Increased access to railroads and the effects on cholera in Prussia

Seminar paper
Date: feb 2023

By Joseph Eriksson Ms.c

Epidemics in economic history

1. Introduction

The 19th century witnessed a dramatic transformation in transportation technology across Europe, particularly with the expansion of railroads. Prussia was at the forefront of this trend, with a rapid and widespread development of rail infrastructure throughout the mid-1800s (Hornung, 2014). This expansion brought about significant changes in the way people, goods, and ideas were transported, resulting in an interconnectedness and globalization that was unprecedented in human history.

At the same time, Prussia was also facing significant challenges in the form of recurring cholera epidemics. Cholera, a waterborne disease caused by the bacterium Vibrio cholerae, had been endemic to Europe for centuries (Cholera - Vibrio cholerae infection, 2022). However, the 19th century saw a number of particularly severe outbreaks that had a devastating impact on communities across the continent which set John Snow in motion to examine the roots of the problem and publish his paper 1855.

It is therefore interesting to explore the relationship between the expansion of the railroads and the spread and severity of cholera outbreaks in Prussia during the 19th century. The growth of the rail network brought about a significant increase in the volume and speed of trade and travel, which had the potential to accelerate the spread of infectious diseases like cholera. It made the world smaller which was good for trade and economic activity overall, but it also made people at the same time more susceptible to contract the disease.

To better understand the impact of rail expansion on cholera outbreaks in Prussia, it is important to examine historical data, use previous research and use statistical analysis to combat endogeneity. This research could shed light on the complex interplay between transportation infrastructure, globalization, economic activity and disease transmission, and could have implications for our understanding of global health challenges today. Research question: *How did the expansion of the railroad infrastructure in 19th century Prussia shape the spread and severity of cholera outbreaks?*

2. Channels through which railroads effect incidence of cholera

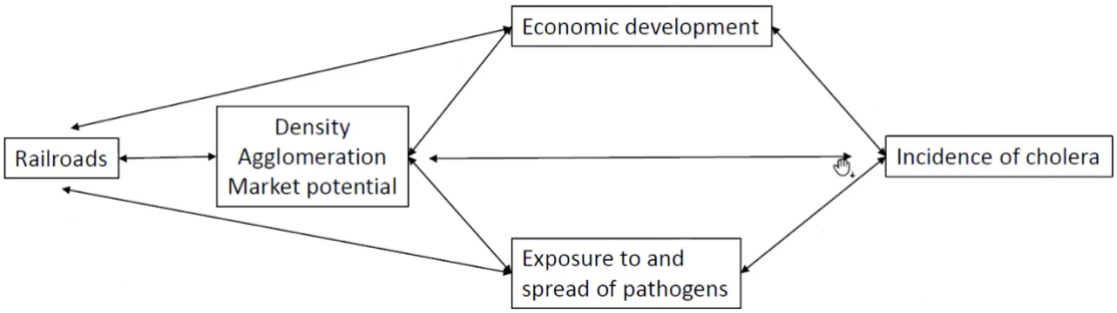

Figure 1: showcase channels that will be explored in this paper connecting expansion of railroads to incidences of cholera.

Railroads - Agglomeration

The railroads to agglomeration channel links the expansion of the railroad to congregation into urban cities and market potential. In (Hornung, 2014) he found a stable and positive effect of having a railroad station on urban growth both in the short- and long run for different periods during 1840-1871 in Prussia. He argues that the railroad expansion improved the conditions for economic activity leading to population growth in urban areas. This was driven by the growth in the regional industries, emergence of regional labor markets and increased mobility of people and goods that drove immigration into the cities. However, he also stresses that there are heterogenous effects of railroads, with some regions experiencing more significant growth and agglomeration than others. Regions that beforehand had favorable conditions to benefit from the railroads had more agglomeration. For example, regions with natural resources such as coal or iron, could now more easily commercialize it.

Agglomeration – exposure to and spread of pathogens

When specifically discussing agglomeration and its effect on exposure and spread of cholera we must understand exactly how cholera is transmitted and what it is.

Cholera is an acute diarrheal illness caused by infection of the intestine with Vibrio cholerae bacteria (Cholera - Vibrio cholerae infection, 2022) and in laymen terms this means that we get cholera when the cholera bacteria reach the intestines. The cholera bacteria is spread from fecal matter and it is normally transmit through contaminated food or water. In Prussia the importance of hygiene was not well known during the 19th century which made it such a widespread disease at that time.

There are several ways for agglomeration to channel into exposure to and spread of cholera. Firstly agglomeration tend to lead to overcrowding and people living more densely together, where people share more spaces. This makes it more difficult to maintain adequate sanitation and hygiene which is essential for prevention of the spread of cholera. For slum areas this is particularly true, inequality in the cities could exacerbate the problems and increase the risk of transmission through contaminated water that the rich can afford to not drink from and eat at places that are less crowded. So crowding will increase the risk of cholera transmission and inequality can exacerbate the problem in the cities.

Another potential mechanism is the disposal of excrement, in the cities it is harder to dispose of excrement and considering that only 20% of cities with population above 50 000 people had water carriage for disposal of excrement 1892 (Hennock, 2000). People had to carry their excrement by hand to a cesspool where it then would be transported out of town and sold to farmers as fertilizer. This directly makes people come into contact with the bacteria and could therefore be a potential transmission channel.

Lastly with agglomeration increasing and railroads expanding, distance is getting less relevant. Urban areas with high population density and mobility can act as hubs of disease transmission, using the railroads to transfer infected masses of people all around country to the outskirts of the provinces and to other bigger cities nationally and inter-nationally. The development can accelerate the spread of cholera and other diseases. But overall, there are also other very complex outcomes of agglomeration such as how agglomeration effects human behavior compared to if they didn’t move to the city, and this could affect how they interact with other people in the city. Maybe they bring norms from the countryside that reduce transmission or maybe it’s

the opposite. They could also bring beverages such as beer into the city and substitute it to water to get access to an uncontaminated beverage which is what a visit to Aying brewery at the outskirts of Munich taught me. Apparently, there was not access to clean water so instead they brewed beer to solve this problem, which solved that problem but maybe created other problems instead...

Agglomeration – economic development

Increase in the railroad network connecting more people can help people allocate themselves better. If there's a very skilled worker on the countryside, his potential would probably not be captured there compared to a larger city where many other people and businesses reside. Improving railroad network improves access to jobs in other parts and makes it easier to commute longer distances, creating larger and denser cities where businesses are clustering close to each other. This clustering of businesses creates a beneficial environment where businesses have an easier time cooperating and finding expertise in many different fields that would otherwise be unavailable.

Economic development

The main idea through which economic development effects cholera incidences is through improved living conditions. As economic activity increases in society people get higher real wages and can therefore afford a healthier lifestyle with more food, afford to visit doctor, afford better accommodation, afford to take leisure. This would clearly reduce incidence of cholera. But I don't believe it's quite as simple as that. I believe that economic inequality, sanitation infrastructure, the sanitation movement and scientific knowledge in epidemiology also has a role to play. I will discuss Alfani, Snow, Hennock and Gallardo-Albarran and connect it to economic development in Prussia.

Inequality could potentially extract a lot of the gains from the increased economic development caused by railroads, this would negatively bias our estimate. Looking at The World Inequality Database top 10% had roughly 35% of the total income in 1874 (Database, u.d.). If we compare this to France at the time it was closer to 80%

(Alfani, Epidemics, Inequality, and Poverty in Preindustrial and Early Industrial Times, 2022). Taking present day values and France into consideration I'm concluding that inequality is not a significant problem.

Endogeneity through economic development

I would also consider the context of where and when this analysis is created. During the 19th century where London also had issues with cholera outbreaks and most of the population believed this was due to miasma and that it was air borne transmission when it in reality was linked to polluted drinking water filled with fecal matter that ended up in their drinking water (Snow, 1855). From this experiment Snow advocated for the separation of water sources for drinking and waste disposal. He argued that providing clean water supplies and proper sanitation facilities would be critical for preventing the spread of cholera and other waterborne diseases.

My point with my last paragraph is that the railroad expansion in Prussia is happening during this time where the discourse is happening and investment into these sanitation facilities that Snow partly advocated for would reduce incidences significantly. This could potentially cause endogeneity problems because the places where such investments would happen would be at richer cities or provinces that have high railroad expansion. This could overestimate the effect railroads have on incidences through economic development due to some technological shock.

Also, second point is that economic development doesn't cause lower incidence of cholera directly. Economic development is a catch all term for increased money to spend or being able to afford more or/and better goods than before and being able to spend and afford more stuff can potentially decrease cholera incidences if it is spent on things that reduce or eliminate the chance of getting infected. If this extra money is spent on water pipe infrastructure, hygiene facilities or on individual level on goods and services that increase health and hygiene then an increase in economic development will lead to decrease in cholera incidences and mortality rates. However, if we don't even know that better hygiene reduces death then it doesn't matter how much money we earn because we won't spend it on improving our hygiene or sanitary infrastructure etc.

Hennock looked at the sanitary movement in England and Germany between 1838-1914 (Hennock, 2000). Prussia's response to cholera was a bottom-up response from local government that started long after Snows discovery 1855. One of the questions he asked himself was, why wasn't there any interest in sanitary reform during the 50s and 60s. I doesn't make sense that Snow solved the problem and provided a good alternative understanding to the problem and that Prussia doesn't try to apply his solution after having cholera outbreaks for nearly 50 years. The main reasons that Hennock found was the lack of concern over the level of urban deathrates and the high cost of sanitary infrastructure which only richer societies could afford. For Germany only 27% of towns with populations above 10 000 people had central water supply 1876 but it was 86% in 1895. And concerning the disposal of excrement, only 20% of all German towns with population over 50 000 had water carriage for disposal of excrement in year 1892.

What we have learned from the Hennock paper that is relevant in our paper is firstly that because sanitary reform originated from local governments we should use fixed effects to take into account that some governments might introduce reforms at different times. Secondly that we don't expect the sanitary reforms to take into effect before the 1870s but even then, it was still very few places that implemented measures.

Gallardo-Albarran picked up where Hennock left and looked at the development of sanitary infrastructure and how it affected mortality in Germany 1877-1913 (GALLARDO-ALBARRAN, 2020). The main findings from the paper suggest that supply of safe drinking water reduced number of deaths, but it was to a lower extent than previously expected. The reason for the that was due to absence of efficient systems of sewage removal and contact with feces that Hennock discussed. The other reason was water contamination of the water supply. Those three conditions created a favorable environment for the spread of diseases which offset the positive effects of waterworks. Other results they found were substantial heterogenous effects of these sanitary technologies especially for local socioeconomic conditions. Municipalities with high levels of income inequality and employment had less impact from the implementation of these sanitation technologies on mortality rates.

Railroads - economic activity

An issue is that railroads are not just randomly built they are usually built for economic or military reasons. As an economy develops, the need for increased transportation also increases which can cause simultaneity bias which means that the railroad and economic development are not independent of each other and changes in one railroad can both cause and be caused by changes in the economic development. Same thing could be said about agglomeration and economic activity. Where there is economic activity there is economic opportunity and potential gains to be made.

In figure 2 taken from (Hornung, 2014) the railroads that was built in 1845 is connecting the major cities within Prussia, it does not look like some exogenous shock unrelated to economic activity. In figure 3 even more areas are covered by railroads, and the previously unconnected cities to the north, west and south of Prussia are in 1875 connected. They have also expanded into more rural parts of Prussia where not there are no major cities yet from the 1845 map. It connects smaller cities and towns to the major cities to a higher degree than before. In 1875 only a several of the counties haven't built railroads yet, the majority have at least built between 0-50 square km

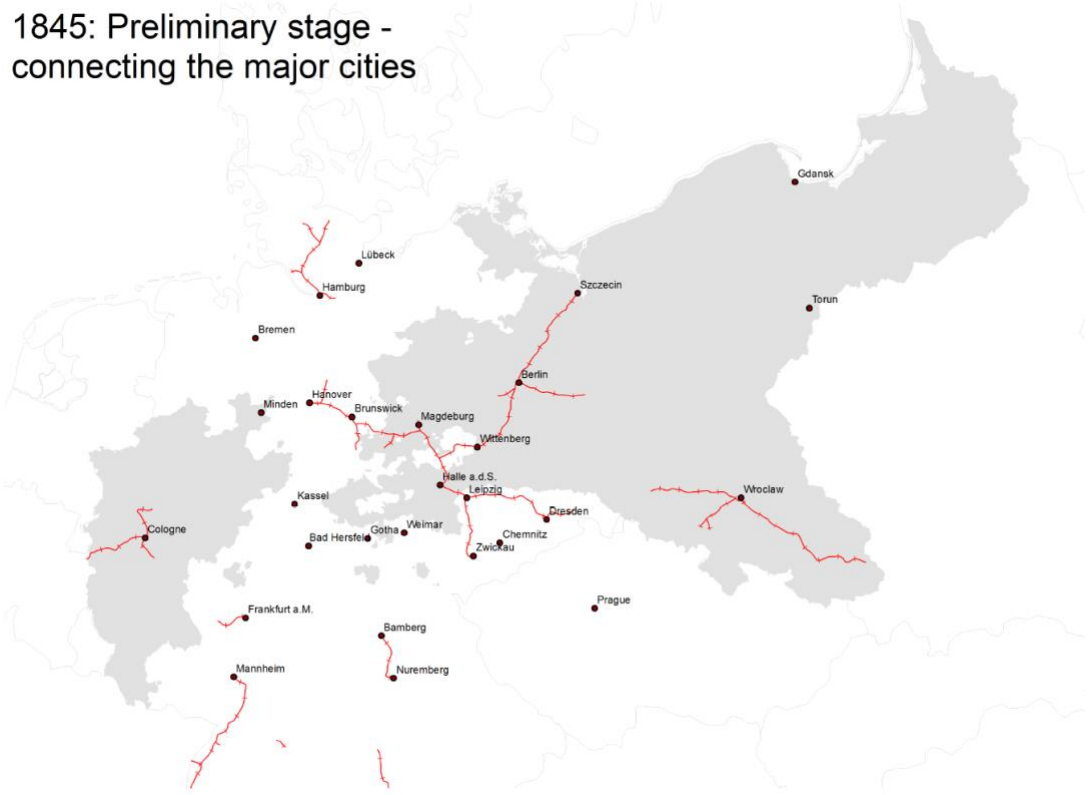

Figure 2: Hash lines indicate railroad routings. Grey area depicts Prussian territory in period 1815-1867. Source (Hornung, 2014)

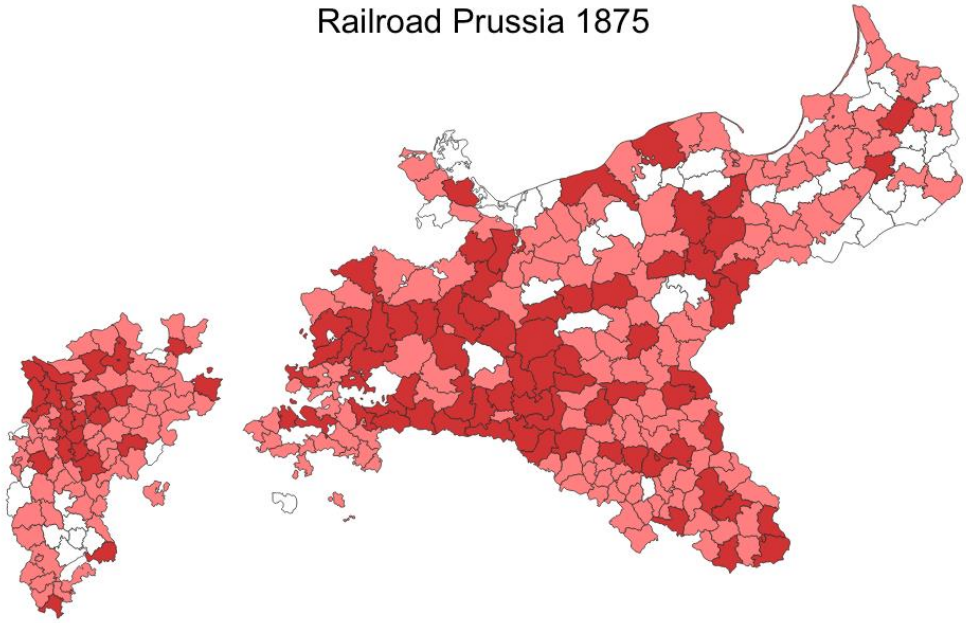

Figure 3: Map area depicts Prussian territory in year 1875 divided by counties. White area has no railroad in county, light red has 0-50 square km railroad and dark red has >50 square km railroad. Source: created by Author.

Increasing railroads improved market efficiency and costs, in (Donaldson, 2018) they found 4 results from expansion of railroad network in colonial India (India, Pakistan, Bangladesh). Prior to this expansion livestock carried most of India's commodity trade on their backs traveling up to 30 km per day along dirt roads. The first result they found was that trade costs and interregional price gaps decreased. Second was that interregional and international trade increased. Thirdly they found increased real income levels and that fourthly that this in turn increased welfare. The results are consistent with economic theory in that reducing trade costs and trading more leads to higher real income and increased product scope that take advantage of gains from Ricardian comparative advantage. I found these results comparable to what to expect in Prussia from expanding railroad network, they don't have other means of transporting other than livestock and boats. This paper is also interesting for another reason, it deals with the endogeneity of railroads being built for economic reason through arguing that the British government built these roads for military reasons, this would mean that railroads would be an exogenous shock

without the simultaneity bias discussed in the previous paragraph that we have to deal with in this paper.

Incidence of cholera – economic activity

The Malthusian framework and Smithian framework both have theories to how population effects economic activity. According to Malthus, population growth tends to outstrip the growth of resources, leading to declining real income levels in the long run (Alfani, Epidemics, Inequality, and Poverty in Preindustrial and Early Industrial Times, 2022). In preindustrial times, this was often exacerbated by epidemics and famines or other population shocks that reduced population levels and temporarily relieved pressure on resources. However, as living conditions improved, the population began to grow more rapidly, putting more pressure on resources and leading to a return to the equilibrium population level where the amount of people that are born are equal to the amount that die.

Alfani's paper (Alfani, Epidemics, Inequality, and Poverty in Preindustrial and Early Industrial Times, 2022) applies unified growth theory to the preindustrial and early industrial periods, arguing that epidemics played a significant role in shaping patterns of economic growth and inequality during this time. In particular, Alfani suggests that epidemics had a leveling effect on income and wealth distribution by reducing population levels and temporarily relieving pressure on resources. However, as the population recovered and resumed its growth, inequality and poverty returned.

A critic to the view that a negative population shock leads to better economic situation is some post-industrialization economies have moved into a Smithian framework where population has positive effects on real income in the long run (Jedwab, Johnson, & Koyama, 2022). The Smithian framework is a new era perspective on economic development that emphasizes the importance of the specialization and trade for promoting economic growth. This perspective suggests that as the size of the market expands mainly in terms of population, the potential for specialization increases which in turn create a series of positive-feedback mechanisms that increase technical advances. Under the Smithian interpretation,

population growth is seen as an important driver of economic growth, as higher population densities expand the size of the market and create opportunities for greater specialization and trade. In this view, the reduction of population and market size due to events such as the Black Death is seen as a hindrance to economic development, as it limits the potential for division of labor and trade.

Ending commentary

Section 2 discusses previous studies and the potential problems with endogeneity we have when answering how railroads affect cholera incidences in Prussia. We cannot get causal estimates by just naively estimating an OLS regression, we have simultaneity problems, biases in both directions and many other things to take into consideration when creating the model. And even if we could get causal effect, we would still have to deal with the effects being heterogenous across counties and provinces (More or less populated) and creating railroads connecting many bigger cities together would probably affect incidences more than building railroads connecting to a coal mine up in a mountain.

However, we have also in this section built up real causal links through which railroads can affect incidences of cholera which is fundamental for estimating causal effects to begin with. Railroads effects agglomeration and it is settled that urban cities grew in Prussia during the 19th century. Through the development of the urban cities and the industrialization the economic condition for people improved for the better even when accounting for inequality. But at the same time as the urban cities developed and the economy flourished, we also observed that it led to overcrowding, worse sanitary conditions and problems with disposal of excrement that increased exposure to cholera. On the one hand we have economic development which overall is expected to reduce cholera incidences and on the other hand the exposure to

cholera has increased which positively effects incidence of cholera.

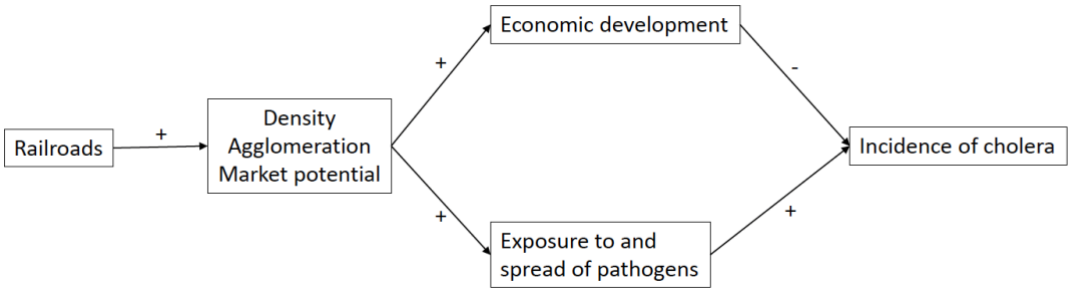

Figure 4: channels railroads and how it affects incidence of cholera with respective sign.

3. Results

My data used for regressions and figures is a strongly balanced panel dataset spanning from 1831 to 1874 with gaps. It has 7304 observations and it is on county level with 332 unique counties. These are the counties of the so-called “old provinces”, i.e. before the annexations resulting from the Austro-Prussian war of 1866. When transforming my dataset into a panel I removed some duplicate variables. The railroad tracks are measured in county-year km. In Prussia there are 8 provinces and all the counties geographically are smaller parts of a province.

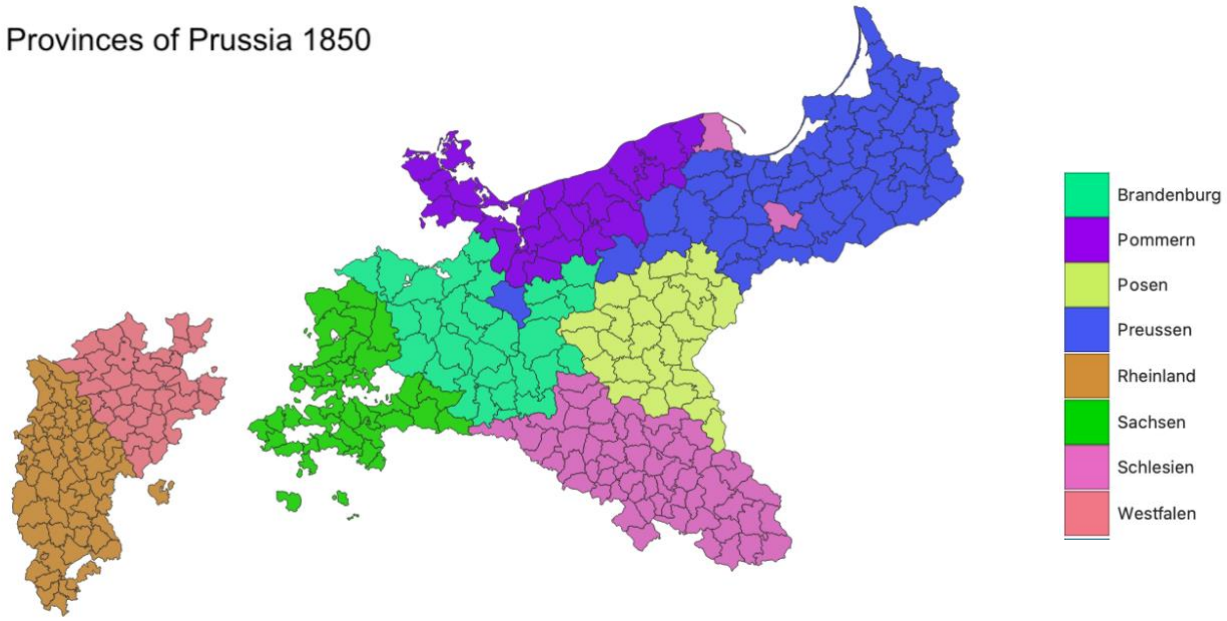

Figure 5: Map shows Prussia in year 1850 where the counties are colored based on which province they belong to.

In figure 6,7 and 9 we have figures that show mortality rate, population and expansion of railroad tracks over the time period 1831-1874 for all provinces. In figure 8 “Landkreis” and “Stadtkreis” can be roughly translated to rural- and urban county where that figure shows us if rural- and urban counties are effected in the same way or if there are systematic differences. We use these figures to get a good overview of Prussia’s provinces, geography, population and how much they expanded their railroad network to study differences between that can help us

understand what are the most important factors that affect cholera mortality rates. But it will also visibly show us the variables that will be used for our models.

Overall mortality rate varies across provinces, which is expected. There are countless things that could affect mortality, many of which we have discussed already in prior section. Taking mortality and population in the province into consideration, Westfalen and Rheinland of west Prussia had best outcome.

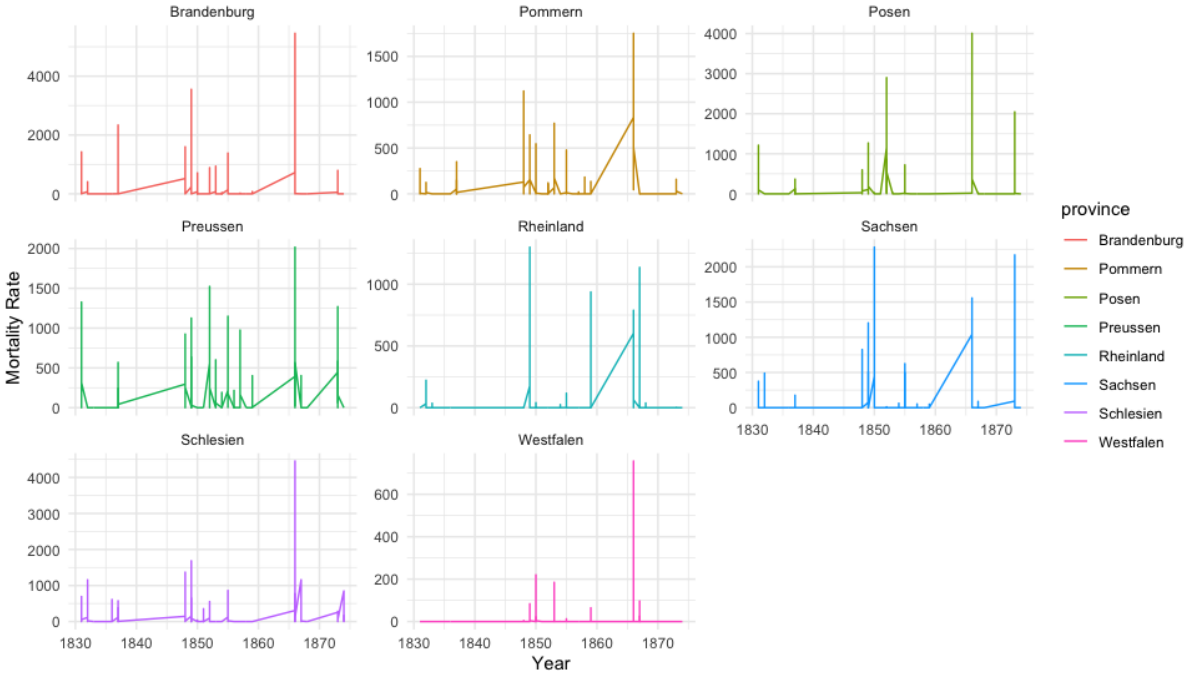

Figure 6: Mortality rate from cholera during 1831-1874 over all provinces.

Figure 6 showcase Westfalen and Rheinland had very few outbreaks and when an outbreak occurred it was associated with low mortality rate when considering total province population in figure 7. Other interesting results from figure 6 is that every province had their worst outbreak around 1850s and 1860s. For Preussen, Brandenburg, Posen and Pommern it was especially bad during the 50s with multiple outbreaks occurring every year. All those 4 are connecting provinces that are in central/east Prussia. For all provinces except for Rheinland and Sachsen the worst cholera outbreak with highest death toll was in the mid 60s. This is consistent with the results from prior section where we discussed the sanitary movement and the lack of a response from Prussia in the 50s and 60s. West Prussia had comparably better outcome than central/East.

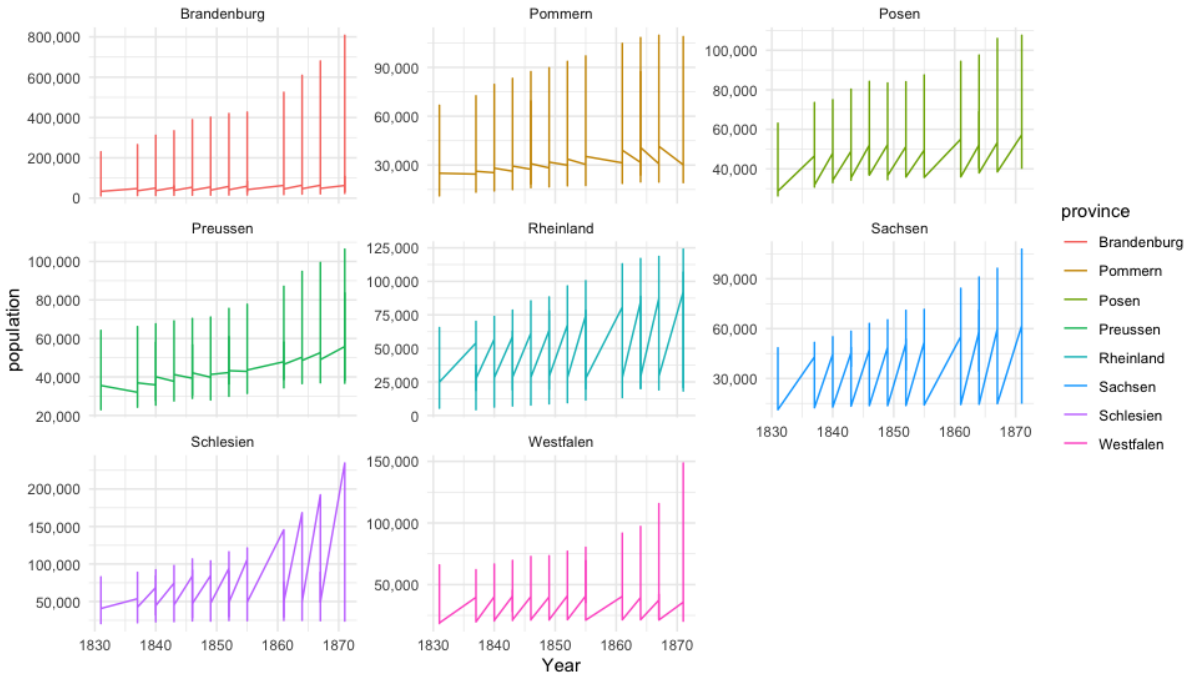

Figure 7: Population development during 1831-1874 for provinces.

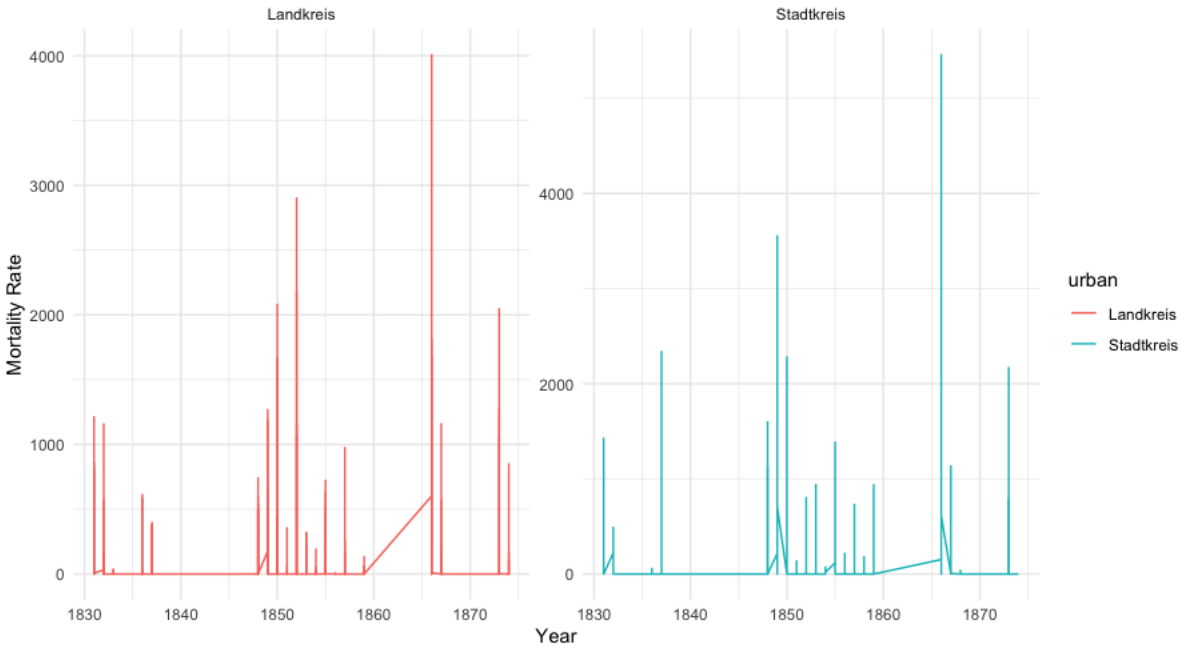

Figure 8: Mortality rate from cholera during 1831-1874 for rural/urban counties.

A little unexpected is the results from figure 8 indicating that rural- and urban counties are similarly exposed to the outbreaks by having very similar mortality rates. It is unexpected partly because of the urbanization during that period and

partly because we expected the cities and its population density to significantly contribute to the increase the exposure- and spread of cholera which in turn also increase incidence of cholera relatively more than in the rural areas.

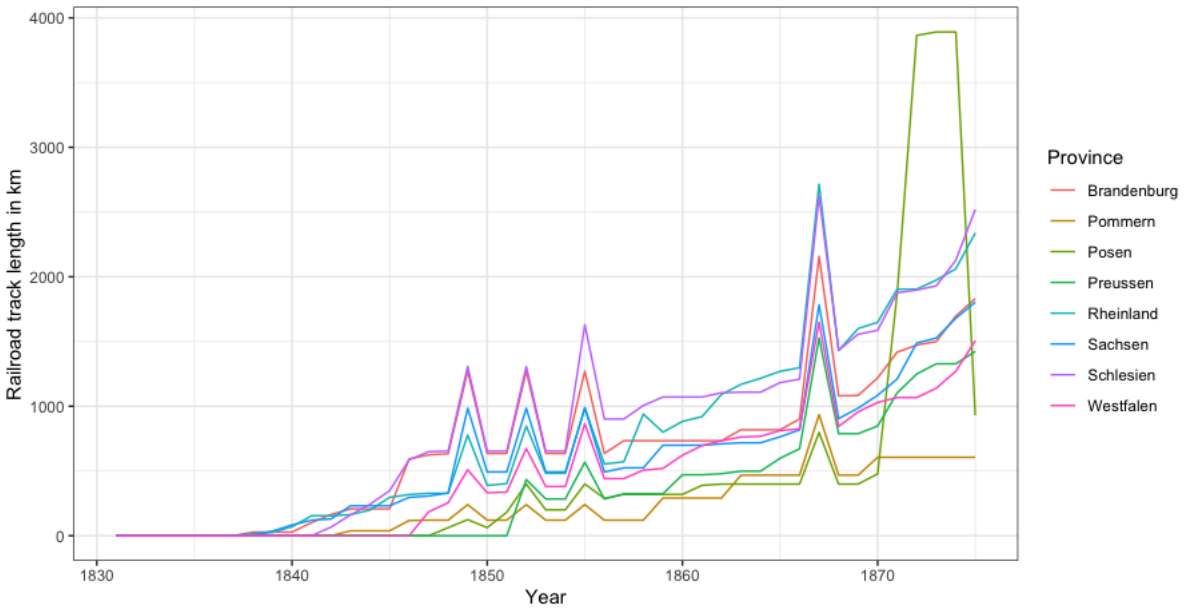

Figure 9: The expansion of railroad track over time for each province, 1831-1874. The counties Friedeberg and Posen incl. Stadt are excluded in this plot.

The expansion of railroads in figure 9 paint a picture that all provinces had similar trends but the differed at the pace the expanded their network over the years. We also see spikes certain years that I interpret as some measurement errors which is also why I removed the 2 counties from the plot and from my models. But I will also post same models with those two counties for transparency in the annex part.

*Table 1: Models created in stata version 16.1, using strongly balanced paneldata with gaps over certain years between 1831-1874. The counties Friedeberg and Posen incl. Stadt are removed from the models. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

VARIABLES	(1) Model1	(2) Model2	(3) Model3	(4) Model4	(5) Model5	(6) Model6	(7) Model7
railroad	0.0302	0.00744	0.0302	0.00744	0.0162	0.0130	0.0676
	(0.0555)	(0.0506)	(0.0555)	(0.0506)	(0.0450)	(0.0436)	(0.0596)
area						0.0225	
						(0.0148)	
Constant	51.75***	52.07***	51.75***	52.07***	35.28***	16.93	53.69**
	(4.680)	(0.724)	(4.680)	(0.724)	(7.838)	(17.84)	(26.22)
County fixed effects	No	Yes	No	Yes	No	No	no

Time fixed effects	No	No	No	No	Yes	Yes	No
County x Time fixed effects	No	No	No	No	No	No	yes
Standard errors	Robust	Robust	Cluster county	Cluster county	Cluster county	Cluster county	Cluster county
Observations	5,940	5,940	5,940	5,940	5,940	5,940	5,940
Number of counties	330	330	330	330	330	330	330

My estimates are not significantly different from zero at 1%, 5% or 10% significance level for any of my models which means that we cannot reject the null that railroads do not affect mortality rates for counties in Prussia during 1831-1874.

Model 1 & 3 and Model 2 & 4 are similar with the difference being the standard errors, clustering the standard errors should show some effect which we don't see which makes me believe something went wrong somewhere. It didn't really matter if I used robust or clustered standard errors for most of my models, with the only exception being model 7 where it worked fine. Maybe it had to do with the panel regression command xtreg in stata. But I anyways used it as I believe it to be the correct standard errors.

In the Hennock paper they discussed that sanitary reforms came from local governments which is why county fixed effects are important. There are also other unobservables that correlate with railroad and mortality rate at the county level such as demographic variables such as population density, income levels and occupation difference, economic development, healthcare, environment and much more discussed. Time fixed effects are bad here because the cholera outbreak themselves are these shocks that mostly effects all counties at the same time meaning that they will be caught up by the model, this is why model 5 & 6 have such small estimates.

I find it however interesting to take year x county fixed effects which consider that sanitary regulation and other county characteristics change over time. I believe it very likely that with urbanization, industrialization, new sanitary regulation and investments into sanitary facilities and other things change counties significantly over time and shouldn't be assumed to be fixed over time. But the time period used might not capture this change as it occurred after 1860s and out latest data point is at

1874. All these changes accelerated over time and became even more significant during the 90s, 1900s and beyond.

Conclusions

The results states that we can't reject the null in favor of the hypotheses that railroads have an effect on mortality rate in Prussia's counties during the period 1831-1874. It is however far from unsettled considering that we still have endogeneity issues. It is difficult to separate whether the effect come railroads or if we are measuring something else such as the agglomeration effect on cholera, the effect of economic development of cholera, the effect of industrialization of cholera, the effect of sanitary technology on cholera. It is difficult because everything is so interconnected, and it changes depending on all the other variables. One way of solving this simultaneity bias could be through creating an instrument like in (Donaldson, 2018) to get railroad expansion as an exogenous random shock instead. Then the building of railroads wouldn't be endogenous with all these other variables, and we would be able to estimate something close to a causal effect.

Looking at section 2 there are so many observed channels through which the building of railroads effect mortality rate. Proving it however is a difficult task due to the nature of the problem, railroads are not directly causing incidences of cholera or if it does it is not by a significant amount, all the big channels are through agglomeration, the increase in globalization of goods and ideas and through the increased exposure to diseases. And those channels are themselves not very concrete things that are easy to capture. The open question for my seminar paper then becomes: How can we create a model that distinguish between agglomeration, economic activity and exposure to cholera caused by railroads compared to everything else?

References

EASTERLIN, R. A. (1999). *How beneficent is the market? A look at the modern history of mortality.* Cambridge university press.

Hornung, E. (2014). *Railroads and Growth in Prussia.* Mac Plack institute for tax law and public finance.

Database, T. W. (n.d.). *Income inequality, Prussia, 1874-1918*. Retrieved from The World Inequality Database: https://wid.world/country/prussia/

Snow, J. (1855). *On the communication of Cholera* . London.

Alfani, G. (2022). *Epidemics, Inequality, and Poverty in Preindustrial and Early Industrial Times.* Journal of Economic Literature .

Hennock, E. (2000). *The urban sanitary movement in England and Germany, 1838±1914: a comparison.* Cambridge UniŠ ersity Press.

GALLARDO-ALBARRAN, D. (2020). *Sanitary infrastructures and the decline of mortality in Germany, 1877–1913.* Economic History Review.

Donaldson, D. (2018). *Railroads of the Raj: Estimating the Impact of Transportation Infrastructure.* American Economic Review.

Cholera - Vibrio cholerae infection. (2022, november). Retrieved from CDC (Center for Disease Control and Prevention): https://www.cdc.gov/cholera/illness.html#:~:text=Cholera%20is%20an%20acute%20diarrheal,be%20severe%20and%20life%2Dthreatening.

Alfani, G. (2022). *Epidemics, Inequality, and Poverty in Preindustrial and Early Industrial Times.* Journal of Economic Literature.

Jedwab, R., Johnson, N. D., & Koyama, M. (2022). *The Economic Impact of the Black Death.* Journal of Economic Literature.

Annex

*Table 2: Models created in stata version 16.1, using strongly balanced paneldata with gaps over certain years between 1831-1874. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

VARIABLES	(1) Model1	(2) Model2	(3) Model3	(4) Model4	(5) Model5	(6) Model6	(7) Model7
railroad	0.0208*	0.00770	0.0208*	0.00770	0.0123	0.0116	0.0341
	(0.0108)	(0.00569)	(0.0108)	(0.00569)	(0.00909)	(0.00882)	(0.0331)
area						0.0238	
						(0.0148)	
Constant	52.91***	53.20***	52.91***	53.20***	36.17***	16.68	53.30**
	(4.671)	(0.127)	(4.671)	(0.127)	(7.855)	(17.78)	(26.08)
County fixed effects	No	Yes	No	Yes	No	No	no
Time fixed effects	No	No	No	No	Yes	Yes	No
County x Time fe							yes
Standard errors	Robust	Robust	Cluster county	Cluster county	Cluster county	Cluster county	Cluster county
Observations	5,976	5,976	5,976	5,976	5,976	5,976	5,976
R-squared		0.000		0.000			0.136
Number of counties	332	332	332	332	332	332	332

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1