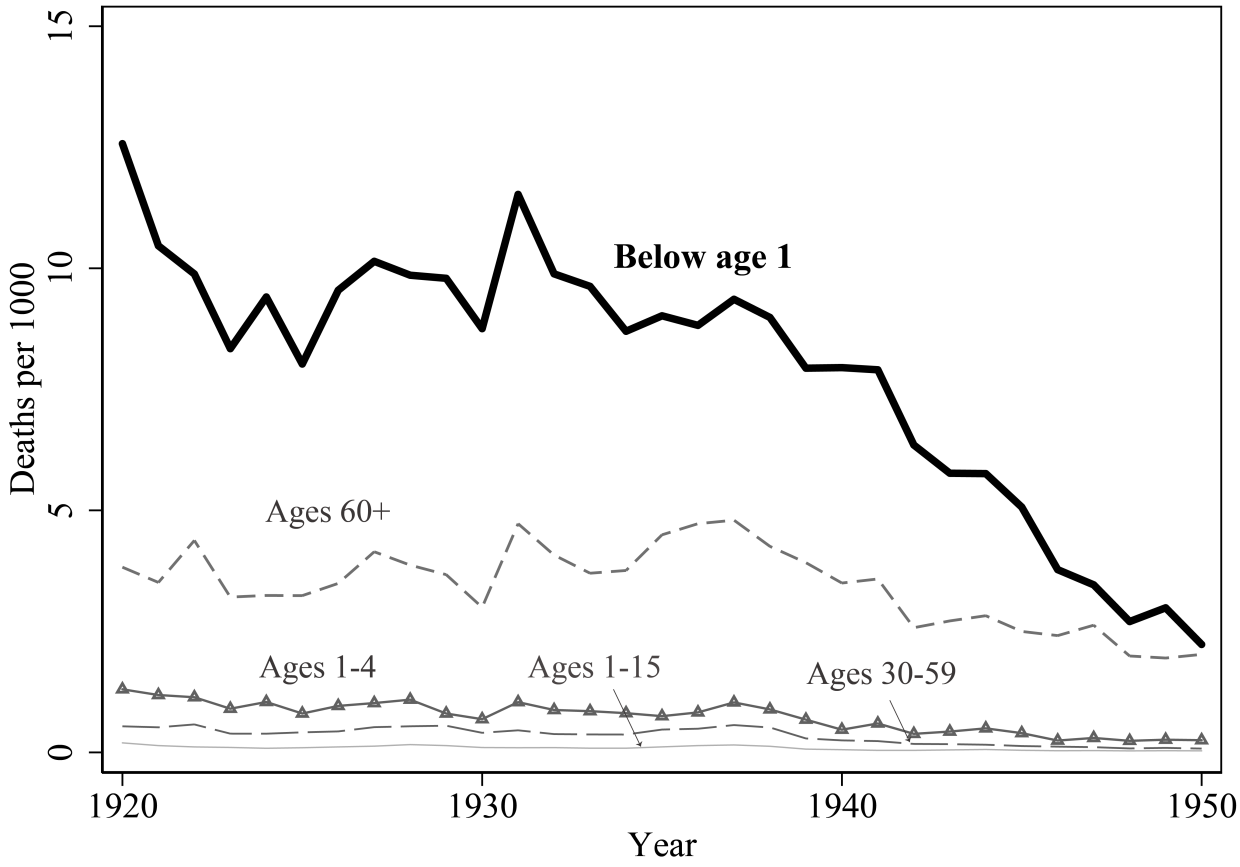
Findings: Mitigation of pneumonia infection in infancy increased labour income in late adulthood by 2.8–5.1 percent. The beneficial effects are strong for health and weaker for years of schooling.

Motivation

- An expanding economic literature finds substantial effects of early-life disease environment on later-life outcomes:
  - Disease outbreaks in early life, e.g. Case and Paxson (2009), Bengtsson and Lindström (2013).
  - Early-life interventions against specific infectious diseases, such as Bleakley (2007) for hookworm infection, Cutler et al (2010) for malaria, Bhalotra & Venkataramani (2011) for pneumonia, Beach et al (2016) for typhoid fever, and Adhvaryu et al (2018) for goitre.
- There is a limited evidence for Europe and Sweden in particular.

Pneumonia mortality rate

In 1939, sulpha antibiotics became available throughout Sweden and launched the decline in pneumonia.



Recall:

The canonical 2 x 2 DD estimator:

$$\texttt{DD} = (\bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{PRE}}) - (\bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{PRE}})$$

or obtain from the regression:

$$\texttt{y}_{\texttt{it}} = \alpha_{\texttt{i}} + \alpha_{\texttt{t}} + \beta^{\texttt{DD}}\texttt{POST}_{\texttt{t}} \times \texttt{TREAT}_{\texttt{i}} + \varepsilon_{\texttt{it}}$$

Methods in Lazuka (2020)

A DD regression equation:

$$\mathtt{y}_{\mathtt{irt}} = \alpha_{\mathtt{r}} + \alpha_{\mathtt{t}} + \beta^{\mathtt{DD}}\mathtt{POST}_{\mathtt{t}} \times \mathtt{BaseRate}_{\mathtt{r}} + \varepsilon_{\mathtt{irt}}$$

where $\mathtt{POST}_{\mathtt{t}}$ equals 1 for cohorts born in 1939-1943, 0 if born in 1934-1938,

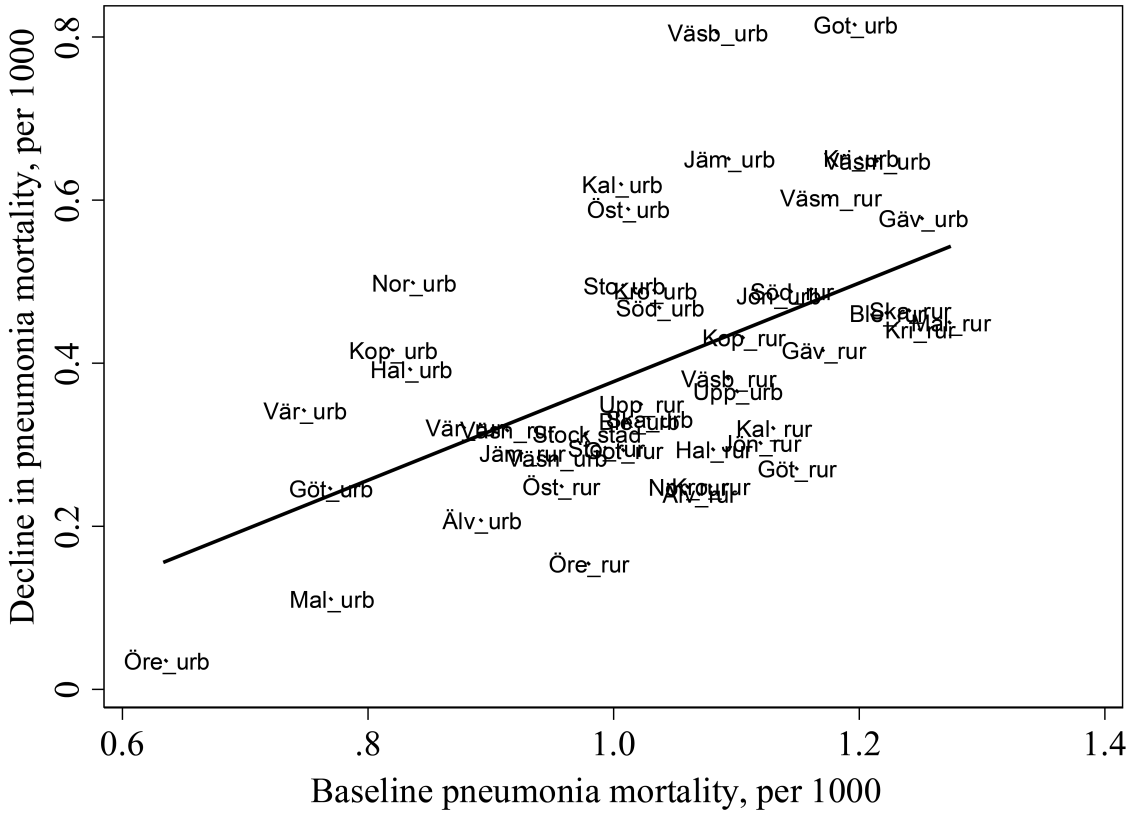
$\mathtt{BaseRate}_{\mathtt{r}}$ is pre-intervention pneumonia mortality in a region of birth $\mathtt{r}$ (i.e. approximates the individual's infection status),

$\alpha_{\mathtt{r}}$ are region-of-birth dummies; $\alpha_{\mathtt{t}}$ are year-of-birth dummies,

$\mathtt{y}_{\mathtt{irt}}$ are the individual's outcomes in adulthood.

Regional convergence in pneumonia

How realistic is the assumption that regions with high infection rates benefited more from the arrival of antibiotics?

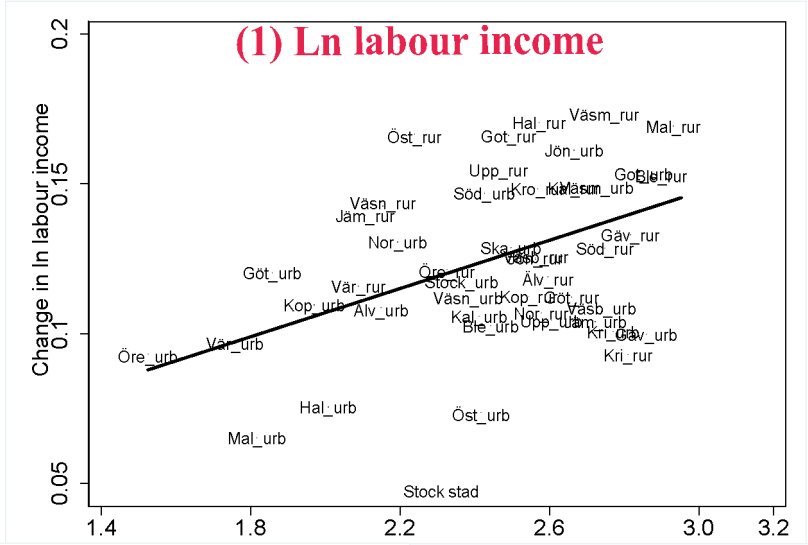


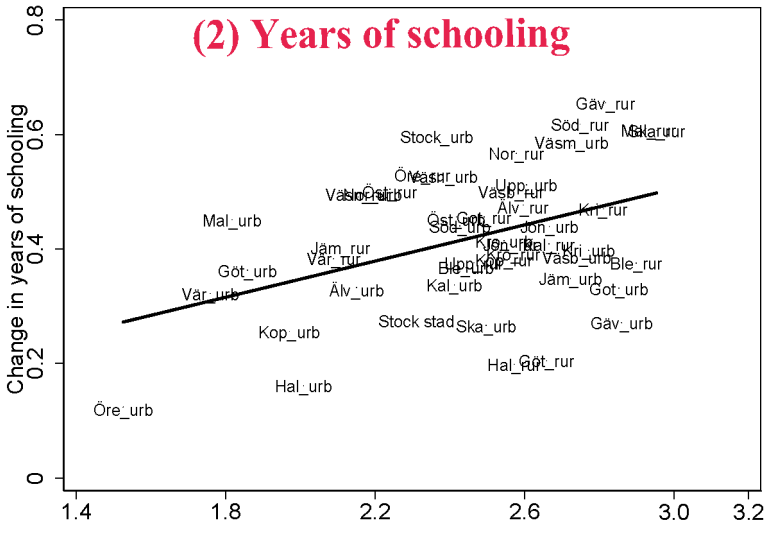
Results: the estimates of β^{DD}

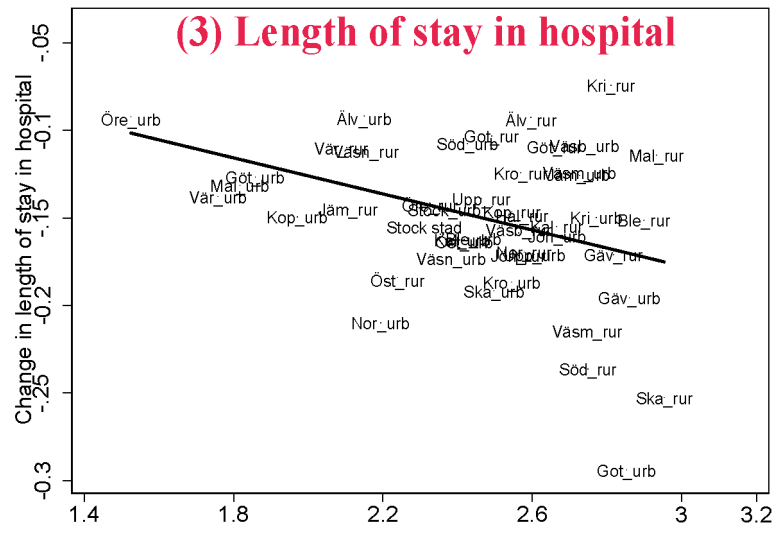
y_{it}: Ln labour income	
$POST_t \times BaseRate_r$	0.043^{***}
	(0.015)
Pre-mean	8.063
y_{it}: Years of schooling	
$POST_t \times BaseRate_r$	0.148^{**}
	(0.055)
Pre-mean	9.271
y_{it}: Length of stay in hospital	
$POST_t \times BaseRate_r$	-0.042^{***}
	(0.013)
Pre-mean	0.770

Note: $BaseRate_r$ is normalized.
$^{***}p<0.01, ^{**}p<0.05, ^{*}p<0.1$.

Results: DD







Baseline pneumonia mortality, per 1000 normalised

Assumptions in DD with a continuous treatment?

Callaway, Goodman-Bacon, and Sant'Anna 2021

- To identify the causal effects in DD with a continuous treatment, one has to impose a Strong Parallel Trends Assumption:

$$\mathbb{E}[\mathtt{Y_t(d)} - \mathtt{Y_{t-1}(0)}] = \mathbb{E}[\mathtt{Y_t(d)} - \mathtt{Y_{t-1}(0)} \mid \mathtt{D} = \mathtt{d}]$$

  where $\mathtt{d}$ is a treatment intensity (dose).

- For all doses, the average change in outcomes over time across units if they had been assigned that amount of dose is the same as the average change in outcomes over time for all units that experienced that dose.

What about covariates?

Introducing covariates with DD

- By introducing controls, you impose the conditional parallel trends assumption.
- Controls must be unaffected by the treatment to avoid bias from "conditioning on a post-treatment variable" Rosenbaum, 1984
- TWFE with controls gives a biased estimate for the ATET (i.e., the effect of interest, Sant'Anna and Zhao, 2020).

NB: Use a proper estimator to estimate β^{DD}, such as the outcome regression, IPW, or a doubly robust DD estimator, following Sant'Anna and Zhao, 2020.

(2) The impact of maternity ward openings

It's a Long Walk: Lasting Effects of Maternity Ward Openings on Labour Market Performance

Review of Economics and Statistics
Forthcoming


The Review of Economics and Statistics

VOL. CIII | JULY 2021 | NUMBER 3

(Continued on back cover)

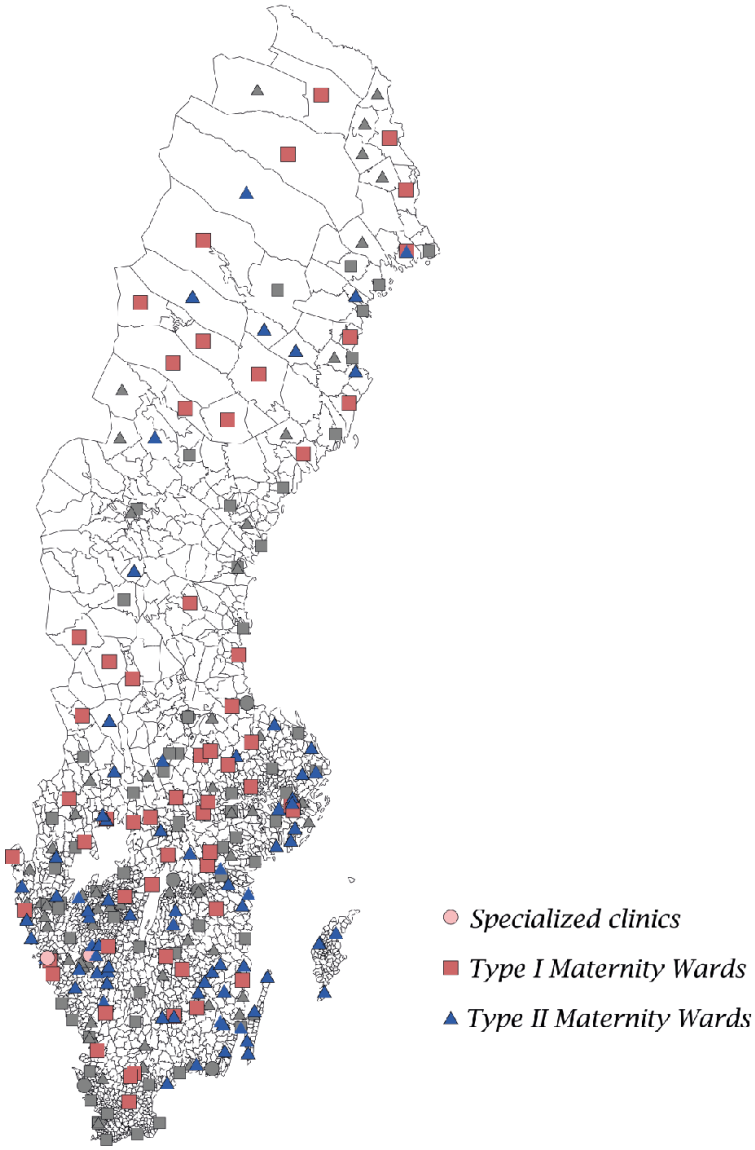
RECSA9
PUBLISHED FOR HARVARD UNIVERSITY BY THE MIT PRESS

Volha Lazuka

Methods: DD approach: it compared outcomes of individuals born in municipalities close to the newly opened maternity wards versus those born in municipalities located further, before and after the maternity ward opening.

Findings: It first finds that the reform substantially increased the share of hospital births and reduced early neonatal mortality. It then shows sizable long-term effects on labour income, unemployment, health-related disability and schooling.

New maternity wards



- In 1931-1946, 170 new maternity wards were opened throughout Sweden, and the share of hospital birth tripled to 90%.
- Expanded access to maternity wards is the most common health care intervention worldwide, but there is a very limited literature on on its potential benefits (e.g., Daysal et al. (2015), Lazuka (2018)).

Recall:

The canonical 2 x 2 DD estimator:

$$\texttt{DD} = (\bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{PRE}}) - (\bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{PRE}})$$

or obtain from the regression:

$$\texttt{y}_{\texttt{it}} = \alpha_{\texttt{i}} + \alpha_{\texttt{t}} + \beta^{\texttt{DD}}\texttt{POST}_{\texttt{t}} \times \texttt{TREAT}_{\texttt{i}} + \varepsilon_{\texttt{it}}$$

Methods in Lazuka, forthcoming

A DD regression equation with differential treatment timing:

$$y_{(i)rmt} = \alpha_m + \alpha_t + \beta^{DD} POST_{mt} \times MW_m + \gamma_{rb} + \varepsilon_{(i)mrt}$$

where $POST_{mt} \times MW_m \equiv MW_{mt}$ equals 1 for cohort t being born during or after the new MW was established $\leq$5.5 km away from the municipality of birth m: treatment turns on starting from different cohorts for each municipality.

α_m are municipality-of-birth dummies, α_t are year-of-birth dummies, γ_{rb} are county by level-of-urbanization by year-of-birth fixed effects.

y_{irmt} are the individual's outcomes in adulthood.

Interpretation of β^{DD} with differential treatment timing

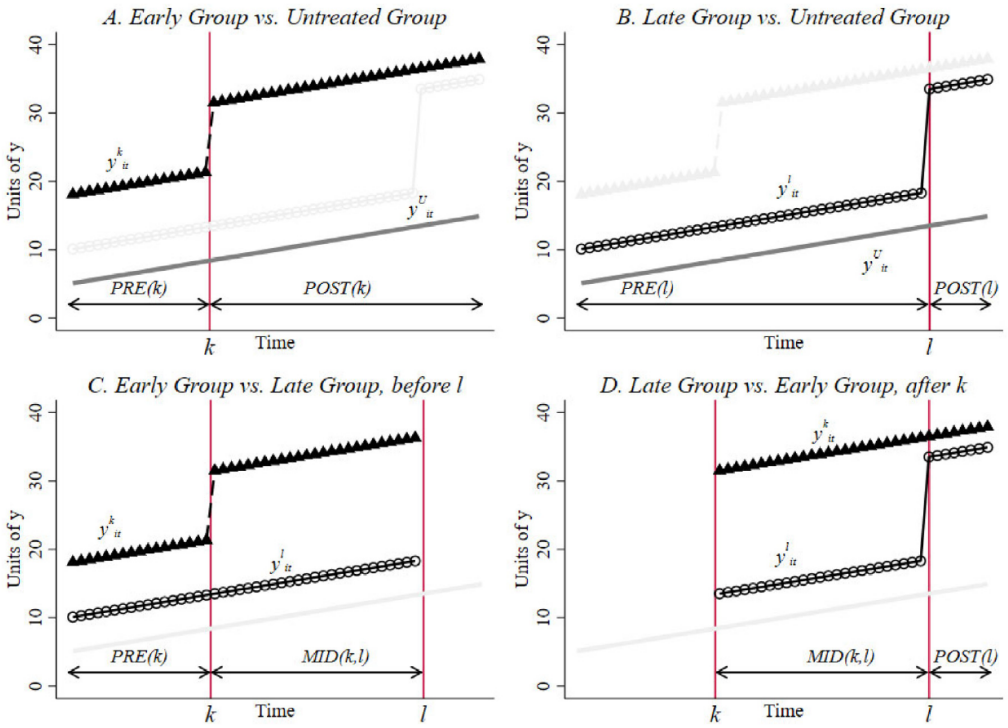


Fig. 2. The four simple (2x2) difference-in-differences estimates in the three group case. Notes: The figure plots outcomes for the subsamples that generate the four simple 2x2 difference-in-difference estimates in the three timing group case from Fig. 1. Each panel plots the data structure for one 2x2 DD. Panel A compares early treated units to untreated units ($\hat{\beta}^{DD}_{kU}$); panel B compares late treated units to untreated units ($\hat{\beta}^{DD}_{\ell U}$); panel C compares early treated units to late treated units during the late timing group's pre-period ($\hat{\beta}^{DD,k}_{k\ell}$); panel D compares late treated units to early treated units during the early timing group's post-period ($\hat{\beta}^{DD,\ell}_{k\ell}$). The treatment times mean that $\overline{D}_k = 0.67$ and $\overline{D}_\ell = 0.16$, so with equal group sizes, the decomposition weights on the 2x2 estimate from each panel are 0.365 for panel A, 0.222 for panel B, 0.278 for panel C, and 0.135 for panel D.

Goodman-Bacon (2021):

The OLS estimate, β^{DD}, in a two-way fixed-effects regression is a weighted average of all possible two-by-two DD estimators:

$$\hat{\beta}^{DD} = \sum_{k \neq U} s_{kU}\, \hat{\beta}^{2x2}_{kU} + \sum_{k \neq U} \sum_{\ell > k} \left[s^{k}_{k\ell}\, \hat{\beta}^{2x2,k}_{k\ell} + s^{\ell}_{k\ell}\, \hat{\beta}^{2x2,\ell}_{k\ell} \right]$$

Results: the estimates of β^{DD}

All treatment groups	Treated & never-treated	Treated & always-treated	Treated only
(1)	(2)	(3)	(4)
	y_{it}: 7-day mortality		
MW	-15.983***	-16.620***	-10.801**	-15.444***
	(3.831)	(3.908)	(4.793)	(5.635)
Pre-mean	28.375	28.542	24.643	24.244
	y_{it}: Ln labour income		
MW	0.042**	0.049***	0.045**	0.050**
	(0.018)	(0.018)	(0.020)	(0.020)
Pre-mean	7.939	7.875	8.002	7.864

Note: ***p<0.01,**p<0.05, *p<0.1.

What about heterogeneity/dynamics?

de Chaisemartin and D'Haultfoeuille, 2020

- In TWFE, weights attached to the treatment effects can be negative.
- β^{DD} from TWFE may not represent the causal effect of interest, ATET, unless we rule out treatment effect heterogeneity over time and across units (i.e.Treatment Effect Homogeneity is an additional assumption).

NB:

1. Conduct decomposition and heterogenuity tests for β^{DD} following de Chaisemartin and D'Haultfoeuille, 2020 and Goodman-Bacon, 2021.
2. Use alternative estimators to estimate β^{DD}, relying on "cohort-average treament effects" Callaway and Sant'Anna, 2020;, de Chaisemartin and D'Haultfoeuille. 2020 or on imputation Borusyak et al., 2021.

Example of the heterogeneity test in Lazuka, forthcoming

Following de Chaisemartin and D'Haultfoeuille, 2020:

- For the 7-day mortality, the negative weights only amount to -0.02, and the minimal standard deviation of ATET across the treated municipalities x cohorts required to revert the sign of β^{DD} is 393 deaths per 1,000 live births, a very large and implausible level of heterogeneity.
- For labor income, the same measure is 0.62 log units, another implausibly large level of heterogeneity.
- Assuming uniform distribution of treatment effects, for positive effects you may compare this standard deviation X sqrt(3) with the plausible value of β^{DD} (`twowayfeweights` in Stata).

(3) The impact of medical innovations

Heterogeneous Returns to Medical Innovations

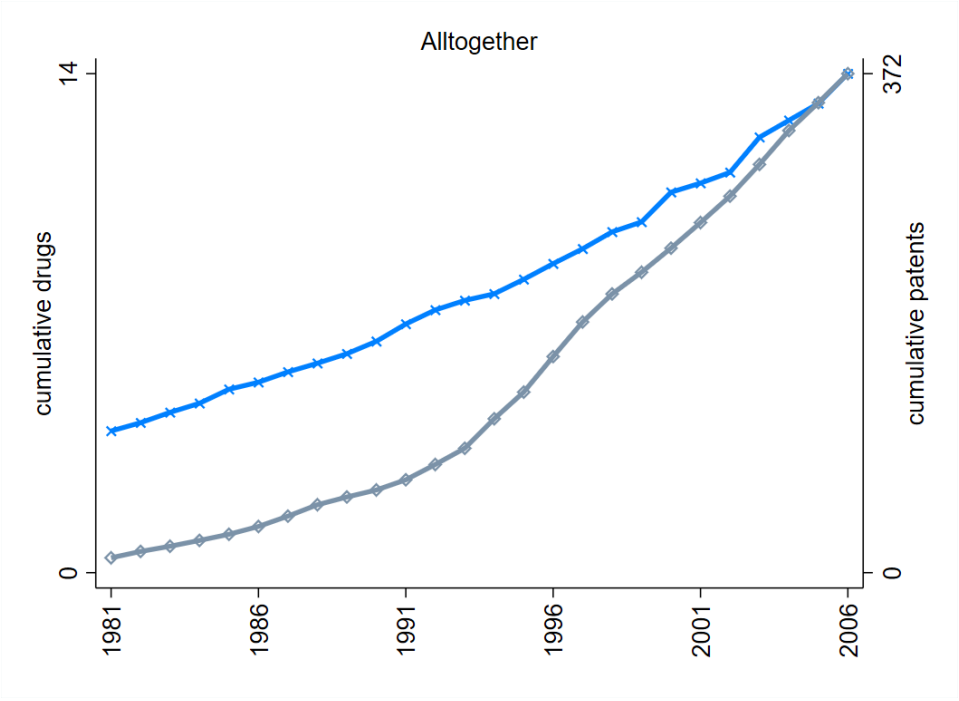
Working Paper

Volha Lazuka

Methods: DDD approach: it estimated the impact of medical innovation on economic outcomes as an innovation-induced reduction in economic loss due to the onset of a specific disease.

Findings: It first finds that the baseline economic loss, which is the impact of a health shock on family income when medical innovations are absent, is at least 28%. It then shows that an increase in medical innovations by one standard deviation raises family income by 15%.

New molecular entities and medical procedures



- Aggregate productivity growth estimates of medical care vary in the extreme from negative to positive yet far from being causal (e.g. Murphy and Topel, 2006; Bloom et al, 2020).
- Several recent studies have examined the impact of single medical innovations or diseases based on credible causal designs (e.g. Garthwaite, 2012; Bütikofer and Skira, 2018; Jeon and Pohl, 2019).

Recall:

The canonical 2 x 2 DD estimator:

$$\texttt{DD} = (\bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{TREAT}}_{\texttt{PRE}}) - (\bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{POST}} - \bar{\texttt{Y}}^{\texttt{CONTROL}}_{\texttt{PRE}})$$

or obtain from the regression:

$$\texttt{y}_{\texttt{it}} = \alpha_{\texttt{i}} + \alpha_{\texttt{t}} + \beta^{\texttt{DD}}\texttt{POST}_{\texttt{t}} \times \texttt{TREAT}_{\texttt{i}} + \varepsilon_{\texttt{it}}$$

Difference-in-difference-in-differences approach

DDD:

- Let units be differentially affected by treatment (e.g. affected $\mathtt{M_i} = 1$ and non-affected $\mathtt{M_i} = 0$), i.e. the estimate of $\beta^{\mathtt{DD}}$ is different across units.
- The simplest way to estimate the triple difference is to estimate separate coefficients in each of sub-sample, $\beta_0^{\mathtt{DD}}$ and $\beta_1^{\mathtt{DD}}$, and compare them.
- Another solution is to put them into a regression on the pooled sample including interactions of $\mathtt{M_i}$ with all variables:

$$\mathtt{y_{it}} = \alpha_\mathtt{i} + \alpha_\mathtt{t} + \beta^{\mathtt{DD}}\mathtt{POST_t} \times \mathtt{TREAT_i} + \alpha_\mathtt{t} \times \mathtt{M_i} + \beta^{\mathtt{DDD}}\mathtt{POST_t} \times \mathtt{TREAT_i} \times \mathtt{M_i} + \varepsilon_{\mathtt{it}}$$

NB: With TWFE, you may need to test for the weighting problem by adding groupXtime fixed effects Goodman-Bacon, 2021.

A matching DDD approach

A DDD regression equation:

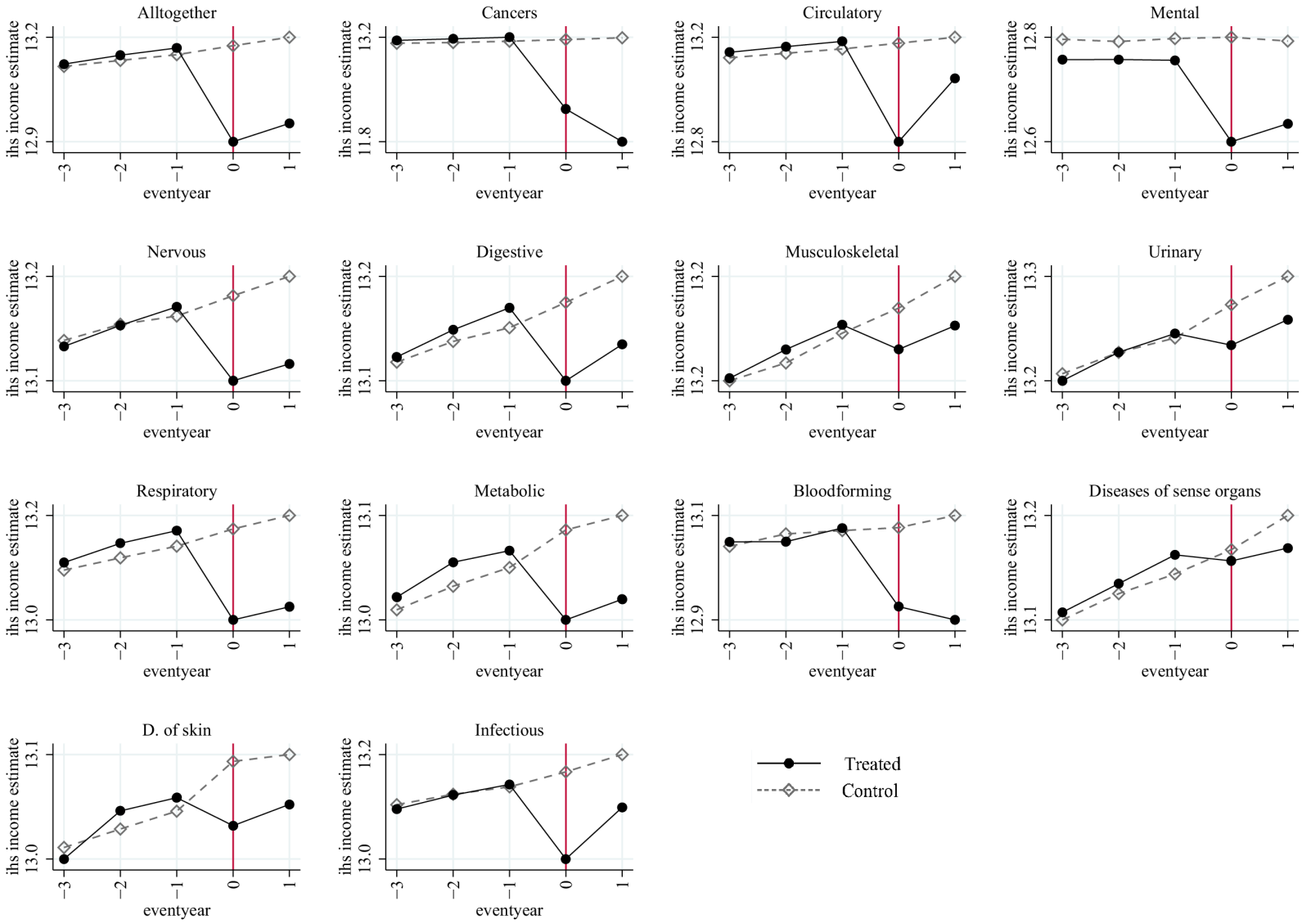
$$y_{it} = \alpha_i + \alpha_t + \beta^{DD} POST_t \times TREAT_i + \alpha_t \times M_i + \beta^{DDD} POST_t \times TREAT_i \times M_i + \varepsilon_{it}$$

where $POST_t \times TREAT_i$ is an indicator for years since a health shock, and M_i are medical innovations available in a year of a health shock s against disease d.

$POST_t \times TREAT_i \times M_i$ is a triple difference.

1. 'Health shock' is identified as an inpatient hospitalization (not 3 y prior), and counterfactuals are individuals hospitalized due to the same disease in the future matched on the propensity score, 91 disease-by-sex groups.
2. M_i are medical innovations: cumulative panels of drugs and patents in diagnostics, therapy and surgery.

$POST_t \times TREAT_i$: Hospitalized vs not-yet-hospitalized



Results: The estimates of β^{DDD}

	ihs family income	ihs family income
	(1)	(2)
DD x l1.drugs	0.00683***	
	(0.00014)	
DD x l1.patents		0.00010***
		(0.00000)
1 SD of l1.drugs /l1.patents	13.74	537.74
1 SD x effect x 100%	9.39% (9.01; 9.76)	5.38% (5.37; 5.39)
Individual FEs	yes	yes

Note: Robust standard errors clustered at individual level are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

What about dynamics?

Event-study specifications are "must-do" to show the plausibility of the identifying assumptions in DD.

Currie et al. (2020):

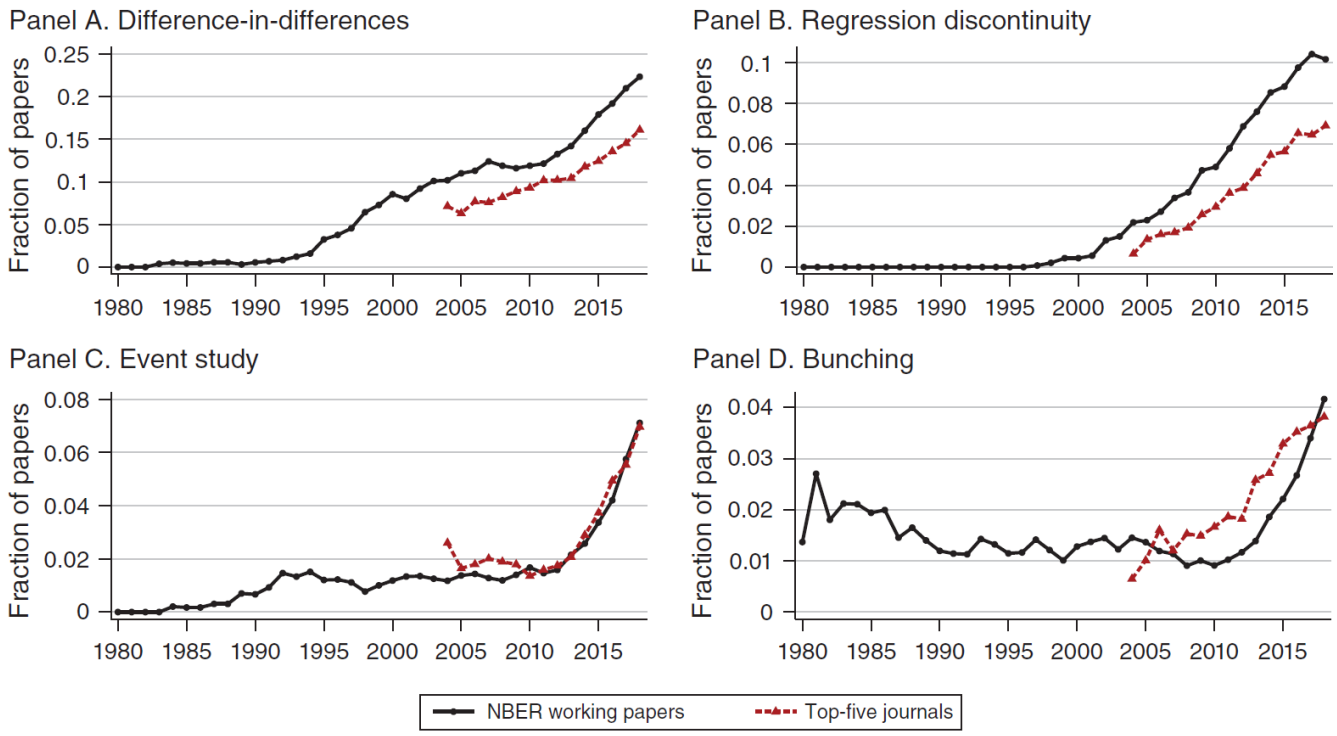
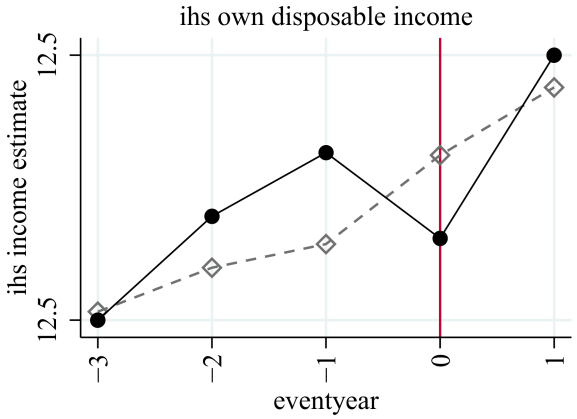


FIGURE 4. QUASI-EXPERIMENTAL METHODS

Notes: This figure shows the fraction of papers referring to each type of quasi-experimental approach. See Table A.I for a list of terms. The series show five-year moving averages.

What about dynamics?



- In TWFE event-study specifications, coefficients on a given lead and lag can be contaminated by effects from other periods, ✓ even if all assumptions hold Sun and Abraham, 2020.
- Pre-trends may arise solely from treament effect heterogeneity!
- NB: Use alternative estimators to estimate event-study coefficients β^{DD}: e.g. Callaway and Sant'Anna, 2020, Borusyak et al., 2021.

Thank you for your attention!

vola@sam.sdu.dk

volhalazuka.com