Colonial Legacies and Health Outcomes in Borneo: Evidence From a Natural Experiment

Marinissen, Luna

Colonial Legacies and Health Outcomes in Borneo:

Evidence From a Natural Experiment

Luna Marinissen

S3074250

BSc Political Science: International Relations and Organisations – Bachelor Project

Leiden University

The Politics of Public Health in Developing Countries

Supervisor: Dr. J. Phillips

Publication Statement: Public

Word Count: 8248

December 20, 2024

Table of Contents

1. Introduction

Most forms of colonial rule ended over half a century ago when the majority of formerly colonised territories gained independence. However, the effects and remnants of this long period of colonialism, that affected a large part of the developing world, remain ever present. Within the literature, general consensus exists that the colonial past has had significant effects on development outcomes. Yet still a debate remains around whether these effects have been negative or positive, from which circumstances these differences stem, and whether these effects persist throughout changes in time. The majority of these studies discuss the effect of colonial legacies on economic development opportunities, and although it is generally agreed upon that this is closely related to public health development, the direct link to health has not been adequately explored (Spruk & Kovac, 2020, p. 854). This is surprising as health is an essential component in the quest to attain equal levels of human development globally. Understanding the relationship between health outcomes and the colonial past will enable a deeper understanding of historical significance in patterns of global public health. Additionally, it will contribute to the debate about the responsibility that former colonial powers bear, particularly in a time where provision of development aid is under pressure (UN GCRG, 2024).

Within the literature a general distinction is made between extractive and inclusive types of colonial rule, however, the argument will be made that specific choices made by colonial powers within extractive systems have lasting effects. The intention of this study is therefore not to address the general implications of colonialism, it is rather to investigate whether different colonial practices have influenced and led to different paths of public health development. Consequently, this research aims to provide an answer to the following question: *"How did differences in colonial policies affect health outcomes in the long-term in former colonial territories?"*

In order to adequately answer the research question, the existing theory on legacies of colonial policies and their effects on health outcomes will first be discussed. Subsequently, it is argued that the case of Borneo presents itself as an opportunity to leverage a natural experiment and investigate the research question. The difference in colonial policies on the British-colonised part of the island will then be compared to the Dutch-colonised side. Thereafter the research design and methodology will be outlined, this includes a discussion of the used data, a spatial analysis of this data, operationalisation of the variables, and the applied method of data analysis. Finally, the results of the regression will be interpreted and corroborated by a discussion of alternative explanations for the observed relationship.

2. Theoretical Framework

Colonialism

Colonialism can generally be defined as a period of forceful domination, oppression and exploitation by a group of people over another group of people, their territory and resources (Horvath, 1972, p. 46). Colonialist practices have a long history; however, this research will focus on the most recent period from the 19^{th} century and onwards, when mainly Western colonial powers dominated and exploited large parts of the world. In many colonised territories the increased state centralisation and institutions building by the colonisers had big political impacts (Ziltener & Künzler, 2013, p. 291). Economically, the colonisers drained wealth from these territories by extracting resources to their own economic benefit, often impoverishing the local population through forced labour and expropriation of their property in this objective (p. 299). On the social aspect, colonialism led to disruptions in indigenous cultures, a disregard to local needs, and forced labour, often in the form of slavery (p. 302). The extent of influence of these factors is dependent on the targets of the coloniser and the characteristics of the colonised territory.

Path Dependency Theory

The theory of path dependency is useful in explaining how colonial legacies have persisted and continue to influence current policies. This theory entails that past political choices, events and institutions influence the trajectory of current and future political decisions (Wilsford, 1994, p. 252). Path dependence can be described as a social process in which the idea of increasing returns is determining; benefits of continuing down a particular path increase over time and make alternative options less attractive due to high costs of reversal as the process progresses (Pierson, 2000, p. 251). In the political realm, short-term goals are often prioritised; however, these can lead to long-term trajectories because established institutions are difficult to change due to legal constraints, vested interests and costs of reform (p. 257). Critical junctures, which are big exogenous shocks, can alter paths of development due to their unexpected nature, colonialism can be seen as such a shock to invaded territories (p. 263). Colonial powers had the coercive power to forcefully impose institutional frameworks centred on resource extraction and elitist control, establishing a trajectory that became increasingly entrenched and eventually leading to a higher chance of persistence into the post-colonial period (Subrahmanyam, 2006, p. 85).

Inclusive and Extractive Institutions

A general distinction between inclusive and extractive types of colonialism is made within the literature. This difference, in part, explains why colonialism has had predominantly positive or negative effects on postcolonial development rates in former colonies (Acemoglu, Johnson & Robinson, 2001, p. 1369). Colonisers set up inclusive institutions in the colonies where Europeans settled, establishing systems that resembled European ones. This form of colonial rule included the introduction of property and liberal rights for the settlers, and efforts were made to strengthen state capacity and its institutions (p. 1370). What type of rule was implemented by colonisers depended on the characteristics of the colony, such as the liveability for Europeans and the military strength and resistance of the local population. Examples of such colonies include former British territories, such as Australia and the United States of America. On the contrary, extractive forms of colonialism are focused on exploiting the resources and local population of the colonised territory to the economic benefit of the coloniser (p. 1370). The oppressive nature of these colonial policies set postcolonial states back in terms of development opportunities, by depleting resources, limiting investments in human capital and technology, and eroding social trust through divisive governance. An example of extractive rule is the forced mining labour system introduced by the colonial power of Spain in Peru and Bolivia, in which individuals of indigenous groups were forced to work in mines and extract valuable natural resources at the benefit of the colonising state (Dell, 2010, p. 1867).

However, it can be argued that not just this broad division between inclusive and extractive colonial systems determine the development paths of former colonised territories, but differences between and within extractive systems matter as well. In the case of the forced mining system, Dell (2010, p. 1899) demonstrates that the affected communities are less wealthy and suffer from worse health compared to the unaffected communities in surrounding areas. Although these unaffected communities were not targeted for this specific extractive institution, they were still subjected to the overall extractive nature of the rule in the country, but to a lesser degree than the affected communities. Similarly, the colonisation of northern China by the Japanese was of extractive and aggressive nature, but at the same time investments were made in state institutions, education, and health, that resulted in better long-term development outcomes in comparison to surrounding regions (Mattingly, 2017, p. 435). Another example in which extractive institutions resulted in better economic circumstances today, are the regions on Java in which Dutch colonisers established sugar factories. Research shows that the immediate surrounding areas of these factories still are more industrialised, have better infrastructure, and health compared to similar untargeted areas (Dell & Olken, 2020, p. 164).

Public Health Determinants

The main threats to public health in the developing states are infectious, water-, food-, and insect-borne diseases, and to a lesser extent noncommunicable diseases that are most prevalent in already developed countries (Soares, 2007, p. 259). Most diseases that pose big threats to health in the developing world are preventable with the existing knowledge and treatment options, but a lack of resources and investments hinder progress. Amongst the most effective primary preventive measures are vaccine-programmes for diseases as measles, smallpox and polio, increased sanitation and safe-water access, and improvements in maternal health (Macinko, Starfield & Erinosho, 2009, p. 154). Additionally, improved access to education on health leads to more awareness amongst individuals on the importance of sanitation, reproductive health and healthcare visits (p. 155). The implementation of these preventative and curative measures mainly rests on policy development and investments in public health by the state. However, effective resource allocation and execution of these policies requires a certain level of state-capacity and bureaucratic autonomy that is not always present in developing states (Burkle, 2006, p. 248). Moreover, democratic systems generally tend to result in more transparency by the government, which could lead to increased public trust and higher compliance rates to health interventions amongst the population (Kosack & Fung, 2014, p. 83). Besides these efforts made by governments and the health sector itself, collective action by communities can also be a channel through which public health improves. Macfarlane, Racelis, and Muli-Musiime (2000, p.842) mention the example of a local community in Kenya that successfully addressed sanitation and health issues through local organisation and mobilisation of members and resources.

Extractive Colonial Policies and Health Outcomes

Several studies investigated how extractive colonial policies affected colonies on these essential political and social factors for public health improvements and establish how this has eventually affected the long-term development trajectory. First, the difference between the use of indirect and direct rule by colonisers proves to have significant long-term development effects. Lee and Schultz (2012, p.9) establish that the use of indirect rule through local governance structures by British colonisers in west Cameroon resulted in better long-term economic and public goods provision compared to the direct rule system that the French applied in the east. The indirect rule strengthened local political organisations, which fostered capabilities to mobilise communities for collective action towards development (p.10). Similar results were found in India where the majority of revenues made by British colonies derived from crash crop production and export. The areas in which these revenues were collected by landlords appointed by British elites, have worse development outcomes in public goods provision, health, and infrastructure, compared to areas in which local government structures were maintained (Banerjee & Iyer, 2005, p. 4).

Second, it has been argued that the introduction of new legal systems under colonial rule have had implications for development outcomes. Dutch rule in Indonesia included the adoption of a civil law system which allowed the colonial administration to claim uncertified lands for economic exploitation, through agricultural production and resource extraction (Fahmi, 2020, p. 105). This new system replaced indigenous land tenure systems, worsening local autonomy and creating structural inequalities that carried over into the post-colonial state when the colonial land-tenure system was inherited (p. 117).

Third, investments in infrastructure and industrial capital have created development opportunities in the long-term, even if they were initiated for colonial extraction. The introduction of the Dutch cultivation system on the island of Java required agricultural and industrial operations to maximise sugar manufacturing (Dell & Olken, 2020, p. 171). In these regions, factories and the required infrastructure were established, and the higher level of industrialisation in the surrounding areas persisted in the post-colonial period. Consequently, these areas continue to be strong economic centres with overall greater wealth, educational attainment, and health, in contrast to surrounding regions that were merely engaged in the extraction of cash-crops, which were later transported to these factories (p. 200).

Finally, forced labour and slavery during the colonial period have had significant effects on the health and wealth of the affected communities. In the south of the United States of America slavery created long-lasting, institutionalised, racial sentiments towards blacks, these sentiments have been passed on through many generations and still lead to inequalities between white and black parts of the population (Reece, 2022, p. 678). Institutionalised prioritisation of white health as a consequence, results in distrust amongst the black population towards the healthcare system, which in turn leaves the black population with limited access and worse health outcomes overall (p. 685). The forced labour mining system in Peru and Bolivia resulted in lower household consumption and higher prevalence of stunting amongst the children in the affected communities, compared to unaffected surrounding areas (Dell, 2010, p. 1899). The focus on mining discouraged investment in other sectors such as agriculture and human capital, which has left the affected groups of people more vulnerable to malnutrition and poverty.

3. Dutch and British Colonial Policies in Borneo

Case Selection

The island of Borneo provides an interesting case to investigate whether differences in past colonial policies have had long-term effects on health outcomes. At present, the island consists of a Malaysian part in the north, and an Indonesian part in the south, often referred to as Kalimantan. During the colonial period the Indonesian part was colonised by the Dutch, whereas the Malaysian part was under British colonial rule. Brunei is the third country situated on the island and has also been under British influence in the past; however, it remained a self-functioning sultanate and was only classified as British protectorate. A comparison between the British- and Dutch-colonised part of the island allows to test whether differences in extractive colonial rule have had diverging health outcomes. As previously discussed, similar studies on development outcomes have been conducted, such as the comparison between colonial rule by British and French colonisers in Cameroon (Lee & Schultz, 2012). Other studies have looked at variation between extractive institutions within one colonial territory, such as study on forced labour in Peru and Bolivia (Dell, 2010). However, up until now the case of Dutch and British extractive colonial rule in Borneo has not been studied in a similar way, and therefore could provide new insights in the relationship between colonial legacies and health outcomes. Additionally, it will serve as an illustrative case in which initial colonial interest sparked because of the presence of valuable natural resources, such as oil and rubber (Eilenberg, 2014, p. 14).

Prior to colonial rule in Borneo, no formal political borders existed on the island. Authority of existing kingdoms was based on the number of subjects under their rule rather than territorial control, and there were large groups of free indigenous people, the Dayaks (Eilenberg, 2014, p. 13). The Anglo-Dutch Treaty of 1824 was the first attempt by British and Dutch colonisers to delineate colonial spheres of influence in the Malay Archipelago (Irwin, 1955). While the agreement did not explicitly partition Borneo, it set the stage for the 1891 Anglo-Dutch Boundary Agreement which formalised the division of the island as a result of concern by the Dutch over increased British interference on the island. The location of the border was based on both economic compromises between the two colonial powers, and social and geographical features like rivers, mountains and conflicts between indigenous subgroups that provided convenient demarcation lines (Amster, 2006, p. 210). At the same time this division did not consider the circumstances of the communities living in the border area, disrupting traditional forms of governance and fracturing indigenous groups by demanding them to adapt to different colonial systems (Eilenberg, 2014, p. 13).

Once Indonesia gained its independence in 1949, and North-Borneo in 1957, the colonial border in Borneo remained in place and was recognised by both states. However, a conflict arose when Indonesia opposed the formation of Malaysia in 1963, which involved the North-Borneo states joining the federation of Malaysia (Abdullah, Anuar & Hara, 2022, p. 200). The conflict, known as the Konfrontasi, lasted for three years. This affected the border area in Borneo on both sides, fostering a stronger sentiment of separate national identities, whereas previously people in the border area shared similar ethnic and cultural identities. As a result of the conflict the two countries started to prioritise defence and security issues in the border area, over economic and social development (p. 205). Resulting in more difficulty for people to trade and maintain social relationships between the regions. Studies have shown, however, that in practice border enforcement is not very effective due to weak security capacity and low prioritisation by the state because of the remoteness of the region. Amster (2006, p. 222) finds that the communities of the Kelabit highlands in the border area continue to interact between the border and that trade, social connections, and cultural events are still common. Since this situation in reality contrasts with the official narrative of a strict boundary, not much is known about the possibility for people to use healthcare facilities on the other side of the border. It is plausible that informal social relationships between the two sides of the border are common, but that access to health facilities requires more official registration and documentation which is restricted by state regulations.

Dutch and British Colonial Policies

Both Dutch and British colonial rule in Borneo were fundamentally extractive types of colonialism. However, they had diverging methods of governance, economic extraction and interaction with the indigenous communities. Where the Dutch implemented direct forms of state control and resource exploitation, the British ingrained more indirect, semi-autonomous forms of governance.

Before 1900 the Dutch regarded Borneo as less of a priority within the big territory that constituted the Dutch East Indies, other areas such as Java and Sumatra had larger strategic and economic importance (Black, 1985, p. 281). In Borneo, the Dutch primarily adopted forms of indirect rule, governing through local leaders of the indigenous Dayaks and establishing economic agreements that benefited the colonisers. However, after 1900, a shift towards more direct administration was made, which involved suppressing local resistance and reducing autonomy of local governments structures (p. 282). This shift was the result of discovery of valuable resources such as oil and rubber, and increasing threats in the forms of other colonial powers that showed interest in the region (p. 284). The British governance in Borneo remained more decentralised. Sarawak, the largest state on the island, had been under rule of the Brookes since the 1840s, a British dynasty operating as the White Rajahs. Their rule blended local customs with administrative structures, through maintaining autonomy of local indigenous Dayak leaders (von Feigenblatt, 2010, p. 117). The east part of the British territory, Sabah, had been largely neglected until the late nineteenth century, for similar reasons as the Dutch. However, from 1870 onwards, the region came largely under control of the British North Borneo company. With the main objective to extract more economic benefits from the region, the company implemented more centralised and formalised administrative structures (Doolittle, 2003, p. 101). This distinction reflects different priorities: the Dutch focused on territorial integration and state control, while the British, especially the Brookes placed more emphasis on building cooperative relationships for local stability.

Both systems prioritised economic returns over local welfare for which large-scale plantations were established, rubber in particular turned out to be a lucrative resource in Borneo. This influenced the local population in the Dutch-colonised part significantly. Under the introduction of the 'domein verklaring' most were stripped of their land rights. This policy declared all native and communal lands without formal recognition under Dutch civil law as state domain, effectively including all indigenous land claims that were based on local law systems (Fahmi, 2020, p. 117). These policies led to institutionalised racial hierarchies, as the Dayak population was marginalised in favour of mostly Chinese migrant labourers who were deemed as more fitting to work in plantation systems (Peluso, 2009, p. 74). This land tenure system had been adopted once Indonesia became independent, resulting in a continuation of disputes between the government and indigenous groups. The Brooke administration in Sarawak, on the contrary, made significant efforts to secure property rights for indigenous groups (Cramb, 1993, p. 292). A formalised land tenure system was put in place in which indigenous land claims could be translated into native titles, designed to integrate indigenous land tenure systems into a legal framework. These rights were still often limited but the administration did make efforts to find a middle ground between acknowledging local rights and ensuring control over the land for economic development and resource extraction (p. 293).

The British North Borneo Company in Sabah, pursued a more extractive policy. While, still trying to recognise some native land claims to avoid large resistance by the local population, their approach was even more so focused on advancing commercial interest (Doolittle, 2003, p. 103). Thus, while the main priority on the British side was still resource extraction for the benefit of the coloniser, the land tenure system did not result in the formation of ethnic hierarchies to the same extent as it did on the Dutch side.

Hypothesis

Based on the preceding theoretical discussion it is likely that the more direct rule by the Dutch, which disturbed local governance structures and introduced ethnic disparities, has had worse long-term effects on health outcomes, as opposed to the British policies. While also exploitative, British rule, especially under the Brookes, ensured better protection of local rights and maintenance of local governance structures. Therefore, the following hypothesis can be established:

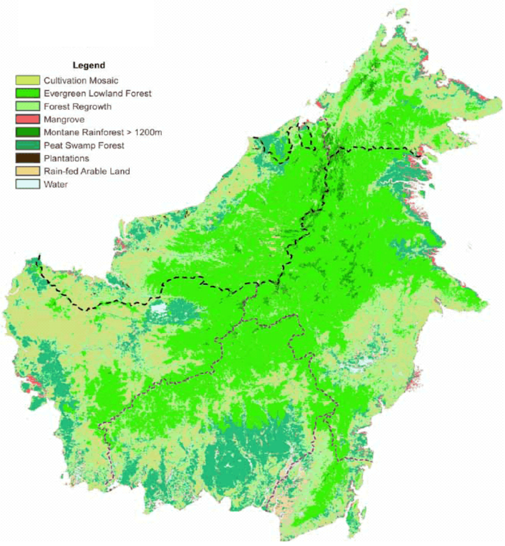
H_1: Health outcomes on the British-colonised part of Borneo are better compared to the Dutch-colonised part

4. Research Design

Methodology

This study takes advantage of a natural experiment created by the geographic border dividing the Dutch- and British-colonised side of Borneo, to evaluate whether health outcomes differ on either side due to the colonial legacies (Posner, 2004, p. 530). The validity of a natural experiment rests on the premise that before the treatment was assigned, in this case the difference in colonial rule, the circumstances of the subjects are similar (Berger, 2009). The earlier discussion on the border in Borneo has demonstrated that prior to colonial intervention, the island lacked formal borders and was for the most part inhabited by indigenous Dayak communities. Furthermore, it is necessary to establish that both sides of the border have similar geographic and climatic characteristics, in order to rule out that any such differences influence health outcomes (Mattingly, 2017, p. 447). *Figure 1* demonstrates that levels of altitude on both sides of the border are rather similar (Feher, n.d.). Similarly, *Figure 2* demonstrates that the vegetation types are similar for both sides of the border, confirming that the two sides of the border share similar climatic characteristics (Langner & Siegert, 2005).

Figure 1 and 2: Geographic and Climatic Characteristics of Borneo



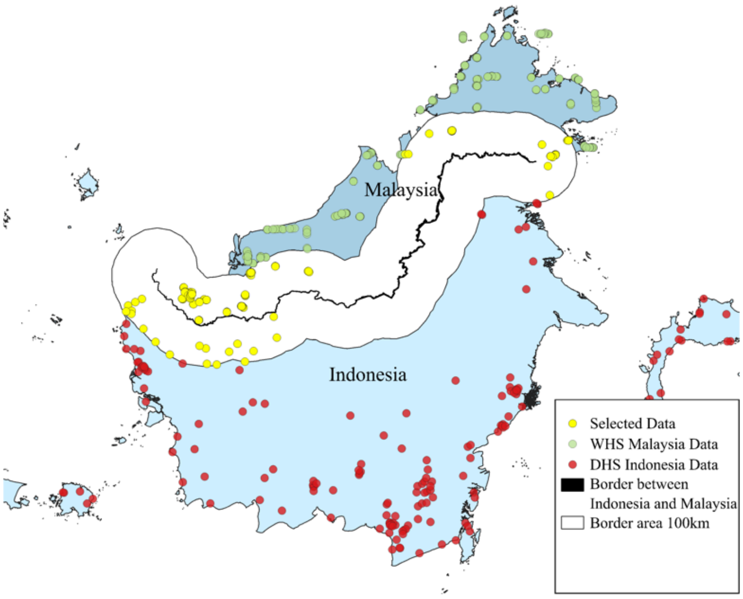
To guarantee that differences in socio-economic and geographic characteristics, that may arise over great distances, do not confound the study, only data within a 100 kilometres threshold from the border is selected. This threshold is based on a balance between having sufficient data for a representative sample on both sides of the border and to still ensure that the characteristics of the territories on both sides remain similar (Mattingly, 2017, p. 447). To estimate causal effects, this study adopts a geographic regression discontinuity design, which rests on the premise that an administrative boundary splits units into a treatment and control group, with the border between Dutch and British colonial rule serving as this key dividing line (Keele & Titiunik, 2015, p. 128). The regions on both sides of the border represent two different "treatments" due to the difference in colonial institutions. This design employs a normal regression analysis of the data, in which the jump between the two sides of the border is a binary variable, and the distance to the border is controlled for (p. 130). Additionally, demographic control variables are included to control for differences in access to the investigated health outcome, ensuring that the effect of the border on the health outcome is isolated.

<u>Spatial Analysis of the Data</u>

The availability of data for the British-colonised side of the border is limited, which restricted the research to use data from the Malaysian World Health Survey (WHS) conducted in 2003 (WHO, 2005). This survey is funded, and under the assistance, of the World Health Organisation. It includes individual-level data gathered through interviews with randomly selected household members, who are asked about their health situation on the individual and household level. The fact that this data was collected at the beginning of the century does not prove to be an issue as the research aims to look at the effect of colonial policies, which is relevant over a long period of time. In order to match the time period, the individual-level data for the Dutch-colonised side derives from the 2002-3 Demographic Health Survey (DHS) of Indonesia (BPS, ORC Macro & ICF, 2003). This survey is carried out and funded by a collaboration of national health institutions and the DHS program. The survey similarly includes data on a randomly selected household members, who are asked about their health situation on the individual and household level.

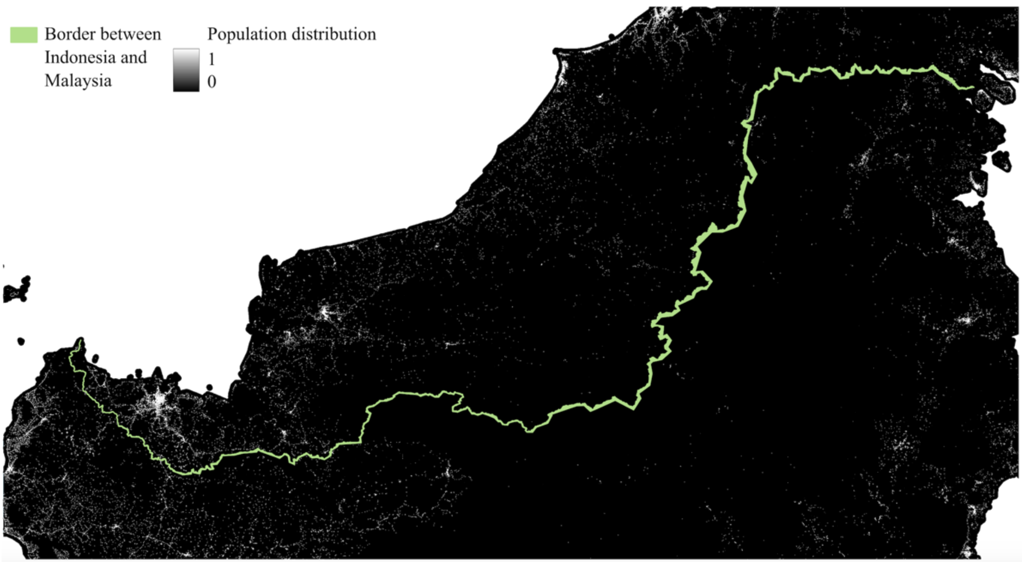
Both surveys include the approximate geographic coordinates of the data, which are used to perform a spatial analysis positioning the data relative to the border, as demonstrated in *Figure 3* below. The DHS datapoints represent a cluster of data within a range of ten kilometres for rural points and two kilometres for urban points, each datapoint for the WHS survey represents a single selected household. The datasets containing the coordinates did not include all of the required variables for the research, and therefore, additional datasets were merged with the coordinate's dataset. Subsequently, the datapoints that fell within the range of 100 kilometres on both sides of the border were selected and merged into an all-encompassing dataset.

Figure 3: Spatial Analysis of the Data



As *Figure 3* demonstrates, the majority of the selected datapoints are located on either the west or east end of the border. To ensure that this will not lead to a substantial bias in the results, a comparison between the datapoints and the distribution of the population is showcased in *Figure 4.* This map is composed through use of the Global Human Settlement population grid dataset of 2005 (Schiavina et al., 2023). The white spots indicate that the area has at least one inhabitant, while the black parts indicate that the area is uninhabited. As desired, the locations of the selected data in *Figure 3* are consistent with the distribution of the population. The large uninhabited area in the centre of the island can be explained by the presence of a mountainous region that is visible in *Figure 1*.

Figure 4: Distribution of the Population Around the Border in Borneo



The Independent and Dependent Variable

The independent variable is based on the country of residence of the respondent, as the research aims to find if health outcomes are significantly different on the British-colonised side of the border compared to the Dutch-colonised side. This variable is binary, with 0 indicating British-colonised and 1 Dutch-colonised.

The dependent variable aims to represent the health outcomes on both sides of the border. Due to the fact that the data derives from two different surveys and different organisations, it is complicated to find suitable and comparable data. Additionally, it requires a lot of work to ensure that similar variables can also be combined and merged into one dataset, because they are likely to be coded in different ways. The only variable that meets these requirements asks about the primary source of water to which the respondent has access. Access to an improved water source is a relevant measure of health; according to the literature a tenth of global diseases could be prevented or reduced in their occurrence through access to safe water (Prüss-Üstun et al., 2008). Amongst the diseases that are known to be more prevalent due to the use of unsafe water are typhoid, polio, intestinal helminths, diarrhoeal diseases, and it is closely related to malnutrition (Afifah et al., 2018, p. 31). However, the limitations of this measure should be acknowledged. A lack of access to safe water is merely an indicator for higher susceptibility to the abovementioned diseases, it does not translate into actual figures of the amount of people that are affected by these illnesses or the number of fatalities they cause. Importantly, it also does not account for any measurement of access to healthcare, it could be that good access to healthcare limits the consequences of waterborne diseases. Lastly, children seem most susceptible to many of these illnesses, therefore the measurement might not be equally representative across the entire population.

Initially, both datasets included a variable which looked at the main source of drinking water for the respondent. However, the options were different in both surveys and therefore these variables have been recoded into a binary variable, where 0 indicates no access to an improved water source and 1 indicates that the respondent does have access. These variables were recoded based on the definition and coding guidance of the Joint Monitoring Program for Water Supply, Sanitation and Hygiene (JMP), initiated by the World Health Organisation and UNICEF (WHO, UNICEF, n.d.). This program differentiates between access to improved, safe, drinking water facilities which broadly include all piped supplies, and unimproved, unsafe, drinking water facilities which are defined as non-piped supplies or surface water. Due to a difference in the coding of rainwater in the original variables an alternative independent variable has been established. Namely, the original data from the DHS for the Dutch-colonised side subdivided rainwater into the category of surface water, which is not deemed as an improved source. However, under the JMP rainwater is considered as improved when it is immediately collected into a tank after it fell. As this is not defined in the dataset of the DHS the first two models of the logistic regression in *Table 1* work with the independent variable where rainwater on the Dutch-colonised is coded as unimproved. Model three uses the alternative independent variable with the most optimistic scenario in which rainwater on the Dutch-colonised side was collected into a tank and would thus be deemed as improved. This issue was not present in the WHS data for the British-colonised side where a distinction was made between surface water and rainwater that is collected into a tank.

Control Variables

Control variables are included to estimate the effect between the independent and dependent variable, while excluding for differences in demographic characteristics on the individual level. As a natural experiment implicates, at the time of colonisation these differences did not exist, and any differences that now exist are ideally to be attributed to the colonial legacies. However, it should be acknowledged that other influences, such as the changes in regime and policies following the independence of both countries, have had lasting impacts on the observed individuals over time. To exclude the effect of these changes, control variables for wealth, gender, education, and age are included in the analysis. Additionally, a control variable for distance to the border is included as part of the geographic regression discontinuity design. All the data for these control variables derives from the earlier discussed DHS and WHS surveys.

First, education is included as a continuous variable, measured in total years of formal schooling. Higher educational attainment is generally associated with more knowledge on health-related issues, and an increased likelihood of understanding the importance of safe water (Fuchs, Pamuk & Lutz, 2010, p. 181). Second, proxy measures for wealth are included, as wealthier people tend to have better access to healthcare and housing with an improved water source (p. 183). Due to limitations in the data, wealth is measured through a binary variable for television and car ownership, in which ownership of a television reflects moderate levels of wealth and car ownership indicates higher levels given the greater financial resources required. Third, a control variable for gender is included; research emphasises that women are often responsible for household water management, leaving them at greater risk of exposure to negative consequences of unsafe water (Sweetman & Medland, 2017, p. 155). Fourth, age is included as a continuous variable to account for generational differences. Older individuals may have better access to safe water and health due to wealth accumulation over time, whereas younger individuals may value access to safe water more as they are likely to be better educated. Finally, an additional control variable is included that accounts for the distance to the border, addressing the possibility that the observed relationship is influenced by systematic differences in the proximity of data collection points to the border on one side, in comparison to the other side (Keele & Titiunik, 2015, p. 137). In particular, datapoints farther from the border could influence the results disproportionally because differences in socio-economic and geographic characteristics tend to increase over greater distances.

5. Results

Table 1 presents the results of the logistic regression with controls that investigates whether a relationship can be established between the independent variable that differentiates between the British-colonised and the Dutch-colonised side of the border, and the dependent variable of access to an improved water source.

Table 1. Logistic regression analysis on access to improved water source (odds ratios)

	Model 1	Model 2	Model 3 (Alternative coding of rainwater)
(Constant)	11.905***	1.093	2.495
Dutch-colonised	0.009***	0.011***	0.278***
	[0.005; 0.017]	[0.006; 0.022]	[0.165; 0.469]
Distance to border	1.003	0.998	0.986***
	[0.992; 1.014]	[0.986; 1.010]	[0.978; 0.994]
Men		0.764	0.718
		[0.434; 1.346]	[0.452; 1.141]
Age		1.016	1.013
		[0.994; 1.039]	[0.993; 1.033]
Tv (Wealth)		2.661***	6.096***
		[1.524; 4.646]	[4.102; 9.061]
Car (Wealth)		2.290	2.851*
		[0.929; 5.648]	[1.095; 7.419]
Education		1.246***	1.076**
		[1.160; 1.339]	[1.019; 1.136]
-2LL	553.674	455.034	685.610
Cox and Snell's R^2	0.533	0.582	0.306
Nagelkerke's R^2	0.712	0.777	0.451
N	891	891	891

Note: odds ratios with 95% confidence intervals in brackets
***$p < 0.001$, ** $p < 0.01$, *$p < 0.05$

The first model includes the independent and dependent variable, while adjusting for the distance to the border as part of the geographic regression discontinuity design. The results indicate that the probability of having access to an improved water source is significantly lower on the Dutch-colonised side of the border (Wald= 255.477, $p < 0.001$). The odds ratio of 0.009 for the Dutch-colonised side show that the odds of having access to an improved water source on this side of the border are 99% lower than the odds for having access on the British-colonised side.

Model two includes the demographic and geographic control variables. As the results show, the odds ratio for the Dutch-colonised side remained similar, the odds of having access to an improved water source are 0.011 times the odds of having access on the British-colonised side. This result remains statistically significant (Wald= 165.432, $p < 0.001$). Additionally, education has a significant effect on the odds of having access to an improved water source. The odds ratio of 1.246 indicates that the odds of having access to an improved water source are 25% higher for each additional year of education followed (Wald= 36.337, $p < 0.001$). This is consistent with the earlier discussed literature, which indicated that higher educated individuals are more aware of the importance of safe water and health. Furthermore, the proxy control for wealth of owning a tv has a significant positive effect as well (Wald=11.844, $p < 0.001$). This result is consistent with the literature, because more wealth generally indicates better access to healthcare and housing with access to safe water. The other control variables do not significantly affect the probability of having access to an improved water source, and the observed effects align with the general theory. Although the control variables did not change the effect of the independent variable on the dependent variable much, the number for Nagelkerke's R^2 is closer to one which shows that second model fits the data better and has better explanatory power.

The third model includes a separately run regression with the alternative dependent variable to account for the difference in coding of rainwater in the two surveys. The odds ratio of having access to an improved water source for the Dutch-colonised side are 0.278 times the odds of having access on the British-colonised side, this result remains statistically significant, but the effect is smaller (Wald= 22.992, $p < 0.001$). The smaller effect can be explained by the fact that all respondents on the Dutch-colonised side with access to rainwater are now coded as having access to an improved water source, resulting in an increase in the total share of people with access. The results for the control variables do not differentiate substantively from the previous model.

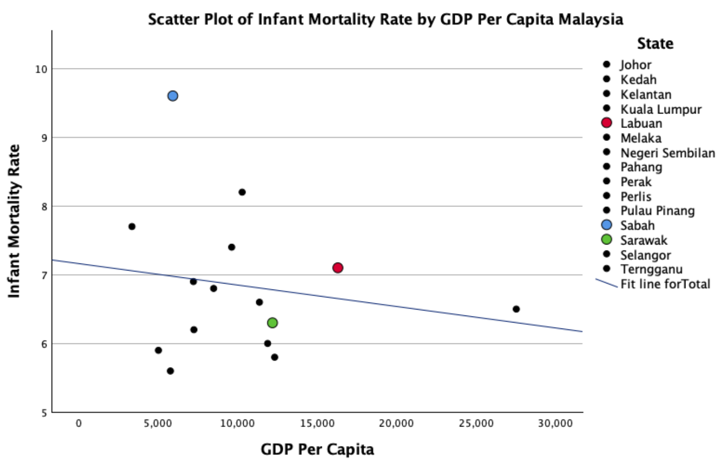
Assumptions

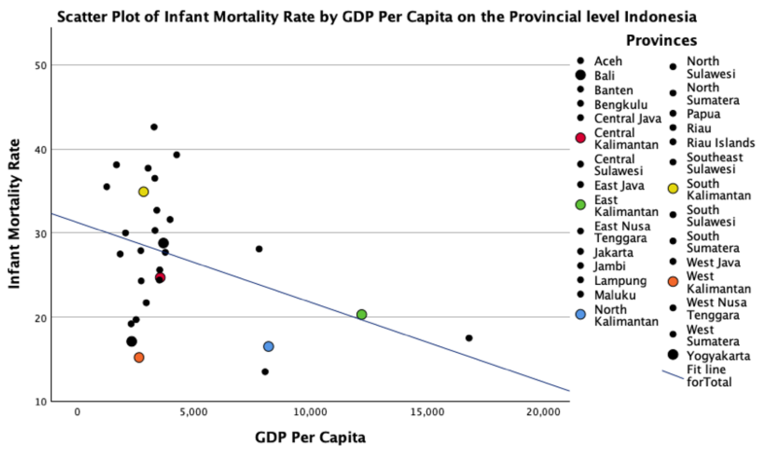
The appropriate assumptions have been tested to ensure that the data is suitable to run a logistic regression and are included in the appendices. The assumption of multicollinearity, which checks that the independent variables are not highly correlated, has been successfully met by testing the VIF and tolerance statistics. The cook's distance and the DFBeta values that test the presence of influential cases are also below the desired values. However, in all models there are a large numbers of cases with high standardised residuals, which could disproportionately affect the results. *Appendix A* and *B* include additional regressions in which the cases with high standardised residuals are taken out. The direction and significance of these results remain similar, and it can therefore be concluded that these cases do not substantially affect the results. Furthermore, *Appendix C* includes an additional regression in which the selected data falls within a 90 kilometres threshold of the border, demonstrating that consistent results are found when a different definition of the border area is applied.

Discussion

All three models show a significant effect of the independent variable on the dependent variable, even when controlling for possible demographic differences that may have developed over time. Taken together, the results are consistent with the hypothesis that health outcomes on the British-colonised side of the border are better compared to the Dutch-colonised side, leading to a failure in rejecting the null hypothesis of no difference between both sides of the border.

Figure 5 and 6: Scatter Plots on GDP per Capita by IMR for Malaysia and Indonesia on the State Level





The scatterplots in *Figures 5* and *6* furthermore illustrate that, based on the relationship between the gross domestic product (GDP) per capita and infant mortality rate (IMR), Borneo as an island is not a notable outlier compared to the rest of both countries. The tables are based on data from the WHO and Malaysia's department of statistics of 2017, due to limited availability of complete data in earlier years (WHO, 2022; OpenDOSM, 2022). IMR is defined as the yearly number of children who die before the age of one per 1000 live births and serves as an important indicator of health, reflecting the notion that structural factors influencing public health also affect infant mortality rates (Reidpath & Allotey, 2003, p. 344). The plots suggest that the findings from the island are generalisable to the rest of Indonesia and Malaysia, indicating that colonial legacies are likely to have persisting effects on the countries as a whole as well.

6. Alternative Explanations

The results of the regression demonstrate a strong, significant effect between the independent and dependent variable. However, it is important to investigate and rule out alternative explanations for this relationship between the side of the border on which the respondent resides and access to safe water (Dunning, 2008, p. 285) This will ensure that the observed relationship is not due to other circumstances, such as the possibility of self-selection in the form of migration and changes in regime and policy after independence (p. 288).

Migration

It is important to examine whether it is easy for inhabitants of the border area to migrate to the other side permanently. If so, it could be that only a select part of the Indonesian population migrates because the living circumstances on the Malaysian side of the border seem better. These individuals may be wealthier and more educated, and therefore better able to get access to safe water on the other side of the border. This could result in a biased sample of people that remain on the Indonesian side who are unable to make this choice. The DHS and WHS surveys did not include any data on migration or birth location which made it difficult to control for this potential issue in the data analysis. However, several studies have looked at migration patterns between both countries revealing that temporary labour migration, rather than permanent migration, is common.

As outlined by Hugo (1993, p. 53), migration is most common from Indonesia to Malaysia, with many individuals seeking economic opportunities on the Malaysian side, where job security is higher due to labour shortages in the market. It should be noted that these shortages are most prevalent in the lower wage sector, and it is thus predominantly the less affluent part of the Indonesian population that temporarily moves to earn income and later return to their families (p. 45). Therefore, it is not likely that these people are more able to obtain access to safe water on the Malaysian side, as they belong to the poorer part of the population. Furthermore, the temporary nature of migration makes it goal-oriented rather than permanent. As earlier discussed the relation between Indonesia and Malaysia has led to policy restrictions on border movement, which limits the ease of migration (Abdullah et al., 2022, p. 205). The weak enforcement of these restrictions by the two states has resulted in a large number of illegal labour migrants on the Malaysian side. As Darmayani (2021, p. 7) emphasises, these illegal movements lead to human security concerns, particularly in the form of human trafficking and exploitation. Based on these studies it can be concluded that it is unlikely that self-selection in the form of migration caused large biases in the data samples, as labour migration is most prevalent amongst the less affluent part of the Indonesian population, is often temporary, and involves significant risks, making migration less attractive in the first place.

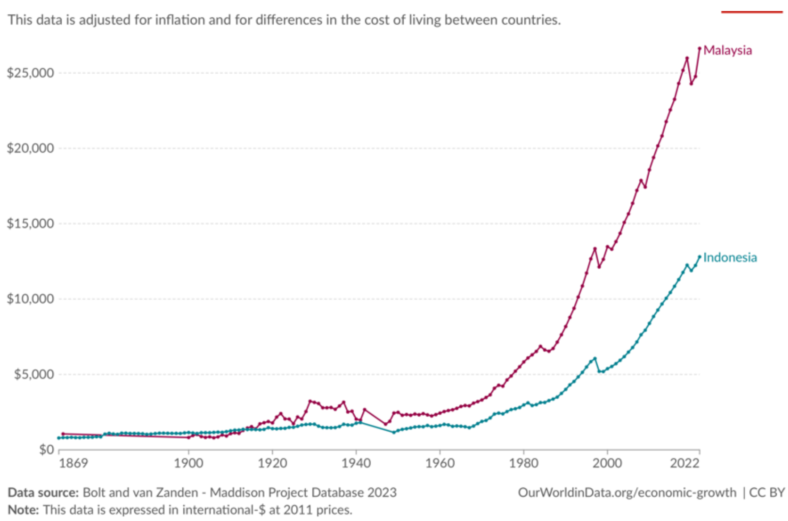
Post-colonial changes

Additionally, it is crucial to establish that changes, unrelated to colonial legacies, in regime and policies following the independence of both countries are not the main drivers behind the observed relationship. Instead, colonial legacies are the primary factor influencing health outcomes, particularly in terms of safe water access. It is important to recognise that these changes after independence have led to differing circumstances on either side of the colonial border. However, significant continuities from colonial rule are likely to explain the results of the data analysis. Not much research has been done specifically on safe water access in Borneo, however, general policy development on the national level is likely to explain the outcomes for Borneo as well.

Indonesia's immediate post-independence governance was largely influenced by the highly centralised nature of Dutch colonial rule, as the majority of republican nationalist leaders that held executive and legislative office were Dutch-educated (Moser, 2008, p. 116). Initial attempts at democracy failed when Sukarno became the first authoritarian president of Indonesia in 1957 and an end to authoritarian rule only came after the fall of the second president, Suharto, in 1998 (Brown, 2003). In the period that followed, Indonesia underwent substantial democratisation and decentralisation reforms. However, the long period of centralised authoritarian rule and the earlier discussed direct rule by the Dutch had weakened and abandoned local government structures, limiting the effectiveness of these decentralisation reforms. The post-independence regime type of Malaysia is officially a parliamentary democracy, formed and introduced towards the end of British colonial rule (Case, 1993, p. 186). However, in practice the Barisan Nasional coalition governed as a one-party state for decades, maintaining power through patronage and suppressing opposition until recent electoral defeat in 2018 (Nadzri, 2018, p. 143). Initially, this coalition had consisted mainly of officials that had worked in the British colonial government. Nevertheless, the strong central coordination and nation-building priorities under this rule did provide the stability necessary for development, as opposed to the political instability in Indonesia that hindered development.

In addition, colonial legacies are reflected in the economic development of both countries. As illustrated in *Figure 7*, Indonesia had a slightly higher GDP per capita at the beginning of the twentieth century under Dutch colonial rule compared to British-ruled Malaysia (Bolt, van Zanden, & Our World in Data, 2023). However, increased colonial investments under British rule in infrastructure and resource extraction in the following decades provoked substantial economic growth, surpassing Indonesia's GDP per capita level. Between 1945 and 1960, around the time that both countries gained independence, Malaysia already had about twice the GDP per capita of Indonesia. Furthermore, it is important to argue that it is unlikely that Malaysian Borneo is just better developed because Malaysia as a whole is richer and has subsidised growth in its Borneo states. As the earlier discussed *Figure 5* demonstrates, the GDP's per capita for the three states on Borneo are not remarkably different from the other states of the country, indicating that their wealth levels are similar. Although the political regime changes after independence likely had substantial impacts on the economic development of both countries, *Figure* 7 shows that the divergence in wealth did not appear after independence. Instead, these differences already existed in the colonial period and have increased ever since.

Figure 7: GDP per Capita for Malaysia and Indonesia, 1869 to 2022



Finally, legacies of colonial rule are present in the development of safe water supply systems as well. During the period of authoritarian rule in Indonesia, water source development was overseen by the national Division of Human Settlement which handled the construction of new water supply systems, and the operational and maintenance responsibilities were delegated to local water enterprises, PDAMs, owned by local governments (Ryu, 2019, p. 7). However, the weak local government structures resulting from direct colonial rule by the Dutch led to suboptimal developments of the water supply systems, as collaboration between the national and local organisations was hindered by weak local leadership and technical capacity. The decentralisation reforms introduced after the fall of the authoritarian regime, delegated more authority to local governments on public goods provision and resource allocation (p. 8). This shift intended to empower local authorities, but instead it highlighted the issues stemming from colonial rule, and many PDAMs suffered from performance issues. Malaysia's efforts towards water supply development are similarly shaped by colonial legacies of British rule. The indirect type of colonial rule made that local government structures were kept in place, which is reflected in the town boards that were established under colonial rule to oversee policies and infrastructure aimed at improving water quality (Ryu, 2019, p. 10). While at the time reflecting the objective of improving living standards to support the colony's economics functions, these boards continued to exist after independence (Abdullah et al., 2024, p. 26). Similar local initiatives did not exist under Dutch colonial rule on the other side of the border. The creation of the National Development Planning Committee and the Economic Planning Unit under the first Malaysian development plan after independence, reinforced centralised water governance and coordinated the water projects on the national level. This body coordinates technical challenges, delays and allocation of resources towards local projects (Ryu, 2019, p. 11). This centralised governance, paired with better-preserved local institutions, enabled Malaysia to develop effective and safe water supply systems, unlike Indonesia.

7. Conclusion

This research aimed to analyse how differences in colonial policies have long-term effects on health outcomes in former colonial territories. To investigate whether a relationship between these two variables could be established, the case of Borneo was studied. The different extractive colonial institutions implemented under Dutch and British colonial rule on the island, along with the establishment of the border between the two territories during this period, presented the opportunity to take advantage of a natural experiment and investigate whether health outcomes on both sides of the border are different as a result of these institutions. Based on theoretical expectations, the difference between more indirect rule by the British as opposed to direct rule by the Dutch, led to the hypothesis that health outcomes on the British-colonised side are better compared to the Dutch-colonised side. The hypothesis was tested using a geographic regression discontinuity design, and the results aligned with the expected hypothesis, showing that health outcomes on the British-colonised side are significantly better than on the Dutch-colonised side. These findings were further corroborated by a discussion of alternative explanations for the observed relationship. Based on these results it can therefore be concluded that colonial policies have long-term effects on health outcomes, and that different colonial institutions lead to differences in these health outcomes.

The main strength of this research lies in the use of a natural experiment and geographic regression discontinuity design to investigate the relationship between colonial legacies and health outcomes in Borneo, a case that has not been studied in this way before. However, the research also has considerable limitations. Due to limited availability of suitable data, the health outcome of access to safe water is not the most representative measure of overall health. While research argues that universal access to safe water could prevent a substantial share of global diseases, the measure does not include specific numbers on prevalence of diseases, mortality rates, or healthcare access. Furthermore, the case of Borneo is not widely generalisable, as colonial powers governed various colonies with distinct objectives. Nevertheless, for cases where the colonised territory is predominantly rural, resource-rich, and the difference between colonial rule lies in the distinction between direct and indirect rule, the study could provide valuable insights.

This research contributes to the existing literature by investigating the effect of different types of extractive colonial institutions on health outcomes for the island of Borneo, an analysis not previously conducted. Furthermore, it finds that colonial legacies persist, and therefore, that history matters for understanding patterns in global health in former colonial territories. Future research on Borneo specifically, could focus on investigating the effect of colonial legacies on a broader scale of health outcomes to enhance the robustness of the findings, if availability of new data would allow to do so. Moreover, this research has proposed that there are important differences between different types of extractive institutions, which could be explored to a greater extent in future research. Overall, recognising the importance of colonial legacies and their persistent effects on public health should be a priority in both academic and political spheres.

Reference list

Abdullah, M. F., Ibrahim, W. K. W. I., Noor, A. M., Hassan, N. M., Ali, I., & Mohamad, A. F. (2024). Water Architecture in Malaysia: A Case Study in Penang, 1957-1985. *Journal of Sustainability Science and Management*, *19*(7), 23-38.

Abdullah, M. K., Anuar, A. R., & Hara, A. E. (2022). Contesting authority discourses in defining relations between Indonesia and Malaysia: A case study in the Kalimantan border areas. *Journal of International Studies*, *18*, 191-217.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, *91*(5), 1369-1401.

Afifah, T., Nuryetty, M. T., Cahyorini, Musadad, D. A., Schlotheuber, A., Bergen, N., & Johnston, R. (2018). Subnational regional inequality in access to improved drinking water and sanitation in Indonesia: results from the 2015 Indonesian National Socioeconomic Survey (SUSENAS). *Global health action*, *11*(sup1), 31-40.

Amster, M. H. (2006). Narrating the Border: Perspectives from the Kelabit Highlands of Borneo. In *Centering the Margin* (Vol. 4, pp. 207–228). Berghahn Books.

Badan Pusat Statistik-Statistics Indonesia (BPS) [Indonesia], ORC Macro & ICF. (2003). *Indonesia Demographic and Health Survey 2002-2003* [Dataset]. Calverton, Maryland, USA: BPS, ORC Macro and ICF.

Banerjee, A., & Iyer, L. (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in India. *American economic review*, *95*(4), 1190-1213.

Berger, D. (2009). Taxes, institutions and local governance: evidence from a natural experiment in colonial Nigeria. *Unpublished manuscript*.

Black, I. (1985). The "Lastposten": Eastern Kalimantan and the Dutch in the nineteenth and early twentieth centuries. *Journal of Southeast Asian Studies*, *16*(2), 281-291.

Bolt, van Zanden, & minor processing by Our World in Data (2023). *GDP per capita – Maddison Project Database – Historical data* [dataset]. Maddison Project Database 2023 [original data]. Retrieved from https://ourworldindata.org/grapher/gdp-per-capita-maddison

Brown, C. (2003). *A short history of Indonesia: The unlikely nation*. Allen & Unwin.

Burkle, F. M. (2006). Globalization and disasters: Issues of public health, state capacity and political action. *Journal of International Affairs*, 241-265.

Case, W. (1993). Semi-democracy in Malaysia: Withstanding the pressures for regime change. *Pacific affairs*, 183-205.

Cramb, R. A. (1993). The evolution of property rights to land in Sarawak: An institutionalist perspective. *Review of Marketing and Agricultural Economics*, *61*, 289-300.

Darmayani, A. (2021). *A Fluid Border? Indonesia-Malaysia Labour Migration and Human Trafficking*. International Institute of Social Studies, The Hague.

Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica*, *78*(6), 1863-1903.

Dell, M., & Olken, B. A. (2020). The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java. *The Review of Economic Studies*, *87*(1 (312)), 164–203.

Doolittle, A. (2003). Colliding discourses: Western land laws and native customary rights in North Borneo, 1881-1918. *Journal of Southeast Asian Studies*, *34*(1), 97-126.

Dunning, T. (2008). Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly*, *61*(2), 282-293.

Eilenberg, M. (2014). Evading Colonial Authority. Rebels and Outlaws in the Borderlands of Dutch West Borneo 1850s–1920s. *Journal of Borderlands Studies*, *29*(1), 11-25.

Fahmi, C. (2020). The Dutch colonial economic's policy on natives land property of Indonesia. *Petitta*, *5*, 105.

Feher, D. (n.d.). *Borneo Map*. FreeWorldMaps. Retrieved from: https://www.freeworldmaps.net/asia/indonesia/borneo.html

von Feigenblatt, O. F. (2009). Some lessons from the White Rajahs of Sarawak: Identifying and strengthening local capacities for peace in the deep South of Thailand. *Journal of Law and Conflict Resolution*, *1*(6), 117-121.

Fuchs, R., Pamuk, E., & Lutz, W. (2010). Education or wealth: which matters more for reducing child mortality in developing countries?. *Vienna Yearbook of Population Research*, 175-199.

Horvath, R. J. (1972). A definition of colonialism. *Current anthropology*, *13*(1), 45-57.

Hugo, G. (1993). Indonesian labour migration to Malaysia: trends and policy implications. *Southeast Asian Journal of Social Science*, *21*(1), 36-70.

Irwin, G. (1955). *Nineteenth-century Borneo: a study in diplomatic rivalry* (1st ed.). Brill.

Keele, L. J., & Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis*, *23*(1), 127-155.

Kosack, S., & Fung, A. (2014). Does transparency improve governance?. *Annual review of political science*, *17*(1), 65-87.

Langner, A., & Siegert F. (2005). *Assessment of Rainforest Ecosystems in Borneo using MODIS satellite imagery.* Remote Sensing Solutions GmbH & GeoBio Center of Ludwig-Maximilians-University Munich.

Lee, A., & Schultz, K. A. (2012). Comparing British and French Colonial Legacies: A Discontinuity Analysis of Cameroon. *Quarterly Journal of Political Science*, *7*(4), 1-46.

Macfarlane, S., Racelis, M., & Muli-Muslime, F. (2000). Public health in developing countries. *The Lancet*, *356*(9232), 841-846.

Macinko, J., Starfield, B., & Erinosho, T. (2009). The impact of primary healthcare on population health in low- and middle-income countries. *The Journal of Ambulatory Care Management*, *32*(2), 150–171.

Mattingly, D. C. (2017). Colonial legacies and state institutions in China: Evidence from a natural experiment. *Comparative Political Studies*, *50*(4), 434-463.

Moser, S. (2008). Performing national identity in postcolonial Indonesia. *National University of Singapore.*

Nadzri, M. (2018). The 14th General Election, the fall of Barisan Nasional, and political development in Malaysia, 1957-2018. *Journal of Current Southeast Asian Affairs, 37*(3), 139-171.

OpenDOSM. (2022). *Annual Early Childhood Deaths by State & Sex* [dataset]. Retrieved from: https://open.dosm.gov.my/data-catalogue/deaths_early_childhood_state_sex?state=sabah&type=infant&sex=both&visual=table

Peluso, N. L. (2009). Rubber erasures, rubber producing rights: Making racialized territories in West Kalimantan, Indonesia. *Development and change*, *40*(1), 47-80.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American political science review*, *94*(2), 251-267.

Posner, D. N. (2004). The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American political science review*, *98*(4), 529-545.

Prüss-Ustun, A., Bos, R., Gore, F., & Bartram, J. (2008). *Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health*. World Health Organization, Geneva.

Reece, R. L. (2022). Slave past, modern lives: An analysis of the legacy of slavery and contemporary life expectancy in the American south. *Journal of Black Studies*, *53*(7), 677-702.

Reidpath, D. D., & Allotey, P. (2003). Infant mortality rate as an indicator of population health. *Journal of Epidemiology & Community Health*, *57*(5), 344-346.

Ryu, S. K. (2019). Variation in Access to Safe Drinking Water across Different Countries: An Explanatory Framework. *Social Sciences*, *8*(2), 68.

Scagliola, S. (2007). The Silences and Myths of a 'Dirty War': Coming to Terms with the Dutch–Indonesian Decolonisation War (1945–1949). *European Review of History: Revue européenne d'histoire*, *14*(2), 235-262.

Schiavina, M., Freire, S., Carioli, A., MacManus, K. (2023). GHS-POP R2023A – GHS population grid multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset].

Soares, R. R. (2007). On the Determinants of Mortality Reductions in the Developing World. *Population and Development Review*, *33*(2), 247–287.

Spruk, R., & Kovac, M. (2020). Persistent effects of Colonial Institutions on long-run development: local evidence from regression discontinuity design in Argentina. *Journal of Empirical Legal Studies*, *17*(4), 820-861.

Subrahmanyam, G. (2006). Ruling continuities: colonial rule, social forces and path dependence in British India and Africa. *Commonwealth & Comparative Politics*, *44*(1), 84-117.

Sweetman, C., & Medland, L. (2017). Introduction: gender and water, sanitation and hygiene. *Gender & Development*, *25*(2), 153-166.

United Nations Global Response Group (UN GCRG). (2024). *Aid under pressure: 3 accelerating shifts in official development assistance*. United Nations Conference on Trade and Development. Retrieved from: https://unctad.org/system/files/official-document/un-gcrg-oda-report_en.pdf

Wilsford, D. (1994). Path dependency, or why history makes it difficult but not impossible to reform health care systems in a big way. *Journal of public policy*, *14*(3), 251-283.

WHO. (2022). *Child Mortality; Subnational region: Indonesia Data* [dataset]. Retrieved from: https://apps.who.int/gho/data/node.searo.NODESUBREGchildmortality-IDN?lang=en

WHO, UNICEF. (n.d.). The WHO/UNICEF Joint Monitoring Programme Facility Types. Retrieved from: https://washdata.org/monitoring/methods/facility-types

World Health Organization (WHO). (2005). *Malaysia World Health Survey 2003*. Geneva, Switzerland: World Health Organization (WHO).

Ziltener, P., & Künzler, D. (2013). Impacts of colonialism–A research survey. *Journal of World-Systems Research*, *19*(2), 290-311.

Appendices

Appendix A

The following output contains the logistic regression that was conducted to investigate the relationship between the independent variable on the two sides of the border, the dependent variable of access to safe water and the control variables for distance to the border, education, age, gender and wealth, within a threshold of a 100 kilometres. Additionally, the descriptives of the cook's distance and DFBetas are included to check the assumptions for influential cases and a linear regression was run to check for the multicollinearity assumptions, both were met. The frequency tables on the standardised residuals show that there might be an issue with outliers, as there is more than one case with a standardised residual above 3.29 and more than 1% of the cases has a value above 2.58. Due to the fact that the level of observation is individual, and the outcome variable is binary, it is unlikely that this will be a big issue. However to ensure that these cases do not have disproportionate effects on the results an additional logistic regression was run, in which these cases were excluded. The results of this regression show that the odds ratio of having access to an improved source of water on the Dutch-colonised side compared to the British-colonised side are even smaller than they were in the original regression, with odds ratios of 0.006 for block 1 and 0.001 for block 2. The direction and significance of the results remain similar, and it can therefore be concluded that these cases do not substantially affect the results.

```
LOGISTIC REGRESSION VARIABLES Safe_water
  /METHOD=ENTER Country
  /METHOD=ENTER Country Gender Age Education TV Car DistBorder
  /SAVE=PRED COOK DFBETA RESID SRESID ZRESID
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	891	96.4
	Missing Cases	33	3.6
	Total	924	100.0
Unselected Cases		0	.0
Total		924	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
NO	0
YES	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Access to improved water source NO	Predicted: Access to improved water source YES	Percentage Correct
Step 0	Access to improved water source	NO	0	421	.0
		YES	0	470	100.0
	Overall Percentage				52.7

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.110	.067	2.692	1	.101	1.116

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Country of residence	584.944	1	<.001
	Overall Statistics		584.944	1	<.001

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	678.555	1	<.001
	Block	678.555	1	<.001
	Model	678.555	1	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	553.674[a]	.533	.712

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted: Access to improved water source NO	Predicted: Access to improved water source YES	Percentage Correct
Step 1	Access to improved water source	NO	390	31	92.6
		YES	54	416	88.5
	Overall Percentage				90.5

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Country of residence	-4.658	.291	255.477	1	<.001	.009	.005	.017

	Distance to border in km	.003	.006	.260	1	.610	1.003	.992	1.014
	Constant	2.477	.298	69.158	1	<.001	11.905		

a. Variable(s) entered on step 1: Country of residence, Distance to border in km.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

     800 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R    600 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I            Y                                                                                Y      I
E    400 +            N                                                                                Y      +
N        I            N                                                                                Y      I
C        I            N                                                                                Y      I
Y        I            N                                                                                Y      I
     200 +            N                                                                                Y      +
         I            N                                                                                Y      I
         I            N                                                                                Y      I
         I            N                                                                                N      I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 50 Cases.
```

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	98.903	6	<.001
	Block	98.903	6	<.001
	Model	777.458	7	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	455.034[a]	.582	.777

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			Access to improved water source		
	Observed		NO	YES	Percentage Correct
Step 1	Access to improved water source	NO	384	37	91.2
		YES	49	421	89.6
	Overall Percentage				90.3

a. The cut value is .500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper

Step 1[a]	Country of residence	-4.504	.350	165.432	1	<.001	.011	.006	.022
	Gender respondent	-.269	.289	.867	1	.352	.764	.434	1.346
	Current age - respondent	.016	.011	1.941	1	.164	1.016	.994	1.039
	Education in single years	.220	.037	36.337	1	<.001	1.246	1.160	1.339
	Has television	.979	.284	11.844	1	<.001	2.661	1.524	4.646
	Has car/truck	.829	.461	3.238	1	.072	2.290	.929	5.648
	Distance to border in km	-.002	.006	.143	1	.705	.998	.986	1.010
	Constant	.089	.636	.020	1	.889	1.093		

a. Variable(s) entered on step 1: Country of residence, Gender respondent, Current age - respondent, Education in single years, Has television, Has car/truck, Distance to border in km.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

      80 +                                                                                                    Y+
         I                                                                                                   YYI
         I                                                                                                  YYYI
F        I                                                                                                  YYYI
R     60 +                                                                                                  YYY+
E        I                                                                                                  YYYI
Q        I  Y  Y                                                                                            YYYI
U        I  N  N                                                                                            YYYI
E     40 + NN  N                                                                                            YYY+
N        I NN  N                                                                                            YYYI
C        I NN NN                                                                                          YYYYYI
Y        I NNNNNN                                                                                        YYYYYYI
      20 + NNNNNN  N                                                                                     YYYYYY+
         I NNNNNNY N   NN                                                                         Y     YYYYYYYI
         I NNNNNNNNNNNNNNYNN                   Y                                                Y YY   YYYYYYYYI
         I NNNNNNNNNNNNNNNNNNNYNYNNNY         YN                        N     YY       Y Y  NYYYYYYYYYYYYYYYYYYI
Predicted ----------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 5 Cases.
```

```
DESCRIPTIVES VARIABLES=PRE_1 COO_1 RES_1 SRE_1 ZRE_1 DFB0_1 DFB1_1 DFB2_1 DFB3_1 DFB4_1 DFB5_1
  DFB6_1 DFB7_1
 /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Predicted probability	901	.01104	.99928	.5319112	.41955627
Analog of Cook's influence statistics	891	.00000	.25031	.0094178	.02896271
Difference between observed and predicted probabilities	891	-.98011	.98501	.0000000	.27197079
Standard residual	891	-2.80432	2.90148	-.0058092	.72040968
Normalized residual	891	-7.01952	8.10617	.0124562	.96020473

DFBETA for constant	891	-.18420	.11625	.0000151	.02036009
DFBETA for Country of residence	891	-.02976	.12667	.0000041	.01300941
DFBETA for Gender respondent	891	-.04627	.07298	.0000047	.01006195
DFBETA for Current age - respondent	891	-.00330	.00246	-.0000003	.00038259
DFBETA for Education in single years	891	-.00959	.00693	.0000000	.00133883
DFBETA for Has television	891	-.06178	.05140	.0000078	.01039767
DFBETA for Has car/truck	891	-.17241	.13041	-.0000218	.01556851
DFBETA for Distance to border in km	891	-.00130	.00157	-.0000002	.00021039
Valid N (listwise)	891				

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10) TOLERANCE(.0001)
  /NOORIGIN
  /DEPENDENT Safe_water
  /METHOD=ENTER Country
  /METHOD=ENTER Country Gender Age Education TV Car DistBorder.
```

Regression

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	.931	.014		67.170	<.001		
	Country of residence	-.809	.020	-.810	-41.220	<.001	1.000	1.000
2	(Constant)	.647	.050		12.956	<.001		
	Country of residence	-.721	.027	-.722	-26.832	<.001	.479	2.089
	Gender respondent	-.024	.022	-.023	-1.134	.257	.873	1.146
	Current age - respondent	.002	.001	.041	1.853	.064	.704	1.420
	Education in single years	.017	.003	.154	6.563	<.001	.627	1.594
	Has television	.106	.024	.099	4.509	<.001	.722	1.385
	Has car/truck	.028	.027	.024	1.021	.308	.629	1.590
	Distance to border in km	7.701E-7	.000	.000	.002	.999	.659	1.518

a. Dependent Variable: Access to improved water source

```
FREQUENCIES VARIABLES=ZRE_1.96 ZRE_2.58 ZRE_3.29
  /ORDER=ANALYSIS.
```

Frequencies

Frequency Table

ZRE_1.96

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	846	91.6	94.9	94.9
	1.00	45	4.9	5.1	100.0
	Total	891	96.4	100.0	

Missing	System	33	3.6		
Total		924	100.0		

ZRE_2.58

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	862	93.3	96.7	96.7
	1.00	29	3.1	3.3	100.0
	Total	891	96.4	100.0	
Missing	System	33	3.6		
Total		924	100.0		

ZRE_3.29

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	871	94.3	97.8	97.8
	1.00	20	2.2	2.2	100.0
	Total	891	96.4	100.0	
Missing	System	33	3.6		
Total		924	100.0		

```
LOGISTIC REGRESSION VARIABLES Safe_water
  /METHOD=ENTER Country
  /METHOD=ENTER Country Gender Age Education TV Car DistBorder
  /SAVE=PRED COOK DFBETA RESID SRESID ZRESID
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	871	100.0
	Missing Cases	0	.0
	Total	871	100.0
Unselected Cases		0	.0
Total		871	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
NO	0
YES	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Access to improved water source NO	YES	Percentage Correct
Step 0	Access to improved water source	NO	0	413	.0
		YES	0	458	100.0
	Overall Percentage				52.6

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.103	.068	2.323	1	.127	1.109

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Country of residence	631.514	1	<.001
	Overall Statistics		631.514	1	<.001

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	749.148	1	<.001
	Block	749.148	1	<.001
	Model	749.148	1	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	455.989[a]	.577	.770

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			Access to improved water source		
			NO	YES	Percentage Correct
Step 1	Access to improved water source	NO	390	23	94.4
		YES	42	416	90.8
	Overall Percentage				92.5

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Country of residence	-5.124	.269	363.320	1	<.001	.006	.004	.010
	Constant	2.895	.214	182.688	1	<.001	18.087		

a. Variable(s) entered on step 1: Country of residence.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

     800 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R    600 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I         Y                                                                                    Y     I
E    400 +         N                                                                                    Y     +
N        I         N                                                                                    Y     I
C        I         N                                                                                    Y     I
Y        I         N                                                                                    Y     I
     200 +         N                                                                                    Y     +
         I         N                                                                                    Y     I
         I         N                                                                                    Y     I
         I         N                                                                                    Y     I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
   Prob:  0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 50 Cases.
```

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	149.319	6	<.001
	Block	149.319	6	<.001
	Model	898.466	7	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	306.670[a]	.644	.859

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			Access to improved water source		
	Observed		NO	YES	Percentage Correct
Step 1	Access to improved water source	NO	384	29	93.0
		YES	37	421	91.9
	Overall Percentage				92.4

a. The cut value is .500

Variables in the Equation

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	Country of residence	-6.549	.560	136.605	1	<.001	.001	.000	.004
	Gender respondent	-.647	.355	3.319	1	.069	.524	.261	1.050
	Current age - respondent	.007	.014	.268	1	.605	1.007	.981	1.034

	Education in single years	.378	.053	50.476	1	<.001	1.460	1.315	1.620
	Has television	1.556	.368	17.906	1	<.001	4.738	2.305	9.739
	Has car/truck	.334	.559	.358	1	.550	1.397	.467	4.176
	Distance to border in km	-.004	.008	.230	1	.631	.996	.982	1.011
	Constant	.483	.807	.358	1	.550	1.621		

a. Variable(s) entered on step 1: Country of residence, Gender respondent, Current age - respondent, Education in single years, Has television, Has car/truck, Distance to border in km.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

     320 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R    240 +                                                                                                    +
E        I                                                                                                   YI
Q        I                                                                                                   YI
U        I                                                                                                   YI
E    160 +                                                                                                   Y+
N        I                                                                                                   YI
C        IN                                                                                                  YI
Y        IN                                                                                                  YI
      80 +N                                                                                                  Y+
         INN                                                                                                YYI
         INNN                                                                                               YYI
         INNNNNNN NN                                     Y                                          YY    YYYYI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 20 Cases.
```

Appendix B

The following output contains the logistic regression that was conducted with the alternative dependent variable that accounts for the difference in coding of rainwater in the two datasets that were merged together for the purpose of this research. Again, it uses the threshold of a 100 kilometres around the border from which the data is selected. The descriptives of the cook's distance and DFBetas are included to check the assumptions for influential cases and a linear regression was run to check for the multicollinearity assumptions, both were met. Similarly to the regression in *Appendix A* the frequency tables on the standardised residuals show that there might be an issue with outliers, as there is more than one case with a standardised residual above 3.29 and more than 1% of the cases has a value above 2.58. To ensure that these cases do not have disproportionate effects on the results an additional logistic regression was run, in which these cases were excluded. The results of this regression show that the odds ratio of having access to an improved source of water on the Dutch-colonised side compared to the British-colonised side are even smaller than they were in the first regression, with odds ratios of 0.058 for block 1 and 0.154 for block 2. The direction and significance of the results remain similar, and it can therefore be concluded that these cases do not substantially affect the results.

```
LOGISTIC REGRESSION VARIABLES Safe_water_recode
  /METHOD=ENTER Country
  /METHOD=ENTER Country Gender Age Education TV Car DistBorder
  /SAVE=PRED COOK DFBETA RESID SRESID ZRESID
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	891	96.4
	Missing Cases	33	3.6
	Total	924	100.0
Unselected Cases		0	.0
Total		924	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
NO	0
YES	1

Block 0: Beginning Block

Classification Table[a,b]

Observed Predicted

			Access to improved water source recoded		
			NO	YES	Percentage Correct
Step 0	Access to improved water source recoded	NO	0	227	.0
		YES	0	664	100.0
	Overall Percentage				74.5

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1.073	.077	194.888	1	<.001	2.925

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Country of residence	162.432	1	<.001
	Overall Statistics		162.432	1	<.001

Block 1: Method = Enter
Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	176.651	1	<.001
	Block	176.651	1	<.001
	Model	176.651	1	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	834.661[a]	.180	.265

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			Access to improved water source recoded		
	Observed		NO	YES	Percentage Correct
Step 1	Access to improved water source recoded	NO	0	227	.0
		YES	0	664	100.0
	Overall Percentage				74.5

a. The cut value is .500

Variables in the Equation

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	Country of residence	-2.361	.209	127.320	1	<.001	.094	.063	.142
	Constant	2.597	.186	194.532	1	<.001	13.419		

a. Variable(s) entered on step 1: Country of residence.

```
               Step number: 1

               Observed Groups and Predicted Probabilities

     800 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R    600 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I                                                       Y                                     Y      I
E    400 +                                                       Y                                     Y      +
N        I                                                       Y                                     Y      I
C        I                                                       Y                                     Y      I
Y        I                                                       Y                                     Y      I
     200 +                                                       N                                     Y      +
         I                                                       N                                     Y      I
         I                                                       N                                     Y      I
         I                                                       N                                     N      I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 50 Cases.
```

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	149.050	6	<.001
	Block	149.050	6	<.001
	Model	325.701	7	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	685.610[a]	.306	.451

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted: Access to improved water source recoded NO	Predicted: Access to improved water source recoded YES	Percentage Correct
Step 1	Access to improved water source recoded	NO	148	79	65.2
		YES	76	588	88.6
	Overall Percentage				82.6

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Country of residence	-1.280	.267	22.992	1	<.001	.278	.165	.469

	Gender respondent	-.331	.236	1.964	1	.161	.718	.452	1.141
	Current age - respondent	.013	.010	1.622	1	.203	1.013	.993	1.033
	Education in single years	.073	.028	6.988	1	.008	1.076	1.019	1.136
	Has television	1.808	.202	79.941	1	<.001	6.096	4.102	9.061
	Has car/truck	1.048	.488	4.607	1	.032	2.851	1.095	7.419
	Distance to border in km	-.014	.004	10.940	1	<.001	.986	.978	.994
	Constant	.914	.530	2.975	1	.085	2.495		

a. Variable(s) entered on step 1: Country of residence, Gender respondent, Current age - respondent, Education in single years, Has television, Has car/truck, Distance to border in km.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

     160 +                                                                                                    +
         I                                                                                                  Y I
         I                                                                                                  Y I
F        I                                                                                                  Y I
R    120 +                                                                                                  Y +
E        I                                                                                                  Y I
Q        I                                                                                                  Y I
U        I                                                                                                  Y I
E     80 +                                                                                                  Y +
N        I                                                                                                  Y I
C        I                                                                                                  Y I
Y        I                                                                                               Y  Y I
      40 +                                                                                              YYYYY +
         I                                                                                              YYYYYYI
         I                        Y         Y                                       Y  Y YY            YYYYYYYI
         I                      NNNNYNYNNN YNNYYNNN  YYN YNN                 YYYYYYYNYYYYNNYYYYY YY YYYYYYYYYYI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 10 Cases.
```

DESCRIPTIVES VARIABLES=PRE_3 COO_3 RES_3 SRE_3 ZRE_3 DFB0_3 DFB1_3 DFB2_3 DFB3_3 DFB4_3 DFB5_3
DFB6_3 DFB7_3
/STATISTICS=MEAN STDDEV MIN MAX.

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Predicted probability	901	.19746	.99411	.7471595	.25592488
Analog of Cook's influence statistics	891	.00002	.29065	.0091388	.02262044
Difference between observed and predicted probabilities	891	-.98283	.78257	.0000000	.35372122
Standard residual	891	-2.85850	1.75516	.0875804	.87862257
Normalized residual	891	-7.56690	1.89717	.0066217	.93936831

DFBETA for constant	891	-.12998	.10927	.0000097	.01804230
DFBETA for Country of residence	891	-.04221	.06401	-.0000005	.00961391
DFBETA for Gender respondent	891	-.03997	.03828	.0000041	.00793244
DFBETA for Current age - respondent	891	-.00230	.00194	-.0000002	.00034697
DFBETA for Education in single years	891	-.00497	.00513	.0000000	.00092083
DFBETA for Has television	891	-.02736	.02327	.0000029	.00700606
DFBETA for Has car/truck	891	-.20562	.14018	-.0000155	.01704631
DFBETA for Distance to border in km	891	-.00068	.00091	-.0000001	.00014671
Valid N (listwise)	891				

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10) TOLERANCE(.0001)
  /NOORIGIN
  /DEPENDENT Safe_water_recode
  /METHOD=ENTER Country Gender Age Education TV Car DistBorder.
```

Regression

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	.585	.064		9.082	<.001		
	Country of residence	-.169	.035	-.194	-4.871	<.001	.479	2.089
	Gender respondent	-.032	.028	-.034	-1.157	.247	.873	1.146
	Current age - respondent	.001	.001	.034	1.042	.298	.704	1.420
	Education in single years	.008	.003	.078	2.253	.024	.627	1.594
	Has television	.363	.030	.388	11.982	<.001	.722	1.385
	Has car/truck	.004	.035	.004	.110	.913	.629	1.590
	Distance to border in km	-.002	.001	-.088	-2.607	.009	.659	1.518

a. Dependent Variable: Access to improved water source recoded

```
FREQUENCIES VARIABLES=ZRE3_3.29 ZRE3_2.58 ZRE3_1.96
  /ORDER=ANALYSIS.
```

Frequencies
Frequency Table

ZRE3_3.29

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	878	95.0	98.5	98.5
	1.00	13	1.4	1.5	100.0

	Total	891	96.4	100.0	
Missing	System	33	3.6		
Total		924	100.0		

ZRE3_2.58

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	875	94.7	98.2	98.2
	1.00	16	1.7	1.8	100.0
	Total	891	96.4	100.0	
Missing	System	33	3.6		
Total		924	100.0		

ZRE3_1.96

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	858	92.9	96.3	96.3
	1.00	33	3.6	3.7	100.0
	Total	891	96.4	100.0	
Missing	System	33	3.6		
Total		924	100.0		

```
LOGISTIC REGRESSION VARIABLES Safe_water_recode
  /METHOD=ENTER Country
  /METHOD=ENTER Country Gender Age Education TV Car DistBorder
  /SAVE=PRED COOK DFBETA RESID SRESID ZRESID
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	878	100.0
	Missing Cases	0	.0
	Total	878	100.0
Unselected Cases		0	.0
Total		878	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
NO	0
YES	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			Access to improved water source recoded		
	Observed		NO	YES	Percentage Correct
Step 0	Access to improved water source recoded	NO	0	214	.0
		YES	0	664	100.0
	Overall Percentage				75.6

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1.132	.079	207.499	1	<.001	3.103

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Country of residence	187.196	1	<.001
	Overall Statistics		187.196	1	<.001

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	211.286	1	<.001
	Block	211.286	1	<.001
	Model	211.286	1	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	763.905[a]	.214	.319

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			Access to improved water source recoded		
			NO	YES	Percentage Correct
Step 1	Access to improved water source recoded	NO	0	214	.0
		YES	0	664	100.0
	Overall Percentage				75.6

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Country of residence	-2.846	.253	126.156	1	<.001	.058	.035	.095
	Constant	3.086	.235	173.069	1	<.001	21.895		

a. Variable(s) entered on step 1: Country of residence.

```
Step number: 1

Observed Groups and Predicted Probabilities

     800 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R    600 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I                                                        Y                                       Y   I
E    400 +                                                        Y                                       Y   +
N        I                                                        Y                                       Y   I
C        I                                                        Y                                       Y   I
Y        I                                                        Y                                       Y   I
     200 +                                                        N                                       Y   +
         I                                                        N                                       Y   I
         I                                                        N                                       Y   I
         I                                                        N                                       Y   I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 50 Cases.
```

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	162.765	6	<.001
	Block	162.765	6	<.001
	Model	374.052	7	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	601.139[a]	.347	.517

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			Access to improved water source recoded		
			NO	YES	Percentage Correct
Step 1	Access to improved water source recoded	NO	145	69	67.8
		YES	74	590	88.9
	Overall Percentage				83.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Country of residence	-1.869	.311	36.074	1	<.001	.154	.084	.284
	Gender respondent	-.447	.258	2.993	1	.084	.640	.385	1.061
	Current age - respondent	.016	.011	2.131	1	.144	1.017	.994	1.039

	Education in single years	.082	.030	7.561	1	.006	1.086	1.024	1.152
	Has television	2.071	.217	90.726	1	<.001	7.934	5.181	12.150
	Has car/truck	1.118	.618	3.271	1	.071	3.058	.911	10.266
	Distance to border in km	-.016	.005	12.261	1	<.001	.984	.975	.993
	Constant	1.398	.587	5.669	1	.017	4.045		

a. Variable(s) entered on step 1: Country of residence, Gender respondent, Current age - respondent, Education in single years, Has television, Has car/truck, Distance to border in km.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

     200 +                                                                                                   Y+
         I                                                                                                   YI
         I                                                                                                   YI
F        I                                                                                                   YI
R    150 +                                                                                                   Y+
E        I                                                                                                   YI
Q        I                                                                                                   YI
U        I                                                                                                   YI
E    100 +                                                                                                   Y+
N        I                                                                                                  YYI
C        I                                                                                                  YYI
Y        I                                                                                                 YYYI
      50 +                                                                                                 YYY+
         I                                                                                                 YYYI
         I                                                                             Y    Y             YYYYI
         I                   YNNNYY NN   NNYY NY   YYY   Y                        Y YYYYYYYNYYYYYYY    Y YYYYYI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for YES
          The Cut Value is .50
          Symbols: N - NO
                   Y - YES
          Each Symbol Represents 12.5 Cases.
```

<u>Appendix C</u>

The following output contains the logistic regression that investigates the relationship between the independent and dependent variable, within a threshold of a 90 kilometres. This slightly smaller threshold is included to show that the results are not completely dependent on the size of the threshold that was selected. The result of this regression show that the direction and significance of the result stay similar to the results of the regression that worked with a threshold of a 100 kilometres. The effect of the independent variable on the dependent variable is however larger, with an odds ratio of 0.002 indicating that the odds of having access to an improved water source on the Dutch-colonised side is 0.002 times the odds of having access on the British-colonised side. This is stronger compared to the odds ratios of 0.010 (model 1) and 0.011 (model 2), this indicates that the effect becomes bigger the closer a respondent lives to the border.

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES Water_access
  /METHOD=ENTER Country
  /METHOD=ENTER Country Gender Education Age Car Tv
  /SAVE=COOK LEVER DFBETA RESID SRESID ZRESID
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression
Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	687	96.4
	Missing Cases	26	3.6
	Total	713	100.0
Unselected Cases		0	.0
Total		713	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Block 0: Beginning Block
Classification Table[a,b]

	Observed		Predicted: Access to improved water source: No	Yes	Percentage Correct
Step 0	Access to improved water source	No	0	267	.0
		Yes	0	420	100.0
	Overall Percentage				61.1

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.453	.078	33.497	1	<.001	1.573

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Country of respondent	570.297	1	<.001
	Overall Statistics		570.297	1	<.001

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	678.832	1	<.001
	Block	678.832	1	<.001
	Model	678.832	1	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	239.191[a]	.628	.852

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted: Access to improved water source No	Yes	Percentage Correct
Step 1	Access to improved water source	No	252	15	94.4
		Yes	14	406	96.7
	Overall Percentage				95.8

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Country of respondent	-6.189	.380	265.001	1	<.001	.002	.001	.004
	Constant	3.298	.263	157.368	1	<.001	27.067		

a. Variable(s) entered on step 1: Country of respondent .

```
          Step number: 1

          Observed Groups and Predicted Probabilities

     800 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R    600 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I                                                                                                    I
E    400 +                                                                                               Y    +
N        I                                                                                               Y    I
C        I                                                                                               Y    I
Y        I    N                                                                                          Y    I
     200 +    N                                                                                          Y    +
         I    N                                                                                          Y    I
         I    N                                                                                          Y    I
         I    N                                                                                          Y    I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is .50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 50 Cases.
```

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	40.687	5	<.001
	Block	40.687	5	<.001
	Model	719.519	6	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	198.504[a]	.649	.881

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted: Access to improved water source No	Predicted: Access to improved water source Yes	Percentage Correct
Step 1	Access to improved water source	No	252	15	94.4
		Yes	14	406	96.7
	Overall Percentage				95.8

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Country of respondent	-6.357	.581	119.632	1	<.001	.002	.001	.005
	Gender of respondent	-.932	.469	3.943	1	.047	.394	.157	.988
	Education in single years	.315	.065	23.405	1	<.001	1.371	1.206	1.558
	Current age - respondent	.033	.018	3.451	1	.063	1.033	.998	1.070
	Car in household	1.640	.931	3.106	1	.078	5.156	.832	31.950

	Tv in household	-.549	.471	1.356	1	.244	.578	.229	1.455
	Constant	.911	.922	.978	1	.323	2.488		

a. Variable(s) entered on step 1: Country of respondent , Gender of respondent , Education in single years, Current age - respondent, Car in household, Tv in household .

```
              Step number: 1

              Observed Groups and Predicted Probabilities

     320 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R    240 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                   YI
U        I                                                                                                   YI
E    160 +                                                                                                   Y+
N        I                                                                                                   YI
C        I                                                                                                   YI
Y        I                                                                                                   YI
      80 + N                                                                                                 Y+
         I N                                                                                                YYI
         I N                                                                                               YYYI
         INNNNNNN N                                                                                  Y YYYYYYYI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is .50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 20 Cases.
```

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Analog of Cook's influence statistics	687	.00000	.98120	.0105963	.05716126
Leverage value	697	.00012	.11277	.0103179	.01260915
Difference between observed and predicted probabilities	687	-.98889	.98084	.0000000	.19423933
Standard residual	687	-3.00742	2.81958	.0106210	.54328691
Normalized residual	687	-9.43309	7.15402	-.0060579	.87832111
DFBETA for constant	687	-.38265	.21086	.0000773	.03638420
DFBETA for Country of respondent	687	-.04699	.27846	-.0000052	.02432909
DFBETA for Gender of respondent	687	-.09870	.18333	.0000060	.01936724
DFBETA for Education in single years	687	-.02598	.01683	-.0000043	.00228888
DFBETA for Current age - respondent	687	-.00600	.00523	-.0000015	.00066935
DFBETA for Car in household	687	-.78137	.10957	-.0000105	.03223215
DFBETA for Tv in household	687	-.09473	.20224	.0000102	.02008200
Valid N (listwise)	687				

```
LOGISTIC REGRESSION VARIABLES Water_acces_recoded
  /METHOD=ENTER Country Gender Education Age Car Tv
  /SAVE=COOK LEVER DFBETA RESID SRESID ZRESID
  /CLASSPLOT
  /CASEWISE OUTLIER(2)
  /PRINT=SUMMARY CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	686	96.2
	Missing Cases	27	3.8
	Total	713	100.0
Unselected Cases		0	.0
Total		713	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			Recoded variable of access to improved water source		
			No	Yes	Percentage Correct
Step 0	Recoded variable of access to improved water source	No	0	121	.0
		Yes	0	565	100.0
	Overall Percentage				82.4

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1.541	.100	236.665	1	<.001	4.669

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Country of respondent	148.629	1	<.001
		Gender of respondent	4.884	1	.027
		Education in single years	32.169	1	<.001
		Current age - respondent	18.213	1	<.001
		Car in household	63.126	1	<.001
		Tv in household	170.921	1	<.001
	Overall Statistics		212.110	6	<.001

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	221.290	6	<.001
	Block	221.290	6	<.001
	Model	221.290	6	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	417.880[a]	.276	.455

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			Recoded variable of access to improved water source		
	Observed		No	Yes	Percentage Correct
Step 1	Recoded variable of access to improved water source	No	75	46	62.0
		Yes	52	513	90.8
	Overall Percentage				85.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Country of respondent	-1.725	.341	25.616	1	<.001	.178	.091	.347
	Gender of respondent	-.456	.317	2.070	1	.150	.634	.340	1.180
	Education in single years	.071	.039	3.235	1	.072	1.073	.994	1.159
	Current age - respondent	.017	.014	1.483	1	.223	1.017	.990	1.045
	Car in household	2.334	1.050	4.945	1	.026	10.324	1.319	80.812
	Tv in household	1.574	.264	35.647	1	<.001	4.824	2.878	8.086
	Constant	.694	.681	1.039	1	.308	2.002		

a. Variable(s) entered on step 1: Country of respondent , Gender of respondent , Education in single years, Current age - respondent, Car in household, Tv in household .

```
             Step number: 1

             Observed Groups and Predicted Probabilities

     320 +                                                                                                     +
         I                                                                                                     I
         I                                                                                                     I
F        I                                                                                                     I
R    240 +                                                                                                     +
E        I                                                                                                    YI
Q        I                                                                                                    YI
U        I                                                                                                    YI
E    160 +                                                                                                    Y+
N        I                                                                                                    YI
C        I                                                                                                    YI
Y        I                                                                                                    YI
      80 +                                                                                                    Y+
         I                                                                                                 Y  YI
         I                                                                                                YYY YI
         I                                        Y N YY    Y                           YYYYYY           YYYY YI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

          Predicted Probability is of Membership for Yes
          The Cut Value is .50
          Symbols: N - No
                   Y - Yes
          Each Symbol Represents 20 Cases.
```