MASTER THESIS

Does behaviour impact the vulnerability of the Philippines community to Dengue incidence?

The Epidemic Risk and Priority Project

AUTHOR
Ing. C. Meijerink
(Student number: s1894358)

FACULTY OF SCIENCE AND TECHNOLOGY
HEALTH SCIENCES MSc.
INNOVATIONS IN PUBLIC HEALTH

EXAMINATION COMMITTEE

Chair:	Dr. M.M. Boere-Boonekamp
Member UT:	Prof. dr. A. Need

510

AN INITIATIVE OF THE NETHERLANDS RED CROSS

Date: June 22, 2018

UNIVERSITY OF TWENTE.

Abstract

Introduction: Infectious diseases pose a real threat to the human population. Being prepared for the spread of infectious diseases can reduce the mortality rate. Early detection is an important factor regarding preparedness. The Netherland Red Cross (NLRC) and 510.Global aim to facilitate this early detection with the epidemic risk and priority (EPI) project. The EPI project identifies epidemic risk at an early stage at global, national, regional and community level. The EPI project enables possibilities for creating awareness allowing early warning, building resilience through prevention and prediction of the contribution of epidemic risk factors. This master thesis will describe a part of the EPI project and focusses specifically the effect of behavioural indicators, on the dengue incidence in the Philippines.

Methods: The EPI project aims to develop a framework, a model and an index. The EPI framework is the evidence-based backbone of the EPI project based on a literature study. The EPI framework consists of (sub-)components to operationalize risk on dengue incidence which is caused by certain behaviour of the community. Open and confidential data sources were used to provide valid and reliable indicators as input for the (sub-)components. This set of indicators was reduced to a set of statistically independent indicators using correlation coefficients and variance influence factors (VIF) scoring as criteria. The EPI model was then created using the regression coefficients for the independent indicators of dengue incidence as parameters. The model was then used to calculate the EPI index based on regional data on the various components and the index was finally visualised on a map.

Results: The components "Self-Perceived Health", "Healthcare Seeking Behaviour", "Prevention" and "Trust" were used to operationalize "behaviour" in the EPI framework. The demographic health survey (DHS), world value survey (WVS) and pre-disaster indicators were then used to provide valid and operationalized indicators as input. A set of twelve independent indicators was used to measure the (sub-)components. The regression coefficients were calculated with these indicators based on a Poisson regression. The intercept of the model indicated that behaviour influences dengue incidence with a $\boldsymbol{\beta_i} = 6.8519$ and $P = 0.5490$. Even though, the model is a good fit with $R^2 = 0.91$, conclusion should be drawn with caution because of the P-value.

Discussion: Concluding, the EPI model for behaviour can indicate risk for dengue incidence. The EPI model for behaviour is only one part of the EPI project and all models need to be combined and tested as a whole. The set of indicators for behaviour will be reduced by excluding the health seeking behaviour indicators when combining the models. This could increase the validity of the model as a whole.

List of abbreviations

-C-	
CDC	Centres for Disease Control and Prevention
COIN	Competence Centre on Composite Indicators and Scoreboards
CRA	Community risk assessment
-D-	
DHF	Dengue Haemorrhagic Fever
DHS	Demographic and Health Surveys
DOH	Department of Health
DSS	Dengue Shock Syndrome
-E-	
EPI	Epidemic Risk and Priority
-H-	
HDX	The humanitarian data exchange
-I-	
IASC	Inter-Agency Standing Committee
ICRC	International committee of the Red Cross
IFRC	International Federation of Red Cross
IHR	International Health Regulation
INFORM	Index for Risk Management
-K-	
kNN	k-nearest neighbour imputation
-L-	
LGUs	Local Government Units
-M-	
MSc	Master of Science
-N-	
N	Number of
NGO	Non-Governmental Organisation
NLRC	Netherlands Red Cross
NRC	The Netherlands Red Cross
-P-	
PRC	Philippines red cross
-Q-	
qGIS	quantum geographic information system
-R-	
RAV	Relative Added Value
-S-	
SD	Standard deviation
se	Standard error
SQL	Structured Query Language
SSMS	SQL Server Management Studio
-T-	
TPB	Theory of planned behaviour
-U-	
UN OCHA	United Nations Office for the Coordination of Humanitarian Affairs
USAID	United States Agency for International Development
UT	University of Twente
-V-	
VIF	Variance inflation factor
-W-	
WHO	World Health Organisation

Table of Content

1. Introduction

Infectious diseases pose a threat to the human population[1]. Natural disasters and epidemics of infectious diseases are among the most destructive and costly among natural disasters[2]. They are one of the major triggers for emergency responses via the Red Cross since epidemics often follow natural disaster[3,4]. The Red Cross is an independent, neutral organization ensuring humanitarian aid, protection and assistance for victims of armed conflict. They act before, during and after disasters and health emergencies to meet the needs and improve the life of vulnerable people[5]. The Red Cross wants to increase their focus on disaster preparedness, by optimizing and improving existing products and procedures or developing new ones[6].

Disaster preparedness refers to measures taken to prepare for and reduce the effects of disasters[6,7]. One important factor for improving preparedness is the early detection of the spread of infectious diseases[7]. Actions being taken because of the early detection of infectious diseases can reduce the mortality rate for a lot of infectious diseases[8,9]. For example, research shows that early detection and treatment of dengue has reduced the mortality rate from 10-20% to less than 1% [10]. Surveillance of emerging infectious diseases is vital for the early detection since this can increase the ability to respond locally and reduce global risk[11]. For example, a study of the Ebola epidemic in 2007 showed that the lack of surveillance has caused an increase in global risk because of the long response time and underestimation of global risk[12]. Global surveillance aims to rapidly detect changes in incidence, risk or other factors to properly recognize and react to the emerging situation[13].

New technologies based on predictive modelling are becoming increasingly available for the early detection of emerging infectious diseases[14]. One example are web-based surveillance tools[13,15]. These are often used by major public health organisations to facilitate risk assessment and therefore enable the early detection of outbreaks which could lead to possible epidemics[15]. However, most of the predictive models for infectious diseases are highly specific for certain geographic locations and/or target diseases and thus fail in predicting distribution of global risk on infectious diseases[16,17].

Therefore, the Netherlands Red Cross and 510.Global have initiated the EPI project in which they aim to develop an epidemic risk and priority (EPI) tool that can be used globally for a variety of diseases. The EPI project aims to facilitate the identification of epidemic risk factors at an early stage at global, national, regional and community level. Furthermore, the EPI project enables possibilities for creating awareness allowing early warning, building resilience through prevention and determining the contribution of epidemic risk factors. This thesis will focus on the risk factors with regards to vulnerability caused by behavioural aspects. The research question is: “What is the effect per indicator and all indicators as a whole in the category behaviour on the dengue incidence in the Philippines?”

2. Scope of the Epidemics Risk and Priority (EPI) project

510.Global is an initiative of the Netherlands Red Cross (NLRC) to support the NLRC in achieving faster and more efficient humanitarian aid by using data science techniques[18]. 510.Global has developed the community risk assessment (CRA) toolbox which is visualised in Figure 1. The CRA toolbox is a tool that assess house damage after a natural disaster to enable identification of high priority areas[19,20]. This tool was developed in response to a demand to decrease the response time after a natural disaster[20].

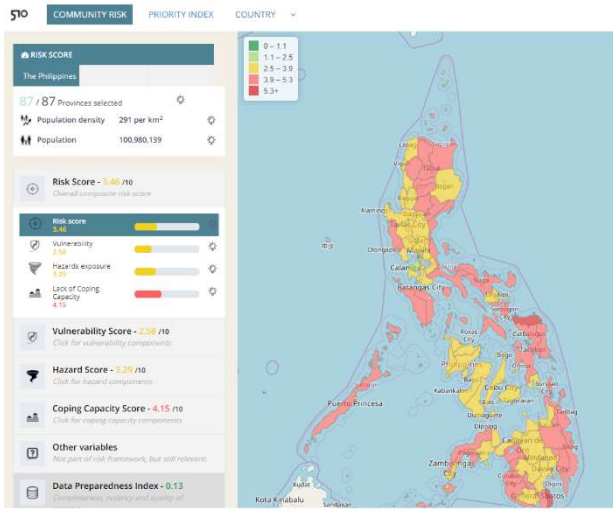

Figure 1: Print screen of the CRA Toolbox, dated 24.04.2018

2.1. INFORM

The CRA toolbox is based on the "index for risk management" (INFORM)[19]. INFORM is a global, objective and transparent methodology for understanding risk during and after disasters. INFORM is used as a methodological guideline for the CRA toolbox to quantitatively assess disaster risk with open-source data. The INFORM guideline is developed by the inter-Agency Standing Committee (IASC) and the European commission in response to an increased demand for evidence-based risk analysis and assessment [21-23].

INFORM is defined in three dimensions[21-23]. The first dimension "Hazard and exposure" measures the probability of physical exposure associated with existing hazards. They are merged because there is only risk when a population is exposed to a hazard. As such, it represents the load that the community must deal with when exposed to a hazard[22]. The second dimension "Vulnerability" measures the intrinsic tendencies of the exposed community to be susceptible to the effects of the hazard. The vulnerability dimension represents economic, political and social characteristics of the community that can be weakened in case of a hazard[22]. And finally, the third dimension "Lack of coping capacity" measures the ability of a country to cope with disasters in terms of formal, organized activities and the effort of the country's government as well as the existing infrastructure which contribute to the reduction of disaster risk[22].

Figure 2 shows the INFORM framework, which is operationalised to defined risk into measurable indicators. The measurable indicators give a value to the components. All these values together are referred to as data. When aggregating the data with equal weights a value is given to the categories and dimensions. The index is the final aggregation of the values of the dimensions. The normalized index is referred to in the framework as risk. When comparing the risk of different countries, it can give an indication of the global distribution of risk. When valid data is used as input for the INFORM framework it supports in evidence-based decision-making as well as developing strategies that build resilience[23].

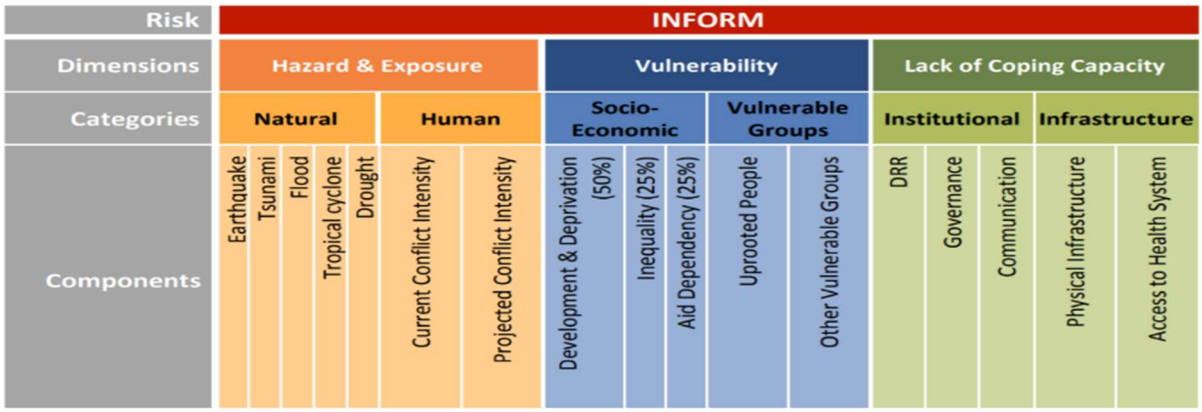

Figure 2: The INFORM model

2.2. The EPI Project

Although INFORM gives a good indication of the global distribution of risk it is not tailored to infectious disease outbreaks and epidemics. Furthermore, most other indexes focussed on infectious diseases are highly specific on disease or region and fail to show the global distribution of risk [16,24,25]. Therefore, the aim of the NLRC and 510.Global is to further extend the INFORM framework and its own CRA toolbox, via the EPI project, to enable the assessment of community level risk on epidemics. The EPI project aims to facilitate the identification of countries, regions and vulnerable communities at high risk for epidemics. This can also enable possibilities to increase awareness, preparedness, mitigation strategies and decrease response time.

The NLRC and 510.Global have decided to focus the development of the first prototype on a case study regarding dengue incidence in the Philippines. The Philippines have been chosen first because of the already existing cooperation between the Philippines Red Cross (PRC) and the NLRC as well as the fact that a great amount of community-based data is available. Dengue has been chosen as a case study, since this is one the most common infectious disease in the Philippines[26,27]. Dengue incidence is also a suitable subject for a global case study since more than 40% of the world's population is living in areas potentially at risk for dengue[28]. With a worldwide incidence of about 50 to 100 million cases per year, dengue fever is a major public health problem[28,29].

The EPI project uses INFORM as a methodological guideline. The developmental process of the EPI project is divided in five general steps, as schematically shown in Figure 3. First, the EPI framework will be created with subcomponents based on a literature review. Open source and/or confidential data will be used to provide indicators for the subcomponents. The set of independent indicators give a value to the subcomponents. The EPI model is calculated with a regression analysis, which will calculate the regression coefficients. The regression coefficient indicates the estimated effect of the indicator on the index. After this, the regression coefficients are used to calculate the predictive value of risk per region. The index is calculated using the regression coefficient and the aggregated values. The index will be visualised with a gradient colour scale on a map. This map will be comparable to one used in the CRA toolbox visualisation. After implementation and optimization of the EPI project in the Philippines the transferability of the EPI model and the compatibility of the framework to other countries and diseases will be assessed.

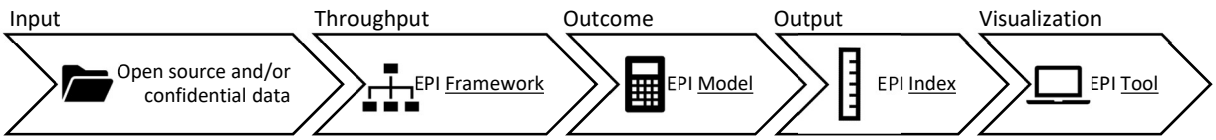

Figure 3: Schematic view of the developmental process of the Epidemic risk and priority tool

The EPI project is an ongoing process in which the development of the entire EPI framework is divided into several master thesis projects (the EPI project overview can be found in 0). The overview of the EPI framework (version 26.03.2018) is shown in Figure 4. In this framework the dimensions are used and defined as by INFORM and the categories and components are changed to fit the EPI framework. The EPI framework is divided in three dimensions namely, infectious hazard and exposure, vulnerability

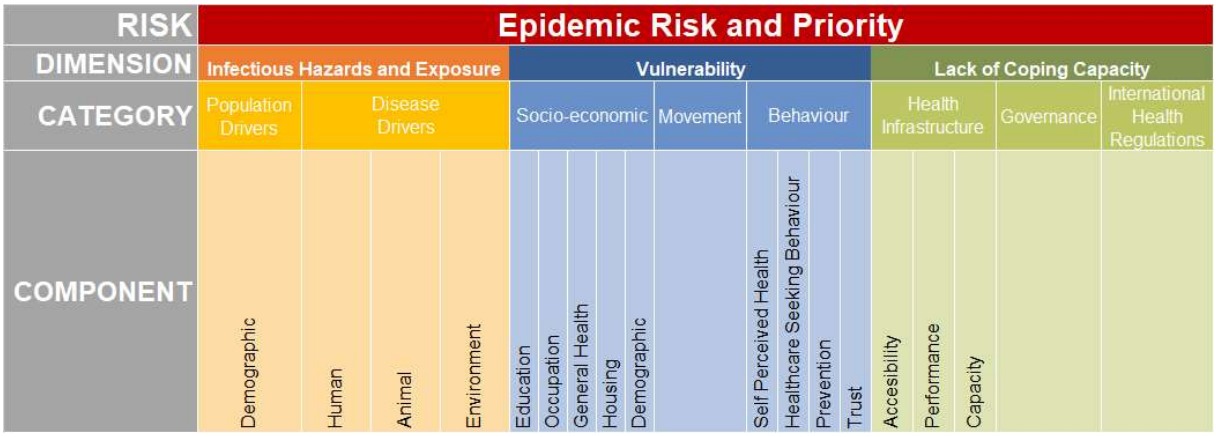

Figure 4: Epidemic risk and Priority Framework version Mai 24, 2018

and lack of coping capacity. Infectious hazard and exposure refers to how communities form a risk area. Vulnerability refers to why communities are at risk. The lack of coping capacity refers to infrastructure and regulation which influence the communities risk.

2.3. Scope of this Master Thesis

This master thesis contributes to a specific part of the EPI project, namely the development of the framework, model, index and visualisation for the category "behaviour" in the dimension vulnerability. Other parts of the EPI project will not be further discussed further in this thesis.

Vulnerability refers to the intrinsic tendencies of the community to change behaviour since communities tend to change behaviour when exposed to an infectious disease[30]. Vulnerability can be divided into three categories namely, socio-economic, movement and behaviour. Where socio-economic vulnerability refers to how the community changes, behaviour refers to the why. These changes in behaviour are measured in the category behaviour[30].

Research Question:
What is the effect per indicator and all indicators as a whole in the category behaviour on the dengue incidence in the Philippines?

Sub questions

1. Which subcomponents in the EPI framework measure the category behaviour for the regions of the Philippines during an infectious disease outbreak?
2. Which open, and/or confidential data sources provide valid and reliable data input for the identified subcomponents in the EPI framework?
3. Which set of statistically independent indicators in the EPI framework measure the category behaviour?
4. What is the estimated effect per indicator in the EPI model on the dengue incidence in the Philippines?
5. What is the validity and the reliability of the EPI model for the category behaviour?
6. What is the EPI index per region for dengue incidence in the Philippines for the category behaviour?

3. Methodology

To calculate the EPI index and the estimated effect per indicator on dengue incidence several steps were taken. These steps will be presented in this chapter. First, the study design will be discussed. Second, the selection of a set of independent indicators will be shown. Finally, the statistical analysis will be discussed.

3.1. Study Design

The geographical ecological study design was used since the purpose of the EPI project is to monitor the population's health by visualising the regional distribution of risk to dengue epidemics in the Philippines. The most optimal study design would be a cohort study with multi-level data which would enable multilevel data analysis with a time component. Since the data for this is currently not available the ecological study design was used. The aggregated measures of a population health have been shown to be a strong measure and often used in comparable studies[15,31,32].

Country profile

The Philippines is an island group in the south-east Asian region located between the South China Sea and the Pacific Ocean as shown in Figure 5**Fout! Verwijzingsbron niet gevonden.**[33]. The Philippines consist of 7107 islands with a total land area of approximately 300.000 km^2 and a coastline of approximately 34.000km. The islands are divided into seventeen administrative regions, 81 provinces and 42036 municipalities[34].

The Philippines is geographically located within the typhoon belt of the western pacific. Because of the geographic location, the Philippines are highly prone to typhoons during the rainy season and endure on average twenty typhoons each year[35]. In addition, they are located along the "pacific ring of fire" which is an area that is highly prone to earthquakes and volcanic eruptions[35].

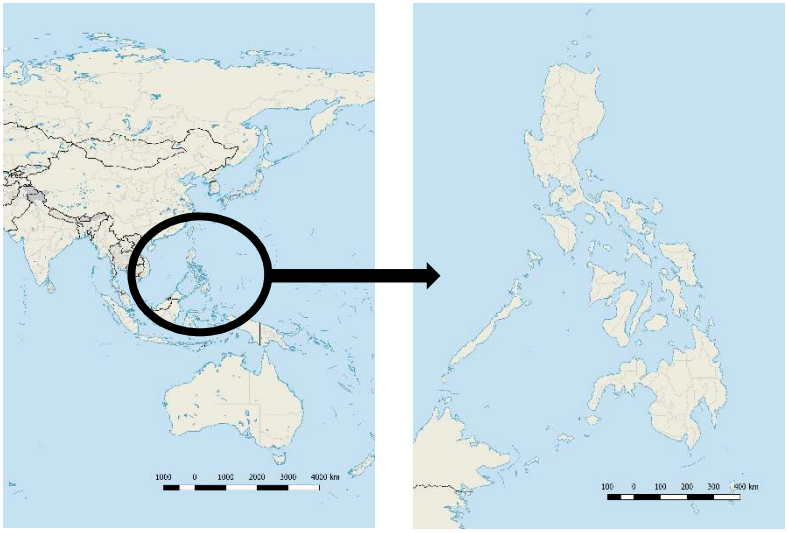

Figure 5: Geographical location of the Philippines

In January 2018 the population of the Philippines reached 100.979.303 persons as shown in **Fout! Verwijzingsbron niet gevonden.**[36]. Although the population count is increasing every year, the annual growth has decreased from 2.4% in the period from 1990-2000 to 1.9% in the period from 2000-2010. According to the Philippines health system review initiated by the WHO, the population growth is linked to a high fertility rate of three children per woman of child bearing age[34].

The Christian religion is most common with 92.5% in the year 2000, of which 81.0% had a Roman Catholic religion. 5.1% of the population had a Muslim religion and are mostly concentrated in the city Mindanao, which is in the south western part of the Philippines. The official language of the Philippines is English and Filipino, which is derived from Tagalog. Both languages are used in governmental businesses, educational systems, business and the media[34].

The Philippines health system is a decentralized system, where the Department of Health (DOH) is serving as the governing agency. The DHO is mandated to provide national policy direction and develop national plans, technical standards and guidelines on health. Local government units (LGUs) and the private sector are providing services to communities and individuals. LGU are autonomous and have the responsibility for their own health services but need to follow the guidance of the DOH. Provincial governments are mandated to provide secondary hospital care. City and municipal administrations need to provide primary care, including maternal and child care. All levels are obligated to provide care in case of both communicable diseases and non-communicable diseases[34].

Dengue Fever

The endemic disease dengue in the Philippines is chosen to be the test case in the EPI project. The incidence of dengue fever per province is plotted in Figure 6**Fout! Verwijzingsbron niet gevonden.** from 2008 to 2016. The year of the highest incidence is 2013 and is plotted per region on a map in Figure 6**Fout! Verwijzingsbron niet gevonden.**. The data from year 2013 is used to perform further analysis. To eliminate a possible bias in future studies the data from other years will also be analysed.

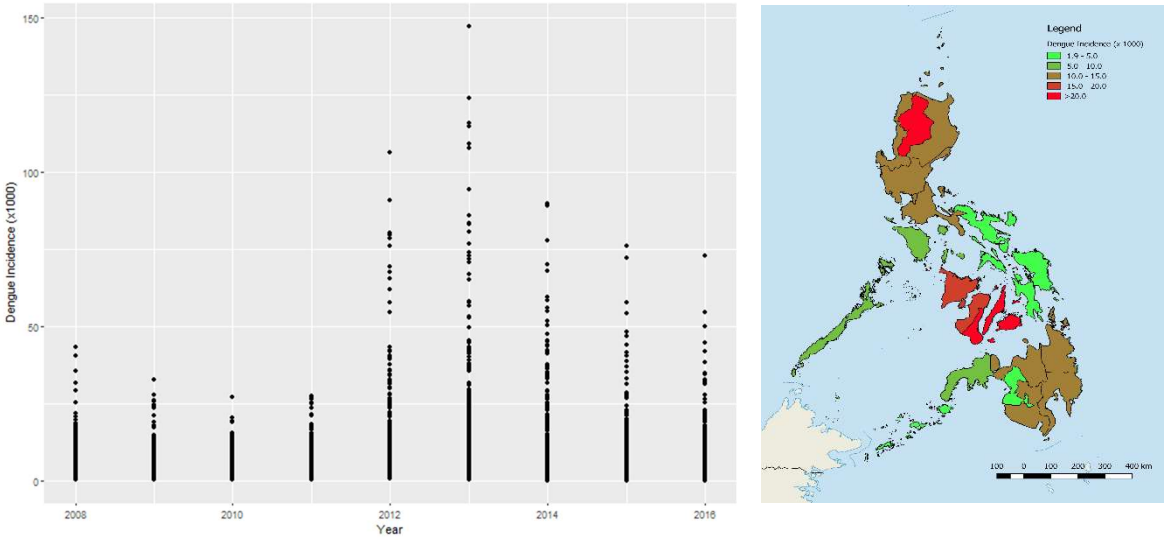

Figure 6: Dengue incidence in the Philippines plotted from year 2008 till 2016 (left) and Dengue incidence summed per Region of year 2013 (right)

Dengue fever is a vector borne disease caused by one of four closely related viruses[37]. It is transmitted by the bite of an Aedes mosquito infected with a dengue virus[28]. The mosquito gets infected when it bites a person or animal who is infected with the dengue virus. It cannot be spread from person to person directly, only indirectly through the mosquito. The incubation period ranges from four to ten days until the first symptoms start. Symptoms can be a sudden high fever, severe headaches, pain behind the eyes, severe joint and muscle pain, general fatigue, nausea, vomiting, skin rash that appear two to five days after the onset of the fever and mild bleedings such as nose bleeds, bleeding gums or easy bruising[38].

Dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) are two complications that can occur after the initial onset of dengue fever. DHF is characterised by high fever, damage to lymph and blood vessels, bleeding from the nose and gums, enlargements of the liver and failure of the circulatory system. DSS is caused when DHF progresses to massive bleeding, shock and death[38].

Where the mosquito is symptom-free, the dengue virus causes symptoms in humans[26,27]. For persons with a first infection the chances are small that they develop DHF or DSS. Younger children and people who never had a first infection tend to have milder symptoms that older children and adults[38]. The dengue vaccination is not effective when used on humans that haven't had a dengue

infection before. Studies show that children can even die when vaccinated before a first infection occurred[39].

Suitable areas for the mosquito to live are areas where the mosquito could lay her eggs, like artificial containers that hold standing water in and around the home. Since there is no effective vaccine available against dengue, general prevention of health is the most important step to avoid a dengue infection according to the CDC.

3.2. Selection of subcomponents in the EPI framework

The subcomponents used in the EPI framework to measure the category behaviour were identified based on a literature study. The search strategy shown in Figure 7 was used to identify the studies included in the literature review. The quality of the articles was assessed in three steps based on the exclusion criteria. The snowball technique and discussion with expert opinions were used to identify other relevant studies and retrospective evaluations that were not covered by this strategy. An overview of the literature is included in Appendix II0 and the consulted experts are listed in Appendix III.

The literature study was used as a guideline to construct the (sub-) components. In comparison to socio-economic indicators, there are no studies that indicate specific indicators to measure behaviour of a community. Therefore, the studies found in the literature study are used as a guideline to create the sub-components, the lessons learned and advises of retrospective analysis of previous epidemics of the last 10 year have been used as a guideline. These sub-components have been combined subjectively into the components.

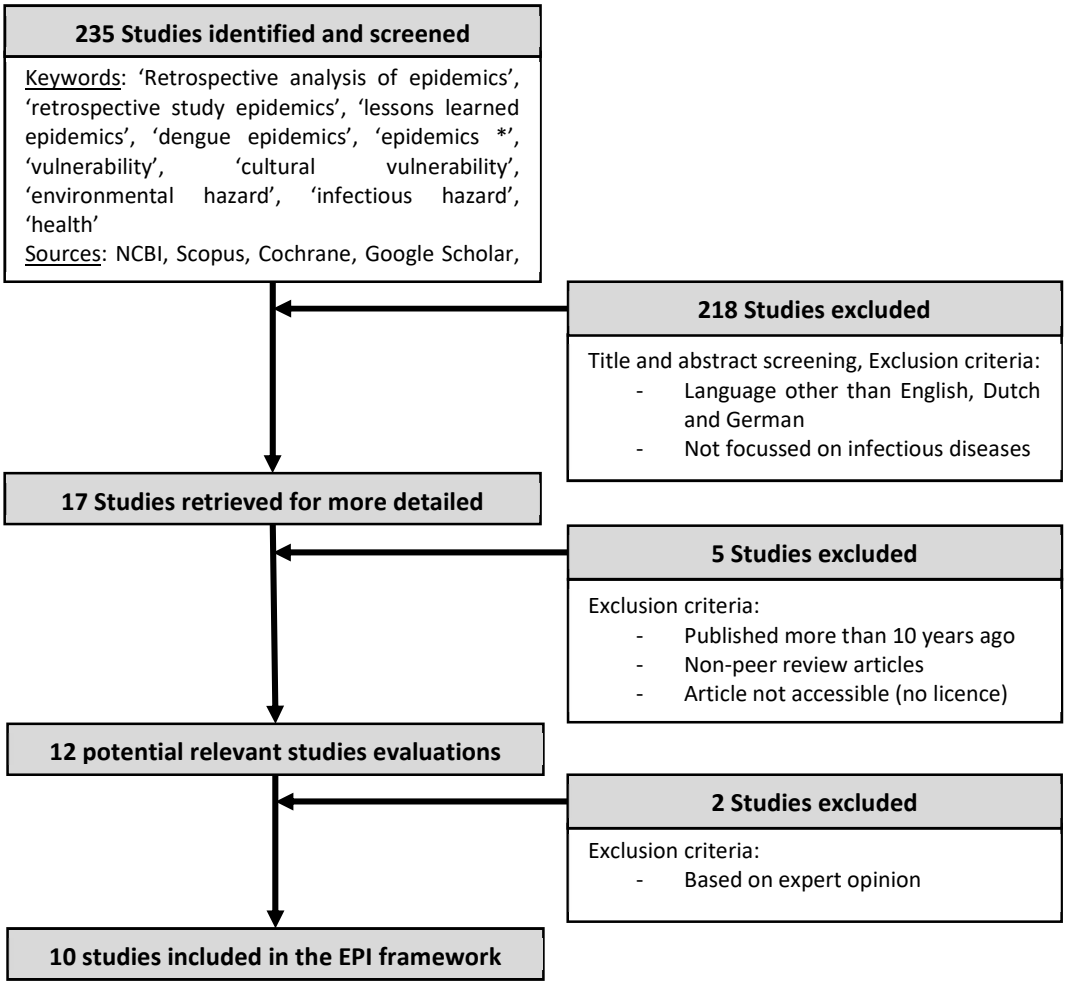

Figure 7: Search strategy of literature research

3.3. Data source and indicator selection

Open and/or confidential data sources were used to provide valid and reliable input for the subcomponents in the EPI framework. This input is based on indicators used in surveys. The selection was based on three criteria namely, the availability of the data, the quality of the data source and to what extent the data represents the (sub-)components in the EPI framework. The assumption has been made that the survey can be generalized to the population if it meets the three criteria. The indicators used in the surveys were used to provide the values of data to the EPI framework. The indicators used for the EPI framework are subjectively selected, where all indicators are either directly measured or used as proxy for a sub-component.

Quality of data was assessed based on three criteria, namely availability, quality of the source and representation. First, the availability of data was assessed based on the current availability as well as future availability. Second, the quality of the data was assessed based on the quality of the data source. Official and global recognized institutions were assessed as reliable sources, whereas research of which the real source was untraceable or the methodology not transparent was assessed as unreliable. Data retrieved from UN OCHA was considered as reliable since data validation is performed by the organisation. Third, the representation of the data for the subcomponents in the EPI framework was assessed based on the indicator providing the data and discussion with expert opinions. The list of experts who were consulted can be found in Appendix III.

3.4. Data storage and preparation

To store all the data in one database, the data warehouse methodology[40,41] was used by means of the software package "MS SQL Server Management Studio (SSMS)" version 17.6[42]. The software package "R Studio" version 3.4.1.[43] (2017-06-30) was used to script the statistical analysis of the data. The "ODBC driver 13 for SQL server" of the R packages "odbc"[44] and "rodbc"[45] was used to enable a real-life data stream from the MS SQL server to R-Studio. The complete R script can be found in Appendix IVAppendix III.

Missing Data

The missing data was identified with the "md.pattern" and "mice_plot" function from the "mice"[46] package. Missing data was handled with two methods based on the amount of missing data[47]. When the amount of missing data was <90% the k-nearest neighbour imputation (kNN)[48] of the R package "VIM"[49] was used with four and six nearest neighbours. The nearest neighbour is determined based on the Gower distance. This measures how different two rows are based on the values and scales the differences on a scale from zero to one. This can be done with both ordinal and numerical values[50]. When the amount of missing data was >90% the indicator was excluded from the analysis since the data is considered unreliable.

Data aggregation

The data was aggregated in one data frame in preparation for the analysis. Per indicator a sum, mean, standard deviation (SD) and standard error (se) as well as the number of entries per region (N) was calculated with the "ddply"[51] function of the R package plyr[52]. The se was calculated with equation 1:

$$se_x = sd_x/\sqrt{N_x} \qquad \textit{Equation 1}$$

3.5. Statistical Analysis

Statistical Significance of indicators

The statistical significance of the indicators in the EPI framework is determined to create a dataset with independent indicators using the correlation coefficients and VIF score as criteria. The linear correlation determines the dependencies between multiple indicators and the collinearity determines whether one indicator explains another. The correlation matrix was used to determine correlation and

the VIF was used to determine collinearity[47,53-56]. The data was normalized with the minmax normalization using equation 2:

$$MinMax\ Normalization\ =\ (\mathrm{m}\ -\ \min(\mathrm{m}))/(\max(\mathrm{m})-\min(\mathrm{m}))$$ *Equation 2*

The correlation matrix and the correlation coefficients were calculated using the "cor"[57] function of the R package stats v3.5.0[58]. The Pearson correlation method is used to measure the linear dependence between the indicators. The correlation matrix was visualised with the "corrplot.mixed" [59]function of the R package corrplot v0.84[60].

In the bottom-left part of the correlation matrix the correlation coefficient is displayed as a number. The correlation coefficient represents the dependence between that indicator with the other indicators. Zero means that there is no correlation, minus one indicates a strong negative correlation and plus one indicates a strong positive correlation. In the top-right part of the correlation matrix the correlation coefficient is displayed as coloured circles. Positive correlations are displayed in blue and negative correlations are displayed in red. The colour intensity and the circle size are proportional to the correlation coefficient. On the right side of the correlation matrix the legend shows the correlation coefficient and the corresponding colour[53,56].

The VIF was calculated using the "vif"[61] function of the R package HH v3.1-34[62]. VIF detects multicollinearity between variables. VIF estimates how much the variance of a regression coefficient is inflated due to multicollinearity with equation 3:

$$VIF-Score_i=(1-{R_i}^2)^{-1}$$ *Equation 3*

R^2 is the coefficient of determination which indicates the variance in the indicator. VIF > 10 are related to R^2 > 0.9. The higher the VIF score of an indicator, the more this indicator correlates with other predictor indicators. Since an often-used threshold for the VIF scores is 10, this threshold has been used to indicate collinearity[56,63].

Estimated effect of indicators in the EPI model

A forward stepwise Poisson regression analysis (with log link) was then performed on the remaining independent indicators to calculate the regression coefficient. The regression coefficient of an indicator indicates the estimated effect of the indicators on the EPI model. The assumption has been made that the dengue data has a Poisson distribution and therefore the Poisson regression is used. To do so, the "glm"[64] function of the R package stats v3.5.0.[58] was used.

The regression model builds a mathematical equation that defines the outcome Y as a function of input X indicators[56]. Since the dependent variable is a count, the Poisson regression is used[65] and the equation for one indicator will look like:

$$Y\sim Poisson(\lambda)$$ *Equation 4*

$$\log(\lambda)_{indicator}=\beta_{i1}x_{v1}$$

In this formula the assumption has been that λ is the output, which is the mean of Y dengue incidence (the dependent variable). The input for this equation is x, which is the indicator (independent variable). The β_1 is the regression coefficient calculated by the regression model.

The stepwise regression consists of adding and removing indicators[56]. The forward selection indicates that the regression starts the process with no indicators in the model and first adds the strongest contributing indicator. After the first "most contributing" indicator has been determined, the second most contributing indicator will be determined and added to the model. This process will continue until all indicators have been processed and the best performing model is created [56]. If an

indicator is not significantly contributing to the model, these indicators can be eliminated in the regression analysis.

A summary of the EPI model was computed by R-studio after running the model. The summary consists of the intercept, regression coefficient, the standard error, the Z-value and the P value. The intercept is the combined measure of all indicators, thus the category behaviour, and indicates the predicted λ when $x_i = 0$. The standard errors of the individual regression coefficients were used for the calculation of the Z-value. The Z-value is a test statistic, if the Z-value is bigger than two (with $\alpha = 0.05$) the indicator is significant, which means that dengue incidence and the indicator are related. The P-value was computed from the Z-values and indicates the strength of the indicator. A small P-value ($P < 0.05$) indicates strong evidence and a large P-value ($P > 0.05$) indicates weak evidence. P-values close to 0.05 are considered marginal contributors.

To estimate the goodness of fit of this model R^2 was calculated with the test parameters null deviance and residual deviance using equation 5:

$$R^2 = 1 - \frac{residual\ deviance}{null\ deviance} \qquad \textit{Equation 5}$$

This formula indicated how much better the model is (residual deviance) compared to the intercept (null deviance). When the ratio is small, to model is the better predicter.

Relative Added value per indicators to the EPI model

The relative added value (RAV) per indicator to the model gave an indication of the importance of the indicator to the model. The RAV is visualized and calculated using Microsoft Excel. The RAV is calculate with equation 6:

$$RAV_{indicator} = \frac{\sum \beta_i}{\beta_i} 100 \qquad \textit{Equation 6}$$

Next, the RAV per indicator and Region was determined to give an indication of the distribution of the RAV over the different regions. This was calculated with Microsoft Excel with the normalized set of twelve independent indicators.

EPI index for dengue

To predict the risk for dengue per region the equation of the EPI model is used. The dengue incidence can be predicted by adding the regression coefficient (β_i) and the corresponding aggregated value (x_i) per indicator to the equation of the EPI model.

$$Y \sim Poisson(\lambda) \qquad \textit{Equation 7}$$

$$\log(\lambda)_{Region} = \beta_{i1}x_{v1} + \beta_{i2}x_{v2} + \beta_{i3}x_{v3} \ldots . + \beta_i x_i$$

This was calculated with the "predict"[66] function of the R package stats v.3.5.0.[58]. Next, the EPI index is calculated by normalizing the predicted dengue incidence on a scale from zero to one. Zero indicates the least prioritized region of the Philippines when compared to the other regions. One indicated the highest prioritized region.

$$Risk\ Index_{\ Region} = (\mathrm{Y} - \min(\mathrm{Y}))/(\max(\mathrm{Y}) - \min(\mathrm{Y})) \qquad \textit{Equation 8}$$

Visualization

The EPI index is visualized per region on a map using the quantum geographic information system (qGIS) software version 2.18.18[67]. Other geographical visualisation was also created with this software. The natural earth quick start kit[68] provided the base map used for the visualizations. The quick start kit provided a small sample of natural earth themes for the software qGIS. Just a small subset of features of the kit were used as a base map, namely the boundary lines of land and of maritime indicator, the lakes, oceans and coastlines, the province lines and the scale ranks with minor islands. Next, a vector layer containing polygons of the regions of the Philippines, was plotted on the base map[69]. The index values per region were then uploaded to qGIS and linked to the polygons. The gradient colour scheme was added by setting the style of the polygons. The legend and the scale bar were added to the map in the printing environment.

4. Results

This chapter will give an overview of the results based on the sequence of the method section. This thesis is focussed on the development of the EPI framework, the EPI model and the EPI index for the category "behaviour" in the dimension "vulnerability" of the EPI project (Figure 3). First will be discussed, the four components, and the sub-components, used to operationalize "behaviour" namely, "Self-Perceived Health", "Healthcare Seeking Behaviour", "Prevention" and "Trust". Second, the selection of data sources which provide valid and reliable indicators as input for the (sub-)components is discussed. Third, the reduction of the set of indicators is discussed. Next, the development of the EPI model with the set of independent indicators and dengue incidence and the calculation of the EPI index is discussed. Finally, the EPI-Index will be visualised on a map.

4.1. Selection of subcomponents in the EPI framework

Based on the literature study a general definition of vulnerability and behaviour is discussed. Secondly, the creation of the sub-components will be discussed. The lessons learned and advises of retrospective analysis of previous epidemics of the last 10 year have been used as a guideline to construct the subcomponents. These sub-components have been combined subjectively into the components. An overview of the EPI framework zoomed in on the behaviour part is shown in Figure 8.

Risk	Epidemics Risk and Priority												
Dimension	Vulnerability												
Category	Behaviour												
Component	Self Perceived Health		Healthcare Seaking Behaviour			Prevention				Trust			
Sub-Component	Subjective Health	Subjective satisfaction	Advice taking behaviour	Consultation seeking	Medication Seeking	General Health	Contracetives	Physical Contact deseased	Vaccination Coverage	Government	Health Care System	Press	

Figure 8: EPI framework zoomed in on the category "behaviour" on component and subcomponent level

The category "behaviour" is one of the three categories in which the dimension "vulnerability" is operationalized. The dimension "Vulnerability" refers to the intrinsic tendencies of the exposed community to be susceptible to the effects of the hazard. Communities tend to change their behaviour in a preventive way with regards to the spread of an infectious disease. Consequently, this causes the prevalence to decrease[30]. These changes in behaviour are measured in the category "behaviour".
The category "behaviour" represents cultural and behavioural aspects of the community that change in case of a hazard[22]. The influence of cultural vulnerability and community behaviour to the spread of infectious diseases is proven in previous epidemics[54,70]. For example, retrospective studies of the Ebola crisis have concluded that causal behavioural aspects which worsened the crisis, were traditional burial rituals, dependence on traditional healers and other cultural practices, secret societies, community resistance by a deep-seated distrust, conspiracy theories by hiding diseased and civil disobedience[54,70,71].

(sub-)components

By successfully changing behaviour in a preventive way, individuals can work or participate in social activities and feel healthy despite limitations[72-74]. The outcome of feeling healthy is measured in the component "self-perceived health" and refers to the adequate adaption to the infectious disease[74]. "Self-perceived health" is measured with the sub-components subjective health and subjective satisfaction. The component "healthcare seeking behaviour" refers to the preference of the community regarding seeking of healthcare. Healthcare in the Philippines can be found in public and private healthcare facilities, traditional medicine, traditional healers, families and friends. Health care from family and friend often appears as medication sharing. Dengue cannot be spread from human to

human but serious complications and death can occur when not treated. The component "Healthcare seeking behaviour" is measured with the sub-components advice taking behaviour, consultation seeking and medication sharing. The component prevention refers to the individual and community actions being taken to prevent disease from spreading.

"Prevention" is measured with the sub-components preventive measures on general health, usage of contraceptive to prevent teenage pregnancies, physical contact with diseased and vaccination coverage. The components "trust" refers to the publics compliance with public policies and preventive measures. When the public has a deep-seated distrust towards the government policies will not succeed. The component "Trust" is measured with the communities' trust is the government, in the healthcare system and the trust in the press.

4.2. Data source and Indicator selection

The data sources used to provide indicators as input for the sub-components are shown in **Fout! Verwijzingsbron niet gevonden.**. All data sources have been selected and assessed as reliable according to the three criteria availability, quality and representation. The DHS and the WVS are both community surveillance surveys that provide a wide range of data. The HDX is an organisation which collects and validates data collection to be used in the humanitarian aid. The dengue incidence data is provided by the Philippine government[75].

Table 1: Description of the data sources

	DATA SOURCE	ORGANISATION	URL	DESCRIPTION
1	Demographic and health survey (DHS)	United States Agency for International Development (US AID)	URL	Nationally representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators sin the areas of population, health and nutrition
2	World Value Survey (WVS)	Institute for comparative survey research	URL	Nationally representative surveys in almost 100 countries using common questionnaires. A time series investigation of human beliefs, values and motivations of people throughout the wold.
3	Pre-disaster indicators	The humanitarian data exchange (HDX) from UN OCHA	URL	The goal of HDX is to make humanitarian data easy to find and use for analysis. HDX has a growing collection of datasets about crises around the world and has been accessed by users in over 200 countries and territories.
4	Dengue incidence data	Philippine Red Cross in cooperation with the Philippine government	-	The dengue data has been released for study purposes only. This should be treated as confidential and cannot be used for other purposes.

The indicators used in the surveys provided the values to the EPI framework. The indicators used for the EPI framework are subjectively selected, were all indicators that could either directly measure or proxy for the sub-component was selected. Table 7 in Appendix V shows the framework with the full set of selected indicators. This full set was reduced with the help of statistical analysis.

No data was found for the sub-component "physical contact with diseased". Since there was no data this sub-component was excluded from the final EPI framework. This data could be collected with a new survey. The sub-component trust in "health care system" was excluded from the final EPI framework since the survey indicator was not part of the Philippine survey and thus there was no data.

4.3. Statistical significance of indicators in the EPI framework

The full set of subjectively selected indicators was reduced with statistical analysis as explained in the method section. Figure 9 shows the final set of independent indicators included in the EPI framework. The indicators within each component have a correlation coefficient smaller than 0.3 which indicates that the indicators are only marginally inter-dependent. Table 8 in Appendix V shows the full list of indicators with the exclusion criteria.

Risk	Dimension	Category	Component	Subcomponents	Indicator Nr.	Indicator
Epidemics Risk and Priority	Vulnerability	Behaviour	Self Perceived Health	Subjective state of health	1	State of health (subjective)
			Health Seeking Behaviour	Advise taking behaviour	3	Was not confined in hospital because of no need
				Consultation seeking	5	Place of first consultation/advice or treatment
					7	Refusal of treatment after consultat doctor because: Symptoms (found to be) harmless
					8	Refusal of treatment after consultat doctor because: Cost
				Medication Seeking	12	Refusal of treatment after consultat doctor because: Medicines bought apart from (medicine / services in) the hospital
			Prevention	General Health	14	Thinks done to keep healthy: Maintain good hygiene
					15	Thinks done to keep healthy: Check-up by docter
					16	Thinks done to keep healthy: Aet fish/lean meat/poultry/ soy beans
					19	Thinks done to keep healthy: None
				Vaccination Degree	21	Vaccination coverage (measles as proxy for willingness)
			Trust	Government	22	Trust in the government

Figure 9: Final List of indicators in the EPI framework

As the correlation matrix in Figure 10 shows, some indicators have a linear correlation coefficient between 0.3 and 0.5. These indicators are assessed with the VIF method and determined to not be multicollinear. Although these indicators are somewhat dependent, they do not explain each other and were therefore included in the framework.

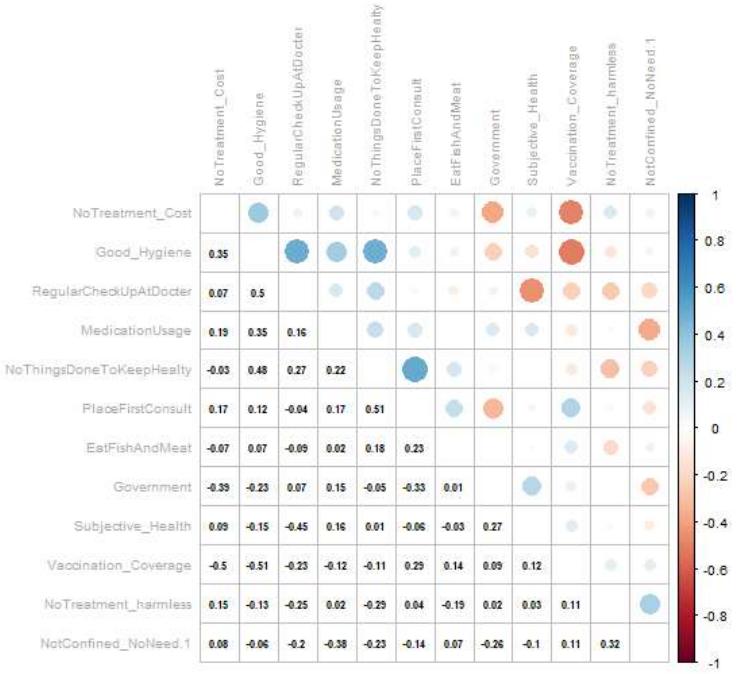

Figure 10 Correlation matrix with correlation coefficients for the final framework

4.4. Estimated effect of indicators in the EPI model

To determine the estimated effect per indicator and for all indicators as a whole (as a whole indicated the category behaviour) the regression coefficients were calculated. To calculate the regression coefficient per indicator for the EPI model a stepwise forward Poisson regression was performed in R studio. Table 2 shows the summery of the regression analysis given by R studio. The indicators names have been abbreviated but the indicator number in the table corresponds with the indicator number shown in Figure 9.

Table 2: EPI model with Poisson regression

INDICATOR		REGRESSION COEFFICIENT ($\boldsymbol{\beta_i}$)	STD. ERROR	Z VALUE	P VALUE
NR.	**(Intercept)**	6.8519	11.4308	0.5990	0.5490
1	**Subjective_Health**	0.2048	5.9233	0.0350	0.9720
3	**NotConfined_NoNeed**	0.9455	3.5658	0.2650	0.7910
4	**PlaceFirstConsult**	0.5775	3.4756	0.1660	0.8680
7	**NoTreatment_Cost**	0.1724	3.9988	0.0430	0.9660
8	**NoTreatment_Harmless**	0.7165	4.0092	0.1790	0.8580
12	**MedicationUsage**	0.7345	1.7060	0.4310	0.6670
21	**Vaccination_Coverage**	8.1055	14.0496	0.5770	0.5640
22	**Government**	2.1554	4.3851	0.4920	0.6230
14	**Good_Hygiene**	1.8205	3.5473	0.5130	0.6080
19	**NoThingsDoneToKeepHealty**	1.6614	3.4599	0.4800	0.6310
15	**RegularCheckUpAtDocter**	1.1365	4.2473	0.2680	0.7890
16	**EatFishAndMeat**	2.1615	3.8409	0.5630	0.5740

All indicators have a substantial effect in the model. The significance of the indicators is higher than wanted. However, compared to the indicators the R^2 of the model is high with $R^2 = 0.91$. This is calculated with equation 5 using $R^2 = 1 - \frac{0.66815}{7.89886}$. This R^2 indicated that the model is a good fit and that the model explains behaviour better than the intercept. The intercept, in Table 2, indicates that the category behaviour influences the dengue incidence with $\beta_i = 6.8519$.

The final mathematic equation as an outcome of the Poisson regression and that will be used for calculating the EPI index is:

$$Y \sim Poisson(\lambda) \qquad \textit{Equation 9}$$

$$\log(\lambda)_{Region} = \beta_{i1}x_{v1} + \beta_{i3}x_{v3} + \beta_{i4}x_{v4} \ldots . + \beta_{i16}x_{i16}$$

4.5. RAV of the indicators to the EPI model

The RAV per indicator to the EPI model is shown in Figure 11. This shows that almost 80% of the model is explained by the indicators of the components prevention and trust. However, when excluding the indicators of the other components, the validity of the model does not change significantly. This means that in future studies or when combining the models in the EPI project these indicators are the first candidates to be excluded.

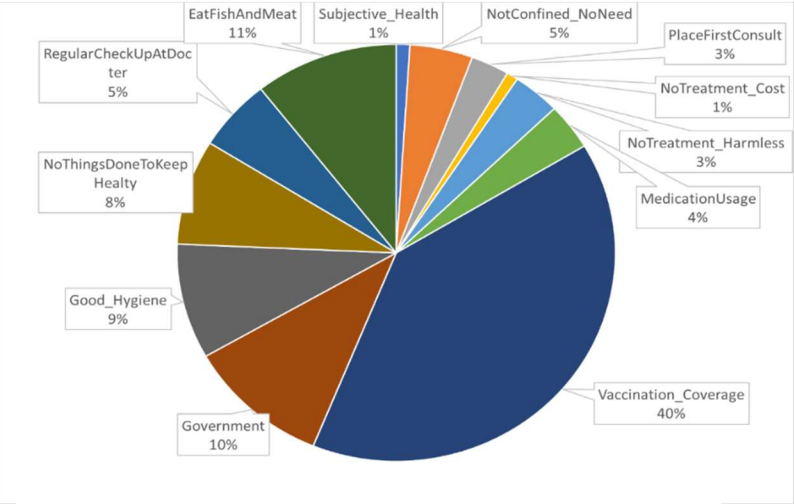

Figure 11: Pie Chart visualising the relative added value per indicator to the model

The distribution of RAV per indicator per region is visualized with a stacked bar chart shown in Figure 12. The shows that not all indicator are equally distributed over the regions.

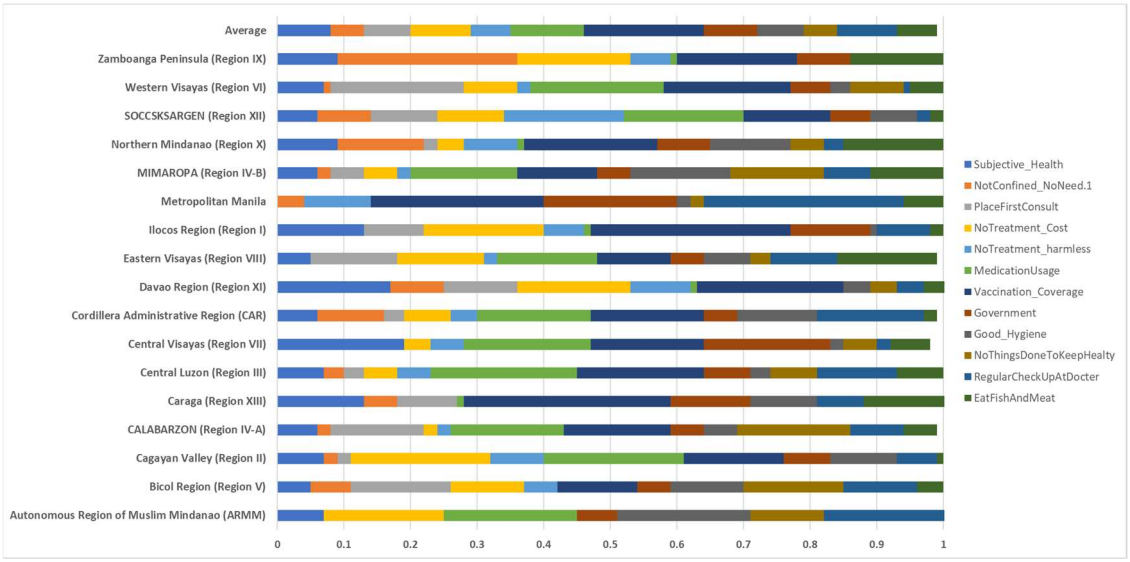

Figure 12:: Distribution of RAV per indicator per Region visualized in a stacked bar chart

4.6. EPI index for dengue

With the EPI model the EPI index per region could be determined. The aggregated values per indicator per region were included in the mathematical equation with the corresponding regression coefficient of the model. The normalized outcome of the combined measures in equation 9 forms the EPI index ass shown in The EPI index is visualized in Figure 13 on the left for comparison to the actual measured dengue incidence in the right figure, clearly showing a good agreement between the two. Which means that even though the indicators by themselves are not that strong, the model as a whole is.
Table 3).

The EPI index indicates the risk a region poses to epidemics compared to other regions given a certain region profile (behaviour indicators). The Cordillera Administrative Region is the highest prioritized region compared to the other regions for risk on dengue incidence with $index = 1$. The Autonomous Region of Muslim Mindanao is the lowest prioritized region for risk on dengue incidence with $index = 0$. Although $Index = 0$ indicates the lowest risk, it should be clearly stated that this does not indicate that dengue does not occur. The EPI index is visualized in Figure 13 on the left for comparison to the actual measured dengue incidence in the right figure, clearly showing a good agreement between the two. Which means that even though the indicators by themselves are not that strong, the model as a whole is.

Table 3: EPI index per region with corresponding predictive value

REGION	PREDICTIVE VALUE	EPI INDEX (ABS)	EPI INDEX (%)
AUTONOMOUS REGION OF MUSLIM MINDANAO (ARMM)	0.00	0	0%
BICOL REGION (REGION V)	0.15	0.148336578	15%
CAGAYAN VALLEY (REGION II)	0.69	0.688159388	69%
CALABARZON (REGION IV-A)	0.39	0.387759756	39%
CARAGA (REGION XIII)	0.57	0.569136427	57%
CENTRAL LUZON (REGION III)	0.51	0.510774279	51%
CENTRAL VISAYAS (REGION VII)	0.90	0.901736175	90%
CORDILLERA ADMINISTRATIVE REGION (CAR)	1.00	1	100%
DAVAO REGION (REGION XI)	0.47	0.473080283	47%
EASTERN VISAYAS (REGION VIII)	0.10	0.101545655	10%
ILOCOS REGION (REGION I)	0.39	0.394499291	39%
METROPOLITAN MANILA	0.51	0.505563377	51%
MIMAROPA (REGION IV-B)	0.18	0.182966712	18%
NORTHERN MINDANAO (REGION X)	0.37	0.374913201	37%
SOCCSKSARGEN (REGION XII)	0.51	0.509247456	51%
WESTERN VISAYAS (REGION VI)	0.70	0.695753807	70%
ZAMBOANGA PENINSULA (REGION IX)	0.29	0.294528013	29%

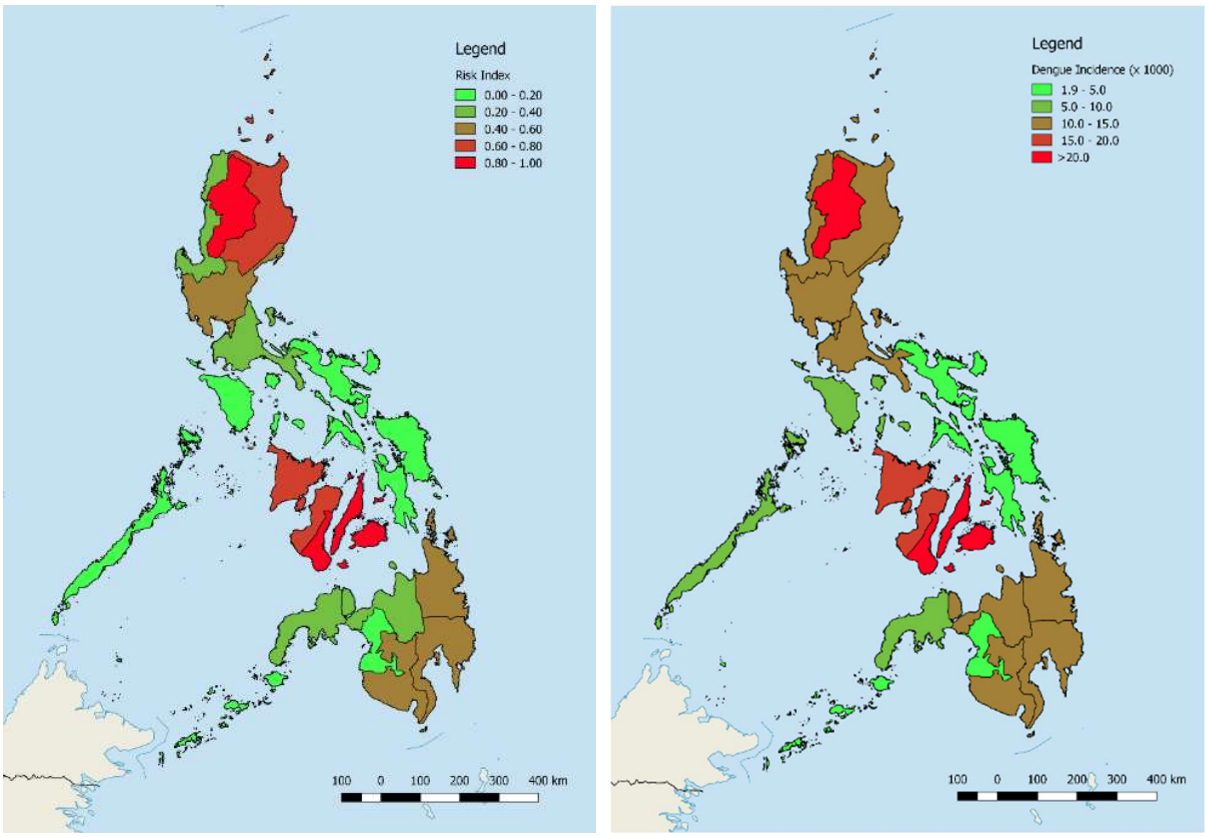

Figure 13: Risk index for dengue incidence plotted per region (left) and dengue incidence in 2013 plotted per region (right)

5. Discussion

In this study, a novel model was developed to identify behavioural risk factors which can make a community vulnerable to epidemics and infectious diseases. As a pilot case study dengue incidence in the Philippines was used. To do this, first a framework was created with the components Self-Perceived Health, Healthcare Seeking Behaviour, Prevention and Trust. The DHS, WVS and pre-disaster indicators were then used to provide valid and operationalized indicators as input. 24 indicators were subjectively included in the full EPI framework. The set of indicators was then reduced by statistical analysis, resulting in a set of 12 independent indicators for the final EPI framework. This set was then used to create the EPI model. The model indicates behaviour to have a relation with dengue incidence with $\beta_i = 6.8519$ and $R^2 = 0.91$, but with $P = 0.5490$ conclusion should be drawn carefully. This demonstrates that the category behaviour has a positive relation with dengue incidence.

5.1. Limitations and strengths

The main limitations of this study regard the data quality. Firstly, there are two subcomponents identified in the literature as possible influences, namely influence of physical contact with diseased and trust on the health care system. However, these subcomponents could not be tested due to a lack of data. Secondly, the geographical resolution was limited to a regional level as the WVS only had data on that level. If both the DHS and WVS had been available on provincial level, the analysis could have been performed with the 81 provinces. Thirdly, not all the surveys are from the same timeframe. The limitations due to the data could limit the internal validity of this study, which may be the cause of the poor P-values found. The P-values could possibly be improved by collecting more standardized data and testing more diseases.

Other limitations are regarding reproducibility and ecological fallacy. Difference in outcome when repeating this study as described here could be found in the literature study since a new methodology was created. The outcome of the literature study could be different when more studies other than retrospective analyses are taken into account. Preferably a complete systematic review should be performed. Ecological fallacy refers to the fact that the results of the study are based on aggregated data or averages of a region. This is a generalisation of a group of people, but that does not mean the results are applicable to every individual in the region.

The major strengths of this study is the evidence-based nature of the model. The model is built on a literature study which is supported by and discussed with many experts. Behavioural vulnerability is a part that has not been taken into account in other models. This model has been built on a white canvas. The major strength of the EPI project is that all dimensions are researched separately. This means that all (sub-)components and indicators are selected individually without bias of the other dimensions. Although this study focuses on the Philippines, the underlying EPI framework is easily transferable to other regions. Furthermore, the possibilities of giving a global policy advise based on a future global model but focussing on a country setting makes the model stronger as the index results are more valid in advising a specific country for policy and intervention improvements.

5.2. Recommendations

Based on the above-mentioned limitations of this study, there are four main recommendations that can be given to further improve this study.

The first recommendation is that all EPI models within the EPI project should be combined to one EPI model to ensure all factors are involved in measuring the risk on epidemics or incidence of infectious diseases. To combine the models, the EPI model for behaviour should be further developed. First, data needs to be collected for the indicators with no data. This can possibly be done through the DOH of the Philippines or the PRC. Second, the model needs to be converted from regional level to provincial level or lower. The regional level has been chosen for this thesis since the data source WVS only had data on region level available. The most optimal study design for the EPI project is a cohort study with multilevel data to add a time component. When multi-level data is available a multi-level regression should be performed to make the EPI model more reliable as a whole. The multi-level regression analysis is appropriate for research designs were data is organized at more than one level[56].

The second recommendation is that the EPI model for behaviour should be tested with other diseases. Dengue is a vector-borne disease that cannot be spread from human to human. By nature, dengue is different from airborne, bloodborne and foodborne diseases which means other factors are important in measuring the risk and preventing the spread.

The third recommendation is to test the robustness of the model with a complete sensitivity analysis. The inclusion and exclusion criteria for literature and therefore the indicators and sub-components. The selection of the data sources since the use of different data sources from different times weakens the correlations as they are not from the same time frame. Different methods for handling missing data can be tested, other normalization methods, different aggregation methods and different regression methods.

The fourth recommendation is that the model should be tested for validity with a qualitative study performed in the Philippines. With a qualitative study the real-life applicability of the model can be tested.

Reference List

1 Myers MF, Rogers DJ, Cox J, Flahault A, Hay SI. Forecasting disease risk for increased epidemic preparedness in public health. Advances in parasitology. 2000;47:309-30, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3196833/

2 Global risk report 2016. World Economic Forum [Internet]. 2016. Available from: http://wef.ch/1QfL09i.

3 Watson JT, Gayer M, Connolly MA. Epidemics after natural disasters. Emerg Infect Dis. 2007;13(1):1-5.10.3201/eid1301.060779

4 Kouadio IK, Aljunid S, Kamigaki T, Hammad K, Oshitani H. Infectious diseases following natural disasters: Prevention and control measures. Expert Rev Anti Infect Ther. 2012;10(1):95-104.10.1586/eri.11.155

5 Mandate and mission of internation federation of red cross 2018 [Available from: http://www.ifrc.org/en/who-we-are/vision-and-mission/.

6 The International Federation of Red Cross and Red Crescent Societies. About disaster management [Available from: http://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management/.

7 International Federation of Red Cross and Red Crescent Societies. Disaster prepardness 2018 [Available from: http://media.ifrc.org/ifrc/what-we-do/disaster-and-crisis-management/disaster-preparedness/

8 Landier J, Parker DM, Thu AM, Carrara VI, Lwin KM, Bonnington CA, Pukrittayakamee S, Delmas G, Nosten FH. The role of early detection and treatment in malaria elimination. Malaria Journal. 2016;15(1):363.10.1186/s12936-016-1399-y

9 Modu B, Polovina N, Lan Y, Konur S, Asyhari A, Peng Y. Towards a predictive analytics-based intelligent malaria outbreak warning system. Applied Sciences. 2017;7(8).10.3390/app7080836

10 Tomashek KM, Gregory CJ, Rivera Sánchez A, Bartek MA, Garcia Rivera EJ, Hunsperger E, Muñoz-Jordán JL, Sun W. Dengue deaths in puerto rico: Lessons learned from the 2007 epidemic. PLoS Neglected Tropical Diseases. 2012;6(4):e1614.10.1371/journal.pntd.0001614

11 Calistri P, Conte A, Natale F, Possenti L, Savini L, Danzetta ML, Iannetti S, Giovannini A. Systems for prevention and control of epidemic emergencies. Vet Ital. 2013;49(3):255-61.10.12834/VetIt.1206.06

12 Milinovich GJ, Magalhães RJS, Hu W. Role of big data in the early detection of ebola and other emerging infectious diseases. The Lancet Global Health.3(1):e20-e1.10.1016/S2214-109X(14)70356-0

13 RodrÍGuez-Prieto V, Vicente-Rubiano M, SÁNchez-Matamoros A, Rubio-Guerri C, Melero M, MartÍNez-LÓPez B, MartÍNez-AvilÉS M, Hoinville L, Vergne T, Comin A, Schauer B, DÓRea F, Pfeiffer DU, SÁNchez-VizcaÍNo JM. Systematic review of surveillance systems and methods for early detection of exotic, new and re-emerging diseases in animal populations. Epidemiology and Infection. 2015;143(10):2018-42.10.1017/S095026881400212X

14 Taylor D, Hagenlocher M, Jones AE, Kienberger S, Leedale J, Morse AP. Environmental change and rift valley fever in eastern africa: Projecting beyond healthy futures. Geospat Health. 2016;11(1 Suppl):387.10.4081/gh.2016.387

15 Christaki E. New technologies in predicting, preventing and controlling emerging infectious diseases. Virulence. 2015;6(6):558-65.10.1080/21505594.2015.1040975

16 Hagenlocher M, Castro MC. Mapping malaria risk and vulnerability in the united republic of tanzania: A spatial explicit model. Population Health Metrics. 2015;13(1):2.10.1186/s12963-015-0036-2

17 Hagenlocher M, Delmelle E, Casas I, Kienberger S. Assessing socioeconomic vulnerability to dengue fever in cali, colombia: Statistical vs expert-based modeling. International Journal of Health Geographics. 2013;12:36-.10.1186/1476-072X-12-36

18 510.Global. Mission and vission 510.Global 2016 [Available from: https://www.510.global/510-mission-vision/.

19 510 Global - Maarten van Veen. Priority index part 1 09.11.2016 [Available from: https://www.510.global/philippines-typhoon-haima-priority-index/.

20 510 Global - Jannis Visser. Priority index 15.12.2016 [Available from: https://www.510.global/surge-informatio-the-philippines/

21 InfoRM is a collaborative project of the Inter-Agency Standing Committee (IASC) and the European Commission. Inform methodology [Available from: http://www.inform-index.org/InDepth/Methodology.

22 Inform methodology in depth [Available from: http://www.inform-index.org/InDepth.

23 INFORM-Index. Guidance note - interpreting and applying the inform global model. 2017, http://www.inform-index.org/

24 Hagenlocher M, Delmelle E, Casas I, Kienberger S. Assessing socioeconomic vulnerability to dengue fever in cali, colombia: Statistical vs expert-based modeling. International journal of health geographics. 2013;12(1):36,

25 Taylor D, Hagenlocher M, Jones AE, Kienberger S, Leedale J, Morse AP. Environmental change and rift valley fever in eastern africa: Projecting beyond healthy futures. Geospatial health. 2016;11(1s),

26 al. R-Re. What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? Tropical Medicine and International Health. 2008;13(8),

27 Buczak. Prediction of high incidence of dengue in the philippines. PLOS Neglected tropical diseases. 2014,

28 Gibbons RV, Vaughn DW. Dengue: An escalating problem. BMJ : British Medical Journal. 2002;324(7353):1563-6, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1123504/

29 Gubler. Epidemic dengue and dengue hemorrhagic fever: A global public health problem in the 21st century. 1998,

30 Cherif A, Barley K, Hurtado M. Homo-psychologicus: Reactionary behavioural aspects of epidemics. Epidemics. 2016;14:45-53.https://doi.org/10.1016/j.epidem.2015.09.003

31 Levin KA. Study design vi - ecological studies. Evidence Based Dentistry. 2006;7:108.10.1038/sj.ebd.6400454

32 Garchitorena A, Sokolow SH, Roche B, Ngonghala CN, Jocque M, Lund A, Barry M, Mordecai EA, Daily GC, Jones JH, Andrews JR, Bendavid E, Luby SP, LaBeaud AD, Seetah K*, et al.* Disease ecology, health and the environment: A framework to account for ecological and socio-economic drivers in the control of neglected tropical diseases. Philos Trans R Soc Lond B Biol Sci. 2017;372(1722).10.1098/rstb.2016.0128

33 WorldAtlas.com. Webimage of the philippines 2015 [updated September 23, 2015. Available from: https://www.worldatlas.com/webimage/countrys/asia/lgcolor/phcolor.htm

34 WHO APO. The philippines health system review. WHO. 2011,

35 Sia-Catedral ML. The philippine disaster risk reduction and management (drrm) system 2016.http://www.adrc.asia/countryreport/PHL/2016/Philippines_CR2016A.pdf

36 OCHA Philippines IM. Un ocha philippines pre-disaster indicators 2018 [updated 2018.03.08. Available from: https://public.tableau.com/profile/ocha.philippinesim#!/vizhome/Pre-disasterIndicatorsDashboard/Dashboard1

37 WHO. Fact sheet: Vector-borne diseases October 2017.http://www.who.int/mediacentre/factsheets/fs387/en/

38 Kularatne SAM. Dengue fever. BMJ : British Medical Journal. 2015;351, http://www.bmj.com/content/351/bmj.h4661.abstract

39 Aguiar M. Dengue vaccination: A more ethical approach is needed. The Lancet. 2018;391(10132):1769-70.10.1016/S0140-6736(18)30865-1
40 Roiger RJ. Data mining: A tutorial-based primer: CRC Press; 2017.
41 Han J, Pei J, Kamber M. Data mining: Concepts and techniques: Elsevier; 2011.
42 Microsoft. Download sql server management studio (ssms) 20.03.2018 [Available from: https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms.
43 R-Studio. Download r-studio 30.06.2017 [Available from: https://www.rstudio.com/products/rstudio/download/#download.
44 RDocumentation. Odbc 06.02.2017 [Available from: https://www.rdocumentation.org/packages/odbc/versions/1.0.1.
45 RDocumentation. Rodbc 13.04.2017 [Available from: https://www.rdocumentation.org/packages/RODBC/versions/1.3-15.
46 Buuren Sv. Package, micr. Oktober 24, 2017, https://cran.r-project.org/web/packages/mice/mice.pdf
47 commission CCCoClaSote. Handbook on contructing composite indicators (an 10 step guide) 2018 [updated 16 january 2018. Available from: https://composite-indicators.jrc.ec.europa.eu/?q=10-step-guide
48 RDocumentation. K-nearest neighbour imputation from vim v4.7.0, 2017 [Available from: https://www.rdocumentation.org/packages/VIM/versions/4.7.0/topics/kNN.
49 RDocumentation. Vim v4.7.0, 2017 [updated 11.04.2017. Available from: https://www.rdocumentation.org/packages/VIM/versions/4.7.0.
50 Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971;27(4):857-71.10.2307/2528823
51 RDocumentation. Ddply of ply v1.8.4, 2016 [Available from: https://www.rdocumentation.org/packages/plyr/versions/1.8.4/topics/ddply.
52 RDocumentation. Plyr v1.8.4 2016 [Available from: https://www.rdocumentation.org/packages/plyr/versions/1.8.4.
53 Correlation coefficient. Evidence-based Obstetrics & Gynecology. 2004;6(2):48-50.https://doi.org/10.1016/j.ebobgyn.2004.04.015
54 Zhu G, Xiao J, Zhang B, Liu T, Lin H, Li X, Song T, Zhang Y, Ma W, Hao Y. The spatiotemporal transmission of dengue and its driving mechanism: A case study on the 2014 dengue outbreak in guangdong, china. Sci Total Environ. 2017;622-623:252-9.10.1016/j.scitotenv.2017.11.314
55 Brézin E, Zee A. Universality of the correlations between eigenvalues of large random matrices. Nuclear Physics B. 1993;402(3):613-27.https://doi.org/10.1016/0550-3213(93)90121-5
56 James G. An introduction to statistical learning : With applications in r. New York, NY :: Springer; 2013. Available from: SpringerLink http://dx.doi.org/10.1007/978-1-4614-7138-7

http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf

Click here to access this title in Ebook Central https://ebookcentral.proquest.com/lib/buffalo/detail.action?docID=1317587

https://login.libproxy.uregina.ca:8443/login?url=https://link.springer.com/openurl?genre=book&isbn=978-1-4614-7137-0.
57 RDocumentation. Cor of stats v3.5.0., x [Available from: https://www.rdocumentation.org/packages/stats/versions/3.5.0/topics/cor
58 RDocumentation. Stats v3.5.0. x [Available from: https://www.rdocumentation.org/packages/stats/versions/3.5.0.
59 RDocumentation. Corrplot.Mixed of corrplot v0.84, 2017 [Available from: https://www.rdocumentation.org/packages/corrplot/versions/0.84/topics/corrplot.mixed.
60 Rdocumentation. Corrplot v0.84 2017 [Available from: https://www.rdocumentation.org/packages/corrplot/versions/0.84.
61 RDocumentation. Vif of hh v3.1-34 2017 [Available from: https://www.rdocumentation.org/packages/HH/versions/3.1-34/topics/vif.
62 RDocumentation. Hh v3.1-34 2017 [Available from: https://www.rdocumentation.org/packages/HH/versions/3.1-34.
63 Dan Liao RV. Variance inflation factors in the analysis of complex survey data. Survey Methodology, June 2012. 2012;Vol. 38, No. 1, pp. 53-62, http://www.statcan.gc.ca/pub/12-001-x/2012001/article/11685-eng.pdf
64 RDocumentation. Glm of stats v3.4.3. 2017 [Available from: https://www.rdocumentation.org/packages/stats/versions/3.4.3/topics/glm.
65 Christian K, Achim Z, Simon J. Regression models for count data in r. journal of statistical software [Internet]. 2008; 27(8).
66 Rdocumentation. Predict.Glm of stats v3.5.0. 2017 [
67 qGIS. Qgis download 2018 [Available from: https://qgis.org/en/site/forusers/download.html.
68 Natural Earth. Natural earth quick start kit (165mb) 2009 - 2018 [Available from: http://www.naturalearthdata.com/downloads/.
69 PhilGIS. Gis data 2018 [Available from: https://www.philgis.org/gis-data.
70 Coltart CEM, Lindsey B, Ghinai I, Johnson AM, Heymann DL. The ebola outbreak, 2013–2016: Old lessons for new epidemics. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017;372(1721):20160297.10.1098/rstb.2016.0297
71 al. Be. Sharing experiences: Towards an evidence based model of dengue surveillance and outbreak response in latin america and asia. BMC Public Health. 2013.10.1186/1471-2458-13-607
72 Huber M, André Knottnerus J, Green L, Horst H, Jadad A, Kromhout D, Leonard B, Lorig K, Loureiro M, W M van der Meer J, Schnabel P, Smith R, Weel C, Smid H. How should we define health?2011. d4163 p.
73 Larson JS. The conceptualization of health (who welness and enviromental model). Medical Care Research and Review. 2016;56(2):123-36.10.1177/107755879905600201
74 Machón M, Vergara I, Dorronsoro M, Vrotsou K, Larrañaga I. Self-perceived health in functionally independent older people: Associated factors. BMC Geriatrics. 2016;16:66.10.1186/s12877-016-0239-9
75 Halstead SB. Dengue: Overview and history. Dengue. Tropical medicine: Science and practice. Volume 5: PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO.; 2008. p. 1-28.
76 McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the theory of planned behaviour: A meta-analysis. Health Psychology Review. 2011;5(2):97-144.10.1080/17437199.2010.521684

Appendix I Project overview

Table 4: Description of the nine projects within the epidemic risk and priority index project

project information			Student	University	Period
1	Dimension: Category:	Infectious hazards and exposure Population drivers and disease drivers	F. Hierink	University of Utrecht	01.2018 – 08.2018
2	Dimension: Category:	Vulnerability Social Economic	A. Teng	Lund University	01.2018 – 11.2018
3	Dimension: Category:	Vulnerability Movement	vacancy		
4	Dimension: Category:	Vulnerability Behaviour	C. Meijerink	University of Twente	02.2018 – 06.2018
5	Dimension: Category:	Lack of Coping capacity Health infrastructure	F. Lammers	Erasmus university	01.2018 – 06.2018
6	Dimension: Category:	Lack of Coping capacity Governance + IHR	B. Veneman		06.2018 – end
7	Timeline of disasters and responses		R. Sunnis		03.2018 – end
8	Economical history of post disaster responses and epidemic responses		M. Pancar		03.2018 – end
9	Creation of database and project lead		K. Arslantas		08.2017 – end

Appendix II Overview Literature Study

Table 5: Overview Literature study

Reference Nr.	DOI	URL	Authors	Year of Study	Study design	Focus
[22]	-	URL	Inter-agency Standing committee and the European commission	Version 2016	Collaborative study	INFORM methodology
[39]	10.1016/S0140-6736(18)30865-1	URL	Aguiar M	2018	Correspondence	Dengue Vaccination
[54]	10.1016/j.scitotenv.2017.11.314	URL	Zhu G, Hao Y, et Al.	2017	Retrospective case study	The Dengue outbreak in Guengdong, China, in 2014
[30]	10.1016/j.epidem.2015.09.003	URL	Cherif A, Barley K, Hurtado M	2016	Epidemiological study	Modelling reactionary behavioural aspects of epidemics
[70]	10.1098/rstb.2016.0297	URL	Coltart C, Heymann D, et Al.	2017	Retrospective study	The Ebola outbreak in 2013 – 2016
[71]	10.1186/1471-2458-13-607	URL	Bardudeen et Al.	2013	Retrospective study	Model based on dengue surveillance and outbreak response
[72]	10.1136/bmj.d4163	URL	Huber M, Smid H, et Al.	2011	Descriptive study	Define Health
[73]	10.1177/107755879905600201	URL	Larson, James S.	2016	Descriptive study	Conceptualization of Health
[74]	10.1186/s12877-016-0239-9	URL	Machon M, Larranaga I, et Al.	2016	Cross sectional study	Self-Perceived Health
[76]	10.1080/17437199.2010.521684	URL	McEachan R, Lawton R, et Al.	2011	Prospective study	Prediction of health-related behaviour

Appendix III List of consulted experts

Table 6: List of consulted experts

Function:	Field of expertise	Discussed with regarding
Organisation lead at 510.Global	Astrophysics and modelling	Overview of the EPI project
Organisation lead at 510.Global	Artificial intelligence	Overview of the EPI project
Scientific lead at 510.Global	Physics and MBA	Overview of the EPI project
Project Lead and data analyst at 510.Global	Econometrics	The CRA toolbox of 510.Global and the methodology
Team member at 510.Global	GIS	Visualisation of the EPI index
Team member at 510.Global	Statistical modelling	The EPI Model
Team member at 510.Global	Statistical modelling	The EPI Model
Communication specialist at 510.Global	Communication and visualisations	Visualisations and presenting
Health advisor at Netherlands Red Cross	Public health and humanitarian aid	The EPI framework
Health advisor at Uganda Red Cross	Public health and humanitarian aid	The EPI framework
Health advisor at Philippine Red Cross	Public health and humanitarian aid	The EPI framework
Team member at Philippine Red Cross	Public health and humanitarian aid	The EPI framework and EPI model
Graduate student at 510.Global. Master of public health at Lund university, Sweden	Public Health	The EPI framework and EPI model
Graduate student at 510.Global. Master of environmental health at Utrecht university	Environmental health	The EPI framework and EPI model
Graduate student at 510.Global. Master of health economic and managements at Erasmus university	Health economics	The EPI framework and EPI model
Project lead (EPI project) at 510.Global	Public health management, statistical analysis and disaster management	Database and overview EPI project
Assistant professor at Erasmus MC, department of public health	Statistical modelling of infectious disease epidemiology and impact of public health interventions	The EPI model and the EPI index
Erasmus MC, department of public health	Clinical aspects of exotic viruses	The EPI framework and Model
Senior scientific researcher at Erasmus MC, department of public health	Epidemiology and tropical diseases	The EPI framework
Associate professor at University of Twente	Behavioural, management and social science	Overview of the EPI project and writing the thesis
Professor in sociology at University of Twente	Sociology of public governance	Overview of the EPI project and writing the thesis
Master in international humanitarian aid	Humanitarian governance	The EPI framework

Appendix IV R-Script used for statistical analyse

```
########################################################## START OF ANALYSIS ###########################################################
#################################### DIMENSION: VULNERABILITY ####### CATEGORY: BEHAVIOUR ######################################
#Start with a clear enviroment
rm(list=ls())

#install the packages you will need throughout the analyses
install.packages("ggplot3")
install.packages("readr")
install.packages("psych")
install.packages("FactoMineR")
install.packages("corrplot")
install.packages("dplyr")
install.packages("tidyverse")
install.packages("odbc")
install.packages("lattice")
install.packages("mice")
install.packages("VIM")
install.packages("clusterSim")
install.packages("RODBC")
install.packages("foreign")
install.packages("dplyr")
install.packages("plyr")
install.packages("Hmisc")
install.packages("HH")
install.packages("raster")


#load the libraries. Once you have installed all the packages
#you only need to load them everytime you start working in R
library(ggplot3)    #For the plotting of data#
library(readr)      #For reading rectangular text data#
library(psych)      #For loading file data#
library(FactoMineR) #For the multivariate data analysis#
library(corrplot)   #For the correlation plot and matrix#
library(dplyr)      #For manipulating data#
library(tidyverse)  #General Package with data analysis options#
library(odbc)       #For connection R to SQL database#
library(lattice)    #Required for MICE#
library(mice)       #For analysing missing values#
library(VIM)        #For imputation of missing data
library(clusterSim) #For normalization methods#
library(RODBC)      #For connection R to SQL database#
library(foreign)    #For reading SPSS data#
library(plyr)      #For data manipulation#
library(Hmisc)      #For data analysis#
library(HH)        #For statistical analysis and data display#
library(raster)     #For geograpical data analysis

########################################################## LOAD DATASOURCES ###########################################################
rm(list=ls())
##connect MS SQL server to R studio
epi <- odbcConnect("EPI", uid="sa", pwd = "Ejz0afm5i69zJdMmxRAB")

#Read and load all the needed data for the analysis
#To connect the SQL database to Rstudio first go the document and make sure you first set up this connection before going through
dengue.1 <- sqlQuery(epi, "SELECT [Month],[Year],[Region],[Dengue_Cases],[1 in dengue patient] FROM [1].[denguecases]")
dhs.2 <-read.spss(file=("D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/Data/PHHR61FL.sav"), to.data.frame=TRUE)
wvs.6 <-read.spss(file=("D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/Data/wvs.6.compressed.sav"), to.data.frame=TRUE)
measl.vac.cov.9 <- sqlQuery(epi, "SELECT [Vaccination],[Region],[Percentage] FROM [9].[measles]")

#################################### Transformation behaviour ##############################################
######create components "self perceived health" for behaviour########
SelfPerceivedHealth_SubjectiveHealth <- data.frame(Region=wvs.6$X048WVS, Subj_Health=cbind(wvs.6$A009), Subj_Happiness=cbind(wvs.6$A008))
md.pattern(SelfPerceivedHealth_SubjectiveHealth)
SelfPerceivedHealth_SubjectiveHealth <- kNN(SelfPerceivedHealth_SubjectiveHealth,variable = "Subj_Happiness", metric=NULL, k=4)
mice_plot <- aggr(SelfPerceivedHealth_SubjectiveHealth, col=c('green','red'),
                  numbers=TRUE, sortVars=TRUE,
                  labels=names(SelfPerceivedHealth_SubjectiveHealth), cex.axis=.7,
                  gap=3, ylab=c("Missing data","SelfPerceivedHealth_SubjectiveHealth"))
SelfPerceivedHealth_SubjectiveHealth <- ddply(SelfPerceivedHealth_SubjectiveHealth, "Region", summarise,
                                              N_Health    = length(Region),
                                              sum_Health  = sum(Subj_Health),
                                              mean_Health  = mean(Subj_Health),
                                              sd_Health    = sd(Subj_Health),
                                              se_Health    = sd_Health  / sqrt(N_Health ))

SelfPerceivedHealth_SubjectiveSatisfaction <- data.frame(Region=wvs.6$X048WVS, Subj_Satisfaction=cbind(wvs.6$A170))
md.pattern(SelfPerceivedHealth_SubjectiveSatisfaction)
SelfPerceivedHealth_SubjectiveSatisfaction <- kNN(SelfPerceivedHealth_SubjectiveSatisfaction,variable = "Subj_Satisfaction", metric=NULL, k=4)
```

```
mice_plot <- aggr(SelfPerceivedHealth_SubjectiveSatisfaction, col=c('green','red'),
                  numbers=TRUE, sortVars=TRUE,
                  labels=names(SelfPerceivedHealth_SubjectiveSatisfaction), cex.axis=.7,
                  gap=3, ylab=c("Missing data","SelfPerceivedHealth_SubjectiveSatisfaction"))
SelfPerceivedHealth_SubjectiveSatisfaction <- ddply(SelfPerceivedHealth_SubjectiveSatisfaction, "Region", summarise,
                                    N_Satisfaction   = length(Region),
                                    sum_Satisfaction  = sum(Subj_Satisfaction),
                                    mean_Satisfaction = mean(Subj_Satisfaction),
                                    sd_Satisfaction   = sd(Subj_Satisfaction),
                                    se_Satisfaction   = sd_Satisfaction / sqrt(N_Satisfaction))

SelfPerceivedHealth <- merge(SelfPerceivedHealth_SubjectiveHealth, SelfPerceivedHealth_SubjectiveSatisfaction, by="Region")
remove(SelfPerceivedHealth_SubjectiveHealth)
remove(SelfPerceivedHealth_SubjectiveSatisfaction)

SelfPerceivedHealth$Region <- as.character(SelfPerceivedHealth$Region)
SelfPerceivedHealth$Region[SelfPerceivedHealth$Region == "PH: NCR"] <- "Metropolitan Manila"
SelfPerceivedHealth$Region[SelfPerceivedHealth$Region == "PH: SOUTH LUZON"] <- "Central Luzon (Region III)"
SelfPerceivedHealth$Region[SelfPerceivedHealth$Region == "PH: VISAYAS"] <- "Central Visayas (Region VII)"
SelfPerceivedHealth$Region[SelfPerceivedHealth$Region == "PH: MINDANAO"] <- "Davao Region (Region XI)"

######create components "health seaking behaviour" for behaviour########
HealthSeakingBehaviour_ConfinementHospital <- data.frame(Region=dhs.2$HV024,
                                    AdviseConfinement=cbind(c(dhs.2$SH211.1, dhs.2$SH211.2, dhs.2$SH211.3, dhs.2$SH211.4, dhs.2$SH211.5, dhs.2$SH211.6,
dhs.2$SH211.7, dhs.2$SH211.8, dhs.2$SH211.9)),
                                    ConfinedInHospital=cbind(c(dhs.2$SH212.1, dhs.2$SH212.2, dhs.2$SH212.3, dhs.2$SH212.4, dhs.2$SH212.5, dhs.2$SH212.6,
dhs.2$SH212.7, dhs.2$SH212.8, dhs.2$SH212.9)),
                                    NotConfined_Distance=cbind(c(dhs.2$SH213A.1, dhs.2$SH213A.2, dhs.2$SH213A.3, dhs.2$SH213A.4, dhs.2$SH213A.5,
dhs.2$SH213A.6, dhs.2$SH213A.7, dhs.2$SH213A.8, dhs.2$SH213A.9)),
                                    NotConfined_NoMoney=cbind(c(dhs.2$SH213B.1, dhs.2$SH213B.2, dhs.2$SH213B.3, dhs.2$SH213B.4, dhs.2$SH213B.5,
dhs.2$SH213B.6, dhs.2$SH213B.7, dhs.2$SH213B.8, dhs.2$SH213B.9)),
                                    NotConfined_Costs=cbind(c(dhs.2$SH213C.1, dhs.2$SH213C.2, dhs.2$SH213C.3, dhs.2$SH213C.4, dhs.2$SH213C.5,
dhs.2$SH213C.6, dhs.2$SH213C.7, dhs.2$SH213C.8, dhs.2$SH213C.9)),
                                    NotConfined_HomeRemidy=cbind(c(dhs.2$SH213D.1, dhs.2$SH213D.2, dhs.2$SH213D.3, dhs.2$SH213D.4, dhs.2$SH213D.5,
dhs.2$SH213D.6, dhs.2$SH213D.7, dhs.2$SH213D.8, dhs.2$SH213D.9)),
                                    NotConfined_NotPhilcheckAccredited=cbind(c(dhs.2$SH213E.1, dhs.2$SH213E.2, dhs.2$SH213E.3, dhs.2$SH213E.4,
dhs.2$SH213E.5, dhs.2$SH213E.6, dhs.2$SH213E.7, dhs.2$SH213E.8, dhs.2$SH213E.9)),
                                    NotConfined_NoNeed=cbind(c(dhs.2$SH213F.1, dhs.2$SH213F.2, dhs.2$SH213F.3, dhs.2$SH213F.4, dhs.2$SH213F.5,
dhs.2$SH213F.6, dhs.2$SH213F.7, dhs.2$SH213F.8, dhs.2$SH213F.9)),
                                    NotConfined_Other=cbind(c(dhs.2$SH213X.1, dhs.2$SH213X.2, dhs.2$SH213X.3, dhs.2$SH213X.4, dhs.2$SH213X.5,
dhs.2$SH213X.6, dhs.2$SH213X.7, dhs.2$SH213X.8, dhs.2$SH213X.9)))
HealthSeakingBehaviour_ConfinementHospital <- na.omit(HealthSeakingBehaviour_ConfinementHospital)
mice_plot <- aggr(HealthSeakingBehaviour_ConfinementHospital, col=c('green','red'),
                  numbers=TRUE, sortVars=TRUE,
                  labels=names(HealthSeakingBehaviour_ConfinementHospital), cex.axis=.7,
                  gap=3, ylab=c("Missing data","HealthSeakingBehaviour_ConfinementHospital"))
HealthSeakingBehaviour_ConfinementHospital <- ddply(HealthSeakingBehaviour_ConfinementHospital, "Region", summarise,
                    N_confinement      = length(Region),
                    sum_AdviseConfinement = sum(AdviseConfinement), mean_AdviseConfinement = mean(AdviseConfinement), sd_AdviseConfinement
=sd(AdviseConfinement), se_AdviseConfinement = sd_AdviseConfinement / sqrt(N_confinement),
                    sum_ConfinedInHospital = sum(ConfinedInHospital), mean_ConfinedInHospital = mean(ConfinedInHospital), sd_ConfinedInHospital =
sd(ConfinedInHospital), se_ConfinedInHospital = sd_ConfinedInHospital / sqrt(N_confinement),
                    sum_NotConfined_Distance = sum(NotConfined_Distance), mean_NotConfined_Distance = mean(NotConfined_Distance),
sd_NotConfined_Distance = sd(NotConfined_Distance), se_NotConfined_Distance = sd_NotConfined_Distance / sqrt(N_confinement),
                    sum_NotConfined_NoMoney = sum(NotConfined_NoMoney), mean_NotConfined_NoMoney = mean(NotConfined_NoMoney),
sd_NotConfined_NoMoney = sd(NotConfined_NoMoney), se_NotConfined_NoMoney = sd_NotConfined_NoMoney / sqrt(N_confinement),
                    sum_NotConfined_Costs = sum(NotConfined_Costs), mean_NotConfined_Costs = mean(NotConfined_Costs), sd_NotConfined_Costs =
sd(NotConfined_Costs), se_NotConfined_Costs = sd_NotConfined_Costs / sqrt(N_confinement),
                    sum_NotConfined_HomeRemidy = sum(NotConfined_HomeRemidy), mean_NotConfined_HomeRemidy = mean(NotConfined_HomeRemidy),
sd_NotConfined_HomeRemidy = sd(NotConfined_HomeRemidy), se_NotConfined_HomeRemidy = sd_NotConfined_HomeRemidy / sqrt(N_confinement),
                    sum_NotConfined_NotPhilcheckAccredited = sum(NotConfined_NotPhilcheckAccredited), mean_NotConfined_NotPhilcheckAccredited =
mean(NotConfined_NotPhilcheckAccredited), sd_NotConfined_NotPhilcheckAccredited = sd(NotConfined_NotPhilcheckAccredited),
se_NotConfined_NotPhilcheckAccredited = sd_NotConfined_NotPhilcheckAccredited / sqrt(N_confinement),
                    sum_NotConfined_NoNeed = sum(NotConfined_NoNeed), mean_NotConfined_NoNeed =
mean(NotConfined_NoNeed),sd_NotConfined_NoNeed = sd(NotConfined_NoNeed), se_NotConfined_NoNeed = sd_NotConfined_NoNeed /
sqrt(N_confinement),
                    sum_NotConfined_Other = sum(NotConfined_Other), mean_NotConfined_Other = mean(NotConfined_Other), sd_NotConfined_Other =
sd(NotConfined_Other), se_NotConfined_Other = sd_NotConfined_Other / sqrt(N_confinement))

HealthSeakingBehaviour_FirstConsultation <- data.frame(Region=dhs.2$HV024, PlaceFirstConsult=cbind(c(dhs.2$SH210.1, dhs.2$SH210.2, dhs.2$SH210.3,
dhs.2$SH210.4, dhs.2$SH210.5, dhs.2$SH210.6, dhs.2$SH210.7, dhs.2$SH210.8, dhs.2$SH210.9)),
                                    ReasonForConsult=cbind(c(dhs.2$SH209.1, dhs.2$SH209.2, dhs.2$SH209.3, dhs.2$SH209.4, dhs.2$SH209.5, dhs.2$SH209.6,
dhs.2$SH209.7, dhs.2$SH209.8, dhs.2$SH209.9)),
                                    NoTreatment_Symp_Harmless=cbind(dhs.2$SH408A), NoTreatment_Cost=cbind(dhs.2$SH408B),
NoTreatment_Distance=cbind(dhs.2$SH408C), NoTreatment_Embarassed=cbind(dhs.2$SH408D), NoTreatment_SelfMedication=cbind(dhs.2$SH408E),
NoTreatment_other=cbind(dhs.2$SH408X))
HealthSeakingBehaviour_FirstConsultation <- na.omit(HealthSeakingBehaviour_FirstConsultation)
mice_plot <- aggr(HealthSeakingBehaviour_FirstConsultation, col=c('green','red'),
                  numbers=TRUE, sortVars=TRUE,
                  labels=names(HealthSeakingBehaviour_FirstConsultation), cex.axis=.7,
                  gap=3, ylab=c("Missing data","HealthSeakingBehaviour_FirstConsultation"))
```

```
HealthSeakingBehaviour_FirstConsultation <- ddply(HealthSeakingBehaviour_FirstConsultation, "Region", summarise,
                                N_Consultation    = length(Region),
                                sum_PlaceFirstConsult          =          sum(PlaceFirstConsult),          mean_PlaceFirstConsult=mean(PlaceFirstConsult),
sd_PlaceFirstConsult=sd(PlaceFirstConsult),  se_PlaceFirstConsult = sd_PlaceFirstConsult / sqrt(N_Consultation),
                                sum_ReasonForConsult          =          sum(ReasonForConsult),          mean_ReasonForConsult=mean(ReasonForConsult),
sd_ReasonForConsult=sd(ReasonForConsult),  se_ReasonForConsult = sd_ReasonForConsult / sqrt(N_Consultation),
                                sum_NoTreatment_Symp_Harmless                                =                                sum(NoTreatment_Symp_Harmless),
mean_NoTreatment_Symp_Harmless=mean(NoTreatment_Symp_Harmless),          sd_NoTreatment_Symp_Harmless=sd(NoTreatment_Symp_Harmless),
se_NoTreatment_Symp_Harmless = sd_NoTreatment_Symp_Harmless / sqrt(N_Consultation),
                                sum_NoTreatment_Cost          =          sum(NoTreatment_Cost),          mean_NoTreatment_Cost=mean(NoTreatment_Cost),
sd_NoTreatment_Cost=sd(NoTreatment_Cost),  se_NoTreatment_Cost = sd_NoTreatment_Cost / sqrt(N_Consultation),
                                sum_NoTreatment_Distance    =    sum(NoTreatment_Distance),    mean_NoTreatment_Distance=mean(NoTreatment_Distance),
sd_NoTreatment_Distance=sd(NoTreatment_Distance),  se_NoTreatment_Distance = sd_NoTreatment_Distance / sqrt(N_Consultation),
                                sum_NoTreatment_Embarassed                                =                                sum(NoTreatment_Embarassed),
mean_NoTreatment_Embarassed=mean(NoTreatment_Embarassed),                sd_NoTreatment_Embarassed=sd(NoTreatment_Embarassed),
se_NoTreatment_Embarassed = sd_NoTreatment_Embarassed / sqrt(N_Consultation),
                                sum_NoTreatment_SelfMedication=                                sum(NoTreatment_SelfMedication),
mean_NoTreatment_SelfMedication=mean(NoTreatment_SelfMedication),          sd_NoTreatment_SelfMedication=sd(NoTreatment_SelfMedication),
se_NoTreatment_SelfMedication = sd_NoTreatment_SelfMedication / sqrt(N_Consultation),
                                sum_NoTreatment_other=          sum(NoTreatment_other),          mean_NoTreatment_other=mean(NoTreatment_other),
sd_NoTreatment_other=sd(NoTreatment_other),  se_NoTreatment_other= sd_NoTreatment_other/ sqrt(N_Consultation))

HealthSeakingBehaviour <- merge.data.frame(HealthSeakingBehaviour_ConfinementHospital, HealthSeakingBehaviour_FirstConsultation, by="Region")
remove(HealthSeakingBehaviour_FirstConsultation)
remove(HealthSeakingBehaviour_ConfinementHospital)

HealthSeakingBehaviour_MedicationUsage  <-  data.frame(Region=dhs.2$HV024,  MedicationBoughtWithoutDoctor=cbind(c(dhs.2$SH228.1,  dhs.2$SH228.2,
dhs.2$SH228.3, dhs.2$SH228.4)))
HealthSeakingBehaviour_MedicationUsage <- kNN(HealthSeakingBehaviour_MedicationUsage, variable="MedicationBoughtWithoutDoctor",
                            metric=NULL, k=1)

mice_plot <- aggr(HealthSeakingBehaviour_MedicationUsage, col=c('green','red'),
          numbers=TRUE, sortVars=TRUE,
          labels=names(HealthSeakingBehaviour_MedicationUsage), cex.axis=.7,
          gap=3, ylab=c("Missing data","HealthSeakingBehaviour_MedicationUsage"))
HealthSeakingBehaviour_MedicationUsage <- ddply(HealthSeakingBehaviour_MedicationUsage, "Region", summarise,
                                N_MedicationUsage    = length(Region),
                                sum_MedicationUsage  = sum(MedicationBoughtWithoutDoctor),
                                mean_MedicationUsage = mean(MedicationBoughtWithoutDoctor),
                                sd_MedicationUsage   = sd(MedicationBoughtWithoutDoctor),
                                se_MedicationUsage   = sd_MedicationUsage / sqrt(N_MedicationUsage))

HealthSeakingBehaviour <- merge.data.frame(HealthSeakingBehaviour,HealthSeakingBehaviour_MedicationUsage, By="Region")
remove(HealthSeakingBehaviour_MedicationUsage)

HealthSeakingBehaviour$Region <- as.character(HealthSeakingBehaviour$Region)
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "ARMM"] <- "Autonomous Region of Muslim Mindanao (ARMM)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "Cordillera Admin Region"] <- "Cordillera Administrative Region (CAR)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "National Capital Region"] <- "Metropolitan Manila"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "I - Ilocos Region"] <- "Ilocos Region (Region I)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "II - Cagayan Valley"] <- "Cagayan Valley (Region II)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "III - Central Luzon"] <- "Central Luzon (Region III)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "IVA - CALABARZON"] <- "CALABARZON (Region IV-A)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "IVB - MIMAROPA"] <- "MIMAROPA (Region IV-B)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "IX - Zamboanga Peninsula"] <- "Zamboanga Peninsula (Region IX)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "V - Bicol"] <- "Bicol Region (Region V)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "VI - Western Visayas"] <- "Western Visayas (Region VI)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "VII - Central Visayas"] <- "Central Visayas (Region VII)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "VIII - Eastern Visayas"] <- "Eastern Visayas (Region VIII)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "X - Northern Mindanao"] <- "Northern Mindanao (Region X)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "XI - Davao"] <- "Davao Region (Region XI)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "XII - SOCCSKSARGEN"] <- "SOCCSKSARGEN (Region XII)"
HealthSeakingBehaviour$Region[HealthSeakingBehaviour$Region == "XIII - Caraga"] <- "Caraga (Region XIII)"

######create components "Prevention" for behaviour########
Prevention_GeneralHealth  <-  data.frame(Region=dhs.2$HV024,  Good_Hygiene=cbind(dhs.2$SH301L),  RegularCheckUpAtDocter=cbind(dhs.2$SH301F),
EatFishAndMeat=cbind(dhs.2$SH301I),                EatFruitsAndVegetables=cbind(dhs.2$SH301J),                DrinkWater=cbind(dhs.2$SH301P),
NoThingsDoneToKeepHealty=cbind(dhs.2$SH301Y))

Prevention_GeneralHealth <- kNN(Prevention_GeneralHealth, variable = "Good_Hygiene",metric=NULL, k=6)
Prevention_GeneralHealth <- kNN(Prevention_GeneralHealth, variable = "DrinkWater",metric=NULL, k=6)
Prevention_GeneralHealth <- kNN(Prevention_GeneralHealth, variable = "NoThingsDoneToKeepHealty",metric=NULL, k=6)
Prevention_GeneralHealth <- kNN(Prevention_GeneralHealth, variable = "EatFruitsAndVegetables",metric=NULL, k=6)
Prevention_GeneralHealth <- kNN(Prevention_GeneralHealth, variable = "RegularCheckUpAtDocter",metric=NULL, k=6)
Prevention_GeneralHealth <- kNN(Prevention_GeneralHealth, variable = "EatFishAndMeat",metric=NULL, k=6)

mice_plot <- aggr(Prevention_GeneralHealth, col=c('green','red'),
          numbers=TRUE, sortVars=TRUE,
          labels=names(Prevention_GeneralHealth), cex.axis=.7,
          gap=3, ylab=c("Missing data","Prevention_GeneralHealth"))
Prevention_GeneralHealth <- ddply(Prevention_GeneralHealth, "Region", summarise,
```

```
N_Good_Hygiene=length(Region),sum_Good_Hygiene=sum(Good_Hygiene),mean_Good_Hygiene=mean(Good_Hygiene),sd_Good_Hygiene=sd(Good_Hygiene
),se_Good_Hygiene=sd_Good_Hygiene/sqrt(N_Good_Hygiene),

N_DrinkWater=length(Region),sum_DrinkWater=sum(DrinkWater),mean_DrinkWater=mean(DrinkWater),sd_DrinkWater=sd(DrinkWater),se_DrinkWater=sd_
DrinkWater/sqrt(N_DrinkWater),

N_NoThingsDoneToKeepHealty=length(Region),sum_NoThingsDoneToKeepHealty=sum(NoThingsDoneToKeepHealty),mean_NoThingsDoneToKeepHealty=me
an(NoThingsDoneToKeepHealty),sd_NoThingsDoneToKeepHealty=sd(NoThingsDoneToKeepHealty),se_NoThingsDoneToKeepHealty=sd_NoThingsDoneToKeep
Healty/sqrt(N_NoThingsDoneToKeepHealty),

N_EatFruitsAndVegetables=length(Region),sum_EatFruitsAndVegetables=sum(EatFruitsAndVegetables),mean_EatFruitsAndVegetables=mean(EatFruitsAndVeg
etables),sd_EatFruitsAndVegetables=sd(EatFruitsAndVegetables),se_EatFruitsAndVegetables=sd_EatFruitsAndVegetables/sqrt(N_EatFruitsAndVegetables),

N_RegularCheckUpAtDocter=length(RegularCheckUpAtDocter),sum_RegularCheckUpAtDocter=sum(RegularCheckUpAtDocter),mean_RegularCheckUpAtDocte
r=mean(RegularCheckUpAtDocter),sd_RegularCheckUpAtDocter=sd(RegularCheckUpAtDocter),se_RegularCheckUpAtDocter=sd_RegularCheckUpAtDocter/sqrt
(N_RegularCheckUpAtDocter),

N_EatFishAndMeat=length(Region),sum_EatFishAndMeatr=sum(EatFishAndMeat),mean_EatFishAndMeat=mean(EatFishAndMeat),sd_EatFishAndMeat=sd(Eat
FishAndMeat),se_EatFishAndMeat=sd_EatFishAndMeat/sqrt(N_EatFishAndMeat))

Prevention_GeneralHealth$Region <- as.character(Prevention_GeneralHealth$Region)
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "ARMM"] <- "Autonomous Region of Muslim Mindanao (ARMM)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "Cordillera Admin Region"] <- "Cordillera Administrative Region (CAR)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "National Capital Region"] <- "Metropolitan Manila"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "I - Ilocos Region"] <- "Ilocos Region (Region I)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "II - Cagayan Valley"] <- "Cagayan Valley (Region II)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "III - Central Luzon"] <- "Central Luzon (Region III)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "IVA - CALABARZON"] <- "CALABARZON (Region IV-A)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "IVB - MIMAROPA"] <- "MIMAROPA (Region IV-B)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "IX - Zamboanga Peninsula"] <- "Zamboanga Peninsula (Region IX)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "V - Bicol"] <- "Bicol Region (Region V)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "VI - Western Visayas"] <- "Western Visayas (Region VI)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "VII - Central Visayas"] <- "Central Visayas (Region VII)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "VIII - Eastern Visayas"] <- "Eastern Visayas (Region VIII)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "X - Northern Mindanao"] <- "Northern Mindanao (Region X)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "XI - Davao"] <- "Davao Region (Region XI)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "XII - SOCCSKSARGEN"] <- "SOCCSKSARGEN (Region XII)"
Prevention_GeneralHealth$Region[Prevention_GeneralHealth$Region == "XIII - Caraga"] <- "Caraga (Region XIII)"

#No Data for Subcomponents: Physical Contact Deseased and Contraceptive Usage

Prevention_VaccinationCoverage <-data.frame(Region=measl.vac.cov.9$Region, Vaccination_Coverage=measl.vac.cov.9$Percentage,
KindOfVaccination=measl.vac.cov.9$Vaccination)
mice_plot <- aggr(Prevention_VaccinationCoverage, col=c('green','red'),
        numbers=TRUE, sortVars=TRUE,
        labels=names(Prevention_VaccinationCoverage), cex.axis=.7,
        gap=3, ylab=c("Missing data","Prevention_VaccinationCoverage"))

Prevention_VaccinationCoverage$Region <- as.character(Prevention_VaccinationCoverage$Region)
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "ARMM"] <- "Autonomous Region of Muslim Mindanao (ARMM)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Cordillera Admin Region"] <- "Cordillera Administrative Region (CAR)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "National Capital Region"] <- "Metropolitan Manila"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region I - Ilocos Region"] <- "Ilocos Region (Region I)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region II - Cagayan Valley"] <- "Cagayan Valley (Region II)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region III - Central Luzon"] <- "Central Luzon (Region III)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region IVA - CALABARZON"] <- "CALABARZON (Region IV-A)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region IVB - MIMAROPA"] <- "MIMAROPA (Region IV-B)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region IX - Zamboanga Peninsula"] <- "Zamboanga Peninsula (Region
IX)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region V - Bicol"] <- "Bicol Region (Region V)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region VI - Western Visayas"] <- "Western Visayas (Region VI)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region VII - Central Visayas"] <- "Central Visayas (Region VII)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region VIII - Eastern Visayas"] <- "Eastern Visayas (Region VIII)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region X - Northern Mindanao"] <- "Northern Mindanao (Region X)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region XI - Davao"] <- "Davao Region (Region XI)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region XII - SOCCSKSARGEN"] <- "SOCCSKSARGEN (Region XII)"
Prevention_VaccinationCoverage$Region[Prevention_VaccinationCoverage$Region == "Region XIII - Caraga"] <- "Caraga (Region XIII)"

Prevention <- merge(Prevention_VaccinationCoverage, Prevention_GeneralHealth, by="Region")
remove(Prevention_VaccinationCoverage)
remove(Prevention_GeneralHealth)

######create components "Trust" for behaviour########
Trust_Government <- data.frame(Region=wvs.6$X048WVS, Conf_Government=cbind(wvs.6$E069_11))#complete
Trust_Government <- kNN(Trust_Government, variable = "Conf_Government",metric=NULL, k=4)
mice_plot <- aggr(Trust_Government, col=c('green','red'),
        numbers=TRUE, sortVars=TRUE,
        labels=names(Trust_Government), cex.axis=.7,
        gap=3, ylab=c("Missing data","Trust_Government"))
Trust_Government <- ddply(Trust_Government, "Region", summarise,
```

```
                N_Government    = length(Region),
                sum_Government  = sum(Conf_Government),
                mean_Government = mean(Conf_Government),
                sd_Government   = sd(Conf_Government),
                se_Government   = sd_Government / sqrt(N_Government))

Trust_HealthCare <- data.frame(Region=wvs.6$X048WVS, Conf_HealthCareSystem=cbind(wvs.6$E069_16))
#All NA's --> No data, not continued#

Trust_Press <- data.frame(Region=wvs.6$X048WVS, Conf_Press=cbind(wvs.6$E069_04))#complete
Trust_Press <- kNN(Trust_Press, variable = "Conf_Press", metric=NULL, k=4)
mice_plot <- aggr(Trust_Press, col=c('green','red'),
            numbers=TRUE, sortVars=TRUE,
            labels=names(Trust_Press), cex.axis=.7,
            gap=3, ylab=c("Missing data","Trust_Press"))
Trust_Press <- ddply(Trust_Press, "Region", summarise,
              N_Press    = length(Region),
              sum_Press  = sum(Conf_Press),
              mean_Press = mean(Conf_Press),
              sd_Press   = sd(Conf_Press),
              se_Press   = sd_Press / sqrt(N_Press))

Trust <- merge(Trust_Press, Trust_Government, by="Region")
remove(Trust_Press)
remove(Trust_HealthCare)
remove(Trust_Government)

Trust$Region <- as.character(Trust$Region)
Trust$Region[Trust$Region == "PH: NCR"] <- "Metropolitan Manila"
Trust$Region[Trust$Region == "PH: SOUTH LUZON"] <- "Central Luzon (Region III)"
Trust$Region[Trust$Region == "PH: VISAYAS"] <- "Central Visayas (Region VII)"
Trust$Region[Trust$Region == "PH: MINDANAO"] <- "Davao Region (Region XI)"

##Dengue Prep###
dengue.1$Region <- as.character(dengue.1$Region)
dengue.1$Region[dengue.1$Region == "ARMM"] <- "Autonomous Region of Muslim Mindanao (ARMM)"
dengue.1$Region[dengue.1$Region == "CAR"] <- "Cordillera Administrative Region (CAR)"
dengue.1$Region[dengue.1$Region == "NCR"] <- "Metropolitan Manila"
dengue.1$Region[dengue.1$Region == "Region.I"] <- "Ilocos Region (Region I)"
dengue.1$Region[dengue.1$Region == "Region.II"] <- "Cagayan Valley (Region II)"
dengue.1$Region[dengue.1$Region == "Region.III"] <- "Central Luzon (Region III)"
dengue.1$Region[dengue.1$Region == "Region.IV.A"] <- "CALABARZON (Region IV-A)"
dengue.1$Region[dengue.1$Region == "Region.IV.B"] <- "MIMAROPA (Region IV-B)"
dengue.1$Region[dengue.1$Region == "Region.IX"] <- "Zamboanga Peninsula (Region IX)"
dengue.1$Region[dengue.1$Region == "Region.V"] <- "Bicol Region (Region V)"
dengue.1$Region[dengue.1$Region == "Region.VI"] <- "Western Visayas (Region VI)"
dengue.1$Region[dengue.1$Region == "Region.VII"] <- "Central Visayas (Region VII)"
dengue.1$Region[dengue.1$Region == "Region.VIII"] <- "Eastern Visayas (Region VIII)"
dengue.1$Region[dengue.1$Region == "Region.X"] <- "Northern Mindanao (Region X)"
dengue.1$Region[dengue.1$Region == "Region.XI"] <- "Davao Region (Region XI)"
dengue.1$Region[dengue.1$Region == "Region.XII"] <- "SOCCSKSARGEN (Region XII)"
dengue.1$Region[dengue.1$Region == "CARAGA"] <- "Caraga (Region XIII)"

#Plot the Dengue data to check if the data is loaded correctly#
PlotDengueHemoragicFever<- ggplot(dengue.1, aes(dengue.1$Year, dengue.1$Dengue_Cases))
PlotDengueHemoragicFever<- PlotDengueHemoragicFever+geom_point() +
 labs(x="Year",y="  Dengue Incidence (million)"
PlotDengueHemoragicFever
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/1. PlotDengueHemoragicFever.png', width=685, height=498)
#No Relevant Missing Data
dengue <- ddply(dengue.1, "Region", summarise,
              N_dengue    = length(Region),
              sum_dengue  = sum(Dengue_Cases),
              mean_dengue = mean(Dengue_Cases),
              sd_dengue   = sd(Dengue_Cases),
              se_dengue   = sd_dengue / sqrt(N_dengue))
write.csv(dengue, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/0. Dengue Incidence.csv")


#Merge ALL
BehaviourFullSet1 <- merge.data.frame(Trust, Prevention, by="Region", all.y = TRUE, incomparables = NA)
BehaviourFullSet1 <- kNN(BehaviourFullSet1, variable = c("N_Press", "sum_Press", "mean_Press", "sd_Press", "se_Press",
                                    "N_Government", "sum_Government", "mean_Government", "sd_Government", "se_Government"),
                metric = NULL, k=6)
BehaviourFullSet2 <- merge.data.frame(BehaviourFullSet1, SelfPerceivedHealth, by="Region", all.x = TRUE)
BehaviourFullSet3 <- merge.data.frame(BehaviourFullSet2, HealthSeakingBehaviour, by="Region", all.x = TRUE)
BehaviourFullSet3 <- kNN(BehaviourFullSet3, variable = c("N_Health", "sum_Health", "mean_Health", "sd_Health", "se_Health",
                                    "N_Satisfaction", "sum_Satisfaction", "mean_Satisfaction", "sd_Satisfaction", "se_Satisfaction"),
                            metric = NULL, k=5)
BehaviourFullSet <- merge.data.frame(BehaviourFullSet3, dengue, by="Region", all.x = TRUE)
```

```
write.csv(BehaviourFullSet, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/1. BehaviourFullSet.csv")

remove(BehaviourFullSet1)
remove(BehaviourFullSet2)
remove(BehaviourFullSet3)

#Create a workingset#
BehaviourFinalSet <- data.frame(Dengue_Incidence= BehaviourFullSet$sum_dengue, Subjective_Health=BehaviourFullSet$mean_Health,
Subjective_satisfaction=BehaviourFullSet$mean_Satisfaction,
                  AdviseConfinement=BehaviourFullSet$mean_AdviseConfinement, ConfinedInHospital=BehaviourFullSet$mean_ConfinedInHospital,
NotConfined_Distance=BehaviourFullSet$mean_NotConfined_Distance,
                  NotConfined_NoMoney=BehaviourFullSet$mean_NotConfined_NoMoney, NotConfined_Costs=BehaviourFullSet$mean_NotConfined_Costs,
NotConfined_HomeRemidy=BehaviourFullSet$mean_NotConfined_HomeRemidy,
                  NotConfined_NotPhilcheckAccredited=BehaviourFullSet$mean_NotConfined_NotPhilcheckAccredited,
NotConfined_NoNeed=BehaviourFullSet$mean_NotConfined_NoNeed, NotConfined_Other=BehaviourFullSet$mean_NotConfined_Other,
                  PlaceFirstConsult=BehaviourFullSet$mean_PlaceFirstConsult, ReasonForConsult=BehaviourFullSet$mean_ReasonForConsult,
NoTreatment_Symp_Harmless=BehaviourFullSet$mean_NoTreatment_Symp_Harmless,
                  NoTreatment_Cost=BehaviourFullSet$mean_NoTreatment_Cost, NoTreatment_Distance=BehaviourFullSet$mean_NoTreatment_Distance,
NoTreatment_Embarassed=BehaviourFullSet$mean_NoTreatment_Embarassed,
                  NoTreatment_SelfMedication = BehaviourFullSet$mean_NoTreatment_SelfMedication,
NoTreatment_other=BehaviourFullSet$mean_NoTreatment_other, MedicationUsage=BehaviourFullSet$mean_MedicationUsage,

Vaccination_Coverage=BehaviourFullSet$Vaccination_Coverage,press=BehaviourFullSet$sum_Press,Government=BehaviourFullSet$sum_Government,
                  Good_Hygiene=BehaviourFullSet$mean_Good_Hygiene, DrinkPlentyWater=BehaviourFullSet$mean_DrinkWater,
NoThingsDoneToKeepHealty=BehaviourFullSet$mean_NoThingsDoneToKeepHealty, EatFruitsAndVegetables=BehaviourFullSet$mean_EatFruitsAndVegetables,
                  RegularCheckUpAtDocter=BehaviourFullSet$mean_RegularCheckUpAtDocter, EatFishAndMeat=BehaviourFullSet$mean_EatFishAndMeat,
                  Region=BehaviourFullSet$Region)

write.csv(BehaviourFinalSet, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/2. BehaviourFinalSet.csv")

remove(BehaviourFullSet)

remove(HealthSeakingBehaviour)
remove(SelfPerceivedHealth)
remove(Prevention)
remove(Trust)
remove(dengue)

remove(measl.vac.cov.9)
remove(dhs.2)
remove(wvs.6)
remove(dengue.1)

################################################## VIEW AND CHECK DATASET ##################################################

### scatterplots ###
splom(BehaviourFinalSet[c(1:31)], data=BehaviourFinalSet)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/2. Scatterplot All.png', width=685, height=498)

splom(BehaviourFinalSet[c(1:5)], data=BehaviourFinalSet)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/2. Scatterplot 2-5.png', width=685, height=498)

splom(BehaviourFinalSet[c(6:10)], data=BehaviourFinalSet)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/2. Scatterplot 6-10.png', width=685, height=498)

splom(BehaviourFinalSet[c(11:15)], data=BehaviourFinalSet)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/2. Scatterplot 11-15.png', width=685, height=498)

splom(BehaviourFinalSet[c(16:20)], data=BehaviourFinalSet)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/2. Scatterplot 16-20.png', width=685, height=498)

splom(BehaviourFinalSet[c(21:24)], data=BehaviourFinalSet)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/2. Scatterplot 21-25.png', width=685, height=498)

splom(BehaviourFinalSet[c(25:31)], data=BehaviourFinalSet)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/2. Scatterplot 21-25.png', width=685, height=498)

### check for outliers by boxplot & histrogram ###
boxplots <- for(i in c(1:31)){boxplot(BehaviourFinalSet[, i], main = colnames(BehaviourFinalSet[i]))}
#Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/3. Bosplot.png', width=685, height=498)

histograms <- for(i in c(1:31)){hist(BehaviourFinalSet[, i], main = colnames(BehaviourFinalSet[i]))}
#Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/4. Histogram.png', width=685, height=498)

######################################## Analysis of data  ####################################################
rownames(BehaviourFinalSet) <- BehaviourFinalSet$Region

### normalisation min-max ###
BehaviourFinalSet_normalized <- BehaviourFinalSet
for(i in 2:30){BehaviourFinalSet_normalized[,i] <- ((BehaviourFinalSet_normalized[,i]-min(BehaviourFinalSet_normalized[,i], na.rm =
TRUE))/(max(BehaviourFinalSet_normalized[,i], na.rm = TRUE)-min(BehaviourFinalSet_normalized[,i], na.rm = TRUE)))}
```

```
write.csv(BehaviourFinalSet_normalized, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/3. BehaviourFinalSet Normalized.csv")

### multicollinearity correlation matrix ###
#correlation matrix
BehaviourFinalSet_normalized_corr <- BehaviourFinalSet_normalized[-c(1,31)]
correlation_matrix <- cor(BehaviourFinalSet_normalized_corr, use = "everything")
write.csv(correlation_matrix, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/4. Correlation_matrix.csv")
corrplot(correlation_matrix, method = "circle", type = "upper", tl.col= 'darkgrey',
      order = "AOE", tl.pos  = 'lt' , tl.cex = .3, res=1200)
corrplot.mixed(correlation_matrix, lower.col = "black", number.cex=.30, title= "Correlation Matrix", tl.col= 'darkgrey',
          order = "AOE", tl.pos  = 'lt' , tl.cex = .40, res=1200)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/5. CorrelationMatrix.png', width=685, height=498)

remove(BehaviourFinalSet_normalized_corr)

### VIF Scoring for selection ###

### VIF score   - First Run ###
VIF_Data_SelfPerceivedHealth <- lm(BehaviourFinalSet_normalized[c(1:3)])
VIF_Score_SelfPerceivedHealth <- vif(VIF_Data_SelfPerceivedHealth)
View(VIF_Score_SelfPerceivedHealth)
#Delete Collum 3 due to high correlation
remove(VIF_Data_SelfPerceivedHealth)
remove(VIF_Score_SelfPerceivedHealth)

VIF_Data_HealthSeekingBehaviour <- lm(BehaviourFinalSet_normalized[c(1, 4, 8, 10, 12:13, 16:20)])
VIF_Score_HealthSeekingBehaviour <- vif(VIF_Data_HealthSeekingBehaviour)
View(VIF_Score_HealthSeekingBehaviour)
#Delete colum 14, 6, 5, 8, 9, 11, 15 due to high correlation
remove(VIF_Data_HealthSeekingBehaviour)
remove(VIF_Score_HealthSeekingBehaviour)

VIF_Data_Prevention <- lm(BehaviourFinalSet_normalized[c(1,22,25:27, 29:30)])
VIF_Score_Prevention <- vif(VIF_Data_Prevention)
View(VIF_Score_Prevention)
#Delete Colum 28 due to high correlation
remove(VIF_Data_Prevention)
remove(VIF_Score_Prevention)

VIF_Data_Trust <- lm(BehaviourFinalSet_normalized[c(1,23:24)])
VIF_Score_Trust <- vif(VIF_Data_Trust)
View(VIF_Score_Trust)
#Keep All colums
remove(VIF_Data_Trust)
remove(VIF_Score_Trust)

### VIF score   - Run selfperceived health, trust and prevention ###
VIF_Data_combined <- lm(BehaviourFinalSet_normalized[c(1,2,21,22,24:25, 27, 29:30)])
VIF_Score_combined <- vif(VIF_Data_combined)
View(VIF_Score_combined)
#Delete colum 23, 26 due to high correlation
remove(VIF_Data_combined)
remove(VIF_Score_combined)

### VIF score   - Run Selection ###
VIF_Data_All <- lm(BehaviourFinalSet_normalized[c(1,2,12:13,16:17,20:22,24:25,27,29:30)])
VIF_Score_All <- vif(VIF_Data_All)
View(VIF_Score_All)
#Delete Colum 18, 4, 23, 10, 19, 8, 17
remove(VIF_Data_All)
remove(VIF_Score_All)

BehaviourFinalSet_Normalized_VIFAdjusted <- BehaviourFinalSet_normalized[c(1,2,12:13,16,20:22,24:25,27,29:31)]
write.csv(BehaviourFinalSet_Normalized_VIFAdjusted, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/5. BehaviourFinalSet_Normalized_VIFAdjusted.csv")

remove(BehaviourFinalSet_normalized)
remove(BehaviourFinalSet)

rownames(BehaviourFinalSet_Normalized_VIFAdjusted) <- BehaviourFinalSet_Normalized_VIFAdjusted$Region

###### Dataset based on VIF Score  ##########

BehaviourFinalSet_normalized_corr <- BehaviourFinalSet_Normalized_VIFAdjusted[-c(1,14)]
correlation_matrix <- cor(BehaviourFinalSet_normalized_corr, use = "everything")
write.csv(correlation_matrix, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/6. Correlation_matrix VIF Adjusted.csv")
corrplot(correlation_matrix, method = "circle", type = "upper", tl.col= 'darkgrey',
      order = "AOE", tl.pos  = 'lt' , tl.cex = .3, res=1200)
corrplot.mixed(correlation_matrix, lower.col = "black", number.cex=.45, title= "Correlation Matrix VIF Adjusted", tl.col= 'darkgrey',
          order = "AOE", tl.pos  = 'lt' , tl.cex = .50, res=1200)
Object <- dev.print(png, 'D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/4. Plots/6. CorrelationMatrix VIF Adjusted.png', width=685, height=498)
```

```
remove(BehaviourFinalSet_normalized_corr)

################ Regression  ####################
#Poission Regression#
#Behaviour_poisson <- as.vector(BehaviourFinalSet_Normalized_VIFAdjusted[,-15])
#Behaviour_poisson <- stepAIC(glm(BehaviourFinalSet_Normalized_VIFAdjusted$Dengue_Incidence ~ .,family=poisson(link = "log") ,data= Behaviour_poisson))
#summary(Behaviour_poisson)
#predict(Behaviour_poisson,BehaviourFinalSet_Normalized_VIFAdjusted[c(1:17),], type = "response")



#logistic regression#
Behaviour_Logistic <- BehaviourFinalSet_Normalized_VIFAdjusted[,-14]
normalit<-function(m){
 (m - min(m))/(max(m)-min(m))}
Behaviour_Logistic <- apply(Behaviour_Logistic,2,normalit)
Behaviour_Logistic <- as.data.frame(Behaviour_Logistic)

Behaviour_Logistic1 <- step(glm(Dengue_Incidence ~ .,family=binomial(link='logit'),data=Behaviour_Logistic),direction = "forward")
summary(Behaviour_Logistic1)

Predict_Logistic <- predict(Behaviour_Logistic1, BehaviourFinalSet_Normalized_VIFAdjusted[c(1:17),],  type="response")
Predict_Logistic <- abs(Predict_Logistic)
Predict_Logistic <- as.data.frame(Predict_Logistic)
Predict_Logistic <- apply(Predict_Logistic, 2, normalit)
View(Predict_Logistic)

write.csv(Predict_Logistic, file = "D:/CloudStation Carla en Robert/Carla/Master Thesis/3. Thesis/2. Data/7. Predictive Values Logistic.csv")

#Continue in QGIS 2.18.18, for visualisation of predictive value on a map
```

Appendix V Indicators

Table 7: List of all selected indicators before the statistical analysis.

Risk	Dimension	Category	Component	Subcomponents	Indicator Nr.	Indicator	Data Source	Level	Year
Epidemics Risk and Priority	Vulnerability	Behaviour	Self Perceived Health	Subjective state of health	1	State of health (subjective)	WVS	regional	1981-2014
				Subjective Satisfaction	2	Satisfaction with your life	WVS	regional	1981-2014
			Health Seeking Behaviour	Advise taking behaviour	3	Was not confined in hospital because of no need	DHS	regional/provincial	2013
					4	Was not confined in hospital because of home remedy	DHS	regional/provincial	2013
				Consultation seeking	5	Place of first consultation/advice or treatment	DHS	regional/provincial	2013
					6	Reason for visiting health facility	DHS	regional/provincial	2013
					7	Refusal of treatment after consultat doctor because: Symptoms (found to be) harmless	DHS	regional/provincial	2013
					8	Refusal of treatment after consultat doctor because: Cost	DHS	regional/provincial	2013
					9	Refusal of treatment after consultat doctor because: Distance	DHS	regional/provincial	2013
					10	Refusal of treatment after consultat doctor because: Embarrassed	DHS	regional/provincial	2013
				Medication Seeking	11	Refusal of treatment after consultat doctor because: Self-medication	DHS	regional/provincial	2013
					12	Refusal of treatment after consultat doctor because: Medicines bought apart from (medicine / services in) the hospital	DHS	regional/provincial	2013
			Prevention	Contraceptives	13	Contraceptive prevalence	DHS	regional/provincial	2013
				General Health	14	Thinks done to keep healthy: Maintain good hygiene	DHS	regional/provincial	2013
					15	Thinks done to keep healthy: Check-up by docter	DHS	regional/provincial	2013
					16	Thinks done to keep healthy: Aet fish/lean meat/poultry/ soy beans	DHS	regional/provincial	2013
					17	Thinks done to keep healthy: Eat plenty of fruits/vegetables/root crops	DHS	regional/provincial	2013
					18	Thinks done to keep healthy: Drink plenty of water	DHS	regional/provincial	2013
					19	Thinks done to keep healthy: None	DHS	regional/provincial	2013
				Physical contact with diseased	20	Physiscal Contact with deseased	No Data		
				Vaccination Degree	21	Willingness to vaccinate	DHS	regional/provincial	2013
			Trust	Government	22	Trust in the government	WVS	regional	1981-2014
				Health Care	23	Trust in Health care system	WVS	regional	1981-2014
				Press	24	Trust in The press	WVS	regional	1981-2014

Table 8: List of indicators with inclusion and exclusion argumentation and statistical analysis

Risk	Dimension	Category	Component	Subcomponents	Indicator Nr.	Indicator	Exclusion / inclusion in model
Epidemics Risk and Priority	Vulnerability	Behaviour	Self Perceived Health	Subjective state of health	1	State of health (subjective)	Included in EPI model
				Subjective Satisfaction	2	Satisfaction with your life	Excluded because of correlation to state of health (subjective)
			Health Seeking Behaviour	Advise taking behaviour	3	Was not confined in hospital because of no need	Included in EPI model
					4	Was not confined in hospital because of home remedy	Excluded because of correlation to reason to refuse treatment, medice bought apart from hospital
				Consultation seeking	5	Place of first consultation/advice or treatment	Included in EPI model
					6	Reason for visiting health facility	Excluded because of correlation to reasons for refusal
					7	Refusal of treatment after consultat doctor because: Symptoms (found to be) harmless	Included in EPI model
					8	Refusal of treatment after consultat doctor because: Cost	Included in EPI model
					9	Refusal of treatment after consultat doctor because: Distance	Excluded because of correlation to reason to refuse treatment
					10	Refusal of treatment after consultat doctor because: Embarrassed	Excluded because of correlation to reason to refuse treatment
				Medication Seeking	11	Refusal of treatment after consultat doctor because: Self-medication	Excluded because of correlation to reason to refuse treatment
					12	Refusal of treatment after consultat doctor because: Medicines bought apart from (medicine / services in) the hospital	Included in EPI model
			Prevention	Contraceptives	13	Contraceptive prevalence	Excluded because of correlation to other preventive measures
				General Health	14	Thinks done to keep healthy: Maintain good hygiene	Included in EPI model
					15	Thinks done to keep healthy: Check-up by docter	Included in EPI model
					16	Thinks done to keep healthy: Aet fish/lean meat/poultry/ soy beans	Included in EPI model
					17	Thinks done to keep healthy: Eat plenty of fruits/vegetables/root crops	Excluded because of correlation to trust in government
					18	Thinks done to keep healthy: Drink plenty of water	Excluded because of correlation to other preventive measures
					19	Thinks done to keep healthy: None	Included in EPI model
				Physical contact with diseased	20	Physiscal Contact with deseased	Excluded because of correlation to other preventive measures
				Vaccination Degree	21	Willingness to vaccinate	Included in EPI model
			Trust	Government	22	Trust in the government	Included in EPI model
				Health Care	23	Trust in Health care system	Excluded because of no data in survey, survey could be extended
				Press	24	Trust in The press	Excluded because of correlation to trust in government

Table 9: Final list of indicators for EPI framework

Risk	Dimension	Category	Component	Subcomponents	Indicator Nr.	Indicator	Exclusion / inclusion	Data Source
Epidemics Risk and Priority	Vulnerability	Behaviour	Self Perceived Health	Subjective state of health	1	State of health (subjective)	Included in EPI model	WVS
			Health Seeking Behaviour	Advise taking behaviour	3	Was not confined in hospital because of no need	Included in EPI model	DHS
				Consultation seeking	5	Place of first consultation/advice or treatment	Included in EPI model	DHS
					7	Refusal of treatment after consultat doctor because: Symptoms (found to be) harmless	Included in EPI model	DHS
					8	Refusal of treatment after consultat doctor because: Cost	Included in EPI model	DHS
				Medication Seeking	12	Refusal of treatment after consultat doctor because: Medicines bought apart from (medicine / services in) the hospital	Included in EPI model	DHS
			Prevention	General Health	14	Thinks done to keep healthy: Maintain good hygiene	Included in EPI model	DHS
					15	Thinks done to keep healthy: Check-up by docter	Included in EPI model	DHS
					16	Thinks done to keep healthy: Aet fish/lean meat/poultry/ soy beans	Included in EPI model	DHS
					19	Thinks done to keep healthy: None	Included in EPI model	DHS
				Vaccination Degree	21	Vaccination coverage (measles as proxy for willingness)	Included in EPI model	DHS
			Trust	Government	22	Trust in the government	Included in EPI model	WVS

Appendix VI Disaster and epidemic Indicators

Table 10: Disaster indicators with open data source

Source	Year	Level (national, regional)	Variables included	Type of data	URL
EM-DAT history of natural disasters	1915-ongoing	national/regional	Total damage of disaster	numerical	http://www.emdat.be/database
EM-DAT history of natural disasters	1915-ongoing	national/regional	People affected by disaster	numerical	http://www.emdat.be/database
EM-DAT history of natural disasters	1915-ongoing	national/regional	Nr. of deaths caused by disaster	numerical	http://www.emdat.be/database
EM-DAT history of natural disasters	1915-ongoing	national/regional	Type of disaster (natural, biological)	categorical	http://www.emdat.be/database
EM-DAT history of natural disasters	1915-ongoing	national/regional	Subtype of disaster	categorical	http://www.emdat.be/database
EM-DAT history of natural disasters	1915-ongoing	national/regional	Place where disaster occurred	geographical	http://www.emdat.be/database
EM-DAT history of natural disasters	1915-ongoing	national/regional	Name of disaster	name	http://www.emdat.be/database
EM-DAT history of natural disasters	1915-ongoing	national/regional	Type of epidemic	categorical	http://www.emdat.be/database

Table 11: Dengue incidence with open data source

Source	Year	Level (national, regional)	Variables included	Type of data	URL
Kaggle Dengue	2008-2016	Regional	The recorded number of dengue cases per 100,000 population per region	numerical	https://www.kaggle.com/grosvenpaul/dengue-cases-in-the-philippines
DHS Philippines Standard 2013	2013	regional/provincial	Illness: Dengue fever		https://dhsprogram.com/data/dataset/Philippines_Standard-DHS_2013.cfm?flag=0 (household recode)