Healthy communities - Understanding determinants of HIV

12 December 2014

1 Introduction

The expiry date of the Millennium Development Goals (MDGs) is just around the corner, meanwhile the post-2015 agenda is being discussed intensively. In this context, it is important to assess the achievement of the MDGs and try to understand why some goals have not been reached.

Reducing HIV prevalence is an important aim of the MDGs. Target 6.A of the MDGs specifies that countries should "have halted by 2015 and begun to reverse the spread of HIV/AIDS" (United Nations 2014). In most regions of the world this goal has been fulfilled: new HIV infections declined and the overall number of new HIV/AIDS infections per 100 adults (15-49 years old) decreased by 44 per cent between 2001 and 2012 (United Nations 2014). However, this trend cannot be observed in all 189 member states of the United Nations. On the contrary, HIV/AIDS prevalence has even increased in some countries.

2 Aim, Research Question and Hypotheses

This paper aims to provide evidence to assess why some countries struggle to achieve MDG 6A. We believe that one possible explanation for the failure of some interventions in reducing HIV/AIDS may lie in the lack of a full understanding of the determinants of the disease, which can in turn lead to ill-specified interventions and wrongly targeted campaigns.

The literature reviewed for this paper identifies a myriad of determinants of health. However, there is a gap in the literature when it comes to the determinants of specific diseases. The aim of this paper is to test the applicability of one of the most commonly used theories of determinants of health, to explain the evolution of HIV/AIDS rates. The first hypothesis of this paper is that all general determinants of health (as identified by our reference model) are determinants of HIV/AIDS. By identifying variables that help to explain HIV/AIDS incidence and prevalence, this paper will help move forward the discussion of the determinants of HIV/AIDS .

Furthermore, given that the literature identifies interlinkages between the different determinants of health (see Solar and Irwin 2010), this paper will evaluate to what extent those linkages are reflected on the micro level. The second hypothesis of this paper is that socioeconomic, environmental and cultural factors can be used to explain individual lifestyle factors.

3 Literature Review

Hurrelmann (Hurrelmann 1989, 76) advocates an interdisciplinary framework for analysing what determines health outcomes. He considers it necessary to use a model that integrates all the aspects of the organism, individual and the environment.

One framework that shows the interaction between individual and environmental factors over time is the salutogenic model developed by Antonovsky in 1979. According to Hurrelmann, Antonovsky's model is a great contribution to interdisciplinary theory, but the downside is its complexity (Hurrelmann 1989).

A simpler and more common model on the main determinants of health is the "rainbow model", developed by Dahlgren and Whitehead (Dahlgren and Whitehead 1991, 11). This model gives an overview of the main health determinants, reflecting the relationship between the individual, its environment and different health outcomes. Individuals are at the centre of the model with a set of fixed biological and genetical preconditions. Building upon these, four layers of influence on health can be identified: individual lifestyle factors, social and community networks, living and working conditions and general socio-economic, cultural and environmental conditions.

Figure 1: Main health determinants
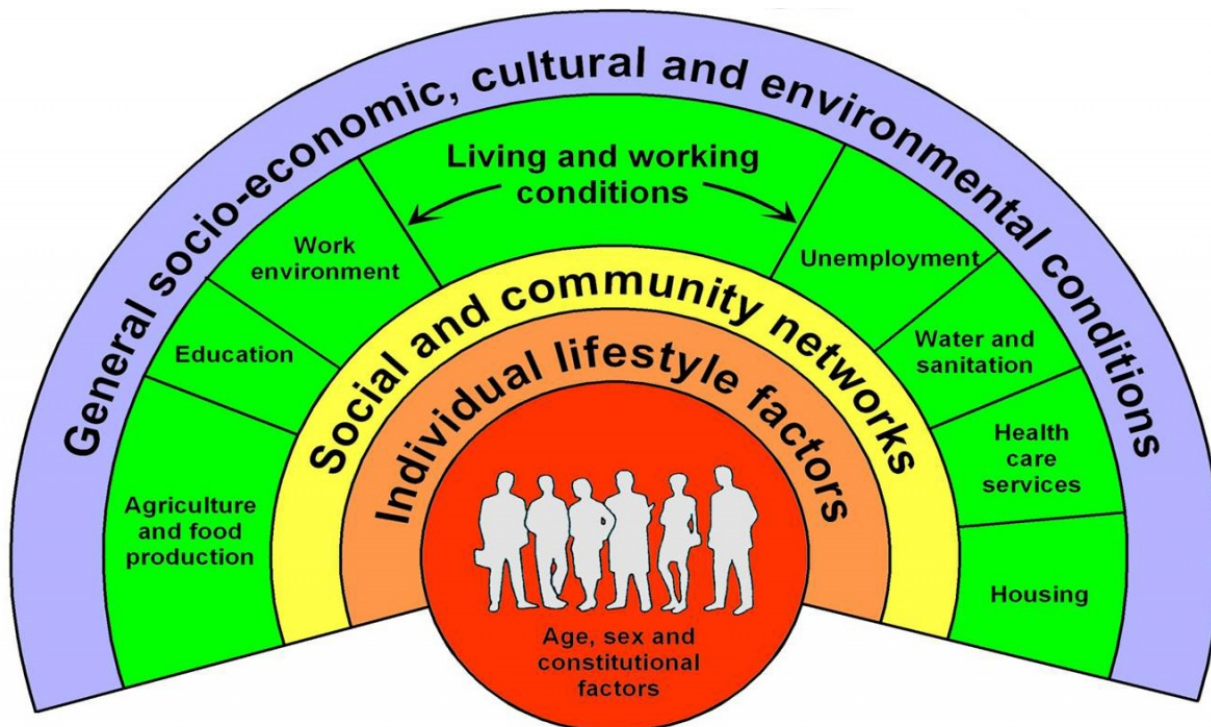
Source: Dahlgren and Whitehead, 1991

4   Data Sources

For this research, two main sources of data were explored. Firstly, databases comprising information for several countries were examined in order to increase comparability of the data. Among the most commonly used databases in the field of this research, it is worth mentioning those from UNAIDS, World Bank, Global Fund for AIDS, Tuberculosis and Malaria, WHO, the Institute for Health Metrics and Evaluation, PEPFA and the AIDS Data Hub. Secondly, country-specific data was used to check the trends of some of the most important variables of this research and to enable micro level analysis. These databases include those from the National Bureau of Statistics (from each country under analysis), Demographic Health Surveys (DHS) and the AIDS Indicator Survey (AIS), obtained from the United States Agency for International Development (USAID).

4.1   World Development Indicators (WDI)

The WDI database comprises 1342 indicators clustered in 10 thematic areas that range from health and education to infrastructure and public sector data. Information is available for 214 countries and dates back

to 1960. All indicators are available for free at the World Bank website and can be downloaded as an Excel sheet, CSV, tabbed TXT or SDMX. In addition, there is a special R package WDI designed to download and use the data.

WDI have been used in a wide range of fields and HIV/AIDS research is not an exception. For examples of relevant literature that also make use of WDI please see (Haacker 2002), (Talbott 2007) and (Kalemli-Ozcan 2011). A list of all WDI indicators used for this research can be found in the Appendix.

4.2 Demographic Health Surveys (DHS) and AIDS Indicator Survey (AIS)

DHS have a high reputation for collecting accurate and nationally representative household data (World Bank 2014). Relevant HIV/AIDS literature used the DHS data like (Haacker 2002) and (World Health Organization and UNICEF 2014). For the case of Vietnam three datasets for the years 2005, 2002 and 1997 are available in USAID's DHS database. The most recent available DHS dataset will be used for the micro level analysis of this research. The DHS was conducted by the General Statistical Office National Institute for Hygiene and Epidemiology between September and December 2005. There are three core questionnaires in DHS surveys: a household, a women's and a men's questionnaire. The DHS sample for Vietnam counts 6337 households, including 7289 observations for female respondents (women aged 15 to 49) and 6707 observations for male respondents (men aged 15 to 49).

As an extension of the DHS data, the micro analysis also integrates the AIDS Indicator Survey (AIS) for Vietnam, again being conducted in 2005. The AIS provides nationally representative estimates of HIV rates, by collecting blood from representative samples of the population of both men and women in a country. The AIS for Vietnam includes 1675 households. In all households, women and men aged 15-49 are eligible to participate in the survey. The linkage of DHS' HIV test results to the full DHS survey record (without personal identifiers) allows for an in-depth analysis of the socio-demographic and behavioural factors associated with HIV prevalence.

Both datasets can be downloaded as Stata System File, Flat Data, SAS System File and SPSS System File. Individual and AIS datasets can also be downloaded as a hierarchical file format.

It is worth mentioning that special permission was granted in downloading the DHS and AIS databases, representing a potential challenge for the reproducibility of our research.

---

5 Data Gathering and Cleaning

This section focuses on the process of gathering the data and cleaning the databases to prepare the variables for the data analysis.

The first step in this process was uploading the databases to R Studio. The first dataset consists of 29 World Development Indicators and it was downloaded from World Bank's website. These indicators represent the independent variables used for this research plus the population indicator that is used to filter small countries. Provided that the focus of this research is on country level data, all regional data was dropped. Further, 169 rows that contained only NA values were deleted.

After dropping empty rows, the data frame was alphabetically (ascending) ordered, rows were grouped by iso2c code and variables were renamed.

The dataset was further cleaned preparing the data for imputation using the AMELIA package. The imputation will be conducted however at a further stage of the research. This process requires that the panel is as balanced as possible, as it feeds from all variables to predict values for the missing observations. The next step was thus dropping variables for which more than 80% of the observations (552) were missing. In addition, countries with a population smaller than one million inhabitants were dropped from the database. 59 countries

fell in that category: 46 islands, 5 European countries (Andorra, Liechtenstein, Luxemburg, Monaco and Montenegro), Bahrain, Bhutan, Belize, Djibouti, Equatorial Guinea, Guyana, Qatar and Suriname. Dropping these countries does not affect the research as the remaining database still contains a highly heterogeneous sample both in geographic and socio-economic terms. Furthermore, deleting these countries improves the dataset as most of these countries lacked information for most of the studied variables.

The second database used for this research was downloaded from UNAIDS' website and it provides information on HIV/AIDS incidence rates (as well as prevalence and deaths caused by HIV/AIDS). The data is publicly available. All columns except the country and the incidence rate were dropped. After renaming the variables, a unique identifier was created and missing values were recoded as NAs. Moreover, some observations in the database were not specific numbers; instead, it was indicated that for that year, prevalence was below a certain threshold (0.01%). In those cases, these observations were replaced by 0.009. The final step in the cleaning of the UNAIDS database consisted of deleting missing values for the dependent variable and deleting the regions with an iso2c equal to a country's iso2c (NA and ZA) to avoid problems in the merging process.

Once both databases were cleaned, the next step was to merge the datasets using the combination of iso2c and year as unique identifier. In the merging process, only observations that were present in both datasets were kept. It is worth noticing that UNAIDS' dataset included observations from 1990 to 2012 so all observation between 1990 and 1999 were dropped. Finally, unnecessary columns from the new database were eliminated.

6 Descriptive Statistics

The descriptive statistics part consists of analysing the main variables of interest and preparing them for the data analysis. Tables, plots and histograms are shown to understand the distribution of the variables.

Figure 2 shows that in most countries of our dataset HIV/AIDS incidence rates decreased between the period of 2000 to 2015 (see Figure 2). The blue line of Figure 2 is however only representing the general downwards sloping trend of HIV/AIDS incidence over time in our sample. The black dots in the above the blue line show that there are still some outliers to this trend.

**Figure 2: HIV/AIDS Incidence Rate over Time**
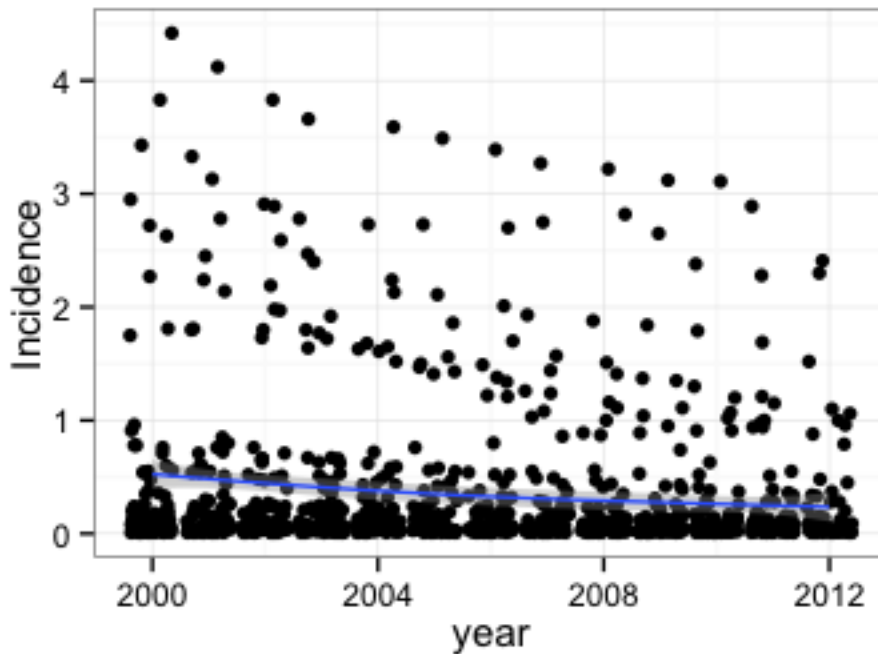
From Figures 3 and 4 (see Appendix) it can be seen which countries are not following this general downward sloping trend. Figure 3 plots the range of observations on HIV/AIDS incidence per country. The figure shows that most countries have low HIV/AIDS incidence rates slightly above 0, but some outliers (countries with high incidence rates) can be identified.

Figure 4 is showing the direction of the change of the HIV/AIDS incidence rate compared to the previous year by country. To analyse the change of the incidence rate from one year to the following year the incidence variable was lagged by one year and a new variable, calculating the difference between the lagged and the original incidence variable was created.

When looking at the distribution of the HIV/AIDS incidence variable in a histogram it becomes apparent that the incidence variable is indeed highly skewed to the left and only few incidence rates are higher than 1.

Figure 5: Distribution of HIV/AIDS Incidence Rate
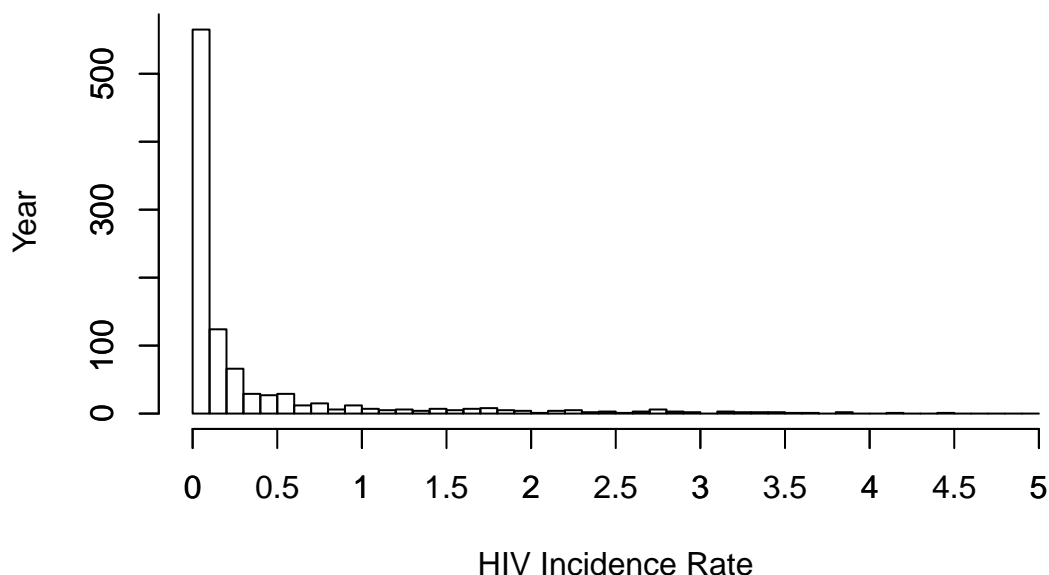
In order to deal with the skewness of this variable, a dummy variable, differentiating between countries with a high HIV incidence and countries with a low HIV incidence is created for the first empirical model. The most accurate measure of the central tendency of a skewed distribution is the median. Therefore, the dependent variable was coded as 1 if the HIV incidence rate is below the median and a value of 0 was assigned to countries with an HIV incidence rate above the median.

Figure 6: Dummy Variable for High and Low HIV/AIDS Incidence

HIV/AIDS Incidence above & below the Median
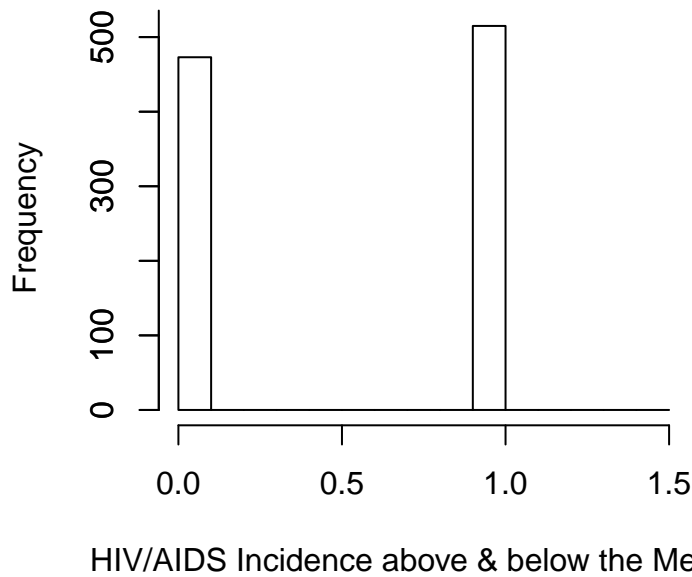
For the second empirical model the sample was reduced and focuses only on countries with a high HIV/AIDS incidence in order to explore, whether the determinants identified in Model 1 hold true when zooming in on the more problematic cases. As the dependent variable was still skewed to the left after restricting the sample to those countries above the mean a log transformation was necessary to approach a normal distribution.

Figure 6: Logged HIV/AIDS Incidence Rate for Countries lying above the Median

HIV/AIDS Incidence Rate
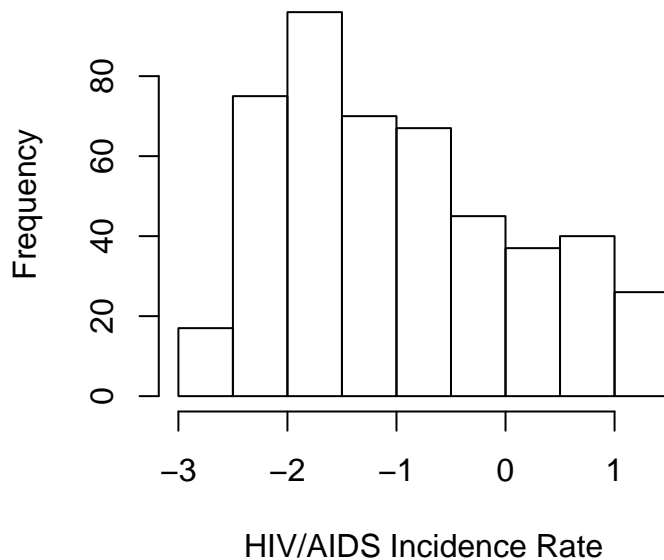
Scatterplots were created for all independent variables by the categories of Dahlgrehn's model in order to see whether further relevant variables are skewed (see Appendix). Due to a high skewness in most of the variables and to ensure a better comparability of the variable units all but one of the independent variables in the sample were logged.

The only variable that was not logged is a variable comparing the share of female unemployment to total unemployment. The distribution of this variable comes already close to a normal distribution, so a log transformation is not urgently necessary. Further, it has some values of zero, which would make it difficult to log the variable.

7 Case Studies - Botswana, Lesotho, Uganda and Malawi

In the following, four case studies are presented to develop a better understanding of best practices and bad practices among countries with high changes in Incidence rates between 2000 and 2012. The cases were selected by the "extreme case selection method" on the basis of peculiar positive or negative changes in incidence rates over time (see Figure 4 in Appendix) (Gerring 2008).

Figure 8 shows the two best practices (green colour), namely Botswana and Malawi and the two bad practices (red colour), Uganda and Lesotho that were selected. In Figure 9 it

Figure 8: Interesting Cases for Change in HIV Incidence Rates
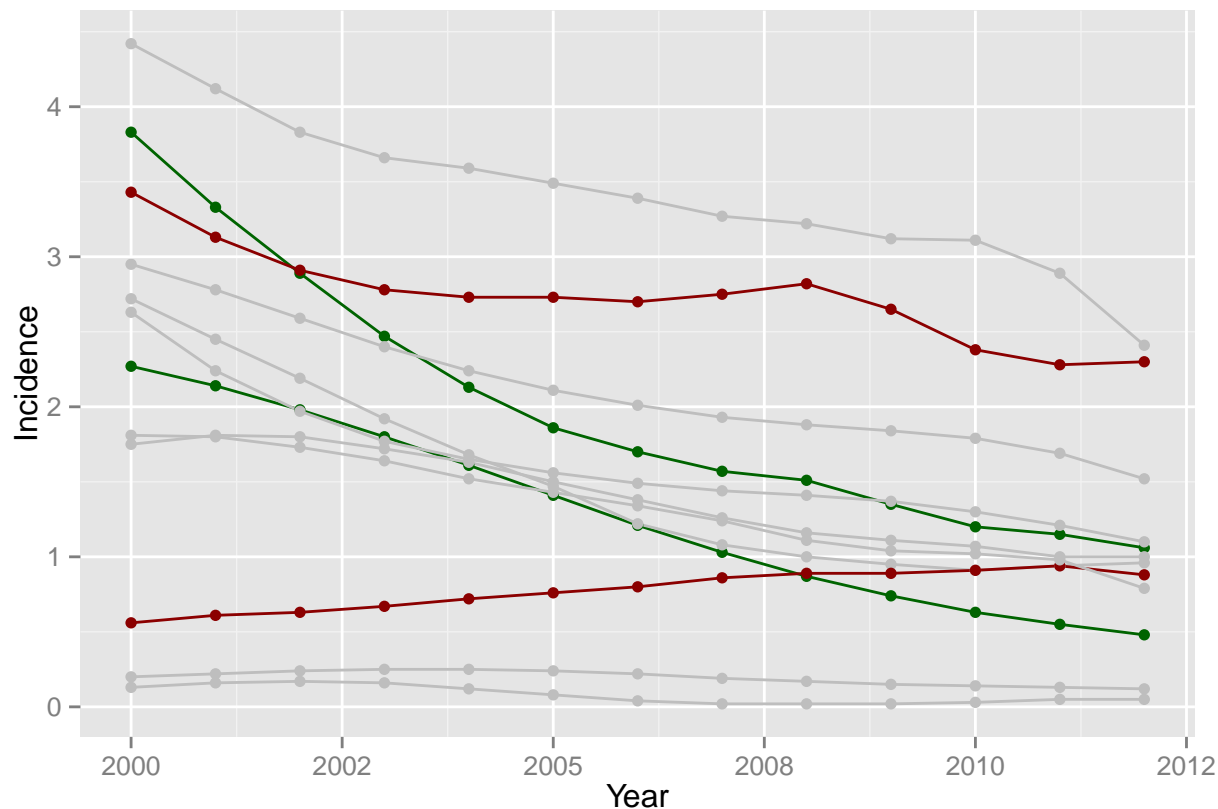
7.1 Case Studies
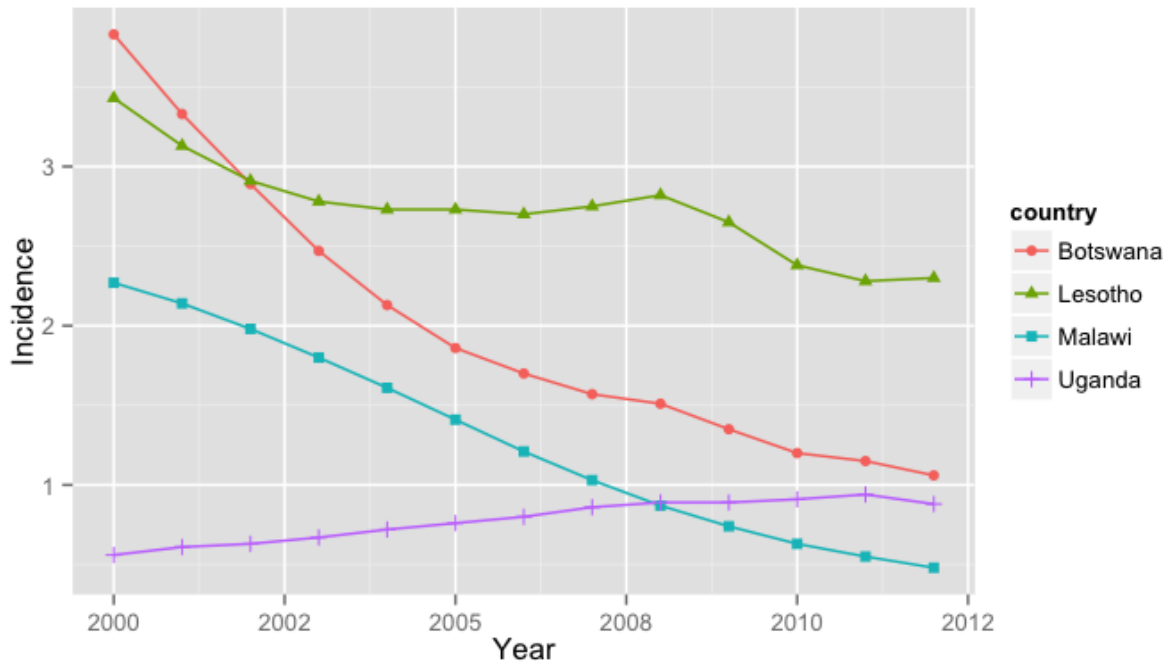
7.2 Female Schooling in Selected Countries

7.3 Female Unemployment in Selected Countries

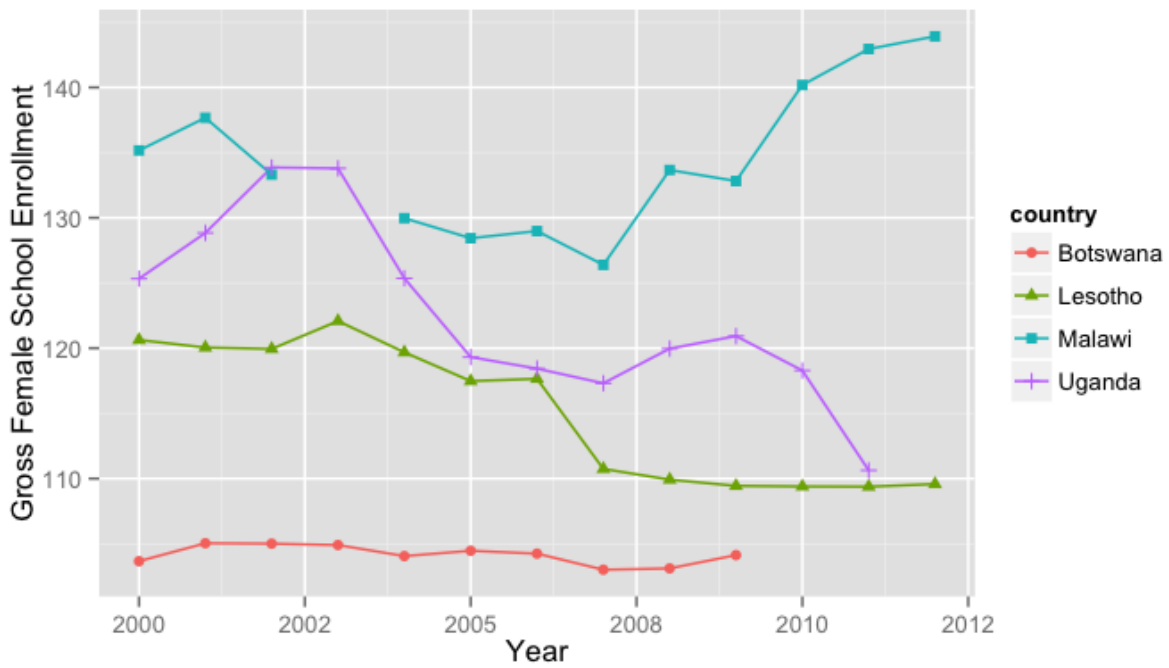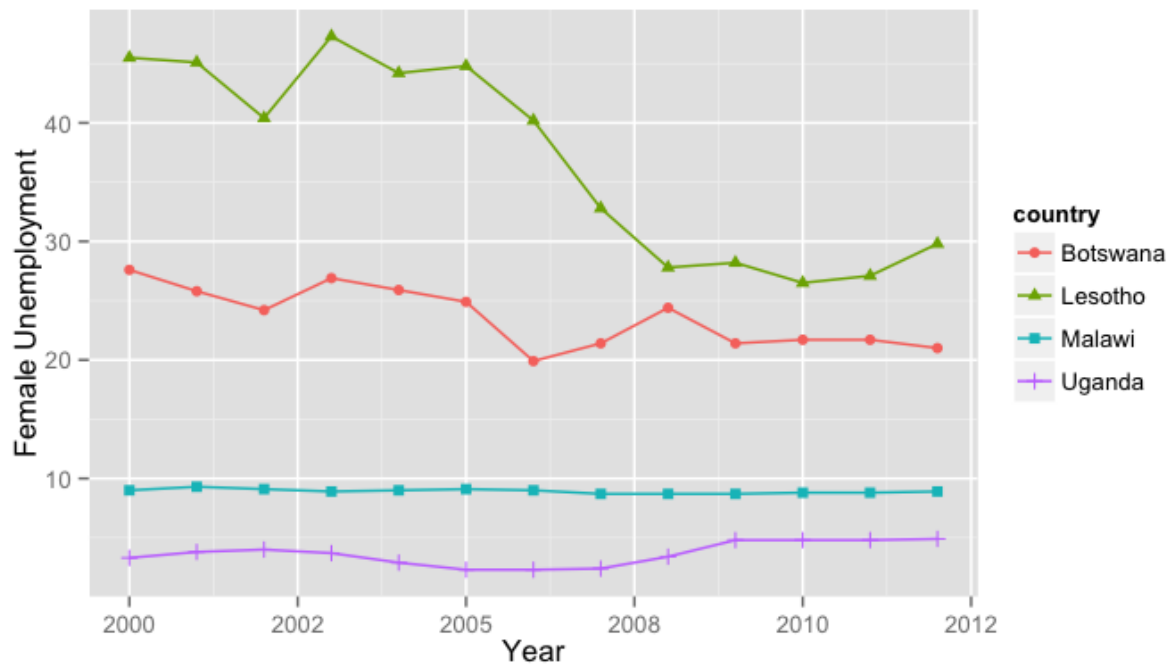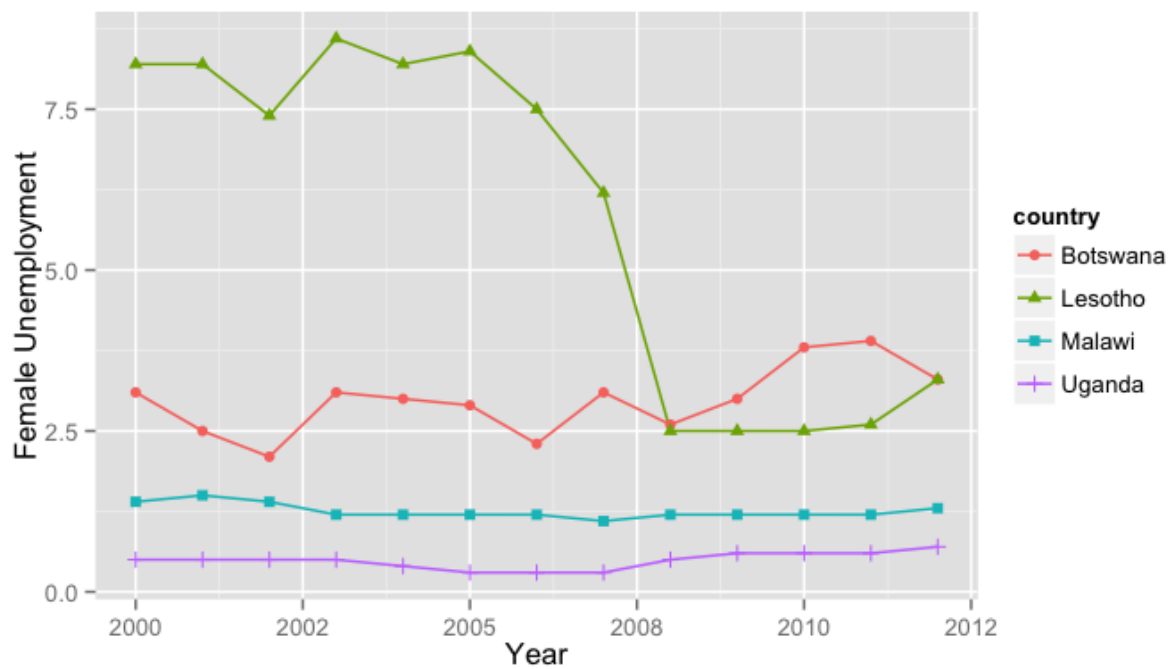## 7.4 Female Unemployment Share in Selected Countries

8 Inferential Statistics

8.1 Data Imputation

The data loaded from the Worldbank contained a lot of missing data because some countries do not report information on specific variables on a yearly basis. A high fraction of missing values produces a variety of problems and limitations to the data analysis. The paper imputed the missing data by using the "Amelia" package to solve this problem. "Amelia" generates by default five multiple, complete datasets that contain estimations of missing data points.

To enable the data imputation highly collinear variables had to be dropped. The test for variance inflation factors showed that in a simple OLS regression model integrating all independent variables, three variables showed high multicollinearity and had a higher variance inflation than the threshold of 10. We tested the multicollinearity between the variables and found that there was high multicollinearity between the GDP and GDP per capital and between Primary Education and Female School Enrollment. In a first step, we excluded one of these multicollinear variables for each group based on their explanatory strength for our research question, Primary Education and GDP. After having further problems of multicollinearity with GDPpc we also excluded Healthcare Expenditure and two variables that were created for the descriptive statistics part only, when showing the difference of HIV/AIDS Incidence over time (Incidence2 and IncidenceDif).

8.2 Model 1 - Comparison of countries with an HIV Incidence Rate below and above the Median

For Model 1 logistic regressions are used for predicting the likelihood that a country has a low HIV incidence rate (the dependent variable Y is equal to 1, rather than 0) given certain values of the HIV determinants (values of X) being hypothetically assumed. As explained in the descriptive statistics part, the dependent variable of Model 1 is a dummy variable, being 1 if the HIV incidence rate is below the median and 0 for countries with an HIV incidence rate above the median.

As Odds and Odds ratios are difficult to interpret and do not allow to make direct statements about predictors in our model, predicted probabilities are calculated after running the logistic regressions. The interpretation of the logistic regression results presented in table 1 will thus focus only on the significance of the variables.

A detailed analysis of the predicted probabilities will be given for the main variable of interest Female School Enrollment. For the predicted probability quintiles of Female School Enrollment are computed and applied to four different scenarios, the case studies that were introduced in the descriptive statistics part.

Table 1: Logistic Regression Results of Model 1

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Constant	-99.57	9.31	-10.69	0.00
GDP per capita	-0.69	0.30	-2.33	0.02
Share of Rural Population	-0.92	0.40	-2.30	0.02
CO2 Emissions per capita	-1.01	0.21	-4.81	0.00
Healthcare Expenditure	0.35	0.36	0.98	0.33
Access to Water	0.04	0.68	0.05	0.96
Access to Sanitation	0.34	0.24	1.43	0.15
Life Expectancy	29.22	2.47	11.83	0.00
Immunisation against DPT	-0.98	1.23	-0.80	0.43
Immunisation against Measles	1.26	1.28	0.99	0.32
Female School Enrollment	-3.36	0.67	-4.99	0.00
Share of Female Unemployment	-0.03	0.03	-0.80	0.42

From the logistic regression results presented in Table 1 it can be seen that the variables GDP per capita, Share of Rural Population, CO2 Emissions per capita, Life Expectancy and Female School Enrollment are showing statistically significant results at the 95% confidence level. The variables Immunization against DPT, Immunization against Measles, Access to Water, Access to Sanitation and Healthcare Expenditure and Share of Female Unemployment are however not statistically significant.

The predicted probabilities were calculated by using the "Zelig"-package as Zelig has the option to handle multiply-imputed data frames. The predicted probabilities plots show that the effect of Female School Enrollment is going in the same direction for all four scenarios even though, the expected values for the predicted probabilities differ. For the Scenario of Uganda the strongest effect can be observed effect.

Predicted Probabilities for Female School Enrollment
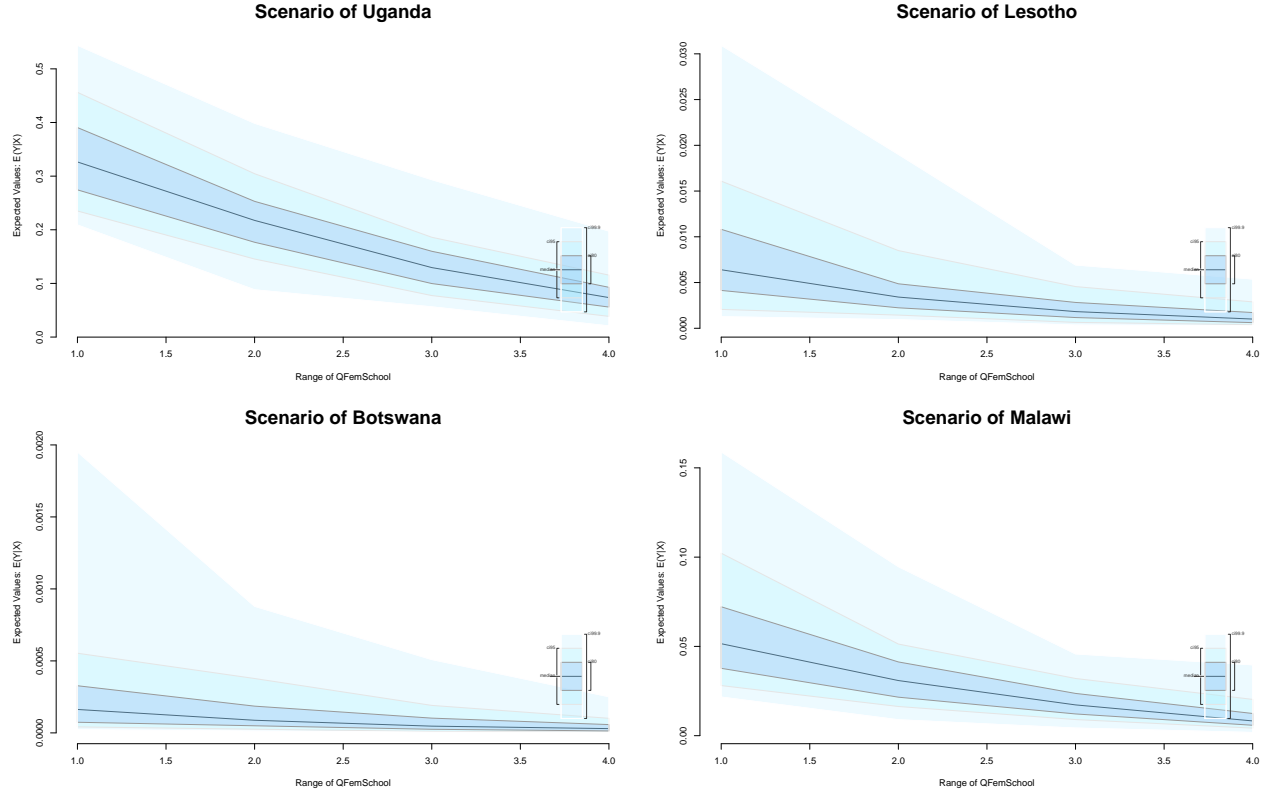
8.3 Model 2 - Focusing on Countries with an HIV Incidence Rate above the Median

As explained in the decriptive statistics part, the depenedent variable for the second empirical model is only considering countries with a higher HIV/AIDS incidence than the median. Further, the dependent variable for this model was logged to approach a normal distribution.

Table 2: OLS Regression Results of Model 2

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Constant	15.26	1.59	9.62	0.00
GDP per capita	0.16	0.08	2.09	0.04
Share of Rural Population	0.53	0.13	4.06	0.00
CO2 Emissions per capita	0.12	0.05	2.60	0.01
Healthcare Expenditure	-0.10	0.11	-0.89	0.37
Access to Water	0.37	0.19	1.97	0.05

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Access to Sanitation	-0.02	0.07	-0.25	0.81
Life Expectancy	-6.94	0.30	-22.98	0.00
Immunisation against DPT	0.13	0.30	0.43	0.67
Immunisation against Measles	-0.02	0.33	-0.05	0.96
Female School Enrollment	1.36	0.14	9.43	0.00
Share of Female Unemployment	0.13	0.02	7.04	0.00

Table 3: OLS Regression Results of Model 2 with robust standard errors

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Constant	15.26	1.59	9.62	0.00
GDP per capita	0.16	0.08	2.09	0.04
Share of Rural Population	0.53	0.13	4.06	0.00
CO2 Emissions per capita	0.12	0.05	2.60	0.01
Healthcare Expenditure	-0.10	0.11	-0.89	0.37
Access to Water	0.37	0.19	1.97	0.05
Access to Sanitation	-0.02	0.07	-0.25	0.81
Life Expectancy	-6.94	0.30	-22.98	0.00
Immunisation against DPT	0.13	0.30	0.43	0.67
Immunisation against Measles	-0.02	0.33	-0.05	0.96
Female School Enrollment	1.36	0.14	9.43	0.00
Share of Female Unemployment	0.13	0.02	7.04	0.00

Table 4: Fixed Effects Regression Results of Model 2

	Value	Std. Error	t-stat	p-value
(Intercept)	-2.2669	2.4491	-0.9256	0.3547
lGDPpc	-0.0315	0.1127	-0.2791	0.7809
lRural	3.1220	0.3577	8.7270	0.0000
lCO2	0.0491	0.0362	1.3564	0.1846
lHCexpend	-0.1548	0.0773	-2.0014	0.0473
lWater	-1.1905	0.3359	-3.5443	0.0005
lSanitation	-0.4486	0.2348	-1.9104	0.0561
lLifeExpect	-1.1540	0.3741	-3.0848	0.0022
lDPT	0.2729	0.1227	2.2245	0.0262
lMeasles	-0.4121	0.1305	-3.1564	0.0016
lFemSchool	-0.0062	0.0897	-0.0688	0.9452
ShFemUnempl	-0.0081	0.0154	-0.5285	0.5972
as.factor(country)Benin	-0.7988	0.4246	-1.8812	0.0600
as.factor(country)Botswana	3.9079	0.2475	15.7914	0.0000
as.factor(country)Burkina Faso	-1.8862	0.3808	-4.9531	0.0000
as.factor(country)Burundi	-0.7273	0.3485	-2.0867	0.0391
as.factor(country)Cambodia	-2.1820	0.3193	-6.8333	0.0000
as.factor(country)Cameroon	1.8073	0.1925	9.3872	0.0000
as.factor(country)Central African Republic	0.3527	0.3383	1.0425	0.2980
as.factor(country)Chad	-1.2207	0.3524	-3.4639	0.0005
as.factor(country)Congo, Dem. Rep.	-0.9546	0.2729	-3.4984	0.0006
as.factor(country)Congo, Rep.	1.4518	0.4481	3.2401	0.0012

	Value	Std. Error	t-stat	p-value
as.factor(country)Cote d'Ivoire	0.6545	0.3440	1.9026	0.0573
as.factor(country)Dominican Republic	2.1132	0.2442	8.6553	0.0000
as.factor(country)El Salvador	1.2584	0.3031	4.1524	0.0000
as.factor(country)Eritrea	-1.9937	0.3849	-5.1799	0.0000
as.factor(country)Ethiopia	-2.4567	0.3642	-6.7448	0.0000
as.factor(country)Gabon	5.4579	0.5295	10.3075	0.0000
as.factor(country)Gambia, The	1.3671	0.2688	5.0867	0.0000
as.factor(country)Ghana	0.1740	0.4346	0.4004	0.6889
as.factor(country)Guatemala	0.6513	0.1867	3.4895	0.0005
as.factor(country)Guinea-Bissau	0.7292	0.3469	2.1017	0.0357
as.factor(country)Haiti	0.3438	0.2849	1.2065	0.2279
as.factor(country)Honduras	0.4574	0.2489	1.8373	0.0667
as.factor(country)Jamaica	1.8066	0.1960	9.2186	0.0000
as.factor(country)Kenya	0.2553	0.2211	1.1548	0.2485
as.factor(country)Lesotho	2.1955	0.3213	6.8340	0.0000
as.factor(country)Liberia	-0.0623	0.4239	-0.1470	0.8832
as.factor(country)Madagascar	-2.2753	0.3723	-6.1110	0.0000
as.factor(country)Malawi	0.4854	0.4769	1.0178	0.3090
as.factor(country)Mali	-1.2762	0.2718	-4.6956	0.0000
as.factor(country)Mauritius	1.2503	0.2241	5.5798	0.0000
as.factor(country)Mozambique	0.8579	0.2991	2.8683	0.0042
as.factor(country)Myanmar	-0.7782	0.2960	-2.6291	0.0096
as.factor(country)Namibia	2.3796	0.2663	8.9375	0.0000
as.factor(country)Nepal	-1.7530	0.3613	-4.8517	0.0000
as.factor(country)Niger	-2.6918	0.4486	-6.0011	0.0000
as.factor(country)Nigeria	0.5419	0.1967	2.7554	0.0059
as.factor(country)Papua New Guinea	-2.3047	0.2767	-8.3306	0.0000
as.factor(country)Rwanda	-0.2379	0.2806	-0.8478	0.3982
as.factor(country)Senegal	-0.4843	0.2184	-2.2176	0.0274
as.factor(country)Sierra Leone	-0.7819	0.3851	-2.0304	0.0425
as.factor(country)Somalia	-2.5405	0.2860	-8.8818	0.0000
as.factor(country)South Africa	4.5206	0.2355	19.1985	0.0000
as.factor(country)Swaziland	2.3213	0.1690	13.7347	0.0000
as.factor(country)Tanzania	-0.2001	0.3935	-0.5085	0.6112
as.factor(country)Thailand	-0.1000	0.2708	-0.3691	0.7121
as.factor(country)Togo	-0.4940	0.3896	-1.2680	0.2048
as.factor(country)Trinidad and Tobago	-0.3994	0.3110	-1.2839	0.1992
as.factor(country)Uganda	0.4963	0.2811	1.7657	0.0786
as.factor(country)Ukraine	2.6254	0.2618	10.0271	0.0000
as.factor(country)Vietnam	-0.7357	0.2929	-2.5121	0.0123
as.factor(country)Zambia	1.8471	0.1670	11.0580	0.0000
as.factor(country)Zimbabwe	2.1824	0.2556	8.5379	0.0000

9 Conclusions

(Sachs and Malaney 2002) (Dupas 2009)

The IncidenceDif variable would actually be the most accurate variable in terms of measuring whether Target 6.A of the MDGs ("have halted by 2015 and begun to reverse the spread of HIV/AIDS" (United Nations 2014) was fulfilled until 2012. A dummy variable on whether MDG6a was fulfilled or not is however not very suitable for identifying the specific determinants of HIV/AIDS at country level as it is not differentiating

between the relative size of the effect (an increase from 0,001 to 0,002 would be as bad as a change from 0,1 to 0,3).

Model 2 - variable is not perfect normal distribution - better beta regression

Famale Labour Market Participation

10   Limitations

The paper had to make some compromises regarding its original aim as outlined in the first research proposal. Due to the significant amount of missing values and the presence of multicollinearity, a considerable number of variables had to be dropped and could ultimately not be integrated in the logistic regression models.

The selection of these variables was not arbitrary but followed instead the theoretical framework guiding this research, i.e. Dahlgren's model. Two levels of Dahlgren's model (Social and Community Networks and Individual Lifestyle Factors) ended up underrepresented after dropping these variables. To deal with this limitation, the research will only use the theoretical framework as an instrument to guide the selection of variables but will not utilise the findings to test the validity of the model.

In terms of the data used to run the regressions, the relative high number of countries that have already halted or reversed the spread of HIV/AIDS in our sample can lead to biased results. In the next stage of the research, the effect of excluding those countries that only halted the spread will be explored.

Another shortcoming faced at this stage was the integration of figures from the descriptive statistics into the final report. A transitory solution was to save those pictures in a subfolder of the repository.

11   Appendix

Descriptive Statistics

Table 5: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
X	988	494.5	285.4	1	988
year	988	2,006.0	3.7	2,000	2,012
GDPpc	945	5,296.0	5,153.0	441.2	30,875.0
Rural	988	57.7	19.5	8.7	91.8
CO2	829	1.7	3.5	0.004	38.2
HCexpend	962	5.9	2.2	1.8	18.4
Water	982	75.5	17.9	23.5	99.8
Sanitation	983	50.4	30.1	6.6	98.9
Unemploym	988	8.7	6.2	0.6	38.7
FemSchool	828	98.1	20.8	20.8	162.4
LifeExpect	988	61.4	9.9	38.1	79.6
DPT	988	79.9	18.0	19	99
Measles	988	79.4	17.5	16	99
Population	988	40,890,589.0	133,140,255.0	1,063,715	1,236,686,732
Incidence	988	0.3	0.7	0.01	4.4
lGDPpc	945	8.1	1.0	6.1	10.3
lRural	988	4.0	0.4	2.2	4.5
lCO2	829	−0.6	1.6	−5.6	3.6
lHCexpend	962	1.7	0.4	0.6	2.9
lWater	982	4.3	0.3	3.2	4.6
lSanitation	983	3.7	0.8	1.9	4.6
lUnemploym	988	1.9	0.7	−0.5	3.7
lHCexpendpc	962	4.1	1.2	0.9	7.0
lFemSchool	828	4.6	0.3	3.0	5.1
lLifeExpect	988	4.1	0.2	3.6	4.4
lDPT	988	4.3	0.3	2.9	4.6
lMeasles	988	4.3	0.3	2.8	4.6
lIncidence	988	−2.4	1.6	−4.7	1.5
QFemSchool	828	2.5	1.1	1	4
ShFemUnempl	988	1.5	2.9	−3.4	23.9
Dummy	988	0.5	0.5	0	1

Figure 3:HIV Incidence Rates per Country over Time
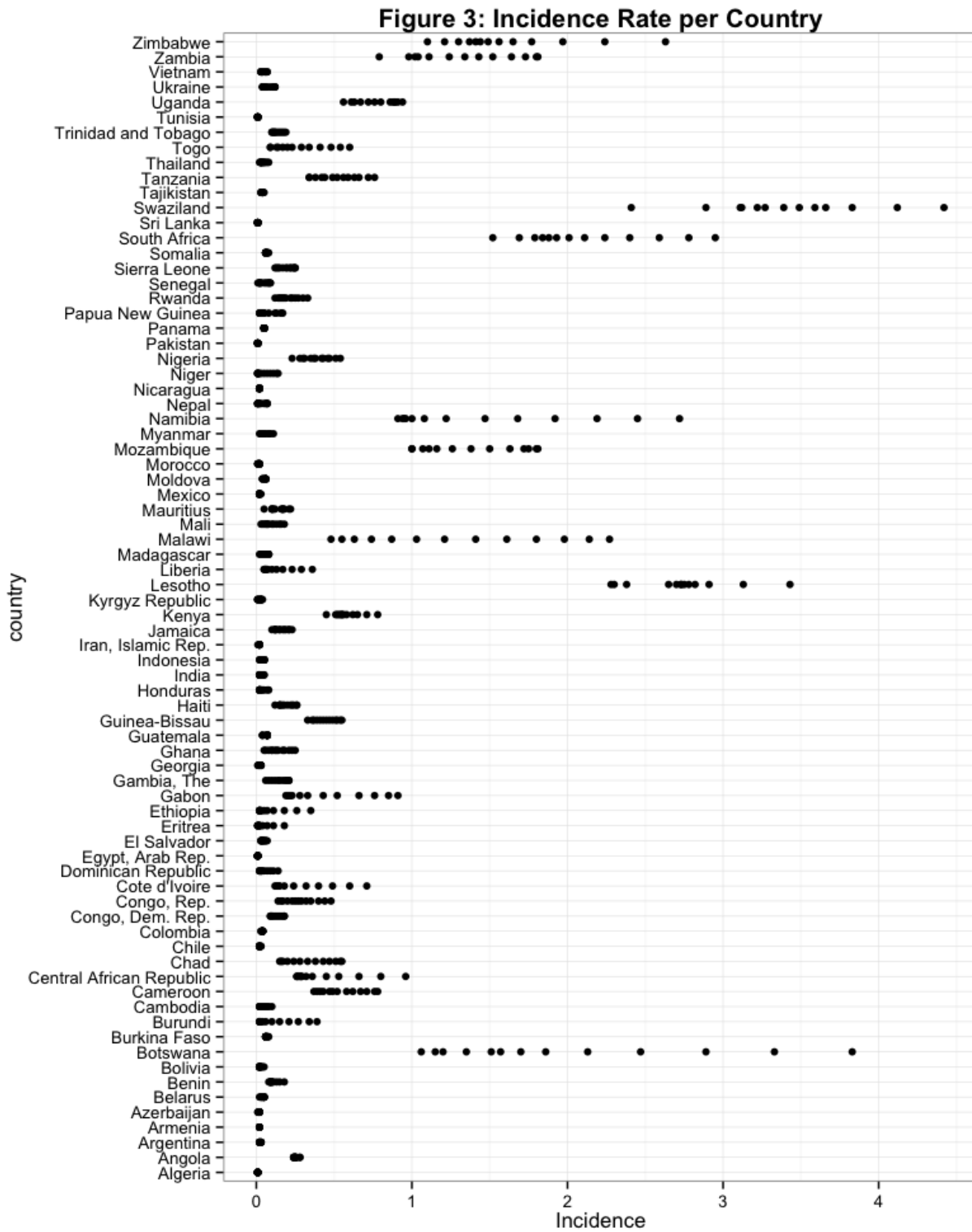

Figure 3: Incidence Rate per Country

Figure 4:Change in HIV Incidence Rate compared to Previous Years per Country

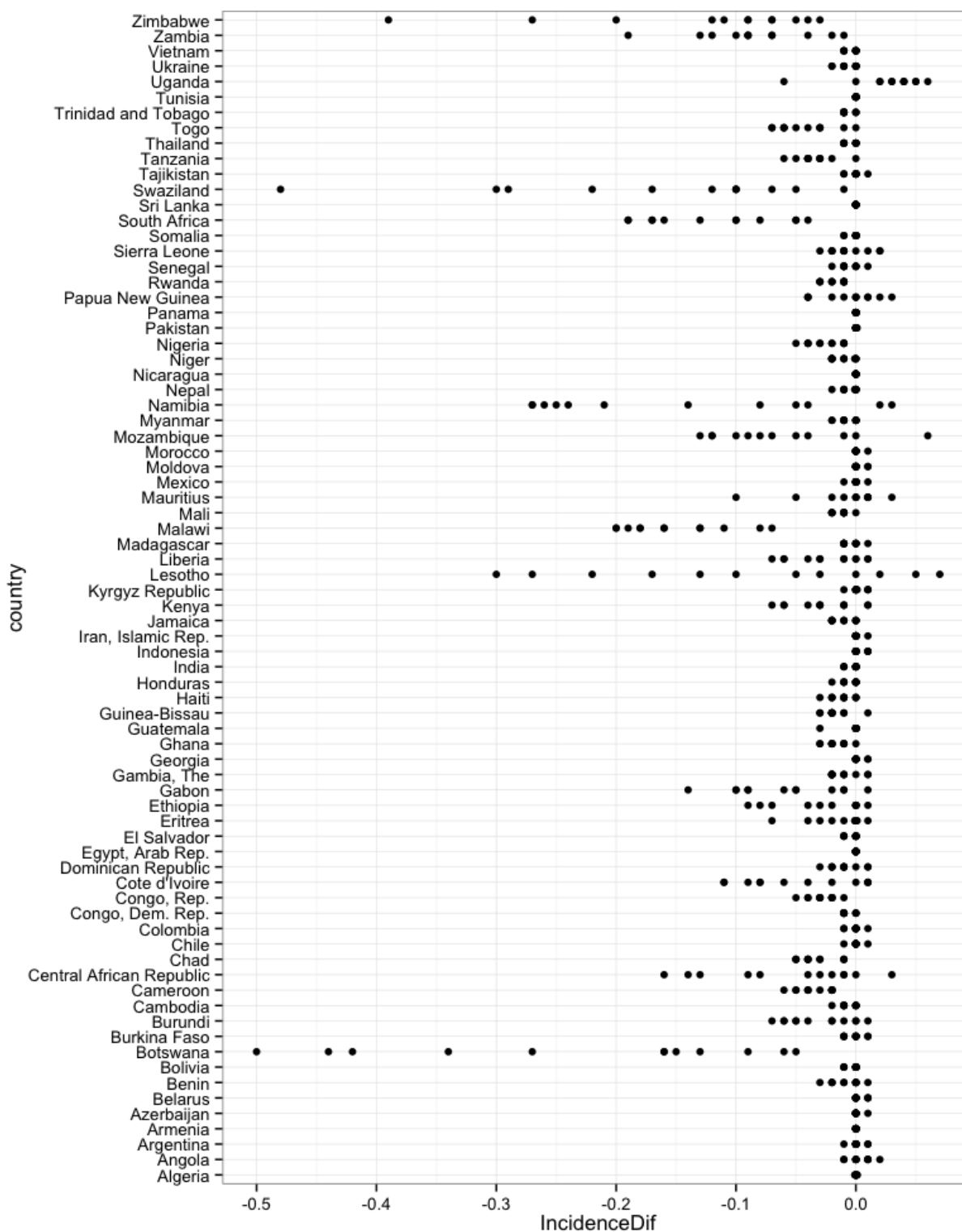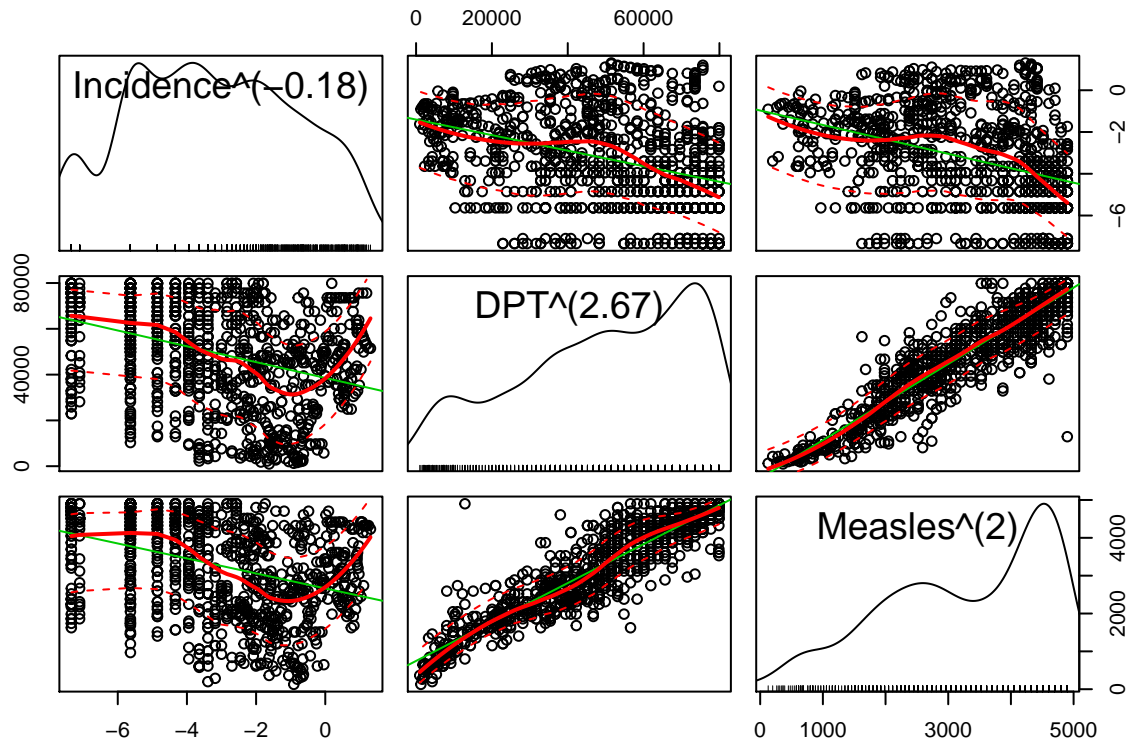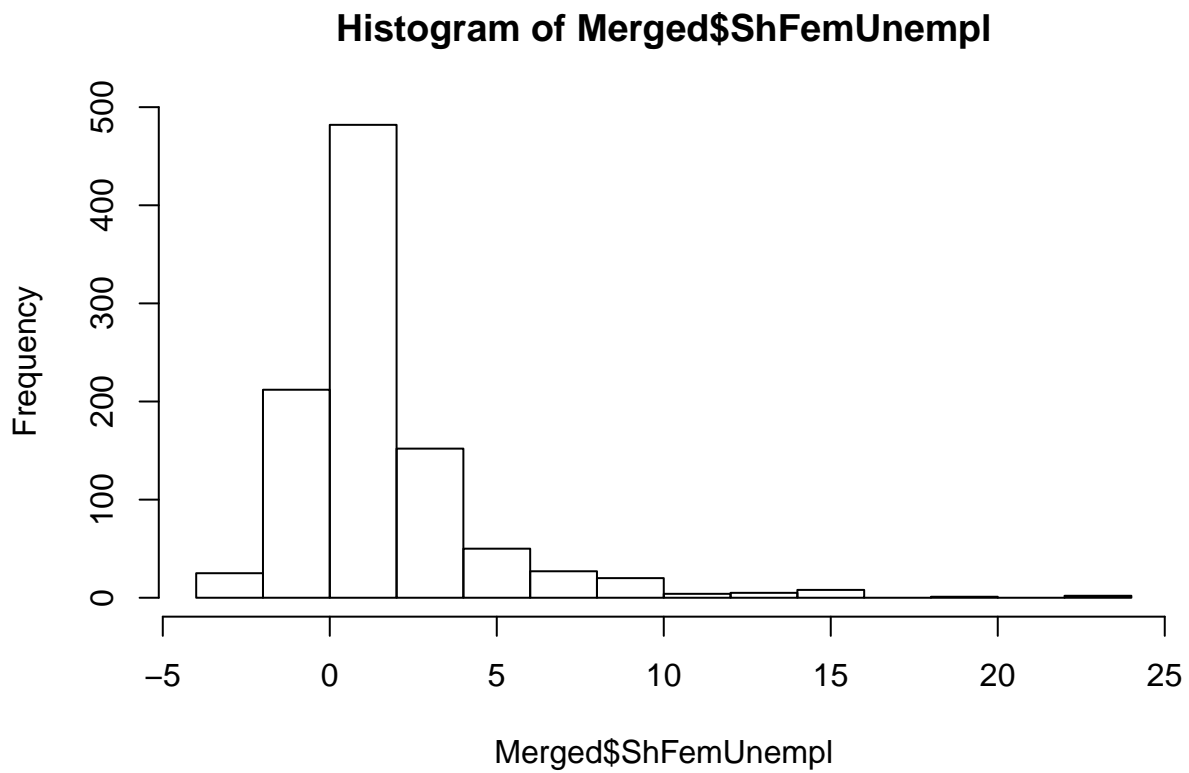**Figure 5:Scatterplot of variables for socio-economic, cultural and environmental conditions**

Figure 6:Scatterplot of variables for living and working conditions

Figure 7:Scatterplot of variables for individual lifestyle factors

Histogram of Merged$ShFemUnempl

Testing for Multicollinearity of the Variables

Table 6: Variance Inflation Factors - Table 1

	vif
GDP	3.22
GDPpc	12.78
Rural	2.71
CO2	3.48
HCexpend	1.68
Primary	51.11
Water	3.94
Sanitation	4.57
Unemploym	2.80
HCexpendpc	7.25
ShFemUnempl	1.80
FemSchool	56.88
LifeExpect	5.55
DPT	8.74
Measles	9.42
Population	3.03
Incidence2	4.88
IncidenceDif	2.73

Table 7: Variance Inflation Factors - Table 2

	vif
GDPpc	5.00
Rural	2.33
CO2	3.14
Water	3.76
Sanitation	4.24
HCexpend	1.20
ShFemUnempl	1.58
Unemploym	1.75
FemSchool	1.35
LifeExpect	3.34
DPT	8.55
Measles	8.97
Population	1.19

Data Imputation Matrix
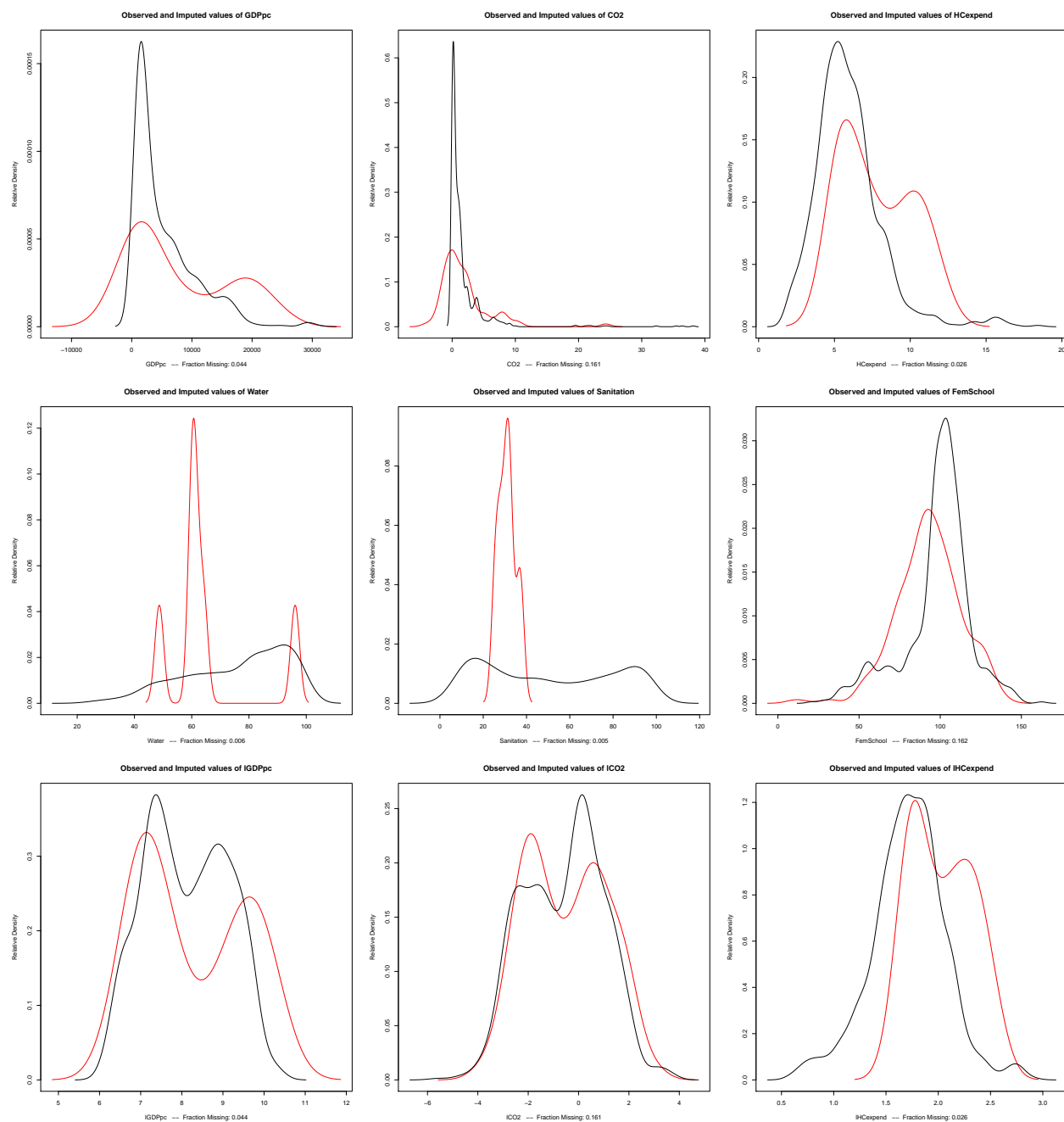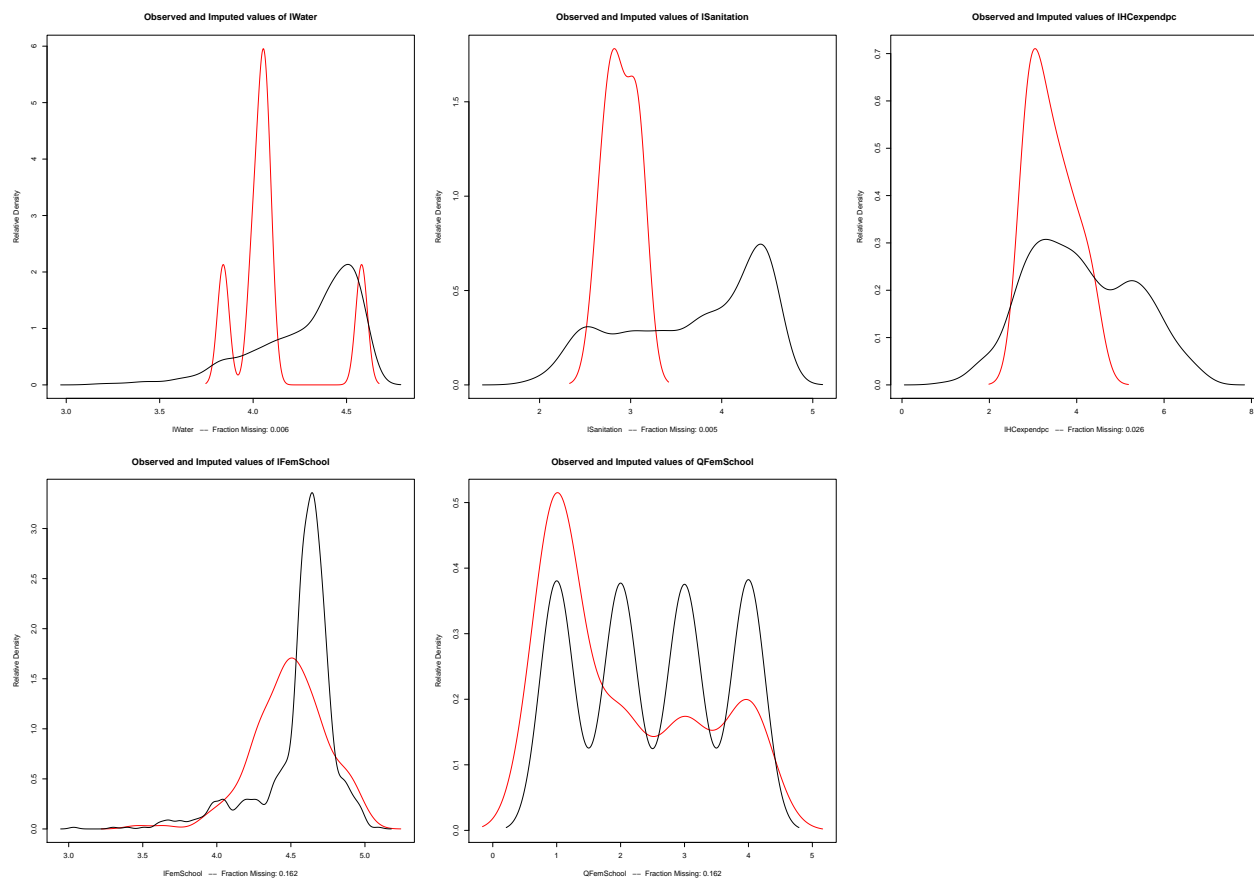
Observed and Imputed values of IWater

Observed and Imputed values of ISanitation

Observed and Imputed values of IHCexpendpc

Observed and Imputed values of IFemSchool

Observed and Imputed values of QFemSchool

Residuals vs Fitted

Fitted values
lm(Incidence ~ lGDPpc + lRural + lCO2 + lHCexpend + lWater + lSanitation + ...

Residuals vs Fitted

Fitted values
lm(lIncidence ~ lGDPpc + lRural + lCO2 + lHCexpend + lWater + lSanitation + ...
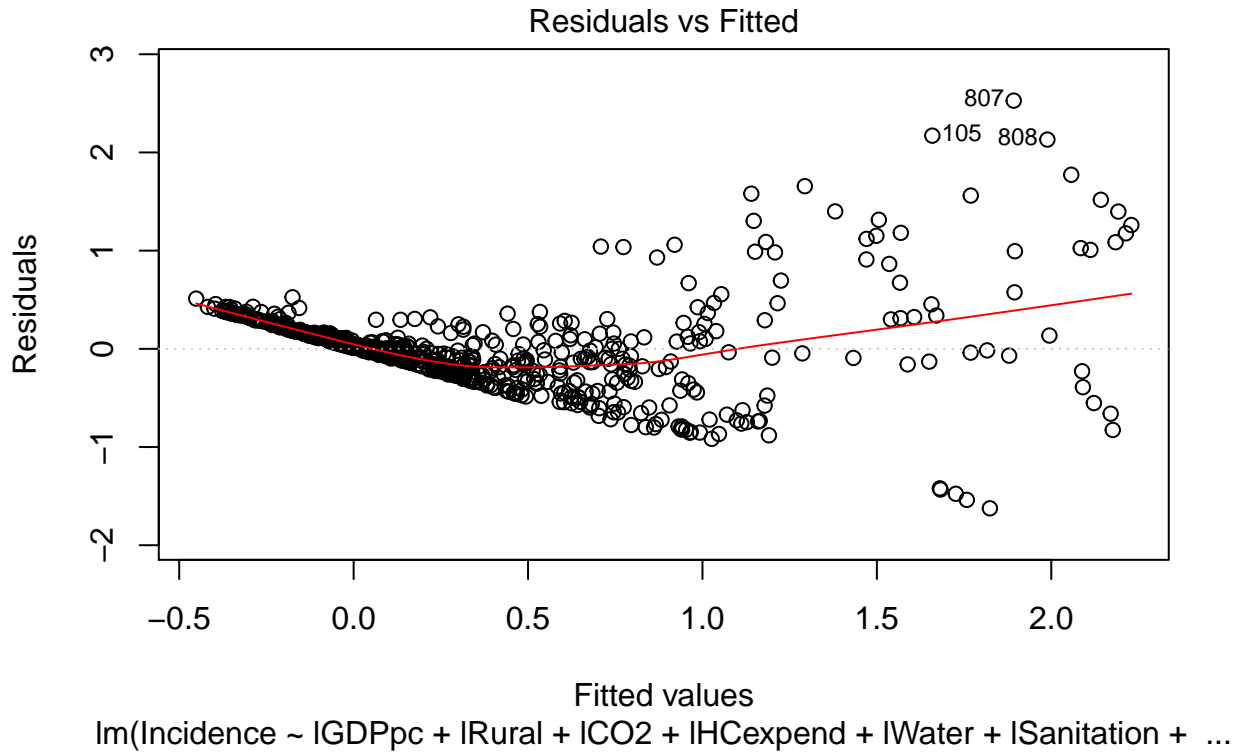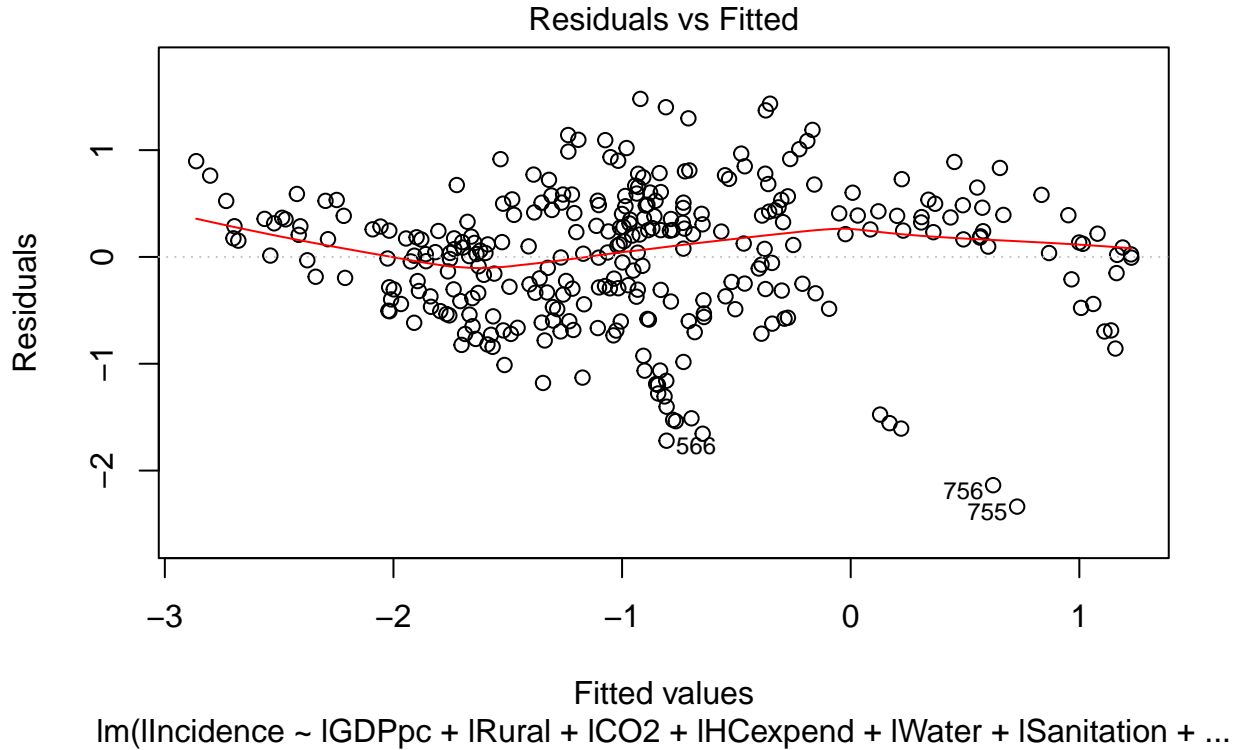
Dahlgren, Göran, and Margaret Whitehead. 1991. "Policies and Strategies to Promote Social Equity in Health." *Institute for Futures Studies.*

Dupas, Pascaline. 2009. *Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya.* National Bureau of Economic Research.

Gerring, John. 2008. "Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques." *J. Box-Steffensmeier, HE Brady, & D. Collier (Reds.), Oxford Handbook of Political Methodology, Oxford University Press, Oxford*, 645–84.

Haacker, Markus. 2002. "The Economic Consequences of HIV/AIDS in Southern Africa." *IMF Working Paper* 02 (38).

Hurrelmann, Klaus. 1989. "Human Development and Health." *Springer-Verlag.*

Kalemli-Ozcan, Sebnem. 2011. "AIDS, "Reversal" of the Demographic Transition and Economic Development: Evidence from Africa." *Springer-Verlag.*

Sachs, Jeffrey, and Pia Malaney. 2002. "The Economic and Social Burden of Malaria." *Nature* 415 (6872): 680–85.

Solar, O, and A Irwin. 2010. "A Conceptual Framework for Action on the Social Determinants of Health. Social Determinants of Health Discussion Paper." *Policy and Practice*, no. 2.

Talbott, John R. 2007. "Size Matters: The Number of Prostitutes and the Global HIV/AIDS Pandemic." *PLoS ONE.*

United Nations. 2014. "Millennium Development Goals." http://www.un.org/millenniumgoals/aids.shtml.

World Bank. 2014. http://microdata.worldbank.org/index.php/catalog/dhs/about.

World Health Organization, and UNICEF. 2014. "Joint Monitoring Programme for Water Supply and Sanitation." http://www.wssinfo.org/definitions-methods/data-sources/.