Wageningen University - Department of Social Sciences

Developments Economics

Determinants of Ebola intensity in rural Sierra Leone

A panel-data analysis of district-intensity

Remi van Balen

August, 2016

International Development Studies
Economics of Development
Supervisors: Maarten Voors & Jeroen Klomp
Thesis code: DEC-80433

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Determinants of Ebola intensity in rural Sierra Leone

A panel-data analysis of district-intensity

R. Z. van Balen

Development Economics Group
Postbus 8130
6700 EW Wageningen
www.dec.wur.nl

Author
R. Z. (Remi) van Balen, MSc
remivb91@gmail.com

Supervisors
Dr. ir. MJ (Maarten) Voors
Assistant Professor Development Economics
maarten.voors@wur.nl

dr. JG (Jeroen) Klomp
Assistant Professor Development Economics
jeroen.klomp@wur.nl

Acknowledgments

Writing this master thesis was a bittersweet run that continued quite a bit longer than expected. Among others; ongoing Ebola, minimal data availability, poor data quality, motivational lows and the interference of a very interesting internship caused delays. But, as a famous thinker once said; 'all good thinks come to an end'[a]. And because this end reached, I want to show my gratitude to all the people who have helped me climbing this small academic mountain. First of all, I want to thank Maarten Voors, my initial and first supervisor, for his patience, endless feedback sessions and support when I was in need of navigational guidance, feeling down, or racing ahead and committing one or another scientific cardinal sin. I am feeling thankful for the opportunity I was given to work with Maarten, and learn from his endless knowledge, scientific storytelling and efficiency. I also want to thank Jeroen Klomp for his useful feedback and his open door policy when I needed input for the econometric marathon I performed to make sense of my data. Furthermore, Jeroen became a most welcome and appreciated replacement for Maarten Voors in the last stretch of the thesis when Maarten worked at the University of California, Berkeley. I want to show my gratitude to Paul Hofman, who supported and motivated me during the process, and offered his feedback and help, regardless whether it was ten in the morning or twelve at night. Kevin Griecco, you were not only a useful sparring partner because your critical feedback on my work, but also because of the pleasant squash breaks. Last but not least, special thanks goes out to family and friends, who have been supporting me during my entire study in so many different ways, your love and friendship was invaluable.

Thank you all.

Remi

[a] Nelly Furtado, All good things, 2006

Abstract

This thesis tries to unravel some of the mechanism underlying the 2014-2015 Ebola outbreak in rural Sierra Leone. We investigate the relation between the rapid spread and high intensity of the Ebola outbreak in rural Sierra Leone and the socioeconomic determinants in this region. We test whether, controlled for local outbreak characteristics, there is a relation between the spread of the virus and population density, treatment intensity, income levels and cultural practices and norms. Due to the lack of quantitative testing it is unclear what factors determined the intensity of the Ebola outbreak. Unravelling the black box around the determinants of the Ebola intensity will help to understand why and how this outbreak turned into the epidemic the world has seen. This knowledge would enable a more adequate response to future outbreaks in Africa. We use a large panel data set on district levels, and analyse this using a combination of random-effects, fixed-effect and Tobit estimators to estimate our parameters of interest. We find strong and stable effects of space and time transmission effects within and between districts. We find no effect of treatment capacity or safe burials on Ebola intensity. Furthermore, from our data set we cannot derive any effect of socioeconomic or district characteristics on Ebola intensity. These results and non-results are largely driven by problems in our data set. The quality and quantity of the data available for the public on the Ebola outbreak is insufficient to draw any further conclusions. This study recommends further research into the response, socioeconomic and geographic mechanisms driving the spread of Ebola.

Contents

1. Introduction

West-Africa is recovering from an epidemic of the Ebola Virus Disease (EVD) which was several times larger in magnitude and length than all previous Ebola outbreaks together (WHO, 2015). Following WHO figures, Sierra Leone, Liberia and Guinea have had over 28.000 Ebola cases and 11.000 Ebola deaths during the course of the outbreak. It is clear without a doubt that this epidemic has been a "public health emergency of international concern'' (WHO, 2015). The overall impact of the epidemic on Sierra Leone has gone far beyond case counts and deaths (Alexander, 2015). The West-African countries faced a big drop in GDP estimates, and also the long term estimates are considerably lower than before the epidemic (Ippolito, Puro & Piselli, 2015; UNPD, 2015). Quarantine measures and fear stalled social and economic movement in the country. Furthermore, shortages in treatment capacity led to increased mortality for endemic diseases like malaria, pneumonia, and diarrhea (Alexander, 2015). Last but not least, the expected but not yet fully researched stigmatization of Ebola survivors might harm the livelihoods of survivors and affected families. Although it is too early to jump to conclusions with so little research done, the Ebola outbreak, and the following humanitarian crisis might have reversed years of economic progress in Sierra Leone.

Although this was certainly not the first outbreak of the Ebola virus in Africa, the intensity and fast spread has raised many questions in the world of researchers and policy makers. Due to the lack of quantitative testing it is unclear what factors determined the intensity of the Ebola outbreak. This hampers the learning that should take place after this outbreak. Unravelling the black box around the determinants of the Ebola intensity will help to understand why and how this outbreak turned into the epidemic the world has seen. This knowledge would enable a more adequate response to future outbreaks in Africa (Bausch & Schwarz, 2014). This thesis tries to unravel some of the mechanism underlying this Ebola outbreak. This increased knowledge might increase efficiency of the response in future outbreaks. The underlying research question of this thesis is*: 'What are the main determinants of the Ebola intensity in rural Sierra Leone?'* We answer this question by analysing the short-term causal effects of possible determinants on the district-time variety in Ebola intensity in rural Sierra Leone. The most accurate measure of Ebola intensity, our dependent variable in the model, is case-counts per district per day. However, because this data was not available for us, we used the number of Ebola alerts to proxy for Ebola intensity. The determinants we regress on Ebola intensity are deducted from the quantitative and qualitative literature already written on the Ebola outbreak. Based on data availability we included several outbreak, response, individual socio-economic and district determinants in our model. We have excluded Western-Area from our sample, because it was a clear outlier in our data, suggesting that very different mechanism were determining the course of the outbreak in this urban area.

We use a combination of random-effects, fixed-effect and Tobit estimators to estimate our parameters of interest. We find strong and stable effects of space and time transmission effects within and between districts. We find no effect of treatment capacity or safe burials on Ebola intensity. Furthermore, from our data set we cannot derive any effect of socioeconomic or district characteristics on Ebola intensity. These results and non-results are largely driven by problems in our data set. The quality and quantity of the data available for the public on the Ebola outbreak is insufficient to draw any further conclusions. This study recommends further research into the response, socioeconomic and geographic mechanisms driving the spread of Ebola.

The remainder of this thesis is structured as follows: Section 2 reviews the context in which this research is embedded. Relevant factors of Sierra Leone and the Ebola outbreak are covered. In section 3 we review the written literature and formulate hypotheses on the determinants of Ebola intensity. In section 4 the presents our data and in section 5 our empirical strategy is introduced. Section 6 presents our different

panel regression results and a small robustness analysis. Finally in section 7 we take stock, discuss our results, place them in perspective and formulate further research and policy recommendations.

2. Context

Sierra Leone is extremely resource-rich, and the climate is suitable for the production of a wide variety of crops. Even so, widespread exploitation, bureaucracy and corruption have led to a poverty rate of 70% and the country is ranked 181th out of 188 on the 18th place on the Human Development Index (UNDP, 2016). The same apparent contradiction, resource-rich and fertile, and yet extremely poor, can be found by the other three countries hit by this Ebola outbreak. Research suggest that specific characteristics of this weak state of development were an important determinant in this crisis. This section will firstly take a quick glimpse at the history and current state of the institutions, economy and living standards in Sierra Leone. Secondly, we will study the course, response and impact of the 2014 Ebola outbreak.

Sierra Leone

Sierra Leone, officially The Republic of Sierra Leone is a West-African country bordering Guinee, Liberia and the Atlantic Ocean. It has three provinces, which are subdivided into 14 districts. The 5.7 million Sierra Leoneans are mostly young (40% is below 15 years old), live in the rural area (62%) and either Muslim (70%) or Christian (25%). In 1961 Sierra Leone became independent from Great Britain. And although in 1962 the first democratic elections were held, the system of indirect rule introduced by the British colonial rulers stayed in place. This system used existent power structures at the local level to handle the day-to-day ruling. Several theories suggest that this system paved the path for rulers mainly seeking to exploit resources and safeguard their positions instead of invest in public goods or enforce law and order (Hobsbawm & Ranger, 1983 in Acemoglu et al., 2014). These and other (political) factors played a role in the start of the civil war between rebels of the Revolutionary United Front (RUF) and the Sierra Leonean Army (SLA). After a big and bloody rebel attack on Freetown in 1999, a UK and UN peace force intervened in the conflict and brought an end to the war (Bellows and Miguel, 2009). In the civil war about 50.000 were killed, and thousands more were brutally wounded, raped or in a different way assaulted. Following rough estimations over a million people became refugee (Bellows & Miguel, 2009). Several studies try to connect this weak institutional quality and the war history with the current development status of Sierra Leone.

The economy of Sierra Leone grows with double digits since the end of the war. But still, 50% of the population lives below the $2.00 a day, and 80% below $3.00 a day (World Bank, 2016). Before the Ebola crisis, Sierra Leone was slowly starting to become less donor-dependent but finding strategies to fight the high youth unemployment, failing fiscal policies, poor infrastructure and earlier mentioned institutional problems (Heritage Economic Freedom Index, 2015) continues to be a challenge. Due to low school enrollments, 5% of the population above than 15 years old is illiterate and as today, only 45% of children follows any secondary education (World Bank, 2016). The civil war destroyed most of the country's infrastructure, the roads, bridges and railways remaining were still dated from before the independence. In the past years, Sierra Leone started to rebuild their infrastructure. Supported by international partners, new roads, railway and bridges are developed, but there is a long way to go. Based on estimations published in the 2011 report on Sierra Leone's infrastructure by the World Bank the efficiency of the logistical system scored a 1.5 out of 5. Only 1% of the rural population has access to piped water or electricity only 2.1% of the total population has regularly access to the internet, and large parts of the rural area of Sierra Leone have no phone coverage (Foster, 2013). The low phone coverage and internet access limit the access to information for the rural population, combined with the weak infrastructure, this places them far out of reach of information or help in case of an emergency or crisis like the Ebola outbreak of 2014. The health care system in Sierra Leone improved a lot in in the years before the Ebola outbreak, but it was prior to Ebola still in a very weak state. Following the World Bank, in 2014 Sierra Leone had the highest percentage of maternal deaths in the world, there were only 1000

educated nurses and around 130 doctors (1 doctor for every 50.000 inhabitants). This, combined with the fact that only 13% of the population has access to improved sanitation facilities and the earlier mentioned high percentage of people living below the poverty line led to the low life expectancy of 57.39 years (World Bank 2016; Life Fact Book, 2015). Prior to the Ebola outbreak, there were very few functioning ambulances In Sierra Leone. In the Freetown region, 5 ambulances had to serve 2 million people. The low capacity of hospitals, nurses, doctors and ambulances led to the estimation of the WHO in September 2014 that Sierra Leone was able to treat effectively just over 500 Ebola patients nationwide. At the time of this estimation, the outbreak infected around 500 people every week in Sierra Leone, which meant treatment capacity was lacking way behind treatment needs (WHO, 2014).

The 2014 Ebola outbreak

The virus of the 2014 outbreak was the Zaire strain, one of the five variants of the zoonotic virus (Bausch, 2014). The Ebola virus originates from an animal reservoir and is extremely contagious and deadly. Although the appearance of the Ebola virus among humans is quite uncommon, it can be transmitted to humans through the contact with infected animal body fluids like saliva and blood, mainly caused by hunting and eating of bush meat (Fauci, 2014). It is believed that the index case, a boy in Méliandou, Guinea in December 2013 (Saéz et al., 2015), was caused by the bush meat scenario (Richards et al., 2015). After the introduction to humans, the virus spread from human to human through the contact with body fluids, especially taking place during the care of patients in the wet phase of the disease (Bausch & Schwarch, 2014) About twenty outbreaks among humans have been reported since 1976, resulting in a total of around 1600 deaths before 2014. In table 1 the impact of the 2014 crisis is shown compared to earlier outbreaks in Africa. During the 2014 outbreak a total of 17,849 cases was reported (of which 15,277 confirmed and 2,622 probable) of these cases, 11.310 people died. Another 10,767 people became sick or died under suspicious circumstances, but were never tested (WHO, 2016). The second biggest outbreak was in 2000-2001 in Uganda, and 'only' took 425 cases.

*Table 1: Six largest Ebola outbreaks chronological**

Date	Index	Countries affected *	# of Cases***	# of Deaths	Mortality***
1976	Sudan	Sudan	284	151	53%
1976	DRC	DRC	318	280	88%
1995	DRC	DRC	315	250	79%
00-01	Uganda	Uganda	425	224	53%
2007	DRC	DRC	264	187	71%
'14- '16	Guinea	Sierra Leone, Liberia, Guinea, Nigeria, Mali, United States, Senegal, United Kingdom, Italy, Spain	28,616	11,310	40%

** Data retrieved from Alexander et al. (2014)*
*** Countries are ranked according to number of cases*
**** Number of Cases and Mortality rate is based on confirmed, probable and suspected cases*

Course of outbreak

A woman from Kailahun is most named as the first Ebola case in Sierra Leone. After visiting her family in Guinea she travelled back to her village in Sierra Leone. The first big outburst of cases was in late May after a funeral of a traditional healer in Sokoma, Kailahun. Rough WHO estimates state that at least 365 people died from Ebola after attending this funeral. This funeral became the start of multiple chains of infections spreading over the country causing more cases, more funerals and eventually even more cases (Fauci, 2014). After this outburst of the epidemic, the disease started to spread over the country. There are remarkable geographical patterns to find in the spread of cases over the country, which makes sense since the virus spread through human to human contact, and patients had to travel long distances to get treatment (Alexander, 2014). During the course of the outbreak, it became very clear that the spread knew a geographical pattern. In figure 1 we see the cumulative number of Ebola cases per district on 12

august 2014. The darker two districts are Kenema and Kailahun were the Ebola outbreak started for Sierra Leone. During the summer of 2014, the virus spread over a large part of Sierra Leone, to arrive in the capital Freetown in July. In September the virus spread so fast in Freetown that the corpses were left on the streets waiting for the burial teams to collect them (New York Times, 2015).

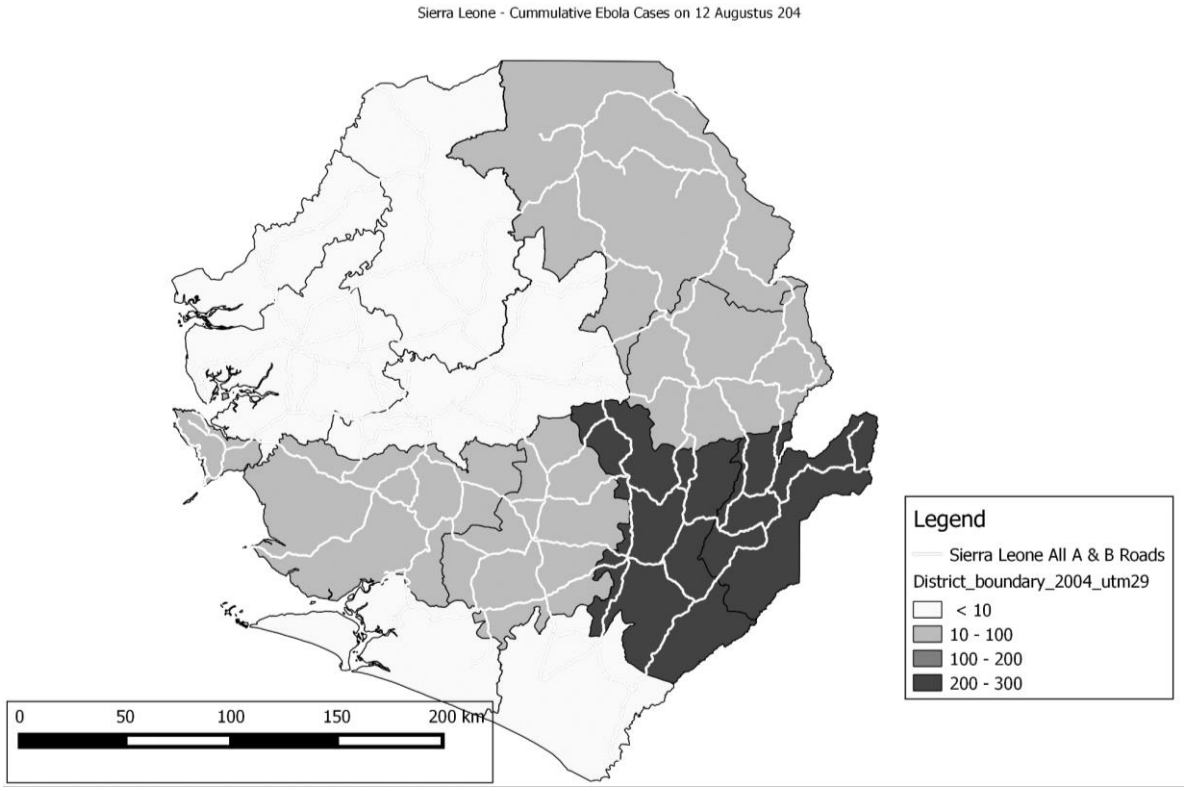

Figure 1 – Cumulative Ebola cases, August 2014

In the autumn of 2014, North-Western districts of Moyamba and Port Loko are reached by the virus. The outbreak continued to gain pace until December 2014. Between October and December, the number of new confirmed cases per day fluctuated between 60 and 120, resulting in the total cases doubling every 20 days (Saéz et al., 2015). In figure 2 we see the cumulative case counts in December 2014, we clearly see that the epicentre of the crisis has shifted from the east to the west. In the first epicentres of the outbreak Kenema and Kailahun, the response system worked efficiently together with local communities and international partners and succeeded in kerbing the disease. Case counts went down, and improvements of reporting, testing contact tracing and treatment supported by local community initiatives helped to stop the uncontrollable spread of the virus (Alexander, 2014). In the unique circumstances of Western-Area, the response system has difficulties to stop the disease from spreading. Just like in the South-Western part of the country.

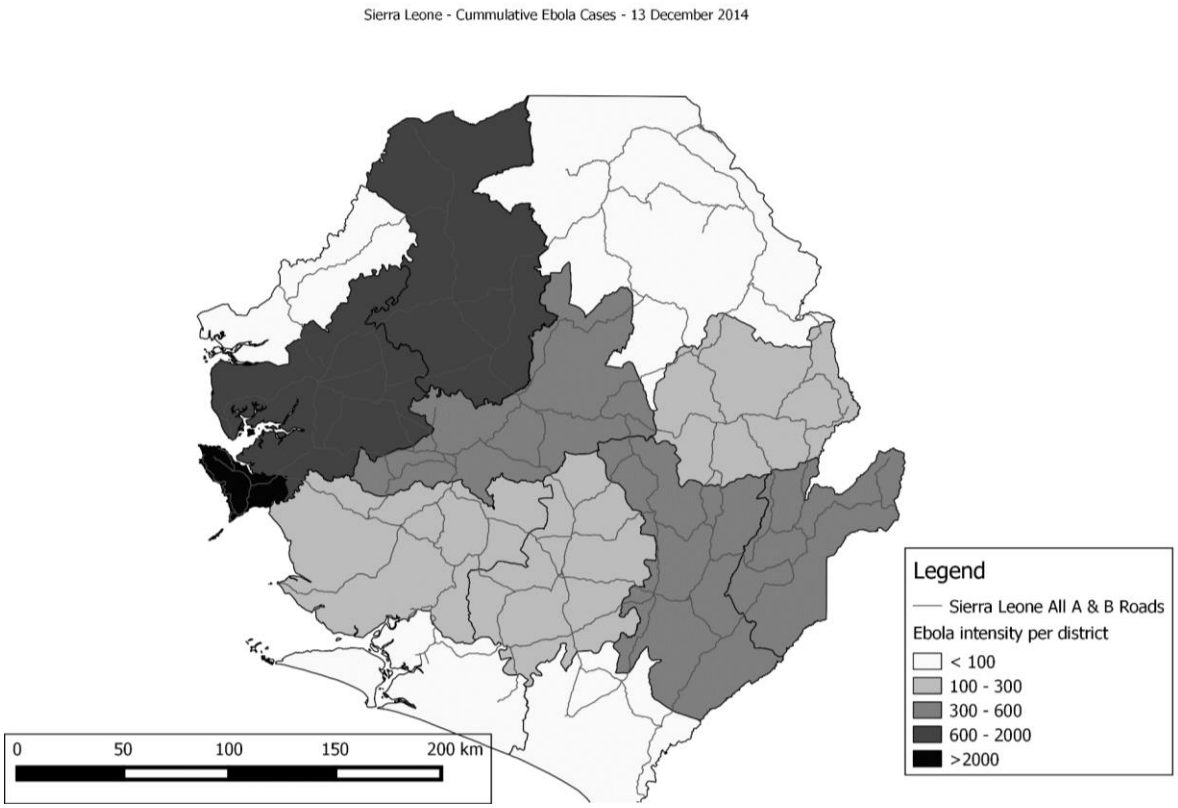

Figure 2 - Cumulative Ebola cases, December 2014

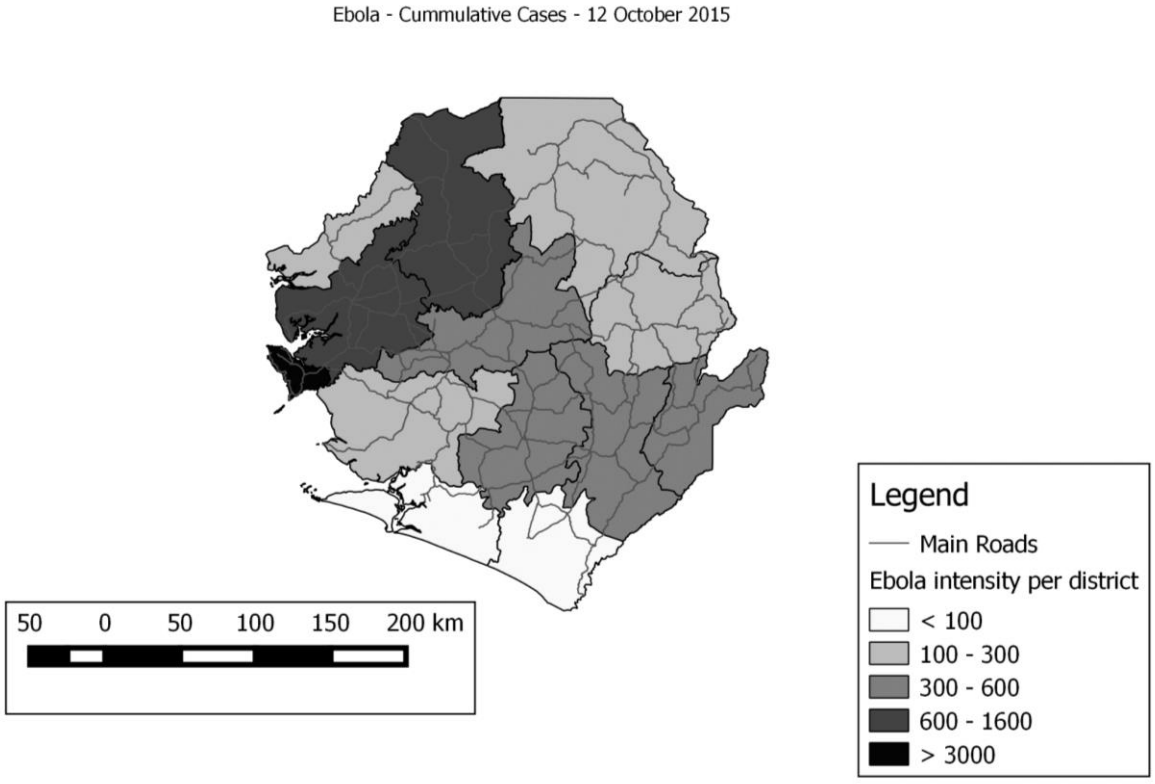

Figure 3 - - Cumulative Ebola cases, December 2014

During the spring of 2015 Sierra Leone witnessed its first days, and later on weeks without new cases. However, the virus had kept showing regular flare-ups during the remainder of 2015, mainly in the Western districts Port Loko, Moyamba and Western-Area. This is shown in figure 3, were we see that these districts remain the most severely hit by the outbreak. It is important to keep in mind that the figures present cumulative case counts. However, from the legends of the different figures, we can see that the case counts in these Western districts keep increasing compared to the eastern districts. In November 2015 Sierra Leone was declared Ebola-free for the first time, but in January 2016 two new, isolated, cases were discovered in the Eastern part of the country (WHO, 2016). At the 17th of March 2016, the government of Sierra Leone declared the country, for the second time, officially Ebola-free, since 42 days had passed since the last person confirmed to have Ebola tested negative for the second time (WHO, 2016). Possible reasons for the geographical patterns can be found in differences between areas in response activities, trade and travel regulations, institutional and community action and other socioeconomic factors. Some of these factors will be tested in this study.

The response

A lot has been said about the national and international Ebola response. The response is often labelled as insufficient and inadequate. Phillips and Markham (2014) argue that the response was a failure of international collective action. Furthermore, the outbreak is often described to be out of the control of national and international response institutions (Otter, 2014). However, the Sierra Leonean government, supported by the WHO, international partners, NGOs and community initiatives initiated a response system. This system was divided into four pillars; case management, each focusing on another dimension of the outbreak. The different response parties each rolled out their own activities in one or more of the pillars to support the Sierra Leonean to curb the spread of the virus. When the outbreak started in Sierra Leone the Emergency Operations Centre (EOC) coordinated by Ministry of Health and Sanitation led the response. But after the big outburst of cases in the summer of 2014 in the Western Area and the Southern districts, this centre was replaced by the National Ebola Response Centre (hereafter NERC) due to inefficiencies and malfunctioning of the Ministry of Health. The NERC was led by the Ministry of Defence and was subdivided in District Ebola Response Centres (DERCs). This new system provided regional and national locations were all response activities were managed and tracked. Within a week a new efficient reporting system was introduced, and daily reports on district level were made to improve the efficiency of the response (Response source).

The first pillar, case management was focussed on breaking the chain of transmission by testing and treatment in isolation. Every patient entering a treatment centre, community care centre or hospital showing symptoms of Ebola was tested for Ebola using a laboratory test and cared for in a special zone with other possible Ebola patients. If the lab tests confirmed an infection with the virus, the patient is moved to a designated treatment area for confirmed cases (Cecchi et al., 2015; WHO, 2015). The second pillar started to come in action once a chain of transmission was identified, partners in pillar two started with case finding and contact tracing to isolate every single case as early as possible. Contact tracing is used to trace down all the human encounters an infected person had during his infectious period. Quarantine is used as a measure during which the contacts get tested regularly for symptoms (Mogelson, 2015). Once a patient is confirmed with Ebola and taken to the hospital, their relatives are forced to spend 21-days at their home, to be monitored for signs of infection (Chandler et al., 2015). This complex operation was led by the DERC of the specific district, assisted by local governmental agencies and NGOs. Furthermore to limit the propagation of the virus the government introduced travel regulations, market, schools and border closures and district or chiefdom quarantines for areas where the disease showed to be out of control. The facilitation and promotion of safe burial practices was the third pillar of the response system. Cultural traditions in Sierra Leone around dead and funerals consists of many traditions and practices including washing, touching and kissing the dead body before it is buried. These traditions are performed to ensure the deceased an afterlife, and abandoning these practices was very hard for

people. Active promotion to abandon these practices and assisting families with safe burial teams were in place to improve this situation (Chandler et al., 2015). Communication, social mobilisation and community engagement was at the very core of the Ebola Response and therefore the fourth pillar. The principle underneath this strategy was to change risky behaviour, inform people about the causes of the Ebola and dispel myths around the virus. This was done through messages on radio, TV, SMS and printed in cities and social gathering places. A good example is a door-to-door campaign during the *'Ose to Ose Ebola Tok'* initiative. This initiative was a three-day lock-down of the entire country in September 2014 to stop the uncontrollable spread of the disease and search door to door for people showing signs of the virus. During this door to door, campaign people were informed by health workers about the threat and causes of the Ebola virus, and safe practices were advertised. Community driven action was also part of pillar four. In different chiefdoms and districts chiefs, imams, priests and local political leaders worked together to implement strategies and rules to fight the disease. In Kenema district, this gathering, initiated by local doctors and the paramount chief, resulted in 30 laws aimed at limiting the spread of the disease. For failing to report a new case or taking in sick people could be fined or jailed (Economist, 2015). The intensity of the community response differed from district to district and from chiefdom to chiefdom. In some chiefdoms the Ebola task force worked closely together with the NERC, village leaders and representatives of civil society groups to mobilise the community, solve logistical problems and set up efficient reporting systems (WHO, 2014). At the same time, there were other districts or chiefdoms where this ownership of the response was not quite of this level.

The ten biggest international supporters of the Ebola outbreak response together donated over 3.6 billion dollars (USAID, 2016). These ten donators included the governments of the USA, UK, Germany, the World Bank and the European commission. They donated these funds through non-governmental organisations like the Doctors without Borders, Goal and Red Cross, through (inter)governmental bodies like USAID, CDC and the World Health organisation. With this support the surveillance and reporting systems of the NERC and DERC were supported, treatment centres were built, equipped and staffed. Local staff was trained in treatment, prevention and control. Testing blood samples of probable cases was at the start of the crisis very hard, since there were only a few laboratories in the country able to do this tests, but later on in the crisis the international response partners had set up laboratories in virtually every bigger city in all the affected districts (NGO activity source).

Impact

The Ebola outbreak has impacted Sierra Leone through different channels. Firstly there is the death toll; out of 14,122 suspected, probable and confirmed cases 3955 people died (WHO, 2016). Next to the death toll, this large number of patients drained private and public resources and caused large public distress and it led to reduced generation of income and labour supply and thus slowing down of economic activity.

The health care system of Sierra Leone suffered enormously from the heavy burden of the Ebola patients, this led to increased mortality under non-Ebola patients continuing long after the peak of the outbreak (Saéz et al., 2015). Next to that people lost trust in the conventional health care system, and travel regulations kept people away from conventional and traditional treatment possibilities. Morse et al. (2016) find weak effects of socioeconomic factors and supply-side shortage for the reduced health care usage of non-Ebola patients, they argue that the larger part of this decline is caused by "distrust and negative Ebola-related experiences" (Morse et al., 2016). The fear and trust combined with the supply side capacity problems resulted in dramatic setbacks in prevention, control and treatment of HIV, tuberculosis and malaria (CDC, 2016).

The Ebola virus itself affected families and communities hit by the virus in many different ways, but economic effects at the national level were mainly a result of the quarantines and market closures (Davis, 2015). These measures, aimed at reducing the probability of the virus travelling to non-affected areas and

people, forced people into social and economic isolation (Saéz et al., 2015). Travel regulations and market closures had a significant effect on communities dependent on agriculture and trade. Glennester & Suri (2015) show in an on-going research on market participation and prices that the Ebola crises has lowered the number of traders and total supply and increased prices of a variety of goods in some districts. Furthermore, a FAO report on Ebola's impact on agriculture showed that the outbreak led to a decrease in farmers' income due to higher costs and lower prices received for cash crops; cash crop farmers were bothered most by travel restrictions (FAO, 2016). At the national level, the Ebola crises caused a decrease in national and international investments in the region rough World Bank (2015) estimates showed a revision of the sector-specific growth projections of 3.3 percent point of GDP in Sierra Leone, reducing growth from 11.3 percent to 8.0 percent. And the short-term fiscal impacts for Sierra Leone were estimated at US$95 million (2.1 percent of GDP for Sierra Leone). The medium-term impact (2015) on output in Sierra Leone was estimated to be between 1.2 percent point and 8.9 percent point of GDP. The long run effects are yet to be discovered. However, it would be a little premature to attribute all these numbers solely to the Ebola outbreak, just before, and during the crisis two state-owned banks and two multinational mining companies have collapsed (Davis, 2015). The determinants for the economic setback of Sierra Leone are a complex set of different factors, and the real relations between these factors have yet to be found.

There is enough reason to suggest that the context of the Western African countries hit by the Ebola outbreak partly determined the intensity of virus. This study will try to unravel some of these complex relations.

3. Literature Review

The duration, rapid spread and intensity of this outbreak surprised many international actors. Because, even though the Ebola virus can spread extremely fast, it is relatively easy to treat and contain. As long as patients are quarantined as quickly as possible, the spread of the disease can be managed with testing of close contacts, basic medical treatment and hygiene and disinfection in the place of treatment. This aligns with the fact that researchers and policy makers expected that Ebola would never become a large public health problem in Africa (Fauci, 2014). Because of these expectations and the successful and efficient manner in which other outbreaks were contested one might wonder why this outbreak escalated in this manner in West-Africa. In this section, the literature on the determinants of the Ebola intensity is reviewed. Based on the determinants mentioned in the literature, hypotheses for this study are formulated. The determinants found in the literature can be grouped into four categories, each consisting of several factors. The first determinant mentioned in the literature is the intensity of the virus in previous periods. Secondly, the response quality, mainly expressed in treatment capacity. Thirdly, individual-level socioeconomic status. And finally fourth category district-specific characteristics like population density, institutional and infrastructural quality are named as determinants of the spread of the Ebola virus. For all of these determinants, the hypotheses that will be tested in this study are specified.

Ebola effects

The first and most important determinant of Ebola intensity in a specific region should be the number of infected people in that region (WHO, 2016; Alexander, 2015). Since the Ebola virus is spreading through human to human contact, more cases in a specific area increases the risk of being infected for other individuals in that area. Plenty of studies investigate the dynamics of the Ebola outbreak using case count data (i.e. Rainisch et al., 2014; Fallah et al., 2015, Fang et al., 2016; Camacho A, et al. 2015). Where most studies 'just' analysed and calculated reproduction rates and forecasted future transmission and case counts, a few studies try to estimate overall coefficients for the impact of previous case counts on current or future case counts. For example, Rainisch et al. (2014) analyse case count data from Sierra Leone and Liberia. Their sample includes the days between April 2014 and August 2014 and all the by that time affected districts (Sierra Leone) and prefectures (Liberia). They find that the number of cases in an area can be predicted by previous case counts, population data, and distances between affected and not affected areas. Another study that finds similar results is the study of Fallah et al. (2015), they analyse 4,437 cases divided over three hundred communities between March and December 2014. They categorised these communities based on social economic status and analyse the effect of previous Ebola and other factors on new case counts. Fang (2016) adds another spatial dimension to this story, in their study they show a weak effect of an increase in cases in neighbouring districts on case counts. Based on these and other results, we expect that all else being equal, more Ebola in previous periods in a specific districts and in neighbouring districts means more Ebola today. Since Alexander (2015) argues that the risk of getting infected is much higher in densely populated areas, we also expect that the transmission effect depends on population density.

H1: Ebola intensity in a specific area in previous periods is positively correlated with the Ebola intensity in that specific area.

H2: The effect of previous Ebola intensity on current Ebola intensity is stronger in densely populated districts.

H3: Ebola intensity in all districts bordering a specific district is positively correlated with the current Ebola intensity in that specific area.

Treatment capacity

The second category of factors influencing the intensity of Ebola cases in a district is the speed, quality and capacity of the response. The Ebola response was hampered by severe treatment capacity problems (WHO, 2016; add other sources). Shortage of space in treatment centers and hospitals, shortage or non-functionality of ambulances, waiting time for burial and disinfection teams and slow contact tracing were only a few examples of this capacity problems (Alexander et al., 2014; Bausch & Schwarz, 2014; Richards et al., 2014; Kurcharski et al., 2015). Several studies claim that these capacity problems have accelerated the disease. Bausch & Schwarz, (2014) claim that better training of staff and a slight increase in treatment capacity and equipment could have prevented thousands of cases. Kurcharski et al., (2015) showed in a mathematical model that the number of available beds in treatments units averted 60.00 Ebola cases resulting in a lower reproduction rate. This is supported by figure 4 from Fang (2016), which shows that the number of healthcare workers that were infected with Ebola was in absolute and relative terms higher at the start of the outbreak when the response was still inadequate. To proxy for this improvement in quality, we use the number of available treatment beds. We expect that more treatment capacity, and thus a higher treatment quality, will reduce the Ebola intensity.

H4: The treatment capacity before today in an area in is negatively correlated with the Ebola intensity in a district.

92016

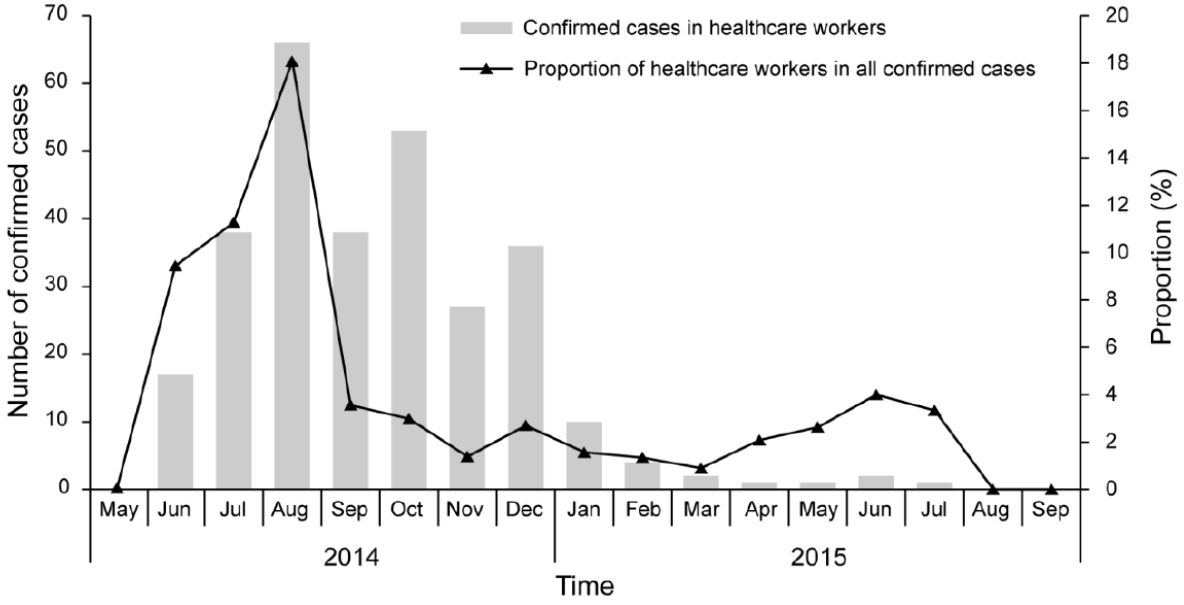

Figure 4 - Absolute and relative number of confirmed cases among healthcare workers. Source: Fang (2016)

Socioeconomic status

The third category of determinants is awareness, human behaviour, cultural practices and socioeconomic factors. The reason most forecasting studies were poorly able to forecast the outbreak might be that they ignored the effect of these factors on transmission rates. The variation over time and space of these factors might be causing the fact that these models outcomes often did not square with the real world data (Butler, 2015). There is a group of studies that point to unsafe burial practices and non-compliance with other safety measures as one of the main reasons for the rapid spread of the disease. The safe burial procedures introduced by the WHO clash with the traditional burial practices in Sierra Leone, which

involves washing, touching and kissing of the corps (Richards, 2015; Mogelson, 2015). Chang (2014) state that traditional burial practices have been linked to the vast majority of the Ebola cases in Guinea. Early on in the crisis, a deep distrust of the government, NGOs and other international response workers started to develop. People were blaming foreign doctors, the government and others from spreading the Ebola virus (WHO, 2016; Allgaier, 2015). This resulted in non-treatment of Ebola patients and in this way to faster transmission of the virus and a higher death toll (Allgaier, 2015). Due to various conspiracy theories, mistrust and fear serious unrest started to develop up to the point that response workers were faced with serious obstruction to doing their job, up to physical violence. In Freetown, just like Monrovia and Conakry, this led to protest marches, riots and mob attacks on isolation and treatment centres (WHO, 2014; Reuters, 2014). Another shocking, and exceptional, example is the killing of a team of health workers and journalist in the village of Womey in South-East Guinea by angry and distrustful villagers (Allgaier, 2015). The extent to which this distrust and non-compliance occur in a region might have been an important factor in the spread of the disease, and it will be tested in this study.

H5: Compliance with the safety regulations in previous periods is negatively correlated with the Ebola intensity.

Next to cultural factors, socio economic factors like income and education might also be related to the number of new cases in a district. Higher educated and wealthier people might be seeking treatment in an earlier stage of their sickness, which increases their chances of survival, and minimises the transmission of the virus to people in their surroundings (Alexander, 2014; Richards et al., 2014; Fang et al., 2016). Fallah et al. (2015) find that infected individuals with a low or middle socioeconomic status are associated with more secondary cases. Therefore, they argue, that poverty status was positively related to rates of transmission. However, the evidence they show is lean, and the black box of unobserved effects influencing both Ebola intensity and socioeconomic status should cause caution in making too strong statements. For example there are district characteristics no one observes leading to both a higher income levels and lower Ebola intensity. Therefore we formulate only relational and no causal hypotheses on these socioeconomic factors.

H6: Income is related to Ebola intensity at district level; direction of relation unclear

H7: Educational level is related to Ebola intensity at district level; direction of relation unclear

District characteristics

The fourth and last category of determinants of the spread of Ebola are country specific characteristics like institutional level and geographical specifics. Alexander et al. (2014) argue population density in the urban areas, and poor infrastructure has contributed to the unprecedented expanse of the virus. Fang (2016) shows a weak result for the positive effect of population density and primary and secondary road proximity on the risk for a chiefdom of getting affected with the virus. However, a higher population density might also mean that hospitals are closer by, and this might decrease the Ebola intensity. Or a hospital, as epicentre of the disease, might have the opposite effect. For remoteness the same holds, one can argue that more remote areas have a smaller risk of getting affected by the virus, but maybe as the virus reach a remote place, it is much harder to kerb the disease. Therefore we are inconclusive in our hypotheses on the effect of population density and remoteness on Ebola intensity. Phone coverage was essential in warning the response teams for new cases. And radio and SMS services were an important tool to inform and educate people about risks and developments of the disease and outbreak (WHO, 2016). Lastly, the Ebola response measures differed per country, and the collaboration between the three countries was limited (Alexander et al., 2014). Since the long borders between the countries are very hard to control, and Ebola tend to travel across borders, we expect that districts bordering Liberia or Guinea had a higher Ebola intensity.

H8: Population density is related to Ebola intensity at district level; direction of relation unclear

H9: Remoteness is negatively correlated with the Ebola intensity.

H10: Phone coverage is negatively correlated with the Ebola intensity.

H12: Bordering Guinea or Liberia is positively correlated with the Ebola intensity

Time effects

In the time span of our sample, the Ebola intensity in all affected countries was non-linear downward sloping. This is caused by both measured, and unmeasured factors. To capture the general downward trend in the data, caused by increased international efforts, awareness and knowledge, we will include for a time trend in the model. We expect this trend to be downward sloping, so the relation between time and Ebola intensity to be negative. The Ebola data in our model is reported to the National Ebola Response Centre (NERC) by smaller district, chiefdom, and even village level health clinics, and response centres. There might be a reporting error in this data, based on the day of the week. To capture this error, we measure the effect of weekend days and Mondays on the Ebola intensity. It might be that some alerts in weekend days are only reported in the week after the actual alerts. Therefore we expect the relation between weekend day and Ebola intensity to be negative.

H13: There is a non-linear negatively time trend in the data

H14: There is a negative weekend reporting effect in the data

H15: There is a positive monday reporting effect in the data

Plenty of studies describe possible relations between socioeconomic factors and the spread and intensity of the Ebola crisis. Many argue that including context variables to the models may lead to models that are better in explaining the intensity of epidemics. However, very few studies tested all these factors simultaneously on a large data set. This study will test the hypotheses specified in this section using a larger sample than most of the earlier mentioned studies, we try to build on and verify the work of Fallah (2014), Rainish et al. (2015) Krauer et al. (2016) and Fang et al. (2016).

4. Data

This section will present and discuss the variables we use in our models. In annex, I, in table 5 the exact variable definition is presented. Furthermore, table five gives the sources of our raw data and summarizes our hypotheses by showing the direction of the effect that was expected before estimating the model.

Dependent variable

The dependent variable in our study is relative Ebola intensity (E_{it}), this is the number of new Ebola alerts reported to the different District Ebola Response Centers (DERCs) at day t in district I relative to (so divided by) the total population of a district as measured in the 2004 census. Every Ebola alert is a sick individual reported to a DERC by a health clinic, treatment centre, or through the Ebola hotline. This data was retrieved from the NERC website in December 2015 (NERC, 2015) The panel data set we created has a t (day) of 352, spanning the period from the 28th of October 2014 till the 25th of October 2015 and i = 13, being the thirteen[b] districts of Sierra Leone. The use of Ebola alerts to proxy for Ebola intensity, the aggregation on district level, and the timespan of our dataset is based on data availability at time of analysis. Although some papers use case count data at chiefdom level for a longer time span, we could not acquire this data. Our data set starts only in October 2014 because data on the first part of the outbreak is very limited since efficient reporting systems were not in place yet. The dataset ends in October 2015 because the outbreak was virtually over at that point in time. We use the number of alerts to proxy for Ebola intensity because our preferred proxy, the number of confirmed cases, was not available for us in the right aggregation level at the time of data analysis. Case count data we could obtain was either at day level, but aggregated for the whole country. Or it was at district level but aggregated to month- or entire outbreak totals. Ebola alerts are not the same as Ebola intensity, intensity suggest a kind of heat indicator, however larger districts will have more Ebola alerts just because they are more people living there. This effect will end up in the error term, if we do not adjust the dependent variable. Therefore we take the number of alerts relative to the population size in the specific district. Another problem with Ebola alerts is the high chance of double reporting and false reporting in the data set. To test whether Ebola alerts are a good proxy for Ebola intensity we test whether the total number of alerts in a district is a good proxy for the total number of cases in that district, relative to population size. The total number of cases is explained very well by the total number of alerts in a district, the regression estimated with a simple OLS yields a very significant coefficient for E_i;

$$REL\_cases_{i} = \beta_0 + \beta_1 * E_i \qquad (\beta_1 = 0.17^{***}; S.E. = 0.006) \quad i = 1....13$$

In figure 5 the distribution of E_i and of $total\_cases_i$ over the different districts is shown. Clearly, Western Area is overrepresented in the number of alerts and the number of cases. At the same time, the characteristics of the Western Area are completely different compared to any other district in the sample. Where other districts are mainly rural areas with small towns and a few cities scattered around, Western Area is a big urban area. The socioeconomic and health circumstances and specific slum city characteristics make it an outlier district, and because of that, it will be excluded from the sample for the analysis. After excluding Western Area from the sample of our study, the average number of alerts was of 4.03 alerts per day and a fairly large standard deviation of 7.86. However, after correction for population size, the variable E_{it} has a mean of 0.001 and a Standard deviation of 0.002. In table 2, at the end of this section, the descriptive statistics are shown. Figure 6 shows how the total number of daily alerts went down during the time span of our sample, as we in October 2014 there are above 150 Ebola alerts per day. This shows that our dataset is not covering the full outbreak. In October 2014 the outbreak was well on its way, and specific transmission waves, like the one the outbreak started with in the districts of

[b] The districts Western Area urban and Western Area rural are combined to one in this data set.

Kenema and Kailahun, will not be visible in the data. This might impact our results. In annex I exact variable definitions for every variable are presented.

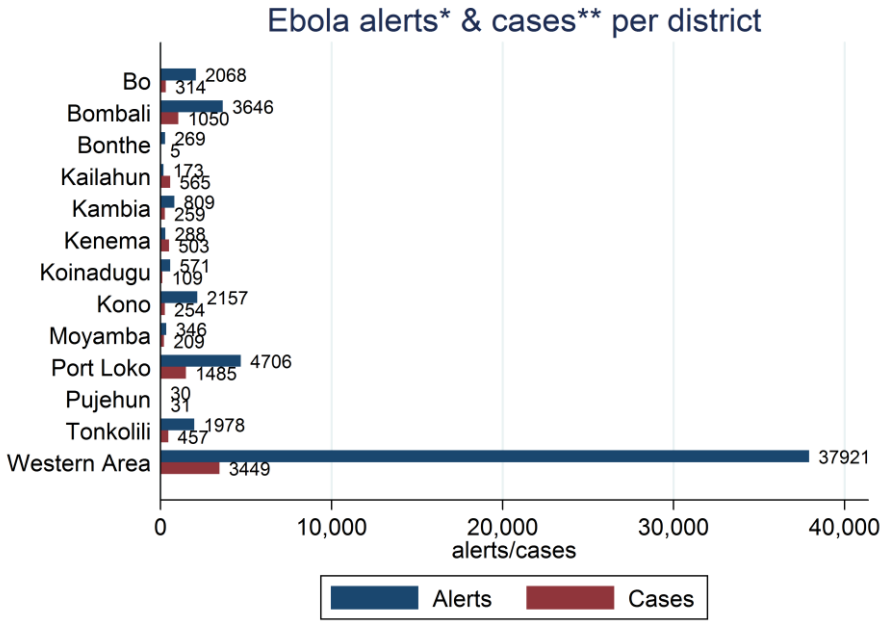

Figure 5 - Ebola alerts and cases per district
** Alerts from October 2014 - October 2015*
*** Cases from May 2014 – December 2015*

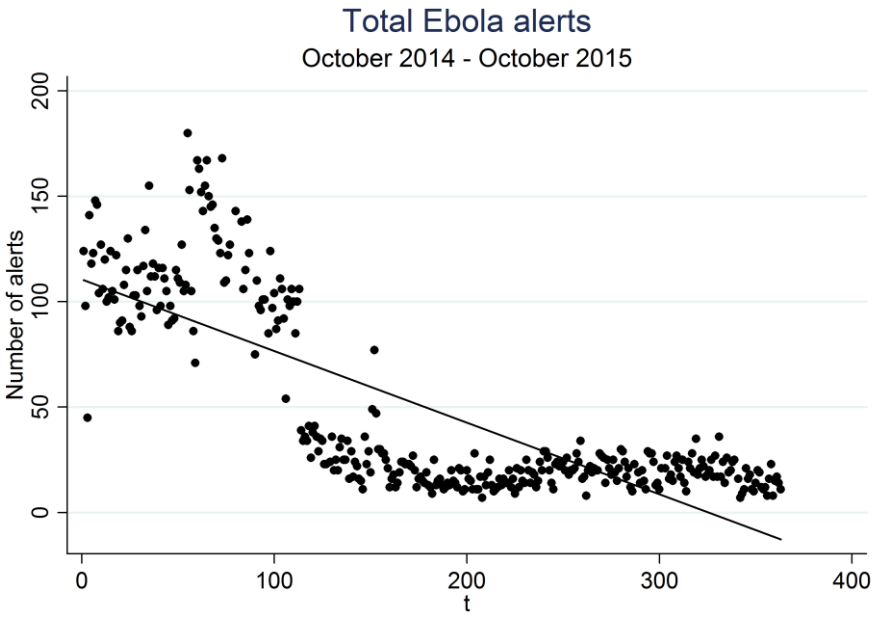

Figure 6 - Total Ebola Alerts in Sierra Leone from October 2014 - October 2015, Data obtained by NERC(2016)

Explanatory variables

As theorised, Ebola is an infectious disease, and therefore the single most important predictor of Ebola intensity in a district should be the Ebola intensity in the same districts in earlier periods and in neighbouring districts in previous periods. In the analysis, we use a lagged summation of E_{it} as a proxy for previous *Ebola intensity* (PE_{it}) and *Previous Neighbouring Ebola* (PNE_{it}).

$$PE_{it} = \sum_{t-2}^{t-23} E_{it}$$

$$PNE_{it} = \sum_{t-1}^{t-22} PE_{bt}^{*}$$

* *b indicates all the districts bordering district i.*

PE_{it} is constructed by taking the sum of $E_{i}t_{-2} - E_{t-23}$. Because of this summation, the first 23 observations of every district are missing. PNE_{it} is constructed by summing the PE_{it} of all the districts bordering district i. For the creation of the variable PNE the Ebola cases in Western Area are used to calculate PNE for the districts Kono and Port Loko, bordering Western-Area. The variables PE and PNE, just like their natural logarithms are relative to the total population size. Furthermore, treatment and safe burial practices were at the very core of the Ebola response; therefore we include *previous treatment capacity* (PTC_{it}) and *previous safe burial practices* (PSB_{it}). PTC_{it} is the relative number of functional bed capacity in Ebola treatment and care centres for (t-2 – t-23) compared to the number of alerts in that period. In this way, we see how many alerts there per functional treatment spot. The lower this number, the higher the treatment capacity. PTC_{it} has 3795 observations, a mean of 8075.91, so on average there were 8075 alerts per treatment spot. The variance in this number is not so high, with a standard deviation of 137.042. PSB_{it} is the fraction of reported funerals that has been coded as safe in the last 23 days by the NERC; it has the high average of 0.94 and a very low variance, the standard deviation is 0.07. Due to these summations in the creation of these variables, many missing observations are created. Therefore the variables PE_{it}, PNE_{it}, PTC_{it}, and PSB_{it}, will only have 3960 observations, and not 4224 like the dependent variables. The observations were one or more variable is missing will be ignored by Stata in the regressions, So, therefore, the regressions will only have 3960 observations, instead of on the full 4224 observations.

Based on data availability and hypotheses deducted from the literature review, a set of socioeconomic context variables is included in the analysis. These independent variables are time-invariant and at district level, so there are 12 observations per variable (all districts minus Western Area). To measure the *remoteness* of a district we use the average travel time by road to District headquarters, which is measured in minutes. The variable *Phone coverage* is measured by asking the question "whether there was phone coverage at a village yes or no" and taking the average. *Population density* is the outcome of *total number of inhabitants* divided by the *size of the district in acres*. *Average expenditure* is used as a proxy for average income in a district. *Educational level* is the share of people that has followed any education. The variable *border* is a dummy variable indicating whether a district is bordering Guinea and/or Liberia. *Densely populated* is a dummy variable indicating whether a district populated above the average population density.

Table 1: Summary statistics, October 2014 - October 2015

Variable	Obs	Mean	Std. Dev.	Min	Max
Dependent					
E	4224	0.001	0.002	0	0.02
lnE	4224	0	0	0	0.002
Independent					
t (days)	352	-	-	-	-
i (districts)	12*	-	-	-	-
PE	3960	0.02	0.04	0	0.18
PE(ln)	3960	0.001	0	0	0.003
PNE	3960	0.22	0.53	0.003	5.89
PNE(ln)	3960	0.002	0.001	0.001	0.005
PSB	3959	0.94	0.07	0.52	1.00
PTC	3960	80.92	82.36	0.00	454.7
PTC(ln)	3960	3.91	1.08	0.00	6.12
Population density	12	64.81	22.67	22.79	95.43
Remoteness	12	274.74	261.10	100.24	1089.2
Expenditures	12	48.82	12.19	29.86	70.21
Phone Coverage	12	0.60	0.19	0.11	0.87
Educational level	12	0.27	0.08	0.15	0.43
Institutional quality	12	0.53	0.12	0.29	0.69
Dens_pop	4224	0.42	0.49	0	1
Border	4224	0.58	0.49	0	1
weekend	4224	0.43	0.49	0	1
monday	4224	0.14	0.35	0	1

*Western Area is excluded from the sample but included as "Neighbour" in the variable PNE and PNE (ln)

There is a number of variables mentioned in the literature but not included in the model. Partly this can be explained by the earlier described problem of data availability. However, the dataset we retrieved from the NERC website contained many more response and Ebola characteristics. But often these variables had an extremely high number of missing observations or were highly correlated with other variables already included in the model. For example, the number of active ambulances is highly correlated with our variable for treatment capacity. Including both in the model would cause collinearity, and thus biased estimates. If the correlation is too high, Stata will automatically omit one of the problematic variables from the regression. Therefore, we excluded variables with a correlation higher than 0.7 with any other variable in the model.

Censoring

The time varying variables E, PE, PNE and PTC used in this study are left censored caused by a high number of observations with a zero. Because of this, these variables are not meeting the normality criterion. Often, non-normal data can be transformed by taking the natural logarithms of the values. In figure 5 in annex IV is shown that taking the logarithms is solving the normality issue for the variables PE, PNE and PTC convincingly. For the variable E the solution is not entirely cogent, so in our analysis, we will also use a Tobit estimator to correct for the left censored data.

5. Estimation Framework

Panel specifications

In this section, the specifications used to test the hypotheses formulated in section three are introduced. Correctly estimating these specifications enables us to explain the variation in Ebola intensity (Ebola alerts) across districts and time and provides us with the estimates of the true parameters for our variables of interest. There are many variables at district level that might have influenced the Ebola intensity but are not included in our model. If we do not specify district-fixed effects they will end up in the error term. However, if these unobserved effects in the error term correlate with any of the independent variables in our model, the exogeneity assumption for the error term is violated (Baltagi, 2008). This will lead to biased OLS estimates. But panel-data contains, by nature, another solution for this unobserved heterogeneity problem. Since the data contains multiple observations per district (i), we can use these different observations to estimate the unobserved district effects (Woolridge, 2010). There are two standard panel-data models that can estimate the district specific time-invariant unobserved effects; random-effects models (REM) and fixed-effects models (FEM). Depending on the severity of the correlation between the error term and our variables of interest we need to make a decision which one to make based on the trade-off between loss of efficiency (FEM) and potential bias (REM). The Hausman test can be used to make this decision. However, the Hausman test only functions on models without time-invariant variables. That's the reason we start with estimating the impact our time-variant independent variables in specification (1), we will perform a Hausman test on the outcomes of specification 1 estimated with a fixed-effects and a random-effects estimator.

$$(1) \quad E_{it} = \alpha_0 + \alpha_1 * PE_{it} + \alpha_2 * PNE_{it} + \alpha_3 * PTC_{it} + \alpha_4 * PSB_{it} + \alpha_5 * weekend_t + \alpha_6 * monday + T^2 + t + \varepsilon_i + \varepsilon_{it}$$

In this specification is E_{it} the Ebola intensity in district i at period t and ε_{it}, is a random error term. In the case that the Hausman test or other evidence suggest that a fixed-effect model should be used to estimate the coefficients, a problem arises. The standard fixed-effect estimator cannot be used to estimate coefficients for time-invariant observed variables. The unit-specific intercept controls for all the observed and unobserved variation of time-invariant variables, and adding extra time-invariant variables in the model will cause perfect collinearity with the district specific intercept. But the socioeconomic factors and district characteristics in specification (2) are time-invariant. To correct for unobserved heterogeneity, and estimate coefficients for parameters of observed time-invariant variables Hausman and Taylor (1981) developed a model. This Hausman-Taylor model can estimate coefficients for independent variables that are time varying and time-invariant regardless whether they correlate with the time-invariant unobserved effect (Woolridge, 2010). In specification (2), we measure the impact of past Ebola intensity, treatment capacity, socio-economic factors and district characteristics on E_{it}, this specification will be estimated with both a random-effects, a Hausman-Taylor and a Tobit estimator;

$$(2) \quad E_{it} = \beta_0 + \beta_1 * PE_{it} + \beta_2 * PNE_{it} + \beta_3 * PTC_{it} + \beta_4 * PSB_{it} + \beta_5 * income + \beta_6 * educational\ level + \beta_7 * remoteness + \beta_8 * phone\ coverage + \beta_9 * population\ density + \beta_{10} * institutional\ quality + \beta_{11} * border_i + \beta_{12} * (PE_{it} * dens\_pop_i) + \beta_{13} * weekend_t + \beta_{14} * monday + T^2 + t + \varepsilon_i + \varepsilon_{it}$$

β_1 - β_4 and α_1 - α_4 capture the impact of the time-varying variables on Ebola intensity; Ebola intensity in previous periods (PE_{it}) and in bordering districts (PNE_{it}), treatment capacity (PTC_{it}) and safe burial practices (PSB_{it}). β_{12} is the coefficient for the interaction term between previous Ebola intensity and population density. β_{13} and α_5 measure the impact of the day being a weekend day and β_{14} and α_5. Furthermore, both specifications have included a squared and normal time trend, to control for the expected non-linear time trend in the Ebola intensity. The unobserved effects on district level are

specified in ε_i. Lastly, the idiosyncratic error term (ε_{it}) in the model should be independent and homoscedastic. By including district effects the independence of the error term has increased, but it is very likely that there is autocorrelation left in the error term due to the fact that the lagged values of the dependent variable are used as an explanatory variable in the model. Furthermore, also the homoscedasticity assumption might be violated. Districts not severely hit by the Ebola outbreak have low values and variance for Ebola intensity (E_{it}), this will result in little variance in independent time-variant variables, and little variance in the error term, while districts with high Ebola intensity, have a high variance in the independent variables, and in the error term. The standard errors, if not corrected, will be inefficient and no longer valid for inference. Therefore, we use robust standard errors to both autocorrelation and heteroscedasticity (HAC).

Robustness

In the last part of the analysis, we test the robustness of our results. Firstly we test for remaining autocorrelation in our error term, after correcting them with the HAC-robust standard errors. We perform the Woolridge test for autocorrelation (Woolridge, 2010) on specification one (1) and two (2). Secondly, we test whether the found effects of the variables of interest are stable over time. In panel-data coefficients might change over time because the environment of the model (the Ebola outbreak) changes considerably. To test whether the found dynamics of the spread of Ebola where stable over time or changing due to changing unmeasured circumstances, we analyse specification one (1) in a rolling window approach (Woolridge 2011). We test equation one multiple times on a rolling subset of the data. We start with the first 60 days of the model and shift with an interval of t=1. In this way, the following 303 regressions will be performed;

(1) $E_{it} = \alpha_0 + \alpha_1 * PE_{it} + \alpha_2 * PNE_{it} + \alpha_3 * PTC_{it} + \alpha_4 * PSB_{it} + \alpha_5 * weekend_t + \alpha_6 * monday + T^2 + t + \varepsilon_i + \varepsilon_{it,}$ for t=1 – t=6

(2) $E_{it} = \alpha_0 + \alpha_1 * PE_{it} + \alpha_2 * PNE_{it} + \alpha_3 * PTC_{it} + \alpha_4 * PSB_{it} + \alpha_5 * weekend_t + \alpha_6 * monday + T^2 + t + \varepsilon_i + \varepsilon_{it,}$ for t=2 – t=61

(3 … 302) …

(303) $E_{it} = \alpha_0 + \alpha_1 * PE_{it} + \alpha_2 * PNE_{it} + \alpha_3 * PTC_{it} + \alpha_4 * PSB_{it} + \alpha_5 * weekend_t + \alpha_6 * monday + T^2 + t + \varepsilon_i + \varepsilon_{it,}$ for t=303 – t=363

We can analyse the variance of the found coefficients by checking the fluctuations of the plotted betas. We expect a little fluctuation for each, but large fluctuations or trends indicate that the parameters we found might be time varying. And thus that Ebola intensity in previous periods, Ebola intensity in bordering districts, treatment capacity and safe burial practices might impact Ebola intensity in different in different phases of the outbreak.

6. Results

In this section, our results are summarized. In table 3 we show the fixed-effects, random-effects and Tobit estimates of the impact of Ebola intensity in previous periods, Ebola intensity in bordering districts, treatment capacity and safe burial effects on Ebola intensity (E_{it}). Table 4 presents the random-effects, Tobit and Hausman-Taylor estimates of various socioeconomic and district characteristics on Ebola intensity, controlled for the time-varying effects found in table 3. Furthermore, in table 3 the results of the Hausman test for model endogeneity are presented, and in the last part of this section, the results of our robustness analysis are shown. The fixed-effects estimates in both the tables are based on within district variation. The random-effects models are using both within and between district variation to calculate the estimates for our parameters of interest. Standard errors in the fixed-effect and random-effect models are clustered at the district level, and robust for heterogeneity and autocorrelation. In the Hausman-Taylor model, the standard errors are calculated with Jack-knife estimation. Unfortunately, in Stata the Tobit model cannot be combined with either a fixed effect nor clustering of standard errors. If (ln) is specified at the explanatory variables, the natural logarithm of the variable is used for the analysis, except in the Tobit regressions.

Ebola effects, treatment capacity and safe burials

In column (1) of table 3 the random-effects, fixed-effects and Tobit coefficients for the various Ebola variables. The Hausman test in Table 3 concludes that the null-hypothesis of no systematic difference between fixed-effect and random-effect estimates is not rejected. This would suggest that the random-effects estimates are non-biased and thus more efficient compared to the fixed-effects estimates. However, this results should be viewed with precaution. Assuming full exogeneity in this data set is a very strong assumption. The non-rejection of the H0 in the Hausman test is not the same as accepting the H0, and might be caused by the low variance in some variables between and within districts (Woolridge, 2010). With this data, and the many non-observed effects, the FE estimator is a much more convincing pick, therefore we print and discuss both models.

In section 3 we formulated three hypotheses on the relation between Ebola intensity and transmission effects. The results in table 3 support our first hypothesis. *H1: Ebola intensity in a specific area in previous periods is positively correlated with the Ebola intensity in that specific area*. Following all three specifications the relative number of Ebola alerts in district i at time t is strongly related to the relative number of alerts in that district in the period $t_{-22} - t_{-1}$. The coefficients are 0.417, 0.410 and 0.043 in the random-effects, fixed-effects and Tobit model respectively. For the random-effect estimate this means that a 1% increase in alerts per capita in the past 21 days (E_{it}) leads to a $1-(1.01^{0.417}) = 0.41\%$ increase in the relative number of Ebola alerts intensity today, this effect is significant at the 1% level. The effect size is hard to compare, in the Tobit regression a 1 unit increase in PE_{it} leads to a 0.043 unit increase in E_{it}. The results of the models are strongly rejecting our third hypothesis; *H3: Ebola intensity in all districts bordering a specific district is positively correlated with the current Ebola intensity in that specific area*, the effects of PNE_{it} is strongly significant, but the effect size is counter intuitive. A higher Ebola in bordering districts in the period between t_{-22} and t_{-1} leads to a lower Ebola intensity at time t. The effect is approximately -0.12% in both the RE and FE model, and smaller than 0.000 in the Tobit regression. In this set of regressions, we find no effect for our hypothesized effects of treatment capacity and safe burials on Ebola intensity.

The models in column (1) and (2) are both able to explain around 60% of the variation in Ebola intensity. But since we are not interested in prediction E_{it} values, the R-square is not very relevant, the found effects change the meaning of the coefficients. However, it is a sign that there are much more factors influencing the variation in $E_{it.}$ Possibly adding district characteristics will improves the fit of the model.

Table 3: Current and previous Ebola Intensity, Treatment Capacity and Safe Burials

Y = Relative Ebola intensity (E_{it})	(1) RE	(2) FE	(3) Tobit
PE_{it} (ln)	0.417***	0.410***	0.043***
	(0.095)	(0.100)	(0.000)
PNE_{it} (ln)	-0.124*	-0.137*	-0.000**
	(0.069)	(0.075)	(0.000)
PTC_{it} (ln)	0.060	0.057	0.000
	(0.075)	(0.077)	(0.000)
PSB_{it}	-0.156	-0.141	-1.049
	(1.077)	(1.097)	(0.832)
$weekend_t$	-0.039**	-0.039*	-0.044
	(0.019)	(0.019)	(0.123)
$monday_t$	0.033	0.033	0.241
	(0.022)	(0.022)	(0.172)
t^2_t	0.000***	0.000**	0.000***
	(0.000)	(0.000)	(0.000)
t_t	-0.007***	-0.008**	-0.018***
	(0.003)	(0.003)	(0.003)
Constant	1.779	2.160	3.184**
	(1.215)	(1.344)	(1.551)
Observations	3959	3959	3959
Number of Clusters	12	12	
R squared	0.586	0.586	-
Fixed-effects	No	Yes	No
Hausman test	$Chi^2(6) = 1.31$ $p = 0.568$		

HAC robust standard errors clustered at district level in parentheses in column (1) and (2). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In table 4 the time-variant variables are estimated again, now in combination with time-invariant variables. For the time varying variables PE_{it}, PNE_{it}, $PTC_{it,}$ and PSB_{it}, table 4 yields the same results as table 3. The same effects are found significant. And the relations are going in the same direction. We find mixed results for our second hypothesis; *H2: The effect of previous Ebola intensity on current Ebola intensity is stronger in densely populated districts,* the estimate of the effect of the interaction between the dummy densly_populated and PE_{it} is not significant in column (1) and weak significant and very small in column (2) of table 4.

Socioeconomic status and District characteristics

After reviewing the effects of the time-invariant variables in our three specifications, we conclude that there is little ground to draw strong conclusions on the found estimates. They are very unstable across estimators, and the found effects are extremely low. Due to the low variance, these low effects can still be very significant, like the -0.001 effect of educational level on the number of Ebola alerts in the random-effect and Tobit estimator, which is significant at the 1 percent level in column (1). But because of the instability of the estimators, and the very strong assumptions of endogeneity on these estimates, we wil not compare these results with our hypothesis. However, the Hausman-Taylor estimator should be able to correct for the district unobserved effects, and still estimating the effects of the time invariant variables. However, none of the time invariant variables is significant in the Hausman-Taylor model, which supports our decision to ignore the found effects of the socioeconomic and district variables in column (1) and (2).

Time effects

Across the two model specifications (table 3 and table 4) and all estimators used, expect for the Tobit regressions, we find a significant weekend reporting effect, showing that less Ebola alerts were being

reported on weekend days. Both the quadratic and the linear time trends in both specifications are significant across all the used estimates. The positive quadratic term indicates that there is a quadratic relation of Ebola alerts reported over time. The effect is positive, so this relation is upward bending. The linear time trend is negative significant, which means that on average the trend is downward sloping in our sample.

Table 4: Ebola Intensity, Socioeconomics and District Characteristics

Y = Ebola intensity (E_{it})	(1) RE	(2) Tobit	(3) Hausman-Taylor
Time-Variants			
PE_{it} (ln)	0.285***	0.046***	0.289***
	(0.050)	(0.003)	(0.028)
PNE_{it} (ln)	-0.114***	-0.000***	-0.106
	(0.044)	(0.000)	(0.132)
PTC_{it} (ln)	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
PSB_{it}	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
$Income_i$	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
$Educational\ level_i$	-0.001***	-0.001**	-0.001
	(0.000)	(0.000)	(0.000)
$Remoteness_i$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$Phone\ coverage_i$	0.001*	0.000	0.000
	(0.000)	(0.000)	(0.001)
$Population\ Density_i$			-0.000
			(0.000
dens_pop	-0.000	-0.000	
	(0.000)	(0.000)	
$Border_i$	-0.000**	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)
$Dens\_pop_i * PE_{it}$	-0.000	-0.000*	
	(0.000)	(0.000)	
$Weekend_t$	-0.000**	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)
$Monday_t$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
t2	0.000***	0.000***	-0.000*
	(0.000)	(0.000)	(0.000)
t	-0.000***	-0.000***	-0.000*
	(0.000)	(0.000)	(0.000)
Constant	0.000	0.001**	0.001
	(0.000)	(0.000)	(0.000)
Observations	3959	3959	3959
Number of Clusters	12		12
R squared	0.682		
Fixed-effects	No	Yes	Yes

HAC robust standard errors clustered at district level in parentheses in column (1). Jackknife standard errors clustered at district level in column (3). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Robustness

To test the robustness of our results we firstly test for remaining autocorrelation. We use a test for serial correlation in the idiosyncratic errors of a linear panel-data model discussed by Wooldridge (2010). The

Woolridge test has the null-hypothesis of *no first-order autocorrelation.* As shown below, for both specifications formulated in the estimation framework in section 5, the p-value of the Woolridge F-statistics is larger than 0.1, so there is no reason to reject the null-hypothesis. This means that after correction of the standard error with the HAC standard errors and adding the time fixed-effects, there is nog serial correlation in the idiosyncratic error term left.

Specification (1):

$E_{it} = \alpha_0 + \alpha_1 * PE_{it} + \alpha_2 * PNE_{it} + \alpha_3 * PTC_{it} + \alpha_4 * PSB_{it} + \alpha_5 * weekend_t + \alpha_6 * monday + T^2 + Д_t + \varepsilon_i + \varepsilon_{it}$

Woolridge F-statistics $(1, 11) = 2.993$

$Prob > F = 0.1115$

Specification (2):

$E_{it} = \beta_0 + \beta_1 * PE_{it} + \beta_2 * PNE_{it} + \beta_3 * PTC_{it} + \beta_4 * PSB_{it} + \beta_5 * income + \beta_6 * educational\ level + \beta_7 * remoteness + \beta_8 * phone\ coverage + \beta_9 * population\ density + \beta_{10} * institutional\ quality + \beta_{11} * border_i + \beta_{12} * (PE_{it} * dens\_pop_i) + \beta_{13} * weekend_t + \beta_{14} * monday + T^2 + Д_t + \varepsilon_i + \varepsilon_{it}$

Woolridge F-statistics $(1, 11) = 3.127$

$Prob > F = 0.1047$

The second test for robustness is the test for stability of the significant coefficients. We performed a rolling window regression on specification (1). With a window of t=60 and a rolling interval of t=1. The stored estimates of this regression are plotted against time in figure 5. We see large fluctuations in the estimates. Although there is no clear pattern. Both the effects of Previous Ebola intensity and bordering Ebola intensity have spikes around t=100 and t=250. There might be unobserved events at country level influencing these fluctuations in these effects. In general, we see an unstable fluctuation of the found effects.

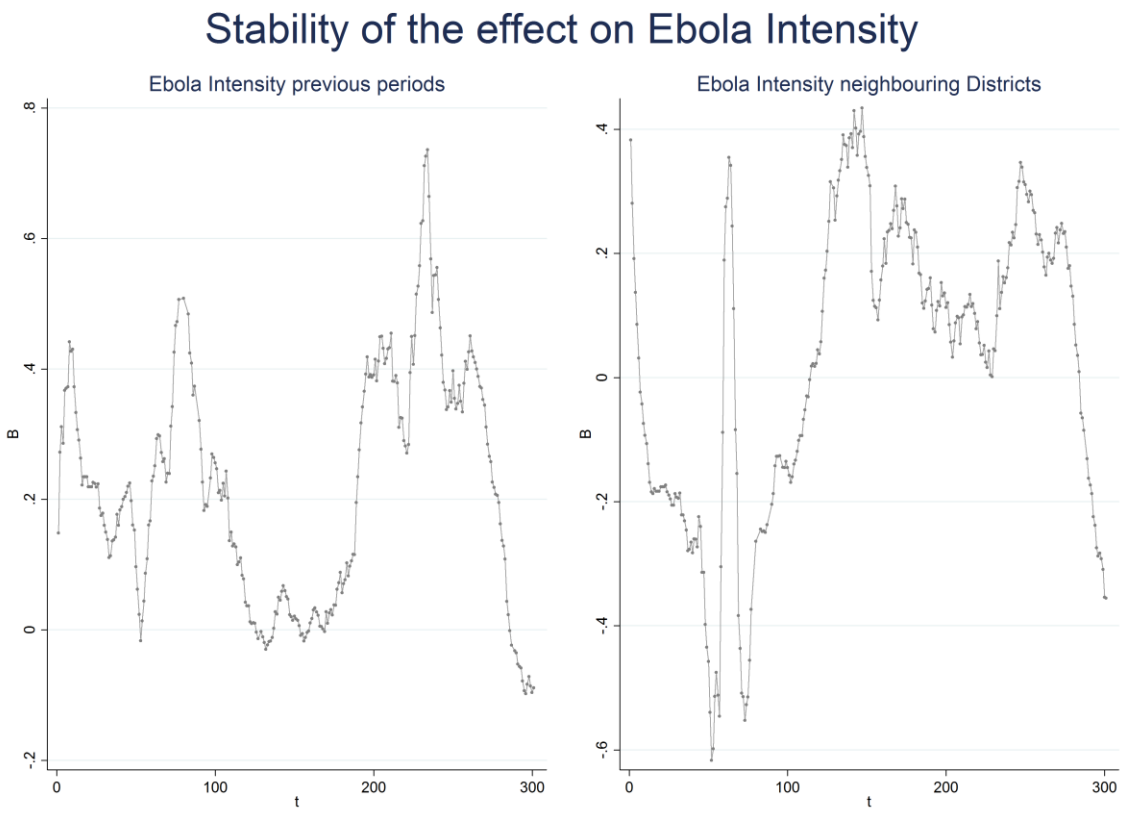

Figure 5 - Stability of the coefficients

7. Discussion and Conclusion

In the aftermath of the Ebola crisis in West-Africa, this study tries to understand some of the mechanisms underlying the fast spread and high intensity of the Ebola outbreak in rural Sierra Leone. The main question we try to answer was '*What are the main determinants of the Ebola intensity in rural Sierra Leone?'* We use Ebola alerts per capita to proxy for Ebola intensity, our main dependent variable. This proxy might be problematic because alerts are not Ebola cases. However, we tested for the validity of this proxy variable by regressing the alerts per capita on total number of cases, and found a strong relation. Still it would be better to use confirmed case count or death data, but this was not available at time of our analysis.

Firstly we tested the transmission effect over time and space of previous numbers of alerts on current number of alerts. In almost all our specifications, we find strong significant positive effects for previous Ebola alerts per capita and neighbouring Ebola alerts per capita on the Ebola intensity at period t. These results coincides with the findings of various studies estimating the effect of Ebola intensity on transmission rates and case counts (i.e. Fang et al., 2016; Krauer et al., 2016; Rainisch et al., 2014 and Fallah et al., 2015). The coefficients for the significant time varying variables are more or less the same over both specifications and all estimators. This suggest that the found effects of PE_{it} and PNE_{it} is a legit effect. However, since PE_{it} and PNE_{it} are constructed from lagged values of the dependent variable, these strong effects might also be caused by autocorrelation. Therefore we performed a test for first order autocorrelation, which indicated that this was not present in our models.

Our findings are not consistent with other studies regarding the impact of treatment capacity and safe burial practices on Ebola alerts, and thus Ebola intensity. Where for example Kurcharski et al., (2015) finds positive effects of treatment capacity on Ebola cases in their quantitative model, and many studies suggest there should be a strong link between Safe burials and Ebola intensity, we don't find any effect for neither of the two. This can be caused by the aggregation level of our data. The effect of treatment capacity on district level is very hard to predict, because treatment capacity is a variable that matters at the small level. Whether there is a hospital or clinic in you village, or at walking distance might influence the Ebola intensity in your community, but the total number of hospitals in your district are not determining the Ebola intensity in your community if none of those hospitals is reachable for members of your community. The same counts for safe burial practices. Regardless of the impact of safe burials on the spread of Ebola, it will not come as a surprise that this aggregated percentage, with a very low variance, is not able to explain the variation in the number of Ebola alerts.

Many studies theorize about the relations between socioeconomic status and Ebola. We tried to test some of these relations. And although some of our estimators showed some significant results, the results of this analysis is very unstable we do not find any robust effects of income, educational level, remoteness, phone coverage, population density or sharing a border with Liberia or Guinea on the number of Ebola alerts in a district.

Very crucial for the interpretation of our results is the type of effects we estimate. In our first fixed-effects model we can only test for district and time varying effects. These within district estimates are unbiased, but might be inefficient because we are not using all the variance in the model to obtain our results. But if we compare these estimates with the more efficient, potentially biased random effects estimator, we find the same effects to be significant. In the robustness analysis, we showed how the found significant effects of previous Ebola intensity, neighbouring Ebola intensity and weekend reporting, are unstable over time. This indicates that there are various unmeasured factors at paly, influencing the data and our model.

The Hausman-Taylor estimator uses internal instruments to instrument for endogenous independent variables. Therefore we are able to estimate the effects of time-invariant variables with district fixed effects. However, the Hausman-Taylor estimates are all but one insignificant. The Hausman-test indicated that we could use the random-effects estimator. But since there is so much reason to assume the exogeneity assumption to be violated, and the found effects of the different random-effects estimators were very unstable, we decided to not draw conclusions from these models.

The reason that the effects we find are unstable and sometimes different from our hypotheses and other literature is at least two fold. The first reason is data availability, at the time of this analysis, there was no comprehensive data set or source were we could get the confirmed case count data at a disaggregated level. Therefore, we had to settle for district level alert data, were chiefdom level case counts were preferred. There were many other factors we expect to influence the Ebola intensity at district and country level, but because of the lack of data, we could not include them in our model. The disaggregation of the data might and the omitting of important factors for determining the Ebola intensity might have caused the non- or counter-intuitive results for so some of the determinants studies. Secondly, the quality of the data is problematic. Several kinds of errors are likely to exist in the data. For example, Fang et al. (2016) argue that increased and improved case finding and contact tracing caused bias in the data because a higher proportion of the total cases was being reported later in the outbreak, caused by increased awareness.

This study contributes to a group of parallel studies testing the effects of various outbreak, socioeconomic, cultural and geographical factors on the spread and intensity of the Ebola outbreak in Western-Africa. Effects of several variables are identified, suggesting that local outbreak characteristics

played a determining role in the 2014 Ebola outbreak. To enable policy makers to learn from this outbreak, the results found in this study need to be validated with a better, more extended and more trustworthy dataset. More response variables should be included in the further analysis to learn what worked in the kerbing of the outbreak. And more disaggregated socioeconomic and district characteristics should be included to identify the determinants on smaller, village or household, level.

8. References

Acemoglu, D., Reed, T., & Robinson, J. A. (2014). Chiefs: Economic development and elite control of civil society in Sierra Leone. *Journal of Political Economy*, *122*(2), 319-368.

Alexander, K. A., Sanderson, C. E., Marathe, M., Lewis, B. L., Rivers, C. M., Shaman, J., ... & Eubank, S. (2015). What factors might have led to the emergence of Ebola in West Africa?. *PLoS Negl Trop Dis*, *9*(6), e0003652.

Allgaier, J., & Svalastog, A. L. (2015). The communication aspects of the Ebola virus disease outbreak in Western Africa–do we need to counter one, two, or many epidemics?. *Croatian medical journal*, *56*(5), 496.

Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.

Bausch, D. G., & Schwarz, L. (2014). Outbreak of Ebola virus disease in Guinea: where ecology meets economy. *PLoS Negl Trop Dis*, *8*(7), e3056.

Bellows, J., & Miguel, E. (2009). War and local collective action in Sierra Leone. *Journal of Public Economics*, *93*(11), 1144-1157.

Butler, D. (2015). How to beat the next Ebola. *Nature*, *524*(7563), 22.

Chan, M. (2014). Ebola virus disease in West Africa—no early end to the outbreak. *New England Journal of Medicine*, *371*(13), 1183-1185.

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M. A., Flasche, S., Baguelin, M., ... & Tiffany, A. (2015). Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study. *PLoS Curr*, *7*.

Chandler, C., Fairhead, J., Kelly, A., Leach, M., Martineau, F., Mokuwa, E., ... & Wilkinson, A. (2015). Ebola: limitations of correcting misinformation. *The Lancet*, *385*(9975), 1275-1277.

Economist 2015: http://www.economist.com/topics/ebola-virus

Fallah, M. P., Skrip, L. A., Gertler, S., Yamin, D., & Galvani, A. P. (2015). Quantifying poverty as a driver of Ebola transmission. *PLoS Negl Trop Dis*, *9*(12), e0004260.

Fang, L. Q., Yang, Y., Jiang, J. F., Yao, H. W., Kargbo, D., Li, X. L., ... & Liu, K. (2016). Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proceedings of the National Academy of Sciences*, *113*(16), 4488-4493.

Fauci, A. S. (2014). Ebola—underscoring the global disparities in health care resources. *New England Journal of Medicine*, *371*(12), 1084-1086.

Foster, T. (2013). Predictors of sustainability for community-managed handpumps in Sub-Saharan Africa: Evidence from Liberia, Sierra Leone, and Uganda. *Environmental science & technology*, *47*(21), 12037-12046.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric Society*, 1377-1398.

Heritage Economic Freedom Index, 2015: http://www.heritage.org/index/

Ippolito, G., Puro, V., & Piselli, P. (2015). Ebola in West Africa: Who pays for what in the outbreak. *New Microbiol*, *38*, 1-3.

Kucharski, A. J., Camacho, A., Checchi, F., Waldman, R., Grais, R. F., Cabrol, J., ... & Edmunds, W. J. (2015). Evaluation of the benefits and risks of introducing ebola community care centers, sierra leone. *Emerging infectious diseases*, *21*(3), 393-399.

Krauer, F., Gsteiger, S., Low, N., Hansen, C. H., & Althaus, C. L. (2016). Heterogeneity in District-Level Transmission of Ebola Virus Disease during the 2013-2015 Epidemic in West Africa. *PLoS Negl Trop Dis*, *10*(7), e0004867.

Mogelson, L. (2015). When the Fever Breaks. *The New Yorker*.

Morse, B., Grépin, K. A., Blair, R. A., & Tsai, L. (2016). Patterns of demand for non-Ebola health services during and after the Ebola outbreak: panel survey evidence from Monrovia, Liberia. *BMJ Global Health*, *1*(1), e000007.

New York times (2015). Retrieved from http://www.nytimes.com/topic/subject/the-ebola-outbreak-in-west-africa

Otter, J. A. (2015). Journal Roundup: Ebola (again), the rise (and rise) and fall of MDROs, and Infection Prevention 2014 Conference, Glasgow. *Journal of Hospital Infection*, *89*(1), 74-76.

Philips, M., & Markham, Á. (2014). Ebola: a failure of international collective action. *The Lancet*, *384*(9949), 1181.

Rainisch, G., Shankar, M., Wellman, M., Merlin, T., & Meltzer, M. I. (2015). Regional spread of ebola virus, west africa, 2014. *Emerging infectious diseases*, *21*(3), 444-447.

Richards, P., Amara, J., Ferme, M. C., Kamara, P., Mokuwa, E., Sheriff, A. I., ... & Voors, M. (2015). Social pathways for Ebola virus disease in rural Sierra Leone, and some implications for containment. *PLoS Negl Trop Dis*, *9*(4), e0003567.

Saéz, A. M., Weiss, S., Nowak, K., Lapeyre, V., Zimmermann, F., Düx, A., ... & Sachse, A. (2015). Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO molecular medicine*, *7*(1), 17-23.

UNDP. (2016) Retrieved from http://hdr.undp.org/en/content/human-development-index-hdi

USAID. (2016). Retrieved from https://www.usaid.gov/ebola/facts

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

WHO | Ebola virus disease outbreak. (2015). Retrieved November/December, 2015, from http://apps.who.int/csr/disease/ebola/en/

World Bank. (2016) Retrieved January 2016, from http://www.worldbank.org/en/country/sierraleone

Annex I: Variable definition

Table 5 Variable definition and hypotheses

Indicator	Variable-name	Operationalisation
Number of new confirmed cases today	E_{it}	Number of new alerts in district (i) at day (t) per capita
Previous number of confirmed cases	PE_{it}	Three week total of E_{it} in district (i) of 21 days before day (t-1)
Previous number of confirmed cases in neighbouring districts	PNE_{it}	Three week total of E_{it} in district (i) of 21 days before day (t-1) of all districts bordering district i
Previous functional treatment capacity	PTC_{it}	Three week average of total functional beds in CCCs and ETUs in district (i) of 21 days before day (t-1)
Share of burials reported safe	PSB_{it}	Three week average of share of burials that were reported save in district (i) of 21 days before day (t-1)
Population Density	$Popdens_i$	Population density in district i is total population/ Area in ha
Remoteness	$Remote_i$	Average travel time to district headquarters in minutes for people in district i
Phone coverage	$Phone_i$	Share of villages with phone coverage out of total number of villages in district i
Expenditures	Exp_i	Average household spending in district i
Educational level	$Educ_i$	Share of adults that followed any education in district i
Weekend reporting	$Weekend_i$	Weekend dummy (to capture weekend reporting bias)
Monday reporting	$Monday_i$	Monday dummy (to capture Monday reporting bias)
Border dummy	$Border_i$	Dummy indicating whether a districts shares a border with guinea or Liberia

Annex II: descriptive statistics

Correlation table

	E	PE	PNE	PTC	PSB	Pop. Dens.	Remoteness	Exp.	Coverage	Educ.	Inst. Qual.
E	1.00										
PE	0.87	1.00									
PNE	0.28	0.33	1.00								
PTC	0.55	0.67	0.23	1.00							
PSB	-0.08	-0.07	0.01	0.07	1.00						
Population density	0.06	0.07	-0.06	0.12	0.25	1.00					
Remoteness	-0.14	-0.17	-0.11	-0.22	-0.04	-0.67	1.00				
Expenditures	-0.11	-0.13	-0.03	-0.45	-0.10	-0.14	0.00	1.00			
Phone Coverage	0.12	0.14	0.15	0.15	0.25	0.84	-0.87	0.04	1.00		
Educational level	-0.23	-0.26	-0.03	-0.26	0.24	0.48	-0.43	0.06	0.56	1.00	
Institutional quality	-0.19	-0.22	-0.07	-0.30	-0.40	-0.52	0.40	-0.13	-0.55	0.04	1.00

Annex III – Graphs

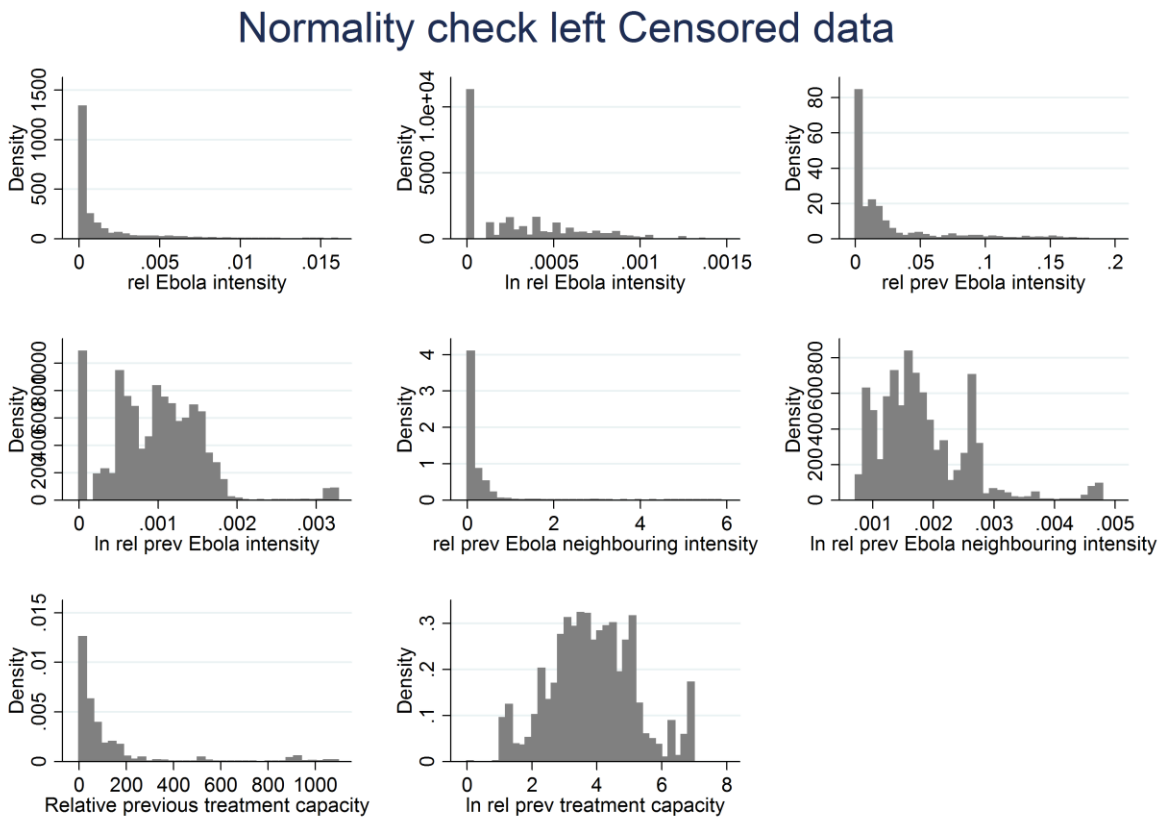

Figure 6 - Normality in the data using logarithms

Annex IV – regression tables of non-relative alert data

Table 3b: Ebola effects, Treatment Capacity and Safe Burials

Y = Ebola intensity (E_{it}) (log)	(1)	(2)	(3)
	RE	FE	Tobit
PE_{it} (log)	0.396***	0.375**	0.043***
	(0.110)	(0.121)	(0.000)
PNE_{it} (log)	-0.217**	-0.261**	-0.000***
	(0.094)	(0.108)	(0.000)
PTC_{it} (log)	0.048	0.039	-0.000
	(0.090)	(0.096)	(0.000)
PSB_{it}	-0.313	-0.338	-0.083
	(1.049)	(1.052)	(0.799)
$weekend_t$	12.211	-0.347*	36.668
	(25.579)	(0.168)	(242.816)
$monday_t$	11.862	-0.283	35.627
	(24.827)	(0.252)	(235.475)
t^2_t	-0.000	-0.000**	-0.000
	(0.000)	(0.000)	(0.002)
Constant	1.779	2.160	3.184**
	(1.215)	(1.344)	(1.551)
Observations	3959	3959	3959
Number of Clusters	12	12	
R squared	0.586	0.586	
Fixed-effects	No	Yes	No
Time dummies	Yes	Yes	Yes
Hausman test	$Chi^2(6) = 2.35$ $p = 0.885$		

HAC robust standard errors clustered at district level in parentheses in column (1) and (2). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4b: Ebola Intensity, Socioeconomics and District Characteristics

Y = Ebola intensity (E_{it})	(1) RE	(2) Tobit	(3) Hausman-Taylor
Time-Variants			
PE_{it} (ln)	0.387***	0.043***	0.435***
	(0.093)	(0.001)	(0.108)
PNE_{it} (ln)	-0.267***	-0.000***	0.022
	(0.044)	(0.000)	(0.066)
PTC_{it} (ln)	0.046	0.000	0.029
	(0.092)	(0.000)	(0.077)
PSB_{it}	-0.450	-0.479	-0.419
	(1.041)	(1.044)	(1.151)
$Income_i$	-0.016***	-0.008	0.005
	(0.006)	(0.006)	(0.026)
Educational $level_i$	-3.757***	-1.807	-0.597
	(0.838)	(1.189)	(4.510)
$Remoteness_i$	0.000	0.000	0.000
	(0.001)	(0.001)	(0.003)
Phone $coverage_i$	2.564***	1.629	2.075
	(0.965)	(1.220)	(4.841)
Population $density_i$	-0.012***	-0.004	-0.016
	(0.005)	(0.008)	(0.050)
Institutional $quality_i$	0.078	-0.040	-0.676
	(0.528)	(0.851)	(4.715)
$Border_i$	-0.337***	-0.234	0.497
	(0.103)	(0.226)	(1.360)
$Dens\_pop_i$*PE_{it}	0.026	-0.000	
	(0.091)	(0.001)	
$Weekend_t$	12.588	38.300	-0.038*
	(25.709)	(242.587)	(0.019)
$Monday_t$	12.228	37.215	0.029
	(24.952)	(235.253)	(0.022)
Constant	3.341**	3.950**	-0.499
	(1.705)	(1.936)	(2.500)
Observations	3959	3959	3959
Number of Clusters	12		12
R squared	0.682		
Fixed-effects	No	Yes	Yes
Time dummies	Yes	Yes	No

HAC robust standard errors clustered at district level in parentheses in column (1). Jackknife standard errors clustered at district level in column (3). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Time-fixed effects added in column (1) and (2).