Examining Potential Socioeconomic and Public Health Factors for Life Expectancy

Giang (Jessie) Le, Zhaoliang (Leon) Zhou, and Calvin Ochi

**Introduction**

Life expectancy at birth is defined by the United Nations Human Development Report as “the years a newborn infant would live if prevailing patterns of age-specific mortality rates at the time of birth were to stay the same throughout the infant’s life” (Human Development Report, 2019). Often life expectancy is used as an indicator of national development and quality of life for the average citizen. Since the 1990’s, life expectancy has increased globally with a drastic increase during the beginning of the Millennium Development Goals (MDGs) era, 2000-2015 (WHO, 2015). Despite this increase, high-income countries still have higher average life expectancies than lower-income countries (WHO, 2015). To determine the cause of this, important socioeconomic and health factors mentioned previously in the literature will be investigated.

We have chosen to include continent as an indicator variable based on a study done by Kpolovie et al. in 2016. Their research suggests there has been significant continental differences in life expectancy. In this study, researchers identified that Africa had significantly lower life expectancies than all other sampled continents. Thus, for our study, we would like to investigate the effect of continents on life expectancy, especially among African countries.

For socioeconomic determinants, previous research indicates that longitudinal contributors to differences in life expectancy between 1970-2004 included income measured by Gross Domestic Product (GDP) per capita, percent expenditure on health, and education represented by years of schooling (Lin et al 2012). However, a study by Kabir (2008) yields a different conclusion: conventional variables for life expectancy (GDP, education, per capita health expenditure, and urbanization) are not significant in accounting for life expectancy on a global scale at the beginning of the twenty-first century. Thus, we would like to investigate those socioeconomic factors and confirm they were not associated with life expectancy from 2000 to 2015.

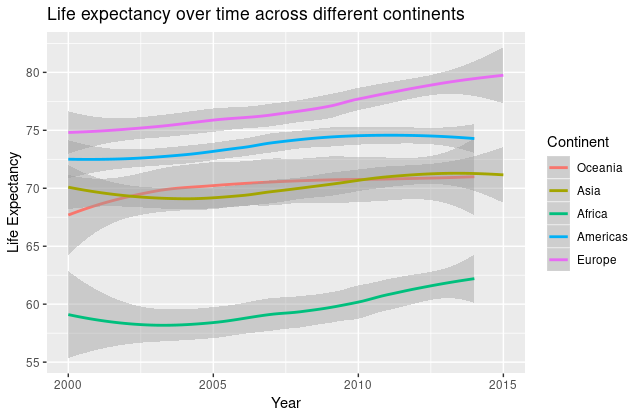
Moreover, for health-related variables, some studies show that life expectancy can be attributed to factors including alcohol consumption, percent immunization, BMI, and HIV/AIDS prevalence (Mondal and Shinta, 2014 and Khan et al., 2016). We are particularly interested in the effects of HIV on life expectancy in Africa compared to other continents. A report from WHO (Andre et al., 2008) states that vaccination increases life expectancy by protecting against diseases and reducing risks of mortality.

Learning more about which factors are associated with differences between life expectancy in developing versus developed countries and between continents might help to inform policy decision-making and provide insights into resource allocation in order to narrow the gap in life expectancy across continents. In this project, we will assess which factors have the greatest effect on life expectancy, and if the factors we identified are consistent with those suggested by the previous studies.

**Materials and Methods**

We obtained our data from Kaggle website (“Life Expectancy (WHO),” 2019) which was originally sourced from the Global Health Observatory (GHO) data repository under World Health Organization (WHO). The socioeconomic information for each country for each year were sourced from United Nation website. The original data had information about life expectancy, socioeconomic variables, and health variables for all 193 countries from year 2000 to 2015.

For original dataset, there were 10 island countries with no information on life expectancy, as well as on the other covariates. Thus, we removed those 10 countries. We added a continent indicator to each country, which had 5 categories: Asia, Africa, Europe, Oceania, and the Americas. From exploratory data analysis, we found that there were differences in life expectancy among different continents between 2000-2015 (Figure 1), and Africa was considerably lower than any other continents. For our modeling process, we used Oceania as the reference continent. To maintain a consistent dataset for model comparison, we omitted any rows that contained any missing values. Thus, our final analysis contained 133 countries in 5 continents. Socioeconomic factors and public health indicators are summarised in Table 1.



**Figure 1.** Average life expectancy over time between 2000-2015 across different continents.

Because of the skewness of the distributions, we used logarithm transformation for the following explanatory variables: Hepatitis B, Polio, and Diphtheria percent immunization among one-year-olds, number of measles cases, and number of HIV deaths per 1,000 live births. In addition, since our study included longitudinal models, we centered year with baseline at 2000. Because percent expenditure on healthcare and income composition of resources were either calculated from or highly correlated with GDP, we removed them from our analysis. Based on exploratory analysis, we found African countries had significantly lower life expectancy than the other continent in 2000; however, they had the fastest growth rate in life expectancy from year 2000 and 2015. Therefore, we created an indicator of African country.

Since our dataset follows a multilevel structure with time as observational units at level 1 and countries as observational unit at level 2, we would like to implement multilevel models for our analysis to identify important factors related to life expectancy. For level 1 variables, we had: year, healthcare factors including alcohol, number of measles cases, BMI, percent hepatitis B, polio, diphtheria immunization among one-year-olds, number of HIV/AIDS deaths per 1,000 live births, and socioeconomic factors including total expenditure, GDP, population, and years of schooling. For our level 2 variables, we had continent and country status.

**Table 1. Descriptive Statistics for Study Data Set**

|  | Mean (SD) or Percentage |
| --- | --- |
| Continent  African  Asian  American  European  Oceanian | 33.0% 23.3% 17.3% 20.3% 0.06% |
| Status  Developed  Developing | 14.3%  85.7% |
| BMI index | 38.1 (± 19.8) |
| Alcohol consumption | 4.53 (± 4.03) liters/person/year |
| Number of Measles cases per 1,000 population | 2,224 (± 10,086) cases |
| Percent Hepatitis B coverage among 1-year-olds | 79% (± 26%) |
| Percent Polio coverage among 1-year-olds | 84% (± 22%) |
| Percent Diphtheria coverage 1-year-olds | 84% (± 22%) |
| Deaths per 1,000 HIV/AIDS live births | 2 (± 6) |
| GDP | 5,566 (± 11,475) USD |
| Percentage of expenditure on health | 690% (±1800%) |
| Years of schooling | 12.1 (± 2.8) years |

For our modeling process, we began with an unconditional means model and an unconditional growth model. Then, our models can be further divided into 4 main categories: socioeconomic models where only prevalent socioeconomic factors were included, health models where prevalent health factors were included, a model with only covariates (no interactions), and a combined model where we had variables from either categories. For each of the models, we began by an unconditional means model, then we added 1 variable at a time for each level. Every time we added a new variable, we used the Likelihood Ratio Test (LRT) comparing the new model to the previous model without the additional term under a 𝜒2 distribution. We only included variables that had significant t-values (t>2) or showed a significant improvement of the model based on the 𝜒2 statistics and p-values from the LRT. For our composite model, which included both socioeconomic factors and health factors, we used parametric bootstrap to compare the composite model with the socioeconomic model and the health model individually, and we expected to see the composite model would perform better.

In addition, we reran each model using Bayesian inference approach, Markov Chain Monte Carlo Sampler for Multivariate Generalised Linear Mixed Models (MCMCglmm)**,**  in order to provide insights into the coefficient estimates, 95% CI, and the significance of each variable.

For model comparison criteria between nested models, we used LRT. For non-nested models, we used BIC since we would prefer simpler models.

**Results**

Multilevel models with different sets of variables and interactions are compared in Table 2. When examining socioeconomic factors, neither total expenditure on health nor percentage of expenditure significantly improved model 1. There was no significant difference in annual increase of life expectancy between developed and developing countries. During exploratory analysis, we discovered that BMI index and number of Measles cases per 1,000 population did not have significant relationships with life expectancy. There was a high correlation between BMI and years of schooling (cor = 0.55). On the other hand, among other public health indicators, only logarithm of HIV/AIDS deaths per 1,000 live births provided significant information about the variability in life expectancy (model 2). Percents of immunization were moderately correlated to one another and showed no statistical significance (t-values <2).

We combined significant socioeconomic and health effects (year, log(GDP), years of schooling, log(HIV deaths per 1,000 live births), continent, and status) from model 1 and model 2 to produce model 3 with only covariates. In model 3, the mean annual increase in life expectancy over 15 years was small (0.18/year). Since BIC of model 3 was smaller than those of model 1 and 2, there was evidence that a combined-effect model better explained the variability in life expectancy. In model 3, Africa had an average life expectancy that is 3.36 lower than that of Oceania, after controlling for the other variables.

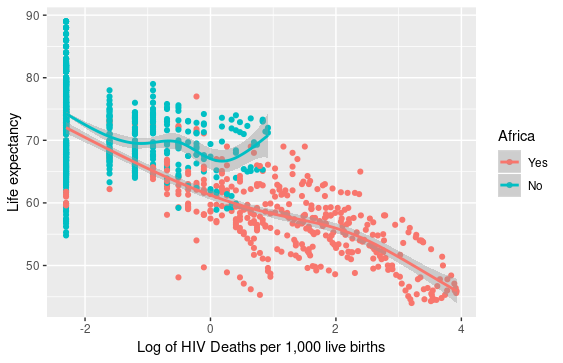
According to BIC, there was statistical evidence that the final model showed superiority in performance than the others. Parametric bootstrap method between model 1 and final model suggested that including HIV deaths per 1,000 live births provided significant extra information about life expectancy (χ2=99, p-value < 0.001). Also, coefficients in model 2 were tested via parametric bootstrap, and the result indicated that including logarithm of percentages of immunization did not significantly improves model performance (χ2=2.48, p-value=0.52). After dropping the insignificant health variables and adding socioeconomic factors (log(GDP) and years of schooling), model performance was significantly improved (χ2=81.7, p-value < 0.001). Examining between model 3 and the final model, the final model had a lower BIC, and parametric bootstrap indicated the significance of level-2 indicators (country status and continent) in explaining the differential change in life expectancy (χ2=41.5, p-value < 0.001). However, there was a lack of evidence that the annual rate of life expectancy changed as African indicator interacted with the annual change. The annual increase of life expectancy in African countries was only 0.012 smaller than that in non-African countries, and there was evidence that this small gap was not significant (t-value <2).

**Table 2. Multilevel Mixed-Effect Models**

|  | Model 1: Socioeconomic factors | | Model 2: Public health indicators | | Model 3: Covariates only | | Final model: Covariates x status and continent | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Predictors | β | SE | β | SE | β | SE | β | SE |
| Intercept | 63.9 | 2.6 | 71.2 | 2.2 | 59.6 | 2.2 | 62.7 | 2.2 |
| Status  Developed (ref.)  Developing | -  -4.26 | -  1.64 | -  -5.35 | -  1.53 | -  -3.95 | -  1.37 | -  -4.3 | -  1.3 |
| Continent  Oceania (ref.)  Africa  Asia  Americas  Europe | -  -10.1  1.73  4.03  4.63 | -  2.01  1.85  1.92  2.05 | -  -7.59  0.85  4.07  5.00 | -  1.88  1.73  1.80  1.91 | -  -3.36  1.24  4.29  3.47 | -  1.57  1.54  1.60  1.71 | -  -5.0  1.1  4.0  3.4 | -  1.7  1.5  1.5  1.6 |
| Log(GDP) | -0.011 | 0.033 |  |  | -0.015 | 0.033 | -0.014 | 0.033 |
| Years of Schooling | 0.617 | 0.10 |  |  | 0.73 | 0.1 | 0.69 | 0.09 |
| Log(HIV death rate) |  |  | -0.46 | 0.23 | -1.58 | 0.17 | -0.59 | 0.23 |
| Log(Hepatitis B) |  |  | -0.03 | 0.08 |  |  |  |  |
| Log(Polio) |  |  | -0.05 | 0.08 |  |  |  |  |
| Log(Diphtheria) |  |  | 0.14 | 0.09 |  |  |  |  |
| Years (centered) | 0.20 | 0.03 | 0.26 | 0.03 | 0.18 | 0.03 | 0.18 | 0.03 |
| African х Years (centered) | 0.255 | 0.055 | 0.05 | 0.06 |  |  | -0.012 | 0.059 |
| African x log(HIV death rate) |  |  | -2.08 | 0.34 |  |  | -2.1 | 0.3 |
| BIC | 6987.53 |  | 6976.93 |  | 6930.27 |  | 6908.72 |  |

After adjusting for socioeconomic variables and HIV deaths per 1,000 live births, developing countries had a mean decrease of 4.3 in life expectancy, compared with that of developed countries (95% CI 1.5-7.1). In 2000, the average life expectancy in Africa was 5 years lower than that in Oceania (95% CI 1.8-8.2), after adjusting for other variables. On the other hand, in 2000, after accounting for the other variables, Asia, Americas, and Europe had 1.1 years higher (95% CI -1.9-4.1), 4 years higher (95% CI 0.9-7.1), and 3.4 years higher (95% CI 0.2-6.7) than Oceania, respectively. The average annual increase in life expectancy among non-African countries was 0.18 years (95% CI 0.12-0.24), while the average annual increase among African countries was slightly smaller, 0.17 years. We are 95% confident that the life expectancy in Africa increased between 0.17 to 0.23 from 2000 to 2015, after adjusting for the other variables. One-year increase of education was associated with an average increase of 0.7 in life expectancy (95% CI 0.5-0.9), after controlling for year, continent, status, GDP, and HIV deaths per 1,000 live births. Surprisingly, a double increase in GDP was associated with an additive decrease of 0.01 in life expectancy (95% CI -0.06-0.03). There was also no significant evidence that log(GDP) could explain extra information about the data (t-value<2). The negative effect of log(HIV deaths) was substantially larger in African countries than the other continents (Figure 2). For non-African countries, a double increase in child deaths by HIV per 1,000 live births was associated with a decrease of 0.4 in life expectancy (95% CI 0.1-0.7). On the other hand, for African countries, a double increase in child deaths by HIV per 1,000 live births was associated with a decrease of 1.9 in life expectancy (95% CI 1.4-2.3).

Results from MCMCglmm (Table A - Appendix) aligned with the results from Table 2, relative to the 95% confidence intervals of fixed effects. MCMCglmm method (Hadfield, 2010) validated the significance of our results.



**Figure 2**. Life expectancy versus log HIV deaths per 1,000 live births between African and non-African countries

**Discussion**

The present study confirms that there was a significant average annual increase of 0.18 in life expectancy from 2000-2015. In any year, life expectancy in Africa was significantly lower than that in other continents. Some common determinants of life expectancy were also verified to be significant: continent, country status, average years of education, and HIV deaths per 1,000 live births. The study also showed that some variables were not significant in explaining the variability in life expectancy among countries over the 15-year period: GDP, total expenditure on healthcare, percentage of expenditure on healthcare, percents of immunization coverage. Additionally, there was no statistical evidence for a difference in annual rate of life expectancy between African and non-African countries. However, we discovered that there was a stronger negative effect of HIV deaths per 1,000 live births on life expectancy in Africa than in the other continent, after controlling for other variables.

Life expectancy was, on average, 4.3 lower in developing countries than in developed countries. There are cumulative reasons for this large difference. Having higher income per capita, developed countries have achieved higher education level, have better and more accessible health facilities, and higher standard of living.

In previous studies, there has been statistical evidence that life expectancy is the lowest in Africa and the highest in Europe (Kpolovie et al., 2016). High disease prevalence, exacerbated by an increase in HIV infection, leads to decreased life expectancy in Africa (The World Bank, 2005). Kpolovie et al. (2016) also concludes that there is also missing evidence that life expectancy in Europe significantly differ from that in North and South America. There is also no difference between Asia and Oceania, which is consistent with our model. It might be because there is a wide variability in socioeconomic and healthcare conditions within large continents of Asia and the Americas. Future studies should sub-divide these continents into groups that are more comparable in politics and economy.

African countries had significantly lower average life expectancy in 2000, but the mean annual increase over the course of 15 years did not differ from the other continents. This was contrary to previous report by Becker et al. (2005), which described a convergence of life expectancy between 1965-1995. That is, countries with lower life expectancy in 1965 tended to have a larger increase in life expectancy. We had expected that African countries, which had the lowest life expectancy in 2000, would experience a greater gain in life expectancy; however, there was no such trend between 2000-2015.

There was evidence that higher HIV deaths among 1,000 live births had an association with decreasing life expectancy, with a stronger effect among African countries. Number of HIV/AIDS deaths among 1,000 live births indicates about newborn healthcare, HIV prevention and treatment policies. A large number of HIV deaths among children reveals that there is a great number of infected adults. More importantly, it also reflects a large percentage of HIV patients who were ill-informed about their conditions or did not receive sufficient education about disease transmission. The number of children born with HIV who could not live past 4 years old also indicates that there might be a lack of healthcare facilities and treatments for them. A report shows that while the increasing usage of antiretroviral drugs to prevent mother-to-child transmission has reduced AIDS death rate in developed countries, the majority of African patients could not receive such treatment because of high expenses (The World Bank, 2005). For these reasons, an increase in the number of HIV deaths per 1,000 live births has a significant association with a decrease in life expectancy, especially in Africa.

Average years of education had a positive relationship with life expectancy. Intuitively, with increase in average education level, population becomes more health-conscious and affords access to high-quality healthcare; therefore, they are able to achieve higher life expectancy (Luy et al., 2019). This is reflected by the strong positive correlation between BMI and years of schooling found from this study. Other studies have shown that education attainment has a positive effect on increasing life expectancy. While there is a rapid rise in life expectancy among highly-educated population, there are little changes in life expectancy among the less educated (Meara et al., 2008).

On the other hand, there was no significant evidence that GDP could explain the variability in life expectancy at a global level. There have been studies at national level indicating that higher income is associated with higher life expectancy (Chetty et al., 2016). On the other hand, a longitudinal study by Lutz and Kebede (2018) including the model with both education and income variables between 1970-2015 concluded while income becomes insignificant, education remains significant. They reasoned that in recent decades, there has been a shift in the disease burden from infectious to chronic diseases, which are largely life-style related. Thus, improving education, instead of income, results in substantial increase in life expectancy.

Our study encountered some limitations. First of all, the data was collected online from an unreliable source. There were multiple missing data and extreme observations that were unaccounted for. For future analysis, information should be compiled by more reliable researchers. Also, there was a large within-continent variability. For example, there were large standard deviations in average life expectancy in Asia and Americas because they are such large continents with many politically and economically distinctive regions. Smaller regions with comparable political, socioeconomic, and health conditions, should be divided as fixed effects. There are also possible confounding factors (e.g. nutrition, air quality, level of urbanization, violent conflicts) that could be influential to life expectancy and worth investigating. But overall, the present study reaffirms significant variables that justify the differences in life expectancy on a global scale. Log(GDP) becomes a less relevant indicator of life expectancy. Those variables include country status, continent, years of schooling, and HIV deaths per 1,000 live births. There is no evidence for any differences in annual increase among continents. In addition, there is a stronger negative effect of HIV deaths per 1,000 live births in Africa than in any other continents.

**References**

Andre, F.E., Booy, R., Bock, H.L., Clemens, J., Datta, S. K., John, T.J., Lee, B.W., Lolekha, S., Peltola, H., Ruff, T.A., Santosham, M., Schmitt, H.J. (2008). Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bulletin of the World Health Organization*, 86(2), 81-160.

Becker G.S., Philipson T.J., Soares R.R. (2005). The quantity and quality of life and the evolution of world inequality. *The American Economic Review*, 95(1), 277-291.

Chetty, R, Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001–2014. *Journal of the American Medical Association*, 315(16), 1750–1766.

Hadfield, J.D. (2010, January). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software,* 33(2), 1-22.

Human Development Report (2019). Human Development Indices and Indicators: 2018 Statistical Update. 15-110.

Kabir, M. (2008). Determinants of Life Expectancy in Developing Countries. The Journal of Developing Areas, 41(2), 185-204.

Khan, A., Khan S., &Khan, M. (2016, November). Factors affecting life expectancy in developed and developing countries of the world (An approach to available literature). International Journal of Yoga, Physiotherapy and Physical Education, 1(1), 04-06.

Kpolovie, P. J., Oshodi, P. O., and Iwuchukwu, H. (2016). Continental Inequality in Life Expectancy. *European Journal of Biology and Medical Science Research*, 4(6), 30-47.

Life Expectancy (WHO): Statistical Analysis on factors influencing Life Expectancy. (2019, May 2). Retrieved from https://www.kaggle.com/augustus0498/life-expectancy-who

Lin, R., Chen, Y., Chien L., and Chan C. (2012). Political and social determinants of life expectancy in less developed countries: a longitudinal study. BMC Public Health, 12(85).

Lutz, W., Kebede, E. Education and Health: Redrawing the Preston Curve. Population and Development Review, 2018; DOI: 10.1111/padr.12141

Luy, M., Zannella, M., Wegner-Siegmundt, C., Minagawa, Y., Lutz, W., Caselli G. (2019). The impact of increasing education levels on rising life expectancy: a decomposition analysis for Italy, Denmark, and the USA. *Genus*, 75(11), doi:10.1186/s41118-019-0055-0.

Meara, E. R., Richards, S., and Cutler, D. M. (2008). The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981–2000.

Mondal, M. N. I, and Shitan, M. (2014). Relative Importance of Demographic, Socioeconomic and Health Factors on Life Expectancy in Low- and Lower-Middle-Income Countries. Journal of Epidemiology, 24(2), 117-124.

The World Bank. (2006). Disease and Mortality in Sub-Saharan Africa. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK2279/pdf/Bookshelf\_NBK2279.pdf.

The World Health Organization. (2015)‎. Health in 2015: from MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals.

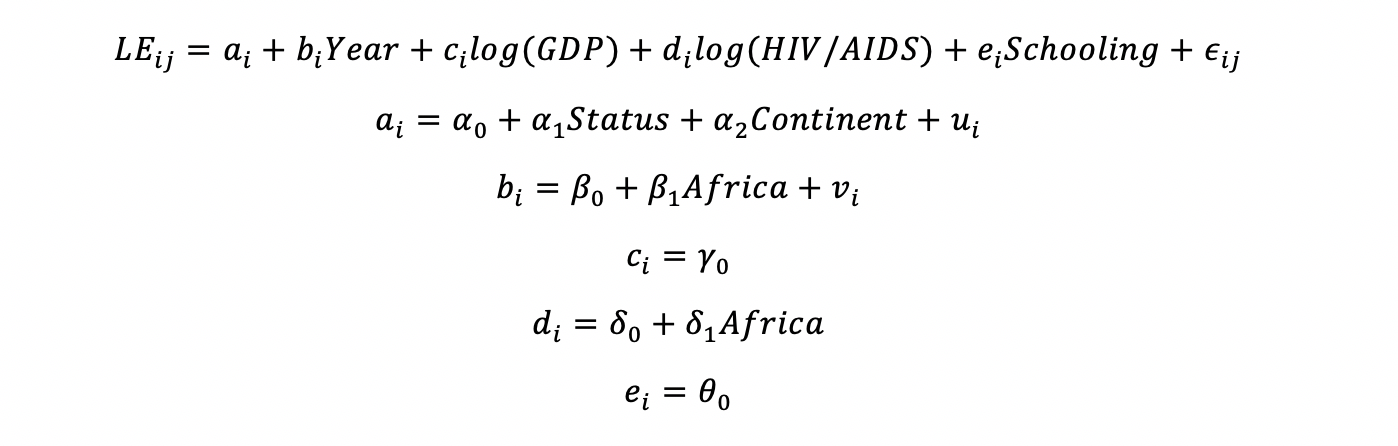
**Annotated Appendix**

**Table A. Markov Chain Monte Carlo Sampler for Multivariate Generalised Linear Mixed Models (MCMCglmm)**

|  | Model 1: Socioeconomic factors | | Model 2: Public health factors | | Model 3: Covariates only | | Model 4: Covariates and country status | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Predictors | β | 95% CI | β | 95% CI | β | 95% CI | β | 95% CI |
| Intercept | 7.18 | (66.2, 77.2) | 72.4 | (68.0, 76.5) | 63.52 | (59.2, 68.0) | 68.9 | (65.02, 73.4) |
| Status  Developed (ref.)  Developing | -  -5.43\* | -  (-9.10, -1.6) | -  -5.84\* | -  (-8.80, -2.68) | -  -4.40\* | -  (-6.98, -1.21) | -  -5.11\* | -  (-7.82, -2.44) |
| Continent  Oceania (ref.)  Africa  Asia  Americas  Europe | -  -13.7\*  -0.83  3.37  1.74 | -  (-17.9, -9.68)  (-4.97, 3.35)  (-1.03, 7.61)  (-2.08, 6.51) | -  -8.43\*  0.10  3.83\*  3.80 | -  (-11.9, -4.9)  (-3.42, 3.32)  (0.44, 7.52)  (3.80, -0.22) | -  -4.45\*  0.42  4.47\*  3.05 | -  (-7.56, -1.34)  (-2.73, 3.40  (1.39, 7.95)  (-0.32, 6.50) | -  -7.37\*  0.30  3.70\*  3.21 | -  (-10.9, -4.24)  (-2.54, 3.69)  (0.63, 7.00)  (0.03, 6.74) |
| Log(GDP) | 0.01 | (-0.07, 0.07) |  |  | -0.02 | (-0.09, 0.05) | -0.006 | (-0.08, 0.06) |
| Years of Schooling | 0.18\* | (0.03, 0.36) |  |  | 0.44\* | (0.30, 0.61) | 0.30\* | (0.17, 0.44) |
| Log(HIV death rate) |  |  | -0.18 | (-0.51, 0.14) | -1.81\* | (-2.07, -1.57) | -0.25 | (-0.66, 0.04) |
| Log(Hepatitis B) |  |  | 0.10 | (-0.05, 0.25) |  |  |  |  |
| Log(Polio) |  |  | -0.02 | (-0.20, 0.12) |  |  |  |  |
| Log(Diphtheria) |  |  | 0.10 | (-0.09, 0.28) |  |  |  |  |
| Years (centered) | 0.11\* | (0.04, 0.19) | 0.26\* | (0.24, 0.28) | 0.21\* | (0.18, 0.24) | 0.22\* | (0.20, 0.26) |
| African х Years (centered) | 0.43\* | (0.33, 0.51) |  |  |  |  | 0.02 | (-0.04, 0.07) |
| African x log(HIV death rate) |  |  | -2.75\* | (-3.25, -2.24) |  |  | -2.71\* | (-3.19, -2.21) |

\* p-value ≤ 0.05

**Final model equation:**

****

**R-codes:**

library(mosaic)

library(tidyverse)

library(ggplot2)

library(gridExtra)

library(dplyr)

library(stats)

library(lme4)

library(GGally)

library(Hmisc)

library(broom)

library(MASS)

library(mnormt)

life <- read\_csv("~/Stats 316 F19/Project/Leon-Jessie-Calvin/Leon-folder/Life-Analysis.csv")

life <- life %>%

filter(!is.na(Lifeexpectancy) & !is.na(Alcohol)) %>%

mutate(cyear = Year - 2000,

log.GDP = log(GDP),

log.pct.exp = log(percentageexpenditure+0.5),

log.HepB = log(HepatitisB),

log.Polio = log(Polio),

log.Diph = log(Diphtheria),

log.HIV = log(`HIV/AIDS`),

africa = ifelse(Continent == "Africa",1,0),

asia = ifelse(Continent == "Asia",1,0),

europe = ifelse(Continent == "Europe",1,0),

america = ifelse(Continent == "Americas",1,0),

oceania = ifelse(Continent == "Oceania",1,0),

african2 = ifelse(africa==1, "African","Non-African"),

developing = (ifelse(Status == "Developing",1,0)),

GDP1000 = GDP/1000) %>%

na.omit()

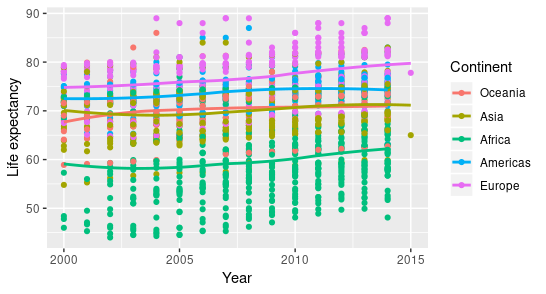
life$Continent <- factor(life$Continent, levels = c("Oceania", "Asia", "Africa", "Americas", "Europe"))

ggplot(data = life, aes(y=Lifeexpectancy,x=Year)) +

geom\_point(aes(color = Continent)) +

geom\_smooth(aes(color=Continent), se=F) +

labs(x="Year", y ="Life expectancy")



```{r} code for the map

#map

WorldData <- map\_data('world') %>% filter(region != "Antarctica") %>% fortify

p <- ggplot() +

geom\_map(data = WorldData, map = WorldData,

aes(x = long, y = lat, group = group, map\_id=region),

fill = "white", colour = "#7f7f7f", size=0.5) +

geom\_map(data = life.map, map=WorldData,

aes(fill=M, map\_id=Country),

colour="#7f7f7f", size=0.5) +

coord\_map("rectangular", lat0=0, xlim=c(-180,180), ylim=c(-60, 90)) +

scale\_fill\_continuous(low="blue", high="red")+

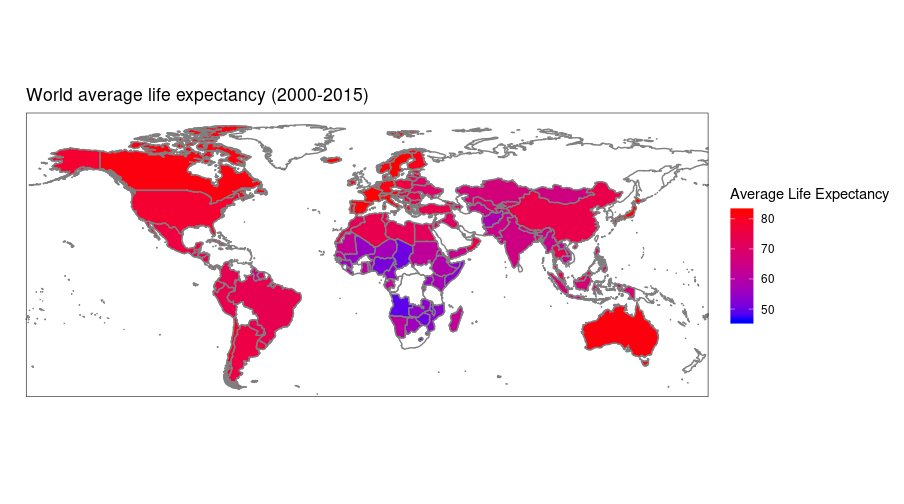
scale\_y\_continuous(breaks=c()) +

scale\_x\_continuous(breaks=c()) +

labs(fill="Average Life Expectancy", title="World average life expectancy (2000-2015)", x="", y="") +

theme\_bw()

p

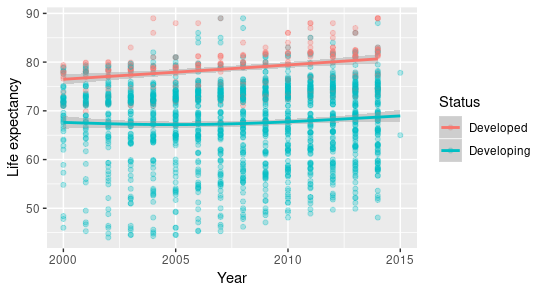


ggplot(data = life, aes(y=Lifeexpectancy,x=Year)) +

geom\_point(aes(color = Status), alpha = 0.3) +

geom\_smooth(aes(color=Status)) +

labs(x="Year", y ="Life expectancy")



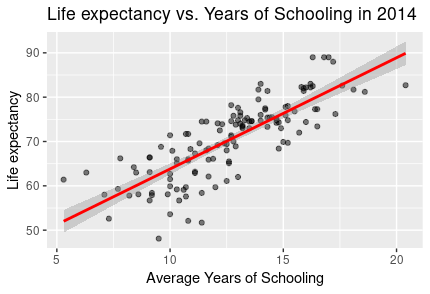
year2014 <- life %>% filter(Year == 2014)

ggplot(data = year2014, aes(y=Lifeexpectancy,x=Schooling)) +

geom\_point(alpha = 0.5) +

geom\_smooth(method = "lm", color="red") +

labs(x="Average Years of Schooling", y ="Life expectancy", title = "Life expectancy vs. Years of Schooling in 2014")



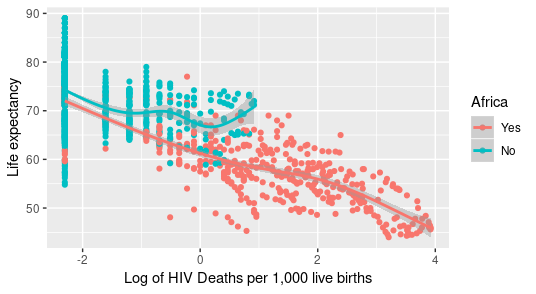
ggplot(data = life, aes(y=Lifeexpectancy,x=log.HIV)) +

geom\_point(aes(color=african2)) +

geom\_smooth(aes(color=african2)) +

labs(x="Log of HIV Deaths per 1,000 live births", y ="Life expectancy") +

scale\_color\_discrete(name = "Africa", label = c("Yes","No"))



theme.1 <- theme(axis.title.x = element\_text(size = 14),

axis.title.y = element\_text(size = 14),

plot.title=element\_text(hjust=.9,face="italic",size=12))

year2014 <- life %>% filter(Year == 2014)

gdp2014 <- ggplot(data = year2014, aes(y=Lifeexpectancy, x=log.GDP)) +

geom\_point() +

geom\_smooth() +

theme.1 +

labs(y="Life expectancy", title= "Life expectancy vs. log(GDP) in 2014")

gdp2010 <- ggplot(data = year2010, aes(y=Lifeexpectancy, x=log.GDP)) +

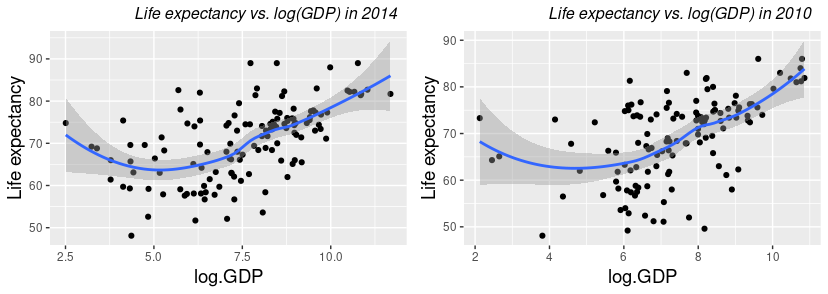
geom\_point() +

geom\_smooth() +

theme.1 +

labs(y="Life expectancy", title= "Life expectancy vs. log(GDP) in 2010")

grid.arrange(gdp2014, gdp2010, ncol=2)



lm\_regression <- life %>%

group\_by(Country) %>%

do(fit = lm(Lifeexpectancy ~ cyear, data=.))

sd\_filter <- life %>%

group\_by(Country) %>%

summarise(sds = sd(Lifeexpectancy))

lm\_regression <- lm\_regression %>%

right\_join(sd\_filter, by="Country") %>%

filter(!is.na(sds))

lm\_info1 <- lm\_regression %>%

tidy(fit) %>%

ungroup() %>%

select(Country, term, estimate) %>%

spread(key = term, value = estimate) %>%

rename(rate = cyear, intercept = `(Intercept)`)

lm\_info2 <- lm\_regression %>%

tidy(fit) %>%

ungroup() %>%

select(Country, term, std.error) %>%

spread(key = term, value = std.error) %>%

rename(se\_rate = cyear, se\_intercept = `(Intercept)`)

lm\_estimates <- lm\_regression %>%

glance(fit) %>%

ungroup() %>%

select(Country, r.squared, df.residual) %>%

inner\_join(lm\_info1, by = "Country") %>%

inner\_join(lm\_info2, by = "Country") %>%

mutate(tstar = qt(.975, df.residual),

intlb = intercept-tstar\*se\_intercept,

intub = intercept+tstar\*se\_intercept,

ratelb = rate-tstar\*se\_rate,

rateub = rate+tstar\*se\_rate)

life\_categorical <- life %>%

select(Country, Status, Continent, african2, developing) %>%

group\_by(Country) %>%

filter(row\_number() == 1)

lm\_estimates <- lm\_estimates %>%

right\_join(life\_categorical, by = "Country")

status.slope <- ggplot(data = lm\_estimates, aes(y=rate,x=Status)) +

geom\_boxplot() +

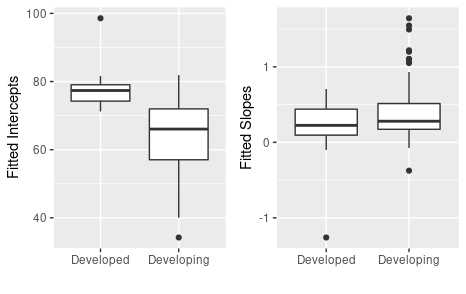
labs(y = "Fitted Slopes", x="")

status.intercept <- ggplot(data = lm\_estimates, aes(y=intercept,x=Status)) +

geom\_boxplot() +

labs(y = "Fitted Intercepts",x="")

grid.arrange(status.intercept,status.slope, ncol=2)



continent.intercept <- ggplot(data = lm\_estimates,

aes(y = intercept, x = Continent)) +

geom\_boxplot() +

labs(y = "Fitted Intercepts", x = "")

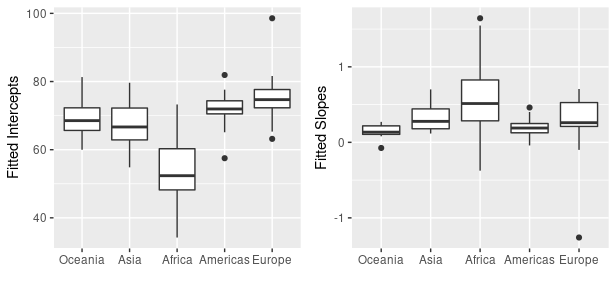
continent.slope <- ggplot(data = lm\_estimates,

aes(y = rate, x = Continent)) +

geom\_boxplot() +

labs(y = "Fitted Slopes", x="")

grid.arrange(continent.intercept, continent.slope, ncol=2)



# Parametric Bootstrap

#m0 is the lmer model under the null hypothesis (i.e. the smaller model)

#mA is the lmer model under the alternative

bootstrapAnova <- function(mA, m0, B=1000){

oneBootstrap <- function(m0, mA){

d <- drop(simulate(m0))

m2 <-refit(mA, newresp=d)

m1 <-refit(m0, newresp=d)

return(anova(m2,m1)$Chisq[2])

}

nulldist <- replicate(B, oneBootstrap(m0, mA))

ret <- anova(mA, m0)

ret$"Pr(>Chisq)"[2] <- mean(ret$Chisq[2] < nulldist)

names(ret)[8] <- "Pr\_boot(>Chisq)"

attr(ret, "heading") <- c(attr(ret, "heading")[1],

paste("Parametric bootstrap with", B,"samples."),

attr(ret, "heading")[-1])

attr(ret, "nulldist") <- nulldist

return(ret)

}

# Unconditional means model

life.0 <- lmer(Lifeexpectancy ~ 1 + (1|Country), REML=T, data = life)

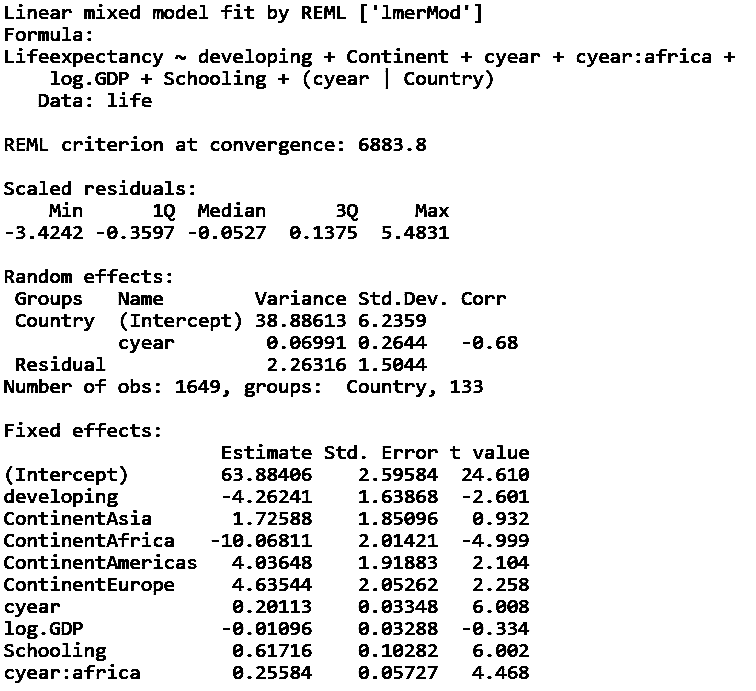
# Unconditional growth model

life.00 <- lmer(Lifeexpectancy ~ cyear + (cyear|Country), REML=T, data=life)

#Socioeconomic model

socioecon1 <- lmer(Lifeexpectancy ~ developing + Continent + cyear + cyear:africa + log.GDP + Schooling + (cyear|Country), data = life, REML = T)

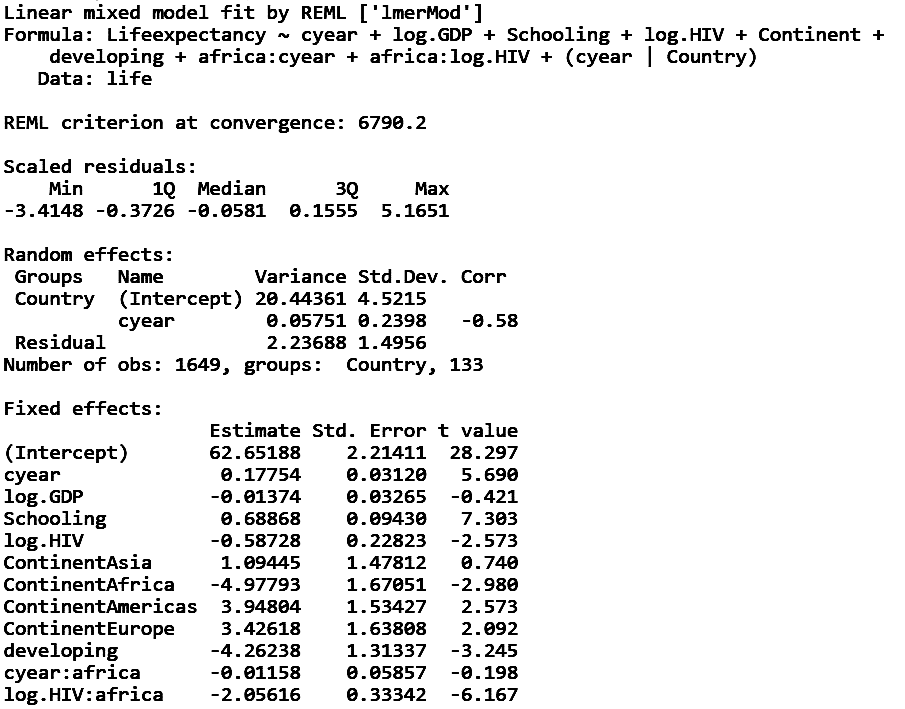
summary(socioecon1)



#Final model

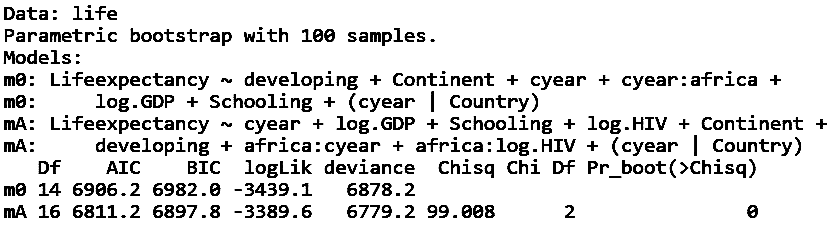
final.model <- lmer(Lifeexpectancy ~ cyear + log.GDP + Schooling + log.HIV + Continent + developing + africa:cyear + africa:log.HIV + (cyear|Country), REML=T, data=life)

summary(final.model)



bRLRT = bootstrapAnova(mA=final.model, m0=socioecon1, B=100)

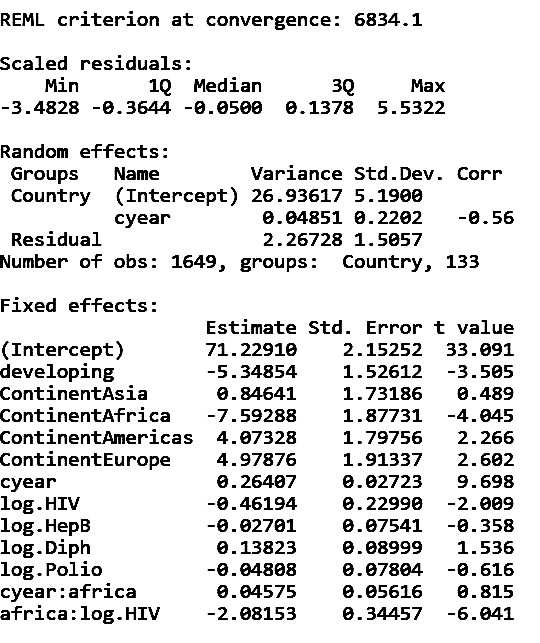
bRLRT



# Health model

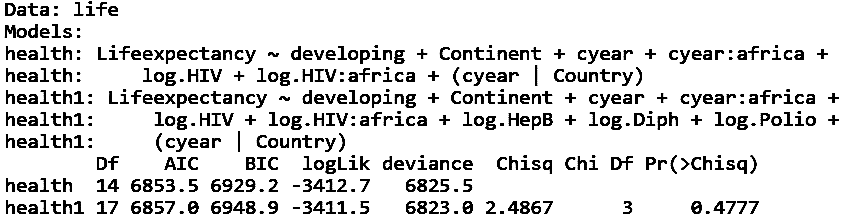
health1 <- lmer(Lifeexpectancy ~ developing + Continent + cyear + cyear:africa + log.HIV + log.HIV:africa + log.HepB + log.Diph + log.Polio + (cyear|Country), data=life, REML=T)

summary(health1)



health <- lmer(Lifeexpectancy ~ developing + Continent + cyear + cyear:africa + log.HIV + log.HIV:africa + (cyear|Country), data=life, REML=T)

anova(health, health1)

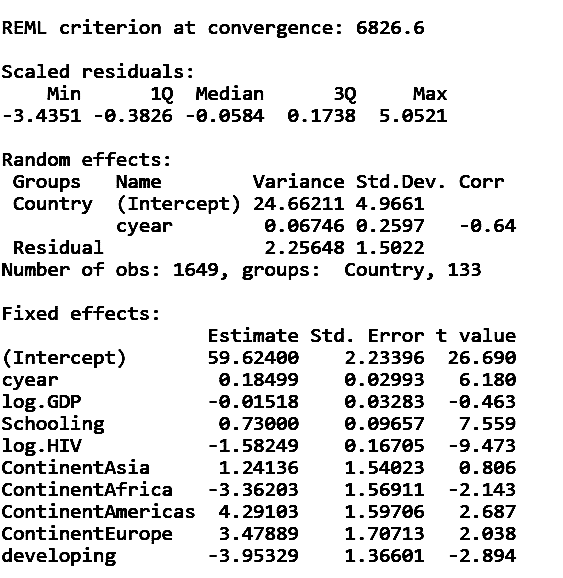


=> log.HepB, log.Diph, log.Polio are not significant

## Covariates Only

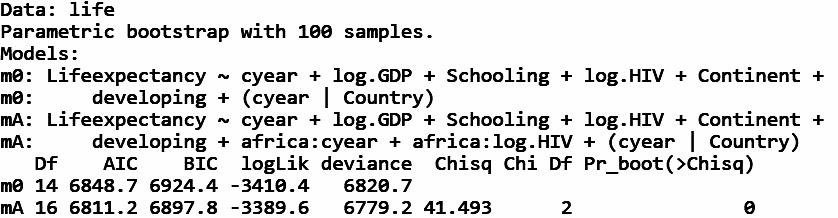
m0 <- lmer(Lifeexpectancy ~ cyear + log.GDP + Schooling + log.HIV + Continent + developing + (cyear|Country), REML=T, data=life)

summary(m0)



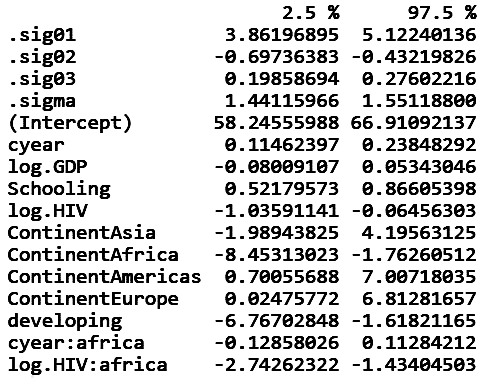
bRLRT = bootstrapAnova(mA=final.model, m0=m0, B=100)

bRLRT



## Confidence Intervals

confint(final.model, method = "boot", boot.type = "perc", nsim = 600)



Bayesian models

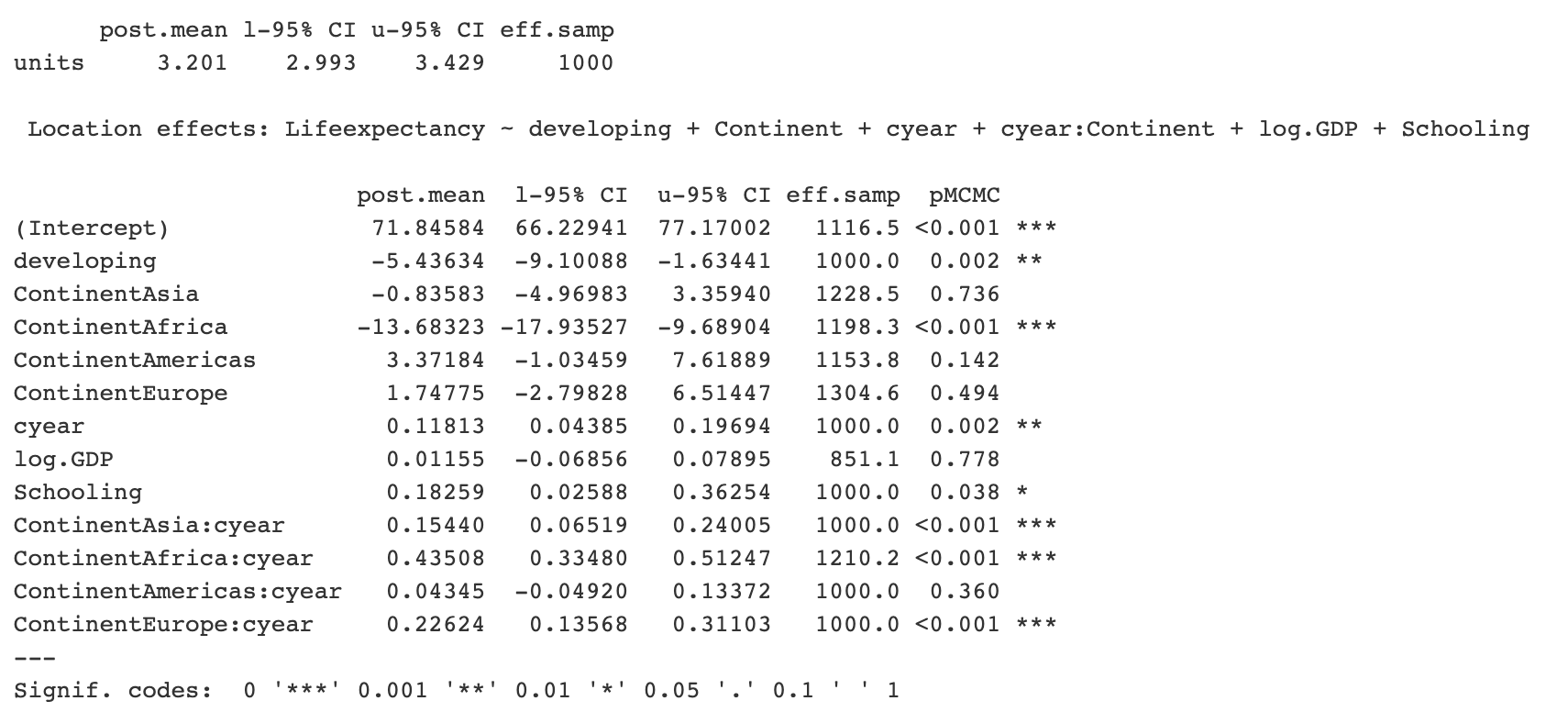
## Socio Econ

socioecon.bayes <- MCMCglmm(Lifeexpectancy ~ developing + Continent + cyear + cyear:Continent + log.GDP + Schooling, random = ~Country + cyear, data=life)

summary(socioecon.bayes)

#autocorr(socioecon.bayes$VCV)

#autocorr(socioecon.bayes$Sol)



## Health

health.bayes <- MCMCglmm(Lifeexpectancy ~ developing + Continent + cyear + cyear:africa + log.HIV + log.HIV:africa + log.HepB + log.Diph + log.Polio,

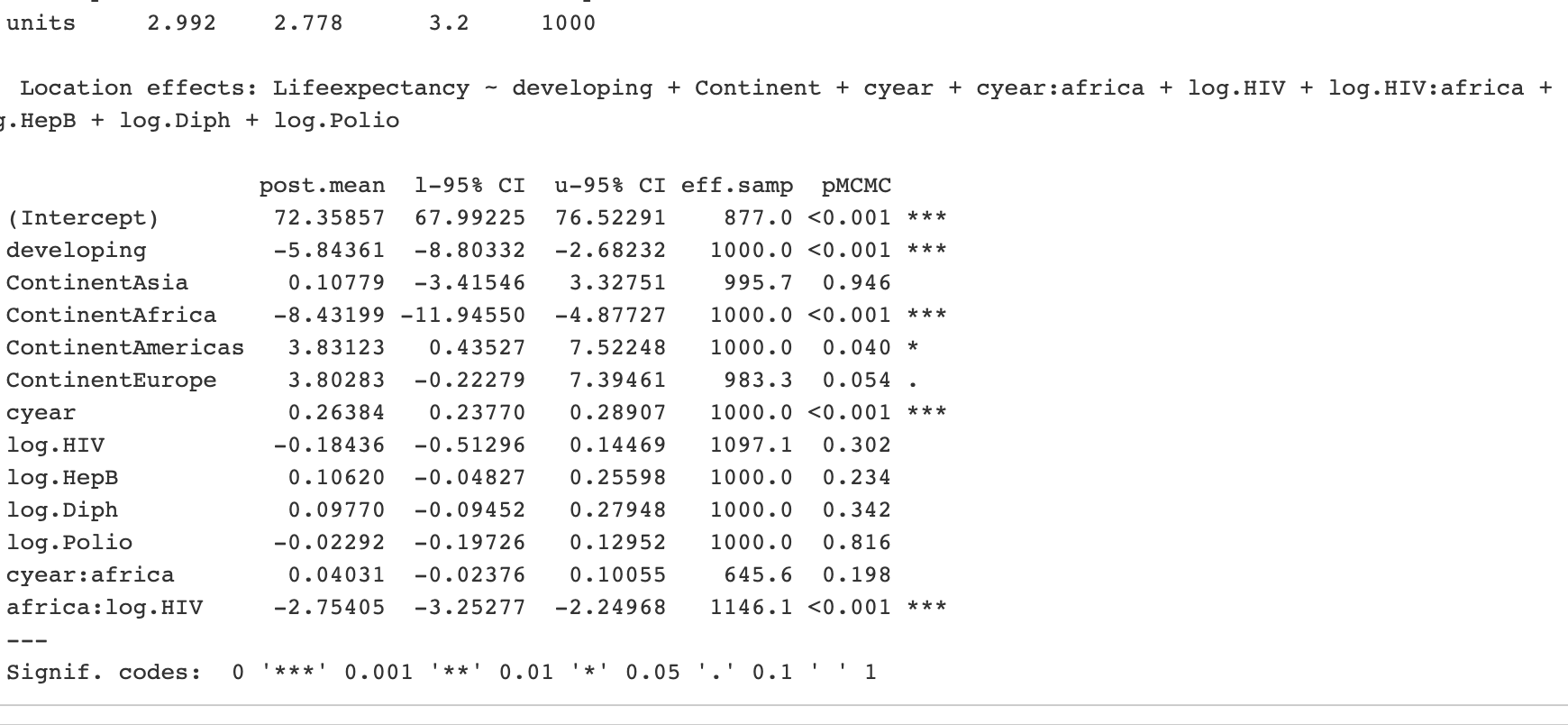
random = ~Country + cyear,

data=life )

summary(health.bayes)

#autocorr(health.bayes$VCV)

#autocorr(health.bayes$Sol)



## Model 3:

bayes.3 <- MCMCglmm(Lifeexpectancy ~ cyear + log.GDP + Schooling + log.HIV + Continent + developing,

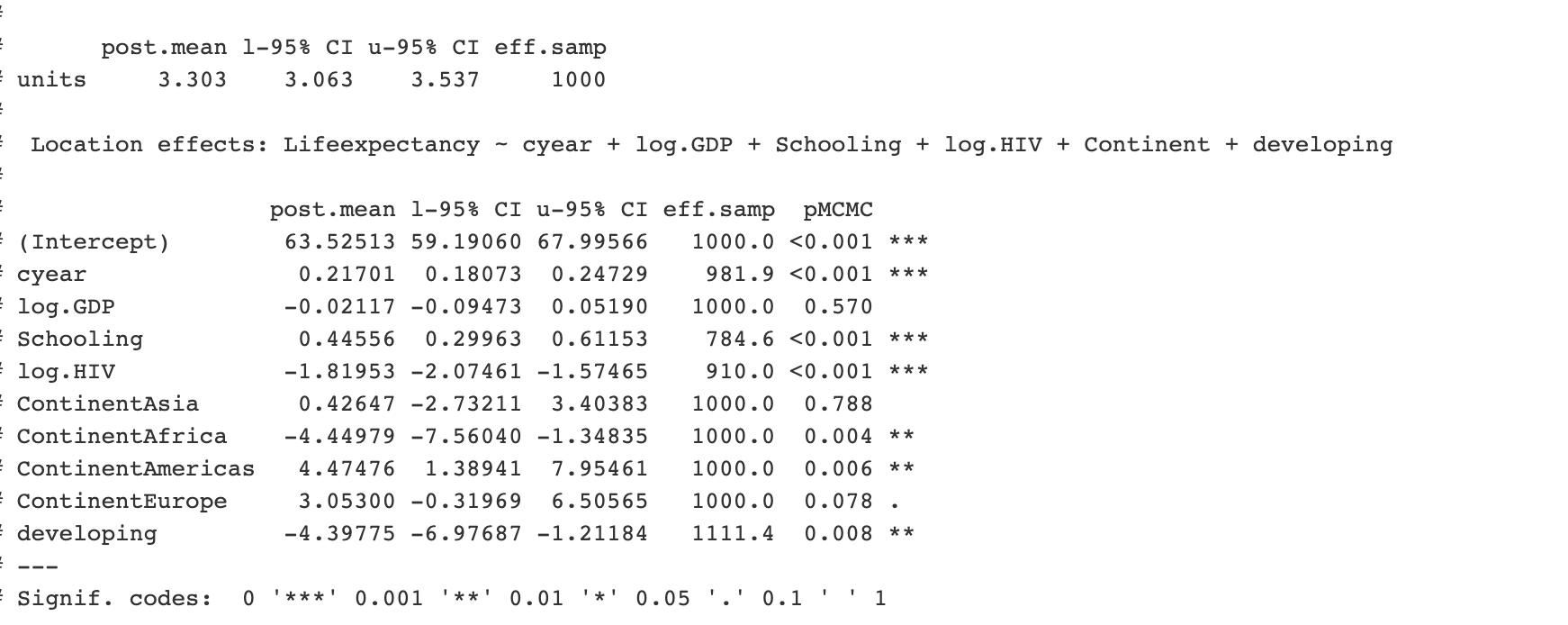
random = ~Country + cyear,

data=life)

summary(bayes.3)

#autocorr(bayes.3$VCV)

#autocorr(bayes.3$Sol)



## Model 4 (final)

bayes.4 <- MCMCglmm(Lifeexpectancy ~ cyear + log.GDP + Schooling + log.HIV + Continent + developing + africa:cyear + africa:log.HIV,

random = ~Country + cyear,

data=life)

summary(bayes.4)

#autocorr(bayes.4$VCV)

#autocorr(bayes.4$Sol)