UNIVERSITÄT SIEGEN

What factors affect excess mortality?

Research Seminar Markets & Strategies II

Supervisor: Prof. Dr. Matthias Hunold, Dr. Jannika Schad

Author: Daria Kharitonova

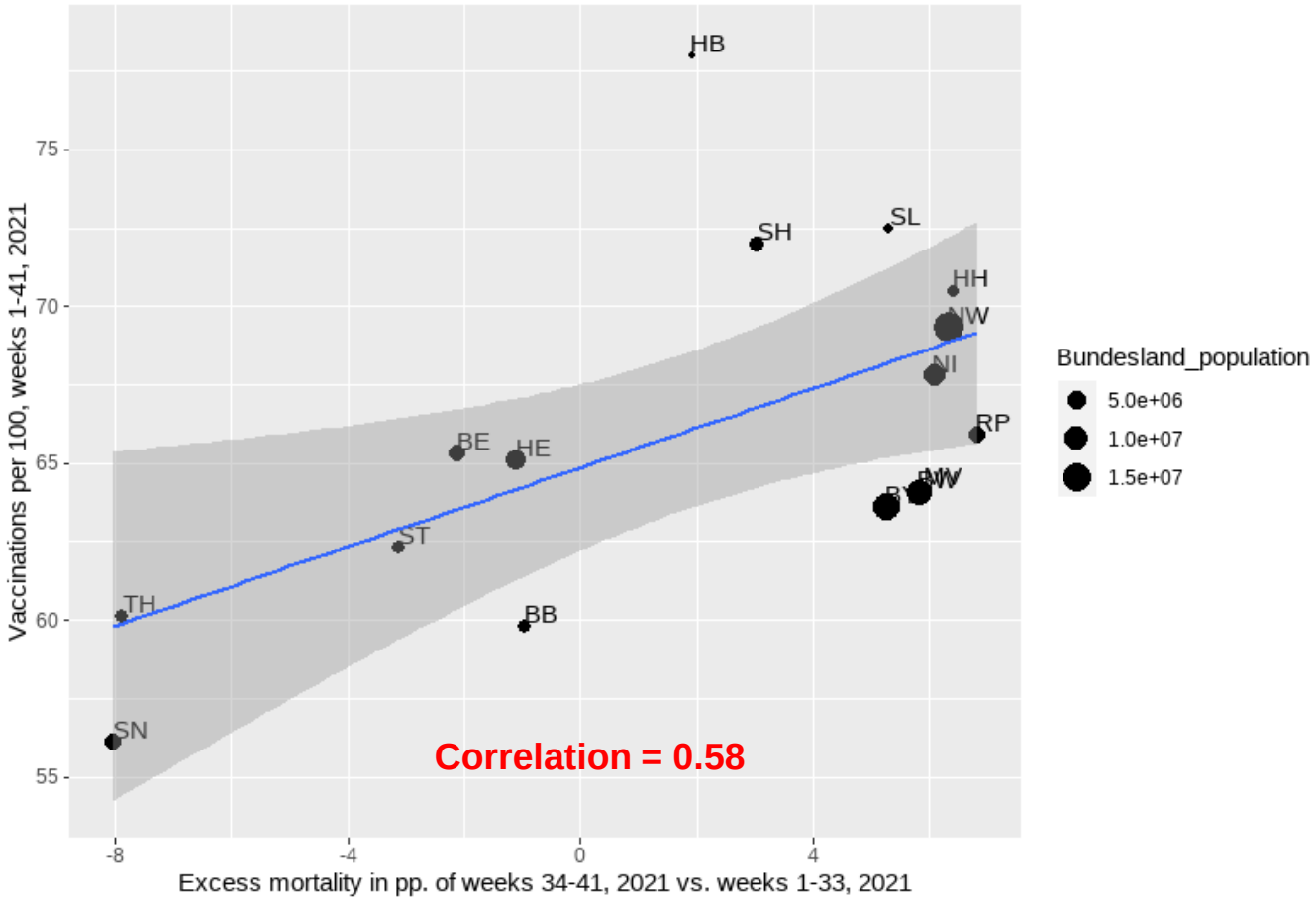
February 15, 2022

Content

1. Motivation
2. Research Question
3. Empirical strategy
4. Models
5. Results and its interpretation
6. Robustness check with lag values
7. Conclusions

Motivation

Correlation between change in Excess mortality and the vaccination rate



On 20.11.2021 Mr. Thorsten Wiethölter claimed in his article:

- *„The higher the vaccination rate, the higher the excess mortality"* referring to the correlation numbers.

How was it calculated? →

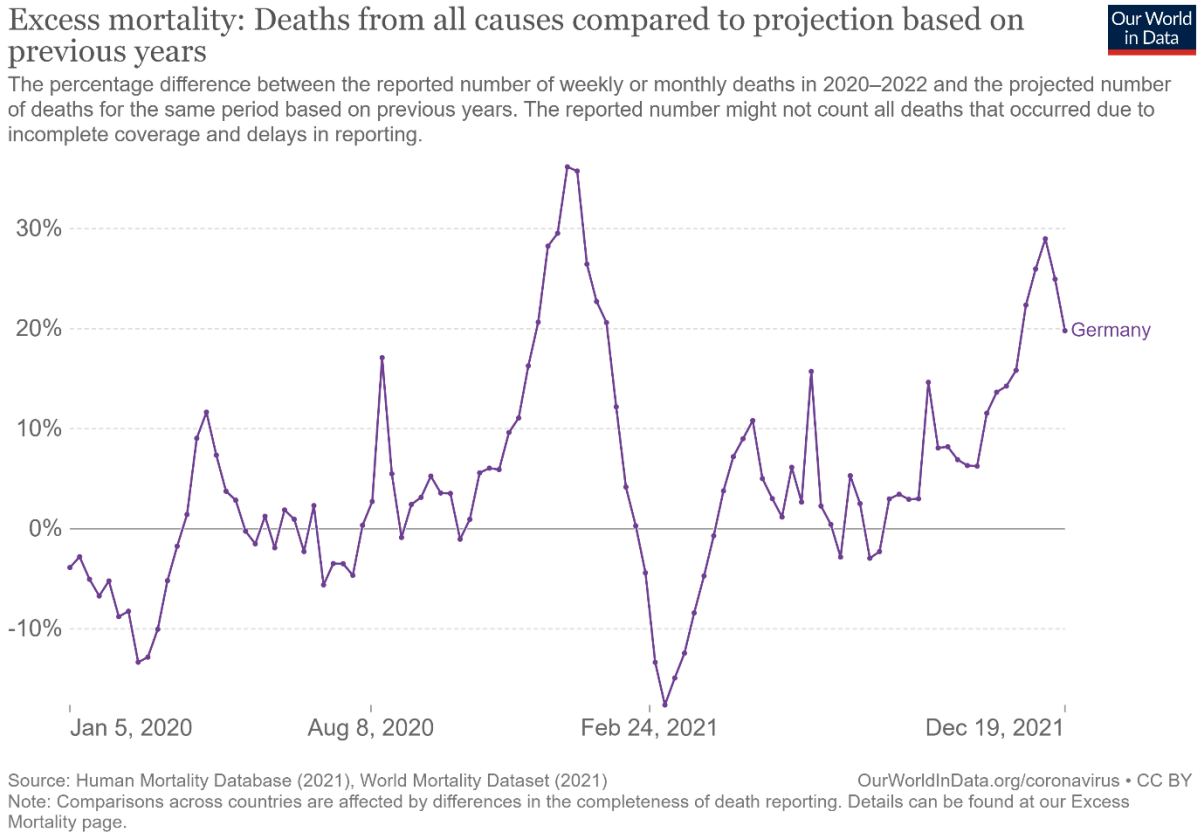
- **Diff. in Excess mortality in pp**. between weeks 34-41, 2021 and weeks 1-33, 2021 by Bundeslands was **set against the fully vaccinated people per 100** in weeks 1-41, 2021
- Excess mortality = Mortality 2021/ Average Mortality 2016-2020

→ Does the vaccination increase the excess mortality?

Source: https://coronakriseblog.wordpress.com/2021/11/20/impfquoten-und-ubersterblichkeit-in-deutschen-bundeslandern/

Research Question

What factors can explain the excess mortality in Germany during COVID-19?



Empirical Strategy

I. COLLECT, AGGREGATE, ENRICH and CLEAN THE DATA

=>

- Data by weeks from COVID-19 start are available on the following levels:
  - ✓ County – 38 095 observations
  - ✓ Bundesland – 2 618 observations
  - ✓ Germany – 208 observations
- Time dimension:
  - ✓ 10th week 2020 – 43th week 2021

**Please refer to the "Additional note" file for the comments on variables*

Empirical Strategy

II. INVESTIGATE THE DATA QUALITY and COMPLETENESS

County level

+ the highest amount of observations
- no data on mortality
- not reliable data on the vaccinations
- no possibility to introduce the age

Bundesland level

+ large amount of observations
- mismatch of the age groups in different data sets
- no possibility to control for demographics change

Germany level

+ the most complete data
- small amount of observations

=> BUNDESLAND and GERMANY LEVELS OF DATA ALLOW TO CONTROL FOR MORE VARIABLES

Empirical Strategy

III. INTRODUCE THE MAIN DEFINITIONS

- Dependent variable

EXCESS MORTALITY	EXCESS MORTALITY ACCOUNTING FOR DEMOGRAPHICS
number of people died in particular week during COVID-19 / average number of people died in 2016-2019 in bundesland i in period t	*number of actual deaths / number of forecasted deaths in Germany in period t in particular age group*
Available at Bundesland and Germany level	Available at Germany level only, allows to account for demographical shift in the age groups over the years

**Please refer to the "Additional note" file for more information on excess mortality calculation*

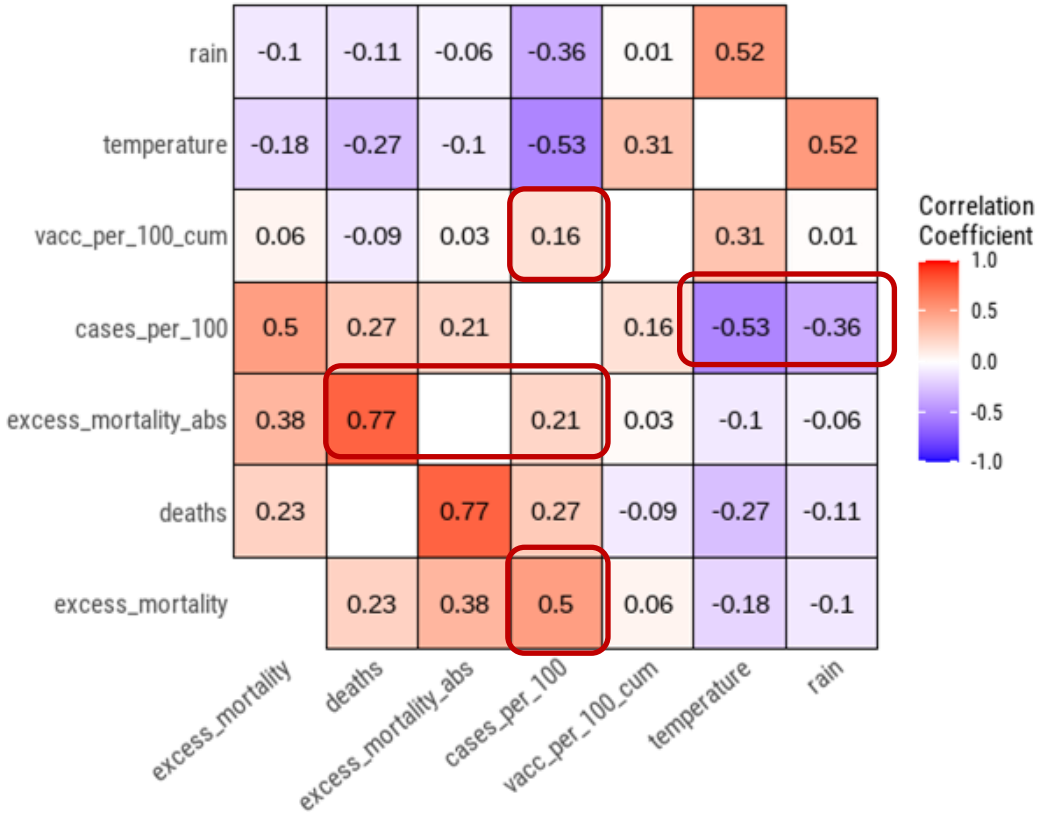
- Independent variables

VACCINATIONS	COVID-19 CASES	AGE GROUP
TEMPERATURE	RAIN	

Empirical Strategy

IV. EXPLORE THE DATA

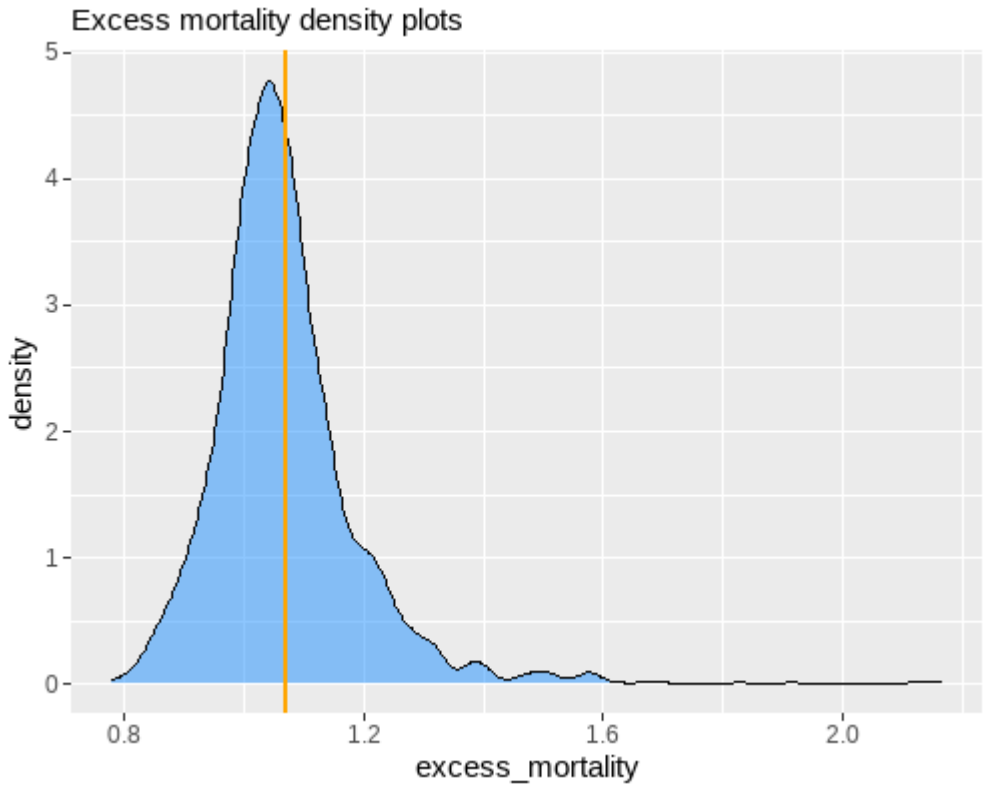
- Correlation matrix of the key variables indicates:
  - Excess mortality strongly correlates with COVID-19 cases and COVID-19 deaths
  - COVID-19 cases per 100 negatively correlate with temperature and rain
  - There is a positive correlation of vaccination rate per 100 and COVID-19 cases per 100

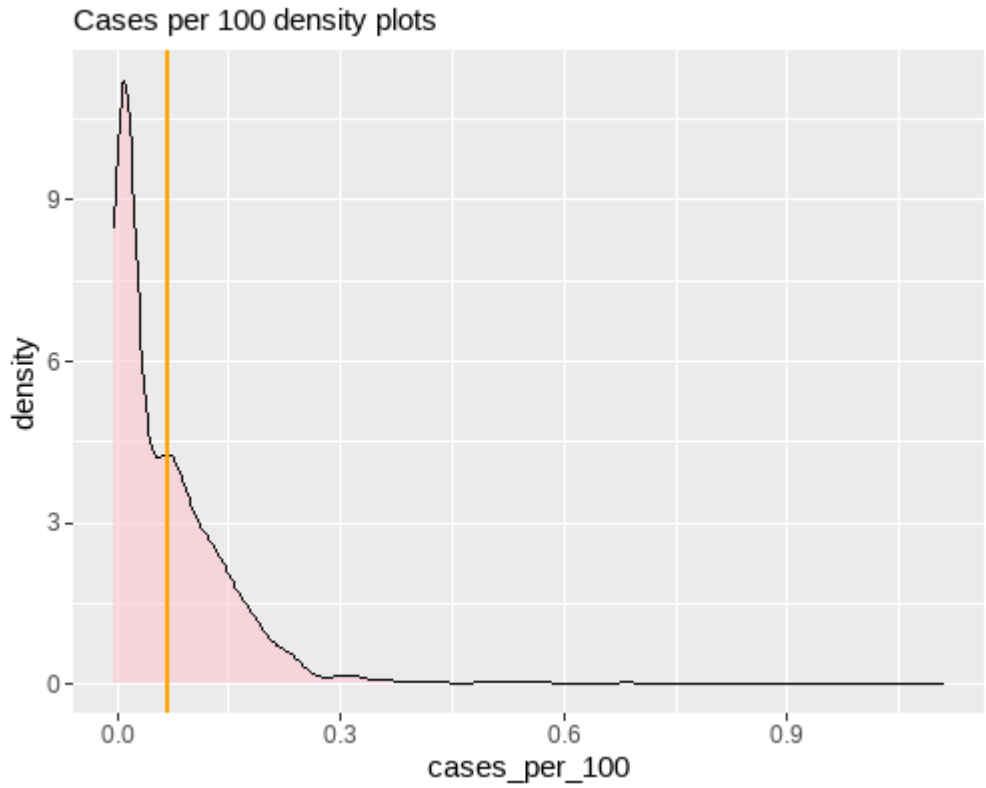


Empirical Strategy

IV. EXPLORE THE DATA

- Excess mortality and COVID-19 cases per 100 are normally distributed with long tail





Period: 10th week 2020 – 43th week 2021

Empirical Strategy

V. ASSUMPTIONS

- The age structure and population of 2021 is the same as in 2020. The demographic data for 2021 were not available at the time of research.
- Mid-year age structure and population is a good proxy for weekly age structure and population.
- Temperature and rain variation within a Bundesland is low.

Model (1): Bundesland level

LINEAR MODEL WITH INTERACTION TERM

$$exc\_mortality_{itk} = \beta_0 + \beta_1\, vacc\_per\_100\_cum_{itk} + \gamma\, cases\_per\_100_{itk-l} + \delta\, age\_group\_60 + \varepsilon\, bundesland_i + \alpha_1\, temperature_{it} + \alpha_2\, rain_{it} + \theta\, cases\_per\_100_{itk} * age\_group\_60 + \epsilon_{itk} \quad (1)$$

- Variables:
  - $exc\_mortality_{itk}$ = mortality in COVID-19 period / average of mortality in 2016-2019 in bundesland i in period t of the age group k
  - $vacc\_per\_100\_cum_{itk}$ = cumulative number of fully vaccinated people (amount of vaccines = 2) per 100 in bundesland i in period t of the age group k
  - $cases\_per\_100_{itk-l}$ = number of COVID-19 confirmed cases per 100 in bundesland i in period t of the age group k, l stands for lag values of COVID-19 cases with $l \in [0, 4]$
  - age_group_60 = dummy variable, equal to 1, if people are older than 60
  - $bundesland_i$ = dummy variable for bundesland i
  - $cases\_per\_100_{itk} * age\_group\_60$ = interaction term of age dummy and cases per 100
  - $temperature_{it}$ = average temperature in bundesland i in period t
  - $rain_{it}$ = average rain l/m² in bundesland i in period t

**Please refer to the "Additional note" file for the comments on variables*

Model (2): Germany level

LINEAR MODEL WITH INTERACTION TERM

$$exc\_mortality_{tk} = \beta_0 + \beta_1\, vacc\_per\_100\_cum_{tk} + \gamma\, cases\_per\_100_{tk-l} +$$

$$\delta\, age\_group\_60 + \alpha_1\, temperature_t + \alpha_2\, rain_t + \theta\, cases\_per\_100_{tk} * age\_group\_60 + \epsilon_{tk} \quad (4)$$

- Variables:
  - $exc\_mortality_{kt}$ = mortality in COVID-19 period / average of mortality in 2016-2019 in period t of the age group k
  - $vacc\_per\_100\_cum_{kt}$ = cumulative number of fully vaccinated people (amount of vaccine = 2) per 100 in period t of the age group k
  - $cases\_per\_100_{kt}$ = number of COVID-19 confirmed cases per 100 in period t of the age group k, , l stands for lag values of mortality with $l \in [0, 4]$
  - age_group_60 = dummy variable, equal to 1, if people are older than 60
  - $temperature_t$ = average temperature in Germany in period t
  - $rain_t$ = average rain l/m² in Germany in period t

**Please refer to the "Additional note" file for the comments on variables*

Model (3): Germany level

LINEAR MODEL WITH INTERACTION TERM AND DEMOGRAPHIC CHANGE IN MORTALITY

$$exc\_mortality\_demo_{tk} = \beta_0 + \beta_1\, vacc\_per\_100\_cum_{tk} + \gamma\, cases\_per\_100_{tk-l} +$$

$$\delta\, age\_group\_60 + \theta\, cases\_per\_100_{tk} * age\_group\_60 + \epsilon_{tk} \quad (5)$$

- Variables:
  - $exc\_mortality\_demo_{tk}$ = number of actual deaths / number of forecasted deaths in Germany in period t of the age group k
  - $vacc\_per\_100\_cum_{kt}$ = cumulative number of fully vaccinated people (amount of vaccine = 2) per 100 in period t of the age group k
  - $cases\_per\_100_{kt}$ = number of COVID-19 confirmed cases per 100 in period t of the age group k, l stands for lag values of mortality with $l \in [0, 4]$
  - age_group_60 = dummy variable, equal to 1, if people are older than 60
  - $temperature_t$ = average temperature in Germany in period t
  - $rain_t$ = average rain l/m² in Germany in period t

**Please refer to the "Additional note" file for the comments on variables*

Results

	Dependent variable:		
	exc_mortality		exc_mortality_demo
	(1)	(2)	(3)
vacc_per_100_cum	-0.0003	-0.0002	-0.0002
	(0.0002)	(0.0004)	(0.0004)
cases_per_100	0.760***	0.550***	0.548***
	(0.078)	(0.196)	(0.192)
age_group60+	0.028***	0.070***	0.027*
	(0.006)	(0.015)	(0.015)
temperature	0.005***	0.004**	0.004**
	(0.001)	(0.002)	(0.002)
rain	0.00000	-0.0002	-0.0002
	(0.0001)	(0.0002)	(0.0002)
cases_per_100:age_group60+	5.824***	4.107***	3.965***
	(0.344)	(0.859)	(0.839)
Constant	0.899***	0.919***	0.929***
	(0.015)	(0.028)	(0.027)
Observations	2,618	154	154
R2	0.393	0.430	0.275
Adjusted R2	0.388	0.407	0.245
Residual Std. Error	0.122 (df = 2595)	0.065 (df = 147)	0.064 (df = 147)
F Statistic	76.516*** (df = 22; 2595)	18.497*** (df = 6; 147)	9.273*** (df = 6; 147)

Note: *p<0.1; **p<0.05; ***p<0.01

Results interpretation

	Dependent variable:	
	exc_mortality (1)	st.dev.
vacc_per_100_cum	-0.0003 (0.0002)	12.87
cases_per_100	0.760*** (0.078)	0.04
age_group60+	0.028*** (0.006)	0.5
temperature	0.005*** (0.001)	5.38
rain	0.00000 (0.0001)	32.64
cases_per_100:age_group60+	5.824*** (0.344)	0.0095
Constant	0.899*** (0.015)	
Observations	2,618	
R2	0.393	
Adjusted R2	0.388	
Residual Std. Error	0.122 (df = 2595)	
F Statistic	76.516*** (df = 22; 2595)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Level – level model (1):

- increase in $vacc\_per\_100\_cum$ **by 1** leads to -**0.0003 pp.** decrease in $exc\_mortality$. The coefficient is neither statistically nor economically significant;
- increase in $cases\_per\_100$ **by 1** leads to **0.760 pp.** increase in $exc\_mortality$. The coefficient is significant at 1% level;
- dummy age_group_60 + **equal 1** increases intercept of $exc\_mortality$ by **0.028 pp**. The coefficient is significant at 1% level;
- Increase in $temperature$ **by 1** lease to **0.005 pp**. increase in $exc\_mortality$. The coefficient is significant at 1% level;
- $rain$ has no effect on excess mortality
- increase in $cases\_per\_100$ within the age_group_60 + **by 1** leads to **5.824** pp. increase in $exc\_mortality$**.** The coefficient is significant at 1% level.

If all statistically significant coefficients will increase by 1 standard deviation $exc\_mortality$ will be:

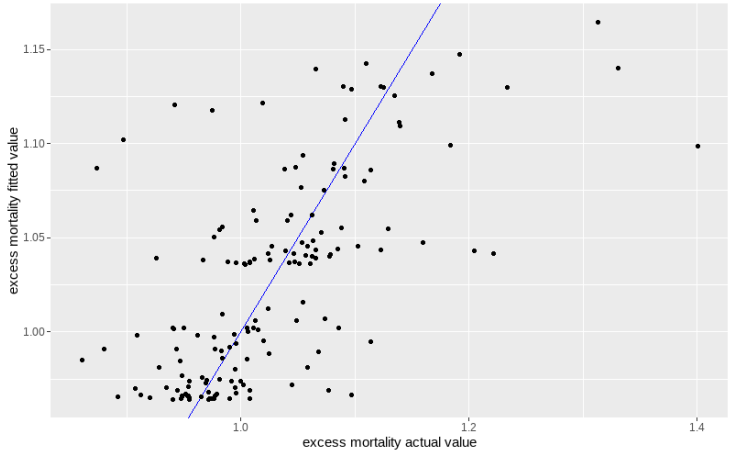
- $\mathbf{0.899} + \mathbf{0.760} * 0.04 + \mathbf{0.028} * 0.05 + \mathbf{0.005} * 5.38 + \mathbf{5.824} * 0.0095 = 0.899 + 0.0304 + 0.014 + 0.0269 + 0.0553 = \mathbf{1.0256}$

→ **2.56% excess mortality**

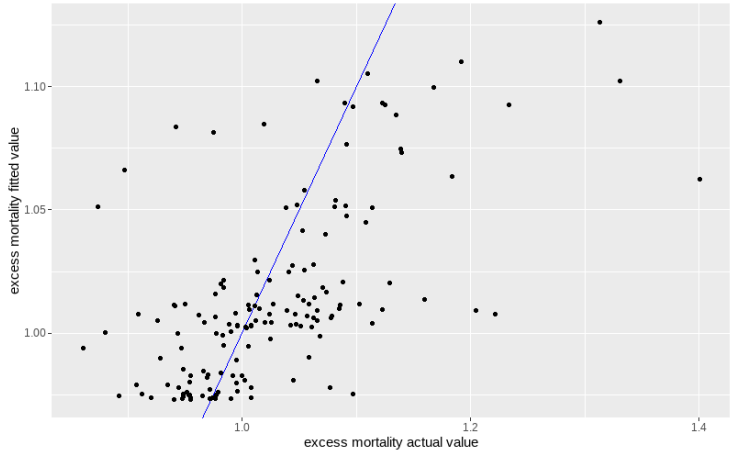
Results: actual vs. fitted values

MODEL (1)

MODEL (2)



MODEL (3



- Prediction errors seem to be of similar size and randomly distributed along the whole value range in all models
- x-axis – actual value, y-axis – fitted value

Robustness check: COVID-19 cases with lags (1)

- **Model 1 – basic model without lags on Bundesland level, Models 2-5 – Bundesland level with different lags**

	Dependent variable:				
	exc_mortality				
	(1)	(2)	(3)	(4)	(5)
vacc_per_100_cum	-0.0003	0.0002	-0.0001	0.0005**	0.0005**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
cases_per_100	0.760***				
	(0.078)				
cases_per_100_lag_1_week		0.557***			
		(0.078)			
cases_per_100_lag_2_weeks			0.797***		
			(0.081)		
cases_per_100_lag_3_weeks				0.280***	
				0.081)	
cases_per_100_lag_4_weeks					0.247***
					(0.083)
age_group60+	0.028***	0.050***	0.033***	0.053***	0.051***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
temperature	0.005***	0.002**	0.005***	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
rain	0.00000	0.00002	-0.00001	0.00000	-0.00005
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
cases_per_100:age_group60+	5.824***				
	(0.344)				
cases_per_100_lag_1_week:age_group60+		2.296***			
		(0.218)			
cases_per_100_lag_2_weeks:age_group60+			5.731***		
			(0.360)		
cases_per_100_lag_3_weeks:age_group60+				1.484***	
				(0.228)	
cases_per_100_lag_4_weeks:age_group60+					2.124***
					(0.375)
Constant	0.899***	0.906***	0.902***	0.966***	0.998***
	(0.015)	(0.016)	(0.015)	(0.016)	(0.016)
Observations	2,618	2,584	2,550	2,516	2,482
R2	0.393	0.369	0.396	0.343	0.339
Adjusted R2	0.388	0.363	0.391	0.337	0.333
Residual Std. Error	0.122 (df = 2595)	0.125 (df = 2561)	0.122 (df = 2527)	0.128 (df = 2493)	0.129 (df = 2459)
F Statistic	76.516*** (df = 22; 2595)	67.996*** (df = 22; 2561)	75.442*** (df = 22; 2527)	59.080*** (df = 22; 2493)	57.229*** (df = 22; 2459)

Please refer to the "Additional note" file for the comments on the results

Robustness check: COVID-19 cases with lags (2)

- **Model 3 – basic model without lags on Germany level accounting for demographic shift, Models 2-5 – Germany level with different lags**

	Dependent variable:				
	exc_mortality_demo				
	(1)	(2)	(3)	(4)	(5)
vacc_per_100_cum	-0.0002	-0.0002	-0.0003	-0.0002	-0.0002
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
cases_per_100	0.548***				
	(0.192)				
cases_per_100_lag_1_week		2.865***			
		(0.928)			
cases_per_100_lag_2_weeks			0.709***		
			(0.198)		
cases_per_100_lag_3_weeks				3.166***	
				(0.960)	
cases_per_100_lag_4_weeks					0.642***
					(0.218)
age_group60+	0.027*	0.031**	0.025*	0.024*	0.028*
	(0.015)	(0.014)	(0.014)	(0.014)	(0.015)
temperature	0.004**	0.004**	0.005***	0.005**	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
rain	-0.0002	-0.0001	-0.0002	-0.0002	-0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
cases_per_100:age_group60+	3.965***				
	(0.839)				
cases_per_100_lag_1_week:age_group60+		-1.931**			
		(0.860)			
cases_per_100_lag_2_weeks:age_group60+			4.895***		
			(0.855)		
cases_per_100_lag_3_weeks:age_group60+				-1.995**	
				(0.874)	
cases_per_100_lag_4_weeks:age_group60+					4.411***
					(0.934)
Constant	0.929***	0.928***	0.909***	0.914***	0.921***
	(0.027)	(0.028)	(0.028)	(0.030)	(0.032)
Observations	154	152	150	148	146
R2	0.275	0.279	0.338	0.328	0.298
Adjusted R2	0.245	0.249	0.310	0.299	0.267
Residual Std. Error	0.064 (df = 147)	0.063 (df = 145)	0.060 (df = 143)	0.061 (df = 141)	0.062 (df = 139)
F Statistic	9.273*** (df = 6; 147)	9.346*** (df = 6; 145)	12.154*** (df = 6; 143)	11.473*** (df = 6; 141)	9.821*** (df = 6; 139)

Please refer to the "Additional note" file for the results comment

Conclusions

- In order to understand, what factors influence excessive mortality other than vaccination I collected the weekly data from COVID-19 start in 2020 till 43th week 2021 on Bundesland and Germany level.
- The main result indicates that **Excess Mortality** can be primarily explained by **COVID-19 cases, Age** and **Temperature**, while the coefficients for **Vaccination rate** are neither economicaly nor statistically significant. The result robustness is confirmed by different model specifications with various lag values of the variable COVID-19 cases.
- → When controlled for other factors, there is **no positive relationship between the Vaccination rate and Excess mortality**. The other question is why there is no negative relationship? The possible explanations are:
  - It is unknown whether excess deaths occured among vaccinated or not vaccinated individuals;
  - The relation between vaccination and excess mortality might be not linear.
- Another question for the further research might be the choice of the lag value for COVID-19 cases. My data indicate that the optimal value of lag ranges from 0 to 2, when the highest R-squared is achieved. However, it should be additionally investigated.

Thank you!

Appendix

Data sources

UNIVERSITÄT SIEGEN

1. Impfquoten und Übersterblichkeit in deutschen Bundesländern: https://coronakriseblog.wordpress.com/2021/11/20/impfquoten-und-ubersterblichkeit-in-deutschen-bundeslandern/
2. Destatis Sonderauswertung Todesfälle https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/Tabellen/sonderauswertung-sterbefaelle.html;jsessionid=5CEEE6B69887285095ADE798EBF52DFF.live712
3. RKI-Impfdaten https://github.com/robert-koch-institut/COVID-19-Impfungen_in_Deutschland
4. COVID-19 deaths and cases: https://github.com/jgehrcke/covid-19-germany-gae
5. Excess mortality due to COVID-19 in Germany https://www.sciencedirect.com/science/article/pii/S016344532030596X
6. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset https://elifesciences.org/articles/69336
7. Temperature https://de.statista.com/statistik/daten/studie/5564/umfrage/monatliche-durchschnittstemperatur-in-deutschland/
8. Rain https://de.statista.com/statistik/daten/studie/576867/umfrage/durchschnittlicher-niederschlag-pro-monat-in-nordrhein-westfalen/
9. Temperature by Bundeslands https://www.wetterkontor.de/wetter-rueckblick/gebietsmittel/monatswerte/thueringen