The Subordination of Women and National Health Outcomes

A Bayesian Hierarchical Regression Model Approach

Kaylee Britton Hodgson

A selected project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Richard Warr, Chair
Gilbert Fellingham
Garritt Page

Department of Statistics

Brigham Young University

April 2019

ABSTRACT

The Subordination of Women and National Health Outcomes
A Bayesian Hierarchical Regression Model Approach

Kaylee Britton Hodgson
Department of Statistics, BYU
Master of Science

How does the treatment of women impact health outcomes? Specifically, how are the health of women, children, and the general public in a country impacted by how systematically and severely women are subordinated in the household in that country? I use a Bayesian hierarchical regression model to test whether poor measures of health outcomes are associated with higher levels of subordination of women. The Bayesian hierarchical regression model is proposed over using multivariate or simple regression analyses because of 1) the non-random missingness in the data, and 2) the relationship between the health outcomes. Before implementing the proposed Bayesian hierarchical regression model, I use a simulation study to test it against these other models. I find that overall the Bayesian hierarchical model tends to estimate with less bias and more accuracy than the other models.

Using the hierarchical model, I find that higher levels of subordination of women in the household are significantly and consistently associated with worse health outcomes for women and children, and that while the results are less consistent for overall health, that the same relationship exists in most of those models. I propose that political and community leaders working to improve health outcomes place higher focus on the empowerment of women to achieve their objectives.

Keywords: hierarchical model, bayesian, multivariate, regression, missing data

ACKNOWLEDGMENTS

I would first like to thank my project chair Dr. Richard Warr of the Statistics Department at Brigham Young University. Dr. Warr mentored and guided my work on this project, while also listening to and encouraging my ideas for it. He was always willing to take time to teach me new concepts, and to help me improve the research and writing.

I would also like to thank Dr. Valerie M. Hudson of the Department of International Affairs at Texas A&M University, Dr. Perpetua Lynne Nielsen of the Department of Statistics at Brigham Young University, and Dr. Donna Lee Bowen of the Department of Political Science at Brigham Young University for their mentorship and their large contributions to this project. The dataset used in this analysis was originally conceptualized by them, and they allowed me to be an active participant in building the dataset. I especially appreciate the countless hours they spent with me teaching me about the theory behind the dataset and analysis, and for the analysis and writing opportunities they gave me that better prepared me for completing this project.

I would like to acknowledge and thank Dr. Gilbert Fellingham and Dr. Garritt Page of the Statistics Department at Brigham Young University, as committee members for this project. I am very grateful for both of their valuable guidance and comments on this project.

Finally, I would like to profoundly thank my husband, James, for the consistent support, love, and encouragement through this degree, and specifically in working on this project. His sacrifices for, and emphatic encouragement of, my ambitions have made this possible. Thank you.

CONTENTS

Contents iv

1 Background and Data 1

1.1 Subordination of Women and Health 1

1.2 Variable Descriptions 4

2 Statistical Methods 15

2.1 Preliminary Approaches 15

2.2 Proposed Model 18

2.3 Simulation Study 21

3 Results and Discussion 27

3.1 Model Diagnostics 27

3.2 Model Results 30

3.3 Discussion and Conclusion 32

Bibliography 35

CHAPTER 1

BACKGROUND AND DATA

"People everywhere have divided themselves into men and women. And almost everywhere men have got the better deal..." (Harari 2011). This dynamic between the sexes is well embedded in society, but has specifically plagued some areas of the world more than others. The variation in this dynamic allows for empirical testing on the impacts of higher levels of women's subordination. I analyze the health and wellbeing impact of this dynamic between the sexes that has consistently left women with the disadvantage. I argue that poor measures of national health are associated with the subordination of women at the household level (quantified as the Syndrome, introduced in Section 1.2).

I specifically test whether women's health, children's health, and the general health of a nation can be explained by the treatment of women in the home. I hypothesize that countries with higher levels of systematic subordination of women will perform worse on health indicators for women, children, and the overall public. I test the hypothesis using a Bayesian hierarchical model, which I expound further on following a discussion of the subject matter and an introduction of the variables.

1.1 Subordination of Women and Health

The subordination of women not only negatively impacts the health of the women themselves, but also their children (Nour 2009), and other members of their communities. When women are disempowered, their ability to care for themselves, care for their children, or impact the community is greatly diminished.

Some of the aspects of the Syndrome scale (conceptualized in Section 1.2), which is the scale used in this study that measures the subordination of women in the home, have more obvious ties to health outcomes, specifically for the women, such as cousin marriage, overall

levels of violence against women in society, legal exoneration for rapists, and occurrences of femicide. I discuss some other aspects of the Syndrome that have less obvious ties to health outcomes.

Child marriage, one aspect of the Syndrome scale measurements, has been found to put women at an increased risk for sexually transmitted disease, depression, cervical cancer, malaria, obstetric fistulas, maternal mortality, and nonuse of maternal health services (Raj and Boehmer 2013; Nour 2006, 2009). Child marriage is clearly detrimental to the physical health of the girls involved. A UNICEF report summarizes the issue: "Child marriage is a violation of human rights, compromising the development of girls and often resulting in early pregnancy and social isolation. Young married girls face onerous domestic burdens, constrained decision-making and reduced life choices" (United Nations Children's Fund 2005). Child marriage also puts the offspring of the child brides at greater risk for premature birth, and therefore infant mortality (Raj and Boehmer 2013; Nour 2009). Wodon (2015) expounds on the health consequences of child marriage for both the girls and subsequent children:

> The practice is also associated with a higher risk of intimate partner violence and other forms of violence, which may lead to severe injuries and even death, as well as losses in earnings and out-of-pocket costs for healthcare. Next, child marriage is associated with higher risks of maternal mortality and morbidity, as well as malnutrition and depression. It is also associated with poor sexual and reproductive health outcomes including through sexually transmitted diseases. The practice also has consequences for children in terms of infant mortality, low birth weight, and stunting. Finally, child marriage also leads to losses in empowerment and decision-making as well as participation more generally (Wodon 2015).

Polygyny is another component of the Syndrome that has been individually associated with poor health outcomes. Polygyny puts a significant strain on the family's resources, which discourages or makes impossible investments in the women and children, with the rare exception of a polygynous family that is wealthy. In polygynous families, there is less investment in, and focus on, the health and nutrition of the wives and their children (McDermott 2018). Polygyny has been significantly tied to higher birth rates, HIV infection, low age of marriage for girls, high maternal mortality, lower life expectancy, higher levels of sex trafficking, and higher levels of domestic violence (McDermott and Cowden 2014). In addition

to the physical woes, studies have also found negative psychological effects associated with polygyny, specifically for the wives and children (Supreme Court of British Columbia 2011). The anxiety and depression felt by the women and children involved in the polygynous relationships is also extended to the women and children in monogamous marriages, who feel insecure in their family circle, worried that their husband or father may follow the same trend and decide to marry additional women (Hudson et al. forthcoming).

If there is a strong preference for sons over daughters in a society, another aspect of the Syndrome, then the sons are valued over the daughters. This leads to families caring for the needs of their sons before, or instead of, putting resources toward their daughters. Intense son preference can also lead to high fertility and high maternal mortality rates in some societies, which negatively impacts the health of both the mother and the children (Afghan Women's Writing Project: Oral Stories Project 2014). A study with 30 focus groups of 250 women in Pakistan emphasized the appalling health effects of son preference on women:

> Women repeatedly become pregnant to deliver as many children as possible preferably sons to become worthier... Women's ability to enforce contraceptive use is very limited because of the unilateral power that their male partners/husbands exercise in fertility decisions. A woman is persuaded to continue bearing children until the family has at least one son; she sometimes delivers 7 or more daughters in order to accomplish the objective... Except the post-delivery period in case of the male baby, when higher allowances are given so that boy can be breastfed, generally meager nutritional allocation and repeated pregnancies make them malnourished (Rizvi et al. 2014).

One consideration in the occurrence of son preference is patrilocality. Patrilocality, again another aspect of the Syndrome, is the cultural practice of a woman moving away from her family when she is married to live with or near her husband's family. Under this practice, because the daughters leave their home to live with their husband's family but the sons stay with their own family, taking care of the sons is the best way for parents to ensure that they will be cared for in their old age. Again, this gives motivation for the families to care for their sons' health before their daughters' (Das Gupta 2009).

In addition to these direct linkages between different forms of women's disempowerment and the specific health outcomes, I propose that there is a direct link between their overall treatment and the health of the women, children, and community. Valerie Hudson, in conversation, has suggested that women are often the main caregivers for both their children and their extended family, and therefore often take on the majority of the responsibilities in the caregiving of their families and communities. Women are often the ones choosing and preparing the food in their home, taking their children and elderly family members to the doctor, choosing whether their children get immunized, treating illness, teaching their children about hygiene and nutrition, along with many other care responsibilities that disproportionately fall on women. As such, when women's voices and interests are subordinated, not only does their own health suffer, but their children and other adults in their family and community who rely on these women suffer negative health consequences as well.

As discussed above, researchers have addressed the linkage between many of the individual indicators of the subordination of women and indicators of health outcomes, an important step. However, in order to determine how the overall disempowering and harming of women affects these health outcomes, we need an indicator that measures these women's overall situation in the home, and we need to evaluate these different aspects of health for women, children, and the community.

1.2 Variable Descriptions

The Syndrome is the main variable of interest in these analyses, and measures the subordination of women in the home. The Syndrome scale is first introduced and its methodology is extensively discussed in Hudson et. al. (forthcoming), which is the first literature to address the linkage between this scale and the health outcomes used in this paper, as well as other outcomes of national wellbeing.

My purpose is to find whether the Syndrome is specifically a good predictor of a country's health and wellbeing, by implementing an analysis method that also deals with

the missing data patterns in the dataset (expounded on in Section 2.1). In this section, I first introduce the Syndrome variable, then the control variables and the reasons for including each in the analysis. Finally, I discuss the indicators of national health and wellbeing that are the dependent values in the analysis.

The Syndrome

The Syndrome scale was created by The WomanStats Project (2017), and combines 11 components that are considered the key indicators of household-level subordination of women, measuring from 2010 to 2015. Hudson et. al. (forthcoming) extensively explain the theory behind the combination of these 11 components in their forthcoming book, but I will only briefly introduce the variable here, accepting their strong theoretical and statistical support for the choice in variable components.

Figure 1.1 is a representation of how the individual components combine to create the overall Syndrome score, again a brief summary of the theory for the variable's components. The combination of these individual components gives an indication of the level that women's interests are subordinated to the interests of men within a country. Figure 1.2 is a mapping of Syndrome scores based on the 11 components, created by The WomanStats Project (2017). Countries which perform worse on these 11 measures have higher Syndrome scores. To give an example, South Sudan has the highest Syndrome score of 16, indicating that this scale identifies South Sudan as the country with the highest levels of systematic subordination of women at the household level. On the other hand, Australia, Netherlands, Norway, Sweden, and Switzerland all have the lowest Syndrome score possible (0), indicating that systematic subordination of women in the household is essentially nonexistent in these countries, based on the scale.

The eleven components of the Syndrome are:

- Prevalence of patrilocal marriage

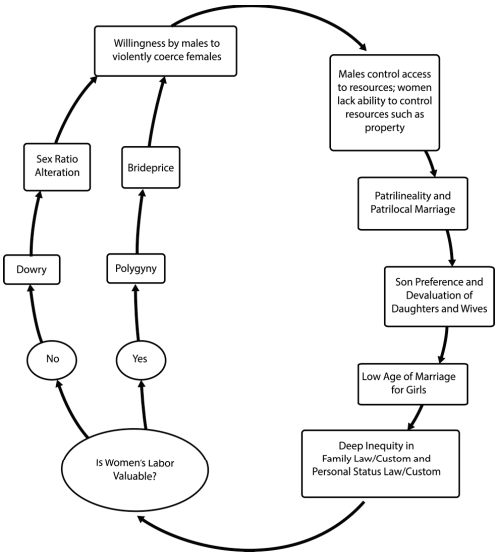

Figure 1.1: Subcomponents of the Syndrome

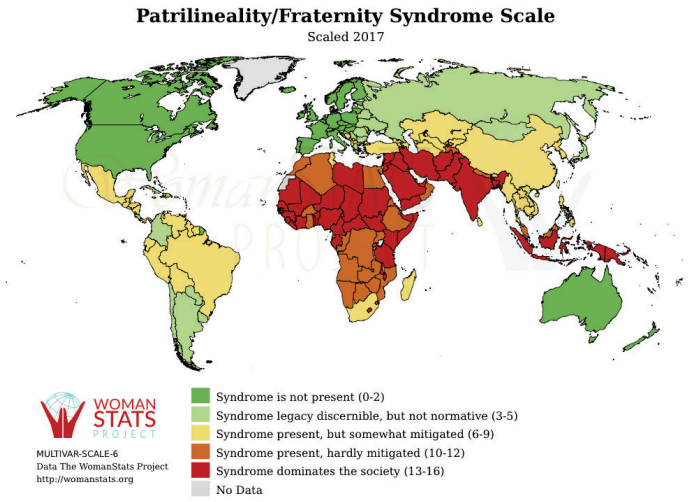

Figure 1.2: Map of Countries' Performance on the Syndrome Scale

- Brideprice/dowry
- Prevalence and legality of polygyny
- Cousin marriage
- Age of marriage for girls
- Women's Property Rights in Law and Practice
- Son preference and sex ratio alteration
- Inequity in Family Law/Custom Favoring Males
- Overall level of violence against women in society
- Societal sanction for femicide
- Legal exoneration for rapists who offer to marry their victims

This scale, to my knowledge, is the most comprehensive national-level scale regarding the subordination of women in the household available, and is therefore a useful variable to answer the question in this analysis.

Control Variables and Evaluation of Multicollinearity

In their forthcoming book, Hudson et. al. propose seven control variables that should be used in regression models that predict national level indicators from the Syndrome scale. In addition to the reasoning for inclusion of each control variable, the rules specified for including a control variable are: 1) the control variables should not be part of the Syndrome scale and should not be hypothesized effects of the Syndrome, 2) the bivariate correlation between any two control variables had to be less than 0.65 in order to avoid multicollinearity issues in the regression models, and 3) the variance inflation factors (VIFs) must be low, (all were lower than 1.5). The variables chosen, and their reason for inclusion, are:

- "Aggregated Civilization Identification", based on the work of Samuel Huntington (Huntington 1996). This is treated as a categorical variable in the analysis. Samuel Huntington identified in his work different categorizations of civilization that he believed undermined peace and stability in a society. The variable included for this analysis is a condensed scale based on his classifications. The scale is condensed into four categories based on Huntington's categorizations: 1) majority Western, Orthodox, and Latin civilizations, 2) majority Muslim civilizations, 3) majority Hindu, Sinic, and Buddhist civilizations, and 4) African countries that are not majority Muslim.

- "Colonial Heritage" , Bowen and Hudson, 2017. This is treated as an ordinal variable in the analysis. This is a dichotomous variable that measures whether a country was colonized (0 if it was not, 1 if it was). "Colonized" is defined as one country sending military to another country to ascertain control over this other country. The value for each country is based on whether a country was colonized for, at minimum, 10 consecutive years from 1700 to 2017. This variable is included with the assumption that, generally, colonizing counties stunts their growth, which negatively impacts their progress in overall health.

- "Percent Urban Population" from The World Bank, 2015 (The World Bank 2015). This is treated as a continuous variable in the analysis. This variable is included because I suspect that higher levels of the proportion of the population that lives in urban areas leads to easier access to health care. Additionally, the percentage of the urban population is a decent indicator of the country's wealth. Countries with more urban areas are often also those with higher GDP. Because the bivariate correlation between this variable and GDP per capita (also from The World Bank) is $r = .662$ (p-value=.001), I do not include GDP as a control variable in this analysis. I keep the urbanization variable instead of GDP because the urbanization variable controls for both the easier access to health care as well as the country's wealth, both of which I believe are indicators of health.

- "Percent Arable Land" from The World Bank (The World Bank 2014). This is treated as a continuous variable in the analysis. This terrain variable measures the percentage of land in a country that is arable, or suitable for growing crops. Multiple studies have identified a link between more extreme environments and the stability of the country (Homer-Dixon 1991; Pickering 2011; Emerson 2011; Fukuyama 2014), which in turn could negatively impact investments and focus on the health of the nation.

- "Number of Unique Land Neighbors" from Wikipedia (Wikipedia 2018)). This is treated as a continuous variable in the analysis. This variable measures the number of other countries whose borders touch a single country. This is included because, as is shown by multiple studies, the number of other countries surrounding one country negatively impacts the security (Starr and Most 1983; Starr and Thomas 2005), which again in turn could negatively impact investments and focus on the nation's health.

- "Ethnic Fractionalization", adapted from Alesina (Alesina et al. 2003)). This is treated as a continuous variable in the analysis. Multiple studies have found that lack of homogeneity in the population negatively impacts the security and stability of a country

(Fearon and Laitin 2003; Blimes 2006). The effects of this likely carry over into the nation's health outcomes.

- "Religious Fractionalization", adapted from Alesina (Alesina et al. 2003)). This is treated as a continuous variable in the analysis. Heterogeneity in religion in a society has been linked to instability and security of a nation (Fox 2004; Juergensmeyer 2017). The effects of this likely carry over into the nation's health outcomes as well.

Global Health Indicators

I consider many indicators of national health and wellbeing to cover the spectrum of national health outcomes that I expect are affected by the Syndrome. These indicators represent a comprehensive look at women's health, children's health, and overall societal health.

While most of the variables considered are continuous, or have enough ordinal levels that they can effectively be treated as continuous, three of the variables considered were ordinal with only three to five levels. While the methods proposed in the next section should also be tested on ordinal response data, I leave that endeavor to future research, and only include continuous variables in this study. The lists of variables below are disaggregated into three groups: women's health (7 indicators), children's health (6 indicators), and overall societal health (16 indicators), for a total of 29 health outcomes variables. The lists include the name of the variable, the year measured, the source of the data, the measurement units of the variable, the directional meaning (higher is better or lower is better), and the sample size once the variable observations are reduced to the countries with full observations from the Syndrome and all control variables. I also include the method or original source for the data collection or coding if available. Additionally, if any special notes are required to describe the variable, I also include that.

- Women's Health (including Maternal Health)

1. Life Expectancy at Birth for Females (2015), World Health Organization, years, lower scores are worse, N=160.

2. Percentage of Pregnant Women Receiving Prenatal Care (2017, or most recent without searching earlier than 2008), The World Bank, percent, lower scores are worse, N=130. This is survey data originally collected by UNICEF's State of the World's Children, Childinfo, and Demographic and Health Surveys.

3. Percentage of Birth Attended by Skilled Staff (2017, or most recent without searching earlier than 2008), The World Bank, percent, lower scores are worse, N=148. This is survey data originally collected by UNICEF's State of the World's Children, Childinfo, and Demographic and Health Surveys.

4. Births per 1000 Women Ages 15-19 (2016), The World Bank, rate per 1000, higher scores are worse, N=160. These data are originally from the United Nations Population Division's World Population Prospects.

5. Lifetime Risk of Maternal Death (2015), The World Bank, percent, higher scores are worse, N=160. The variable represents "the probability that a 15-year-old female will die eventually from a maternal cause assuming that current levels of fertility and mortality (including maternal mortality) do not change in the future, taking into account competing causes of death" (The World Bank 2018).

6. Death by Communicable Diseases, Maternal, Prenatal and Nutrition Conditions as a percentage of Total Population (2015), The World Bank, percent, higher scores are worse, N=160. Because health data is difficult to accurately estimate for developing countries, epidemiological models are used to estimate this variable.

7. Prevalence of HIV Among Women Ages 15+ (2016), The World Bank, percent ("Women's share of population aged 15+ living with HIV (%)", meaning that it is the percentage, out of the total HIV population, of women), higher scores are worse, N=121.

- Children's Health

1. Percentage of Children Under 5 Who are Stunted (2015, or most recent without searching earlier than 2007), The World Bank, percent, higher scores are worse, N=104. This is the percentage of children under 5 who are more than two standard deviations below the median height for their age. The data were collected and recorded under WHO's child growth standards set in 2006.

2. Infant Mortality Rate (2016), The World Bank, rate per 1000 live births, higher scores are worse, N=160.

3. Prevalence of Wasting in Children Under 5 (2015, or most recent without searching earlier than 2007), The World Bank, percent, higher scores are worse, N=103. These estimates are collected from surveys of a sample of the population, censuses, and vital registration systems. This is the percentage of children under 5 who are more than two standard deviations below the median weight for their height. The data were collected and recorded under WHO's child growth standards set in 2006.

4. Percentage of Children Under 5 who are Underweight (2015, or most recent without searching earlier than 2007), The World Bank, percent, higher scores are worse N=103. This is the percentage of children under 5 who are more than two standard deviations below the median weight for their age. The data were collected and recorded under WHO's child growth standards set in 2006.

5. Deaths due to Diarrhea of Children Under 5 (2010), World Health Organization (Accessed from Knoema), percent, higher scores are worse, N=159.

6. Percentage of Children Ages 12-23 Months Immunized Against Measles (2016), The World Bank, percent, lower scores are worse, N=160. The estimates are based on WHO and UNICEF's assessment of national immunization coverage rates, which take both administrative and household survey data.

- Overall Societal Health

1. Percentage of Total Population Using Open Defecation (2015), WHO/UNICEF, percent, higher scores are worse, N=160.

2. Percentage of Population Using Open Defecation in Urban Areas (2015), WHO/UNICEF, percent, higher scores are worse, N=157. Note: if value was missing and Percentage of Total Population Using Open Defecation variable's value was 0, we used a 0).

3. Life Expectancy (2015), World Health Organization, years, lower scores are worse, N=160.

4. Health Expenditure as a percentage of GDP (2015), World Health Organization, percent, lower scores are worse, N=156.

5. Health Expenditure Per Capita (2015), World Health Organization (Accessed from Wikipedia), US dollars, lower scores are worse, N=160.

6. Incidence of Tuberculosis per 100,000 People (2016), The World Bank, rate per 100,000, higher scores are worse, N=160. The World Bank's database warns that the difficulty of obtaining data in developing countries can make the data unreliable, but WHO used "epidemiological models and statistical standards" to estimate the values. (The World Bank 2018).

7. Percentage of Population Between 15-49 with HIV (2016), The World Bank, percent, higher scores are worse, N=121. Epidemiological models are used to estimate this variable when data on HIV is not readily available. The models are updated routinely.

8. Percentage of Adults Ages 15-49 with HIV/AIDS (2016), CIA World Factbook, percent, higher scores are worse, N= 120.

9. Percentage of Population that is Undernourished (2015), The World Bank, percent, higher scores are worse, N=148. The data are originally collected from the Food and

Agriculture Organization of the United Nations. The World Bank database specifies that this variable measures "food deprivation based on average food available for human consumption per person, the level of inequality in access to food, and the minimum calories required for an average person" (The World Bank 2018).

10. Total Alcohol Consumption Per Capita (2015), The World Bank, liters of pure alcohol per capita, higher scores are worse, N=159. The estimates "are produced by summing up the 3-year average per capita (15+) recorded alcohol consumption and an estimate of per capita (15+) unrecorded alcohol consumption for a calendar year" (The World Bank 2018).

11. Cigarette Consumption, number of cigarettes smoked per person per year ages > 15 (2016), The Tobacco Atlas, higher scores are worse, N=159.

12. Access to Improved Water Sources (2016), Social Progress Index, percent of the rural population, lower scores are worse, N=157. This variable measures the percentage of the population with access to an improved water source (that is not more than a 30 minute trip total) for drinking water, originally collected by WHO/UNICEF's Joint Monitoring Programme for Water Supply and Sanitation.

13. Access to Improved Sanitary Facilities (2016), Social Progress Index, percent, lower scores are worse, N=158. The variable measures the percentage of the population that use improved sanitation facilities, originally collected by WHO/UNICEF's Joint Monitoring Programme for Water Supply and Sanitation.

14. Global Hunger Index (2016), Global Hunger Index, scale (min=4, max=46.1), higher scores are worse, N=111. International Food Policy Research Institute (2016).

15. Sustainable Society Index Human Wellbeing (2016), Sustainable Society Index, continuous scale (min=3.1; max=9.0), lower scores are worse, N=145. This variable combines

indicators of 1) basic needs, 2) personal development and health, and 3) well-balanced society.

16. Average Dietary Energy Supply Adequacy (2014-2016), Food and Agriculture Organization of the UN, percent (3-year average), lower scores are worse, N=151.

I hypothesize that there is a significant relationship between the Syndrome's quantification of the subordination of women and these health outcomes.

CHAPTER 2

STATISTICAL METHODS

I implement a Bayesian hierarchical model for the analysis of how the subordination of women affects health outcomes for countries. I first discuss some other possible models for the analysis, and the theoretical reasons for not using those, then introduce and expound on the approach used for the health analysis. Following the introduction and theoretical discussion, I implement a simulation study to verify my analysis approach.

2.1 Preliminary Approaches

One possible option for analysis of this dataset would be to use a multivariate multiple regression model. This approach would account for the possibility that the relationships between the Syndrome and each health indicator are intercorrelated. The multivariate regression model is written as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Xi}, \tag{2.1}$$

where $\mathbf{Y}$ would be a matrix with each country's scores for each health indicator across the rows, $\mathbf{X}$ would be the matrix with the Syndrome and all of the control variables, again with the values for each country across the rows, and $\boldsymbol{\Xi}$ would be the error matrix. The interest would be in estimating all of the Syndrome components of $\boldsymbol{\beta}$, the matrix of coefficients. The multivariate regression model treats each row of $\mathbf{Y}$ as a single observation, which is only considered complete if the estimation for each response variable is given for that observation. Note that if the response data are complete (e.g. there are no missing values in $\mathbf{Y}$), then the parameter estimates in the multivariate model are the same as the

estimates in the individual simple regression models for each response variable, and only the inference on those coefficients changes (Rencher 2003).

The health indicators chosen for the model are collected from many different databases. Because of this, there are different levels of data availability for each of the variables. The sample size (number of countries measured after I reduced the list for each variable down to the country list for the complete observations of the Syndrome variable and control variables) for the health indicators range anywhere from 103 to 160. Because of this large variability in data availability, a multivariate regression approach to this analysis would require either that 1) large numbers of observations be discarded from the analysis, or 2) large numbers of values be imputed. In this specific case, a multivariate regression model would require that every country have an estimated value for every health indicator. Without imputation, 86 out of the 160 countries, more than half of the observations, would be discarded from the multivariate analysis.

Many different methods are used for data imputation, some more sophisticated than others. Expectation-maximization and multiple imputation, perhaps the most sophisticated method, could be used to resolve this issue. Expectation-maximization essentially treats the missing components (or each row with at least one missing value) of the $\mathbf{Y}$ matrix as the responses in a regression model:

$$\hat{\mathbf{y}}_i^{(m)} = \hat{\boldsymbol{\mu}}_i^{(m)} + \mathbf{B}(\mathbf{x}_i^{(c)} - \hat{\boldsymbol{\mu}}_i^{(c)}), \tag{2.2}$$

where (m) indicates that the observation's row has at least one missing value, and (c) indicates that the observation's row has no missing values. At each iteration, using the predicted values of $\mathbf{y}_i^{(m)}$, new regression coefficients are estimated by $\mathbf{B} = \hat{\boldsymbol{\Sigma}}_{(m)(c)}\hat{\boldsymbol{\Sigma}}_{(c)(c)}^{-1}$ and new values of the mean vectors ($\hat{\boldsymbol{\mu}}$) are estimated. This is repeated until convergence. Then, the multiple imputation step accounts for the uncertainty in predicting the missing values (Rencher 2003). The algorithm takes the final estimates from the expectation-maximization

step, $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$, and uses those to find the missing values of $\mathbf{Y}$, with an error term in the calculation:

$$\mathbf{y}_{i,[m]}^{(m)} = \boldsymbol{\mu}_i^{(m)} + \mathbf{B}(\mathbf{x}_i^{(c)} - \boldsymbol{\mu}_i^{(c)}) + \mathbf{e}_{i,[m]}^{(m)}, \tag{2.3}$$

where $[m]$ indicates the missing value in the row with at least one missing observation, and $\mathbf{e}_{i,[m]}^{(m)} \sim$ Multivariate Normal$(\mathbf{0}, \boldsymbol{\Sigma}_{(m)(m)} - \boldsymbol{\Sigma}_{(m)(c)}\boldsymbol{\Sigma}_{(c)(c)}^{-1}\boldsymbol{\Sigma}_{(c)(m)})$ (Rencher 2003).

Although expectation-maximization and multiple imputation provides more accurate imputations, simpler methods are often used for imputation. For example, the mean hot-deck imputation takes the average of each column of $\mathbf{Y}$ and imputes that average value for every missing value.

Both of these methods require that the data be missing at random, and we are not convinced of that in the health dataset. Different databases have varying lists of countries that they measure, so the missingness is likely at least partially reliant on that factor. Additionally, health outcome data collected at the country level is more likely to be reported by countries that have better health outcomes. Because these methods rely on the assumption that the data are missing at random (Rencher 2003), I conclude that this would not be a good approach for the health data.

One possible solution to the multivariate issue of data missingness could be to build individual simple regression models for each health indicator. In this case, we would build 29 individual regression models, then estimate and find the significance of the Syndrome parameter for each model. This would at least partially resolve the data missingness issue because only the number of observations missing for each individual response variable would be excluded in each model, so only between 0 and 57 observations would be excluded from each model depending on the number missing from the response variable. However, we would then lose the ability to take into account the relationship between the health indicators, a benefit of the multivariate regression analysis. And there is still concern that results would be biased because of the likely non-random data missingness.

2.2 Proposed Model

I propose an alternative approach that both avoids the need to impute large numbers of missing data and still considers the relationship between the health indicators: a Bayesian hierarchical approach to linear regression. This approach will allow analysis of each health indicator separately in a linear regression model. However, the relationship between the response variables will be taken into account in the hierarchy of the priors, by assuming that the prior distribution for the coefficients is the same for each model regardless of the response variable. The reasoning is easier to explain in the context of the specific model for this data, so I first introduce the model, then further explain the benefits of the proposed approach.

A hierarchical Bayesian model assumes that the parameters for the prior distribution are not fixed, which avoids the potential issue of overfitting when there are large amounts of parameters being estimated (Gelman et al. 2013). The model essentially "structure[s] some dependence into the parameters" (Gelman et al. 2013) by identifying some distribution among the parameters.

In this project, 29 individual linear regression models are considered, implementing a hierarchical Bayesian method to analyze the data in these models. The same independent variables are used in each model (the Syndrome and the seven control variables), but I analyze a different national-level health indicator as the response variable in each one.

Standardization of Independent and Dependent Variables

Some of the independent variables are categorical and some are continuous, and these continuous variables are not on the same scale, so I scale all of the independent variables. I implement a method proposed by Gelman (2013), to scale each of these independent variables, in order to set the same prior distributions for each (Gelman et al. 2013). I first scale all of the continuous independent variables to have a mean of 0 and a standard deviation

of 0.5. Then I take the colonization status variable, which is dichotomous, and shift it so that the variable also has a mean of 0 and a standard deviation of approximately 0.5. This is done by finding the proportion of 0's (0.13) and proportion of 1's (0.87), then redefining the 0's as 0.87 and the 1's as -0.13. In order to standardize the civilization variable, which is a categorical variable with 4 levels, I created 3 dichotomous variables which indicated 3 of the levels, then standardized each of those in the same way as colonial status. While this type of scaling may be not be appropriate for analysis where we need specific interpretations of the coefficients, this works for the purposes of this analysis because I am only asking whether there is a significant relationship between the subordination of women and health outcomes, so the farther the coefficients are from zero, the more evidence there is to support my hypothesis.

I also make all of the dependent variables commensurate in both directionality and scale. However, I first verify the normality of the response variables, an assumption of the linear model. I find that some of the data are heavily skewed. To resolve this issue, I used the Box-Cox transformation (Box and Cox 1964) method to approximate a transformation for those response variables. The transformation was performed before standardizing the response variables. These transformations helped the data more closely match the normality assumptions of the linear model. The response data are then scaled so that the coefficient of each independent variable has the same prior distribution with the same hyperpriors across all 29 models (the variables are all approximately normally distributed after the transformations, so their distributional assumptions are the same). The scaling is accomplished by simply standardizing all of the dependent variables. This means I subtract the mean and divide by the standard deviation for each value within each variable. In order to make the directionality consistent, I then take all of the standardized versions of the variables whose original definitions specified that higher levels were worse, and multiply their entire vector by negative one. This allows me to select common hyperparameters for the prior distributions of each independent variables' coefficients across the models, because we expect the Syn-

drome and the other control variables to impact the commensurate version of the variables similarly.

Scaling the independent variables simplifies the model by allowing the same priors to be placed on each coefficient within each model, and scaling the dependent variables allows the same priors to be placed on each coefficient across the models. These common hyperparameters and prior distributions for each independent variable across each of the 29 models take into account the expected relationship between, and similar behavior of, the health indicators, by implementing prior information to induce the assumption that, because the health outcomes are related to each other, their relationship to the independent variables will be similar. This is accomplished in the Bayesian hierarchical model without requiring multivariate modeling and thereby ensuring that every individual observation available will be used in the analyses.

The Model

The three-level hierarchical model is structured as follows:

$$\begin{aligned}
&\text{level 1: } y_{ij}|\boldsymbol{\beta}_j, \sigma_{ij} = \beta_{j0} + \sum_{k=1}^{10} x_{ijk}\beta_{jk} + \epsilon_i, \epsilon_i \sim \text{ Normal}(0, \sigma^2) \\
&\text{level 2: } \beta_{j0} \sim \text{ Normal}(0,1),\ \beta_{jk} \sim \text{ Normal}(\mu_k, \phi_k),\ \sigma^2 \sim \text{ Uniform}(0,1) \\
&\text{level 3: } \mu_k \sim \text{ Normal}(0,1),\ \phi_k^2 \sim \text{ Uniform}(0,2),
\end{aligned} \tag{2.4}$$

where $i = 1, ..., 160$ corresponds with the country, $j = 1, ..., 29$ corresponds with the regression model (one model for each health variable), and $k = 1, ..., 10$ corresponds with the independent variables. There are only eight independent variables, but because one of them is a categorical variable with four levels, there will be three coefficients corresponding with that variable.

Because I am fairly certain regarding the behavior of the error term in the model, since the response variables are standardized so that they all have a mean of 0 and a standard

deviation of 1, the error term ϵ_i has a prior distribution that matches the variation of the data. This means that the error term can be as large as it could possibly be if the covariates did not explain any of the variation, but the uniform prior on the error allows the data to drive the size of the error term. The tight prior distribution on β_{j0} is appropriate in this case because, again, all of the response variables are standardized to have a mean of 0 and a standard deviation of 1. The prior for the intercept essentially allows for the β_{j0} coefficient to explain all of the variation in the response even if the other covariates are all equal to zero.

The prior distributions for all of the other covariates are also set to a normal, but with common hyperparameters. At each iteration of the analysis, the same draws of both the error term (ϵ_i) and the hyperparameters (μ_k and ϕ_k^2) are used in all 29 models, which induces a correlation on the $\beta_{.k}$'s, and accounts for the relationship between the various health indicators.

2.3 Simulation Study

Before applying the proposed Bayesian hierarchical method to the dataset in this study, I use a simulation study on multivariate data with values that are missing non-randomly to test the hierarchical method against other possible methods discussed in Section 2.1. I compare the Bayesian hierarchical model to the multivariate regression method and simple multiple linear regression method. For the multivariate model, I use three different versions of the randomly generated data: 1) the data with all the missing values, 2) the data that has been imputed using the mean hot-deck imputation, and 3) the data that has been imputed using expectation-maximization and multiple imputation.

The data missingness in this simulation study closely resembles what I believe to be the missingness patterns in the health data. First, the percentage of observations missing are different for the different response variables, which matches the difference in the observations missing based on where the health data originated from. Once those percentages are chosen,

I find the lowest values for that variable and delete those lowest values that match the percentage that should be missing. For example, if 20% are supposed to be missing for that variable, I delete the lowest 20% of the values for that response variable. This matches the assumption that countries with worse health outcomes are less likely to have recorded values for that outcome variable.

I simulate multivariate data for a regression model, where the $\mathbf{X}$ values are drawn independently and randomly from a standard normal distribution ($x_{ij} \overset{i.i.d.}{\sim}$ Normal$(0, 1)$). The $\boldsymbol{\beta}$ values were chosen so that the coefficient for each explanatory variable (k) across the different response variables (j) were correlated: $\boldsymbol{\beta}_k \sim$ Multivariate Normal$(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ was a j by j positive definite correlation matrix. Then for each iteration,

- $\boldsymbol{\epsilon}$ was randomly generated such that $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim$ Multivariate Normal$(\mathbf{0}, \boldsymbol{\Sigma})$. $\boldsymbol{\Sigma}$ is a positive definite correlation matrix, which induces correlation between the j response variables, but maintains the assumption that the N observations are independent of each other.

- The percentage missing (0-30%) is randomly chosen for each Y_j, and the actual values missing from Y_j are then non-randomly chosen so that the missingness in the simulated data resembles the non-random missingness of the health variables.

- Five different methods are used to estimate the $\boldsymbol{\beta}$ values with the dataset that contains non-random missing values of $\boldsymbol{Y}$.

    1. Multivariate Linear Regression (removing data with missing values)
    2. Multivariate Linear Regression (using data where the missing values have been imputed with the column means of $\mathbf{Y}$)
    3. Multivariate Linear Regression (using data where the missing values have been imputed using expectation-maximization and multiple imputation)
    4. Simple Linear Regression (where each model is fit independently)

5. Bayesian Hierarchical Linear Regression

- For each model, the mean absolute error (MAE) and root mean squared error (RMSE) are calculated to compare the $\hat{\boldsymbol{\beta}}$ values to $\boldsymbol{\beta}$.

This simulation study was repeated with different sample sizes (N=50, N=160, N=300) and different numbers of response variables (m=5 and m=10), each time with 100 iterations. I report the average RMSE and average MAE across the 100 iterations. The RMSE and MAE are calculated each time as follows:

$$\text{RMSE} = \sqrt{\frac{1}{k+1}\sum_{r=0}^{k}(\beta_r - \hat{\beta}_r)^2} \tag{2.5}$$

$$\text{MAE} = \frac{1}{k+1}\sum_{r=0}^{k}|\beta_r - \hat{\beta}_r| \tag{2.6}$$

for each of the five methods listed above, where the $\hat{\beta}_r$ values are estimated using each of the five methods.

These estimate results (the average values across the 100 iterations) are displayed in Table 2.1. The results from the simulation study give multiple interesting insights into the models' performance when fed data with non-random missingness. First, I find that the hierarchical model's (Model 5) average RMSE and average MAE estimates are the lowest for three out of the six simulations. The simple linear regression model (Model 4) has the best, or lowest, average RMSE and average MAE for two of the other simulations, and the multivariate regression model with expectation-maximization and multiple imputation (Model 3) has the best for the final simulation ($N = 300$, $m = 10$).

While the point estimates are useful for comparing the models, because many of the estimates are so close, I also compare the 95% confidence intervals of the true average RMSE and MAE for each method. These intervals are calculated by taking the 100 RMSE and MAE estimates for each method and finding: $\overline{\text{RMSE}} \pm \text{z}_{\alpha/2}\sqrt{\text{var(RMSE)}/100}$ and $\overline{\text{MAE}} \pm \text{z}_{\alpha/2}\sqrt{\text{var(MAE)}/100}$. The intervals results are displayed in Table 2.2. The hierarchical

model's intervals are always very close to the intervals for the simple linear regression model. In the first five simulations, the model with the lowest point estimate for the average RMSE and average MAE (Model 4 or Model 5) does not overlap with any of the first three models, indicating that the lowest estimate was significantly better than for the first three models. In the last iteration, while Model 3 had slightly lower average estimates, the difference was not significant compared to the estimates from Models 4 and 5.

From these results, I conclude that most of the time, the hierarchical regression model estimates the coefficients more accurately than most other models in the simulation. The simple linear regression model performs very similarly with the hierarchical model, with their confidence intervals on the average RMSE and MAE estimates overlapping most of the time. Specifically, the hierarchical model performs significantly better than any of the multivariate models in five out of six simulations, and the simple regression model does the same in four out of the six. In the last simulation, the intervals for models 3-5 overlap. Although the point estimate for the average RMSE and MAE is slightly lower for Model 3 than for the other two, the difference is not significant.

Overall, I conclude from this simulation that most of the time the hierarchical and simple regression models give significantly smaller average RMSE and MAE estimates than any of the multivariate models. The multivariate model with multiple imputation performs similarly to the other two models with a high sample size and a large number of response variables; however, this is the only case where it performs just as well, and the sample size is much larger than the sample size for the health dataset.

Because the simple linear regression model's average bias estimates were often not significantly different than the hierarchical model's, I also compare the average width of the confidence intervals of the simple linear model to the average width of the credible intervals of the hierarchical model, displayed in Table 2.3. I find that in all simulations, the average credible interval width of the Bayesian model is smaller than the average confidence interval

Table 2.1: Simulation Study Model Comparisons - RMSE and MAE

		N=50		N=160		N=300	
		m=5	m=10	m=5	m=10	m=5	m=10
RMSE	Model 1	0.3337	0.2552	0.1153	0.1555	0.0859	0.1056
	Model 2	0.3285	0.1846	0.1455	0.1563	0.1306	0.1264
	Model 3	0.2096	0.1756	0.0976	0.0868	0.0710	**0.0546**
	Model 4	**0.1823**	0.1588	**0.0910**	0.0868	0.0624	0.0565
	Model 5	0.1847	**0.1579**	0.0913	**0.0851**	**0.0623**	0.0566
MAE	Model 1	0.2764	0.2077	0.0953	0.1263	0.0702	0.0858
	Model 2	0.2612	0.1476	0.1131	0.1128	0.0942	0.0882
	Model 3	0.1781	0.1373	0.0796	0.0704	0.0570	**0.0436**
	Model 4	**0.1502**	0.1297	**0.0744**	0.0706	0.0510	0.0449
	Model 5	0.1524	**0.1290**	0.0746	**0.0695**	**0.0509**	0.0449

Table 2.1: This table displays the average root mean squared error (RMSE) and average mean absolute error (MAE) of the estimated coefficient values against the true values for each model (The order of the models is the same as in the algorithm description listed above). The average is across 100 iterations, and is given for different sample sizes (N=50, N=160, N=300) and different numbers of response variables (m=5 and m=10).

Table 2.2: Simulation Study Model Comparisons - Average RMSE and MAE 95% Confidence Intervals

		N=50		N=160		N=300	
		m=5	m=10	m=5	m=10	m=5	m=10
RMSE	Model 1	(.3316,.3358)	(.2534,.2569)	(.1147,.1159)	(.1547,.1562)	(.0854,.0863)	(.0988,.1123)
	Model 2	(.3274,.3297)	(.1839,.1854)	(.1449,.1462)	(.1558,.1569)	(.0937,.0947)	(.1232,.1295)
	Model 3	(.2086,.2107)	(.1734,.1777)	(.0972,.0981)	(.0862,.0869)	(.0707,.0714)	(.0535,.0558)
	Model 4	(.1814,.1833)	(.1580,.1595)	(.0906,.0914)	(.0864,.0871)	(.0621,.0627)	(.0550,.0579)
	Model 5	(.1837,.1857)	(.1571,.1587)	(.0909,.0917)	(.0848,.0854)	(.0620,.0626)	(.0552,.0581)
MAE	Model 1	(.2746,.2782)	(.2061,.2092)	(.0947,.0959)	(.1256,.1269)	(.0698,.0706)	(.0812,.0904)
	Model 2	(.2601,.2621)	(.1469,.1482)	(.1126,.1136)	(.1124,.1132)	(.0937,.0947)	(.0861,.0902)
	Model 3	(.1673,.1690)	(.1362,.1384)	(.0792,.0799)	(.0702,.0707)	(.0567,.0572)	(.0427,.0446)
	Model 4	(.1493,.1510)	(.1291,.1304)	(.0741,.0748)	(.0703,.0709)	(.0508,.0513)	(.0437,.0460)
	Model 5	(.1515,.1533)	(.1283,.1296)	(.0743,.0750)	(.0692,.0698)	(.0507,.0512)	(.0438,.0462)

Table 2.2: This table displays the confidence intervals on the average root mean squared error (RMSE) and average mean absolute error (MAE) of the estimated coefficient values against the true values for each model (The order of the models is the same as in the algorithm description listed above). The intervals are given for different sample sizes (N=50, N=160, N=300) and different numbers of response variables (m=5 and m=10).

width of the simple regression models. This superior inferential property of the Bayesian model indicates that this model more accurately estimates the parameters.

These results indicate that the hierarchical regression and simple regression estimates are statistically significantly less biased, measured by the average RMSE and MAE, than the multivariate regression methods, specifically for the assumed non-random missingness

Table 2.3: Simulation Study Model Comparisons - Credible/Confidence Intervals

	N=50		N=160		N=300	
	m=5	m=10	m=5	m=10	m=5	m=10
Model 4	0.6410	0.6256	0.3471	0.3493	0.2469	0.2448
Model 5	**0.6134**	**0.5972**	**0.3372**	**0.3413**	**0.2416**	**0.2381**

Table 2.3: This table displays the average confidence interval width of Model 4 (Simple Linear Model) and average credible interval width of Model 5 (Hierarchical Bayesian Model) for the estimated coefficient values. The average is across 100 iterations, and is given for different sample sizes (N=50, N=160, N=300) and different numbers of response variables (m=5 and m=10).

patterns of the health dataset. I further find that the hierarchical model consistently has smaller average credible interval widths than the simple linear regression model's confidence intervals. While we should still be cautious and concerned about non-randomly missing data, this method consistently estimates with less bias than imputing data for multivariate regression when the assumptions of random missingness are not met. This corroborates the theory that even the most sophisticated methods for imputation are not effective unless the missingness is random. I do note that the multivariate model, in general, performs better with the mean hot-deck imputed data than with the missing-values data, and better with the multiple imputed data than with the mean hot-deck imputed data.

While the simple linear regression model performs similarly in unbiased estimation to the Bayesian hierarchical model, the Bayesian model 1) has smaller intervals around those estimates which is better for inference, and 2) allows prior information to be used, in this case to impose the assumptions of dependence between the response variables. I therefore continue with the use of the Bayesian model for analysis of the health data.

CHAPTER 3

RESULTS AND DISCUSSION

The hierarchical analysis on the health outcomes was run in JAGS in R with the "R2jags" package (Su and Yajima 2012), which uses Markov chain Monte Carlo (MCMC) methods to sample posterior draws from the posterior distribution for the parameters. For each of the 29 models, I ran five chains of 12,000 iterations, then removed the first 2,000 as burn-in, and finally thinned by 5 to reduce autocorrelation.

I first verify the model with both visual and statistical diagnostics on the posterior draws, then proceed to reporting the results of the analysis.

3.1 Model Diagnostics

I verify the convergence of the posterior draws for the models, and ensure the accuracy of the estimates and quantiles. The results for the convergence diagnostics are displayed for the Syndrome coefficient specifically (since that is the independent variable I am interested in in this analysis), but are fairly representative of the performance of the posterior draws for the other coefficients in the model. Figure 3.1 plots the posterior draws from the MCMC algorithm for the Syndrome coefficient in all 29 models. These plots, which are again representative of the posterior draws of other parameters, display the posterior draws of the MCMC after burn-in and thinning.

These seem to indicate that the posterior draws converged well, because there is no specific pattern in the posterior draws. I therefore proceed to an evaluation of some statistical diagnostics for convergence.

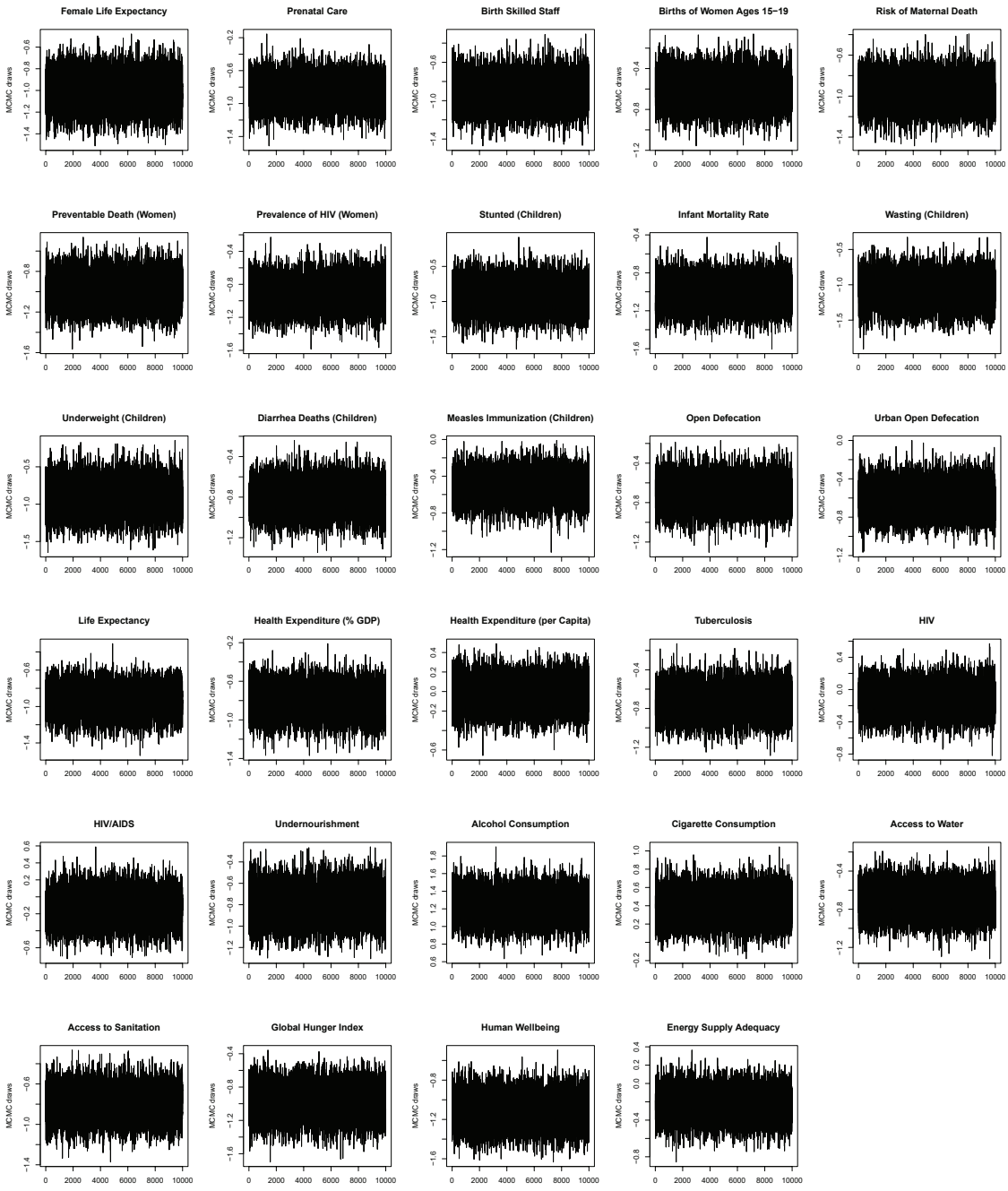

Figure 3.1: These selected convergence plots show that the MCMC algorithm converged.

I find the effective sample size values through a function in the "coda" package in R (Plummer et al. 2006), and the $\hat{R}$ values through the "R2jags" package in R (Su and Yajima 2012). Low effective sample sizes indicate that there is too much autocorrelation between

the draws, so this diagnostic helps to verify that the draws are reasonably independent of each other, and further that there are enough unique posterior draws that inference can reasonably be made on the posterior draws. I find that the lowest effective sample size for the Syndrome coefficient in the 29 models is 8,618 (out of 10,000 posterior draws remaining after burn-in and thinning). The $\hat{R}$ diagnostic values should be around 1 and indicate whether the chains (and within the chains) have converged to the same place as each other. For the Syndrome variable, the $\hat{R}$ values all approximately estimated at 1.001. These values close to 1 indicate that the model is not overly sensitive to different starting values in the MCMC algorithm.

Because I estimate the credible intervals for the Syndrome coefficients in the following section, I also use a diagnostic to check the number of iterations needed to accurately estimate the quantiles: the Raftery-Lewis Diagnostic (Raftery and Lewis 1991). The function used to calculate the Raftery-Lewis diagnostics is also from the "coda" package in R (Plummer et al. 2006). With this function, I find the number of iterations that are required to accurately estimate the quantiles (0.025 and 0.975) with a 95% probability, and an accuracy of 0.005. The minimum number needed for the Syndrome coefficients is 3620, and the max is 3865. Each of these coefficients have 10,000 posterior draws after burn-in, so I conclude that I have more than enough draws to accurately estimate the quantiles for the credible intervals.

Lastly, I examine the plots for the posterior draws of the variance terms in the model. Since the variance terms all have uniform priors, I verify that the draws are not pushing up against the upper bound of the uniform distribution, which would indicate that the bounds of these prior distributions may be inappropriate for this model. These convergence plots are in Figure 3.2. While the variance terms are all lower bounded at 0, a property of the variance term, the higher valued posterior draws do not appear to come close to the upper bound (1 or 2) in any of the plots. I therefore conclude that the uniform priors on the variance terms are fitting for this model.

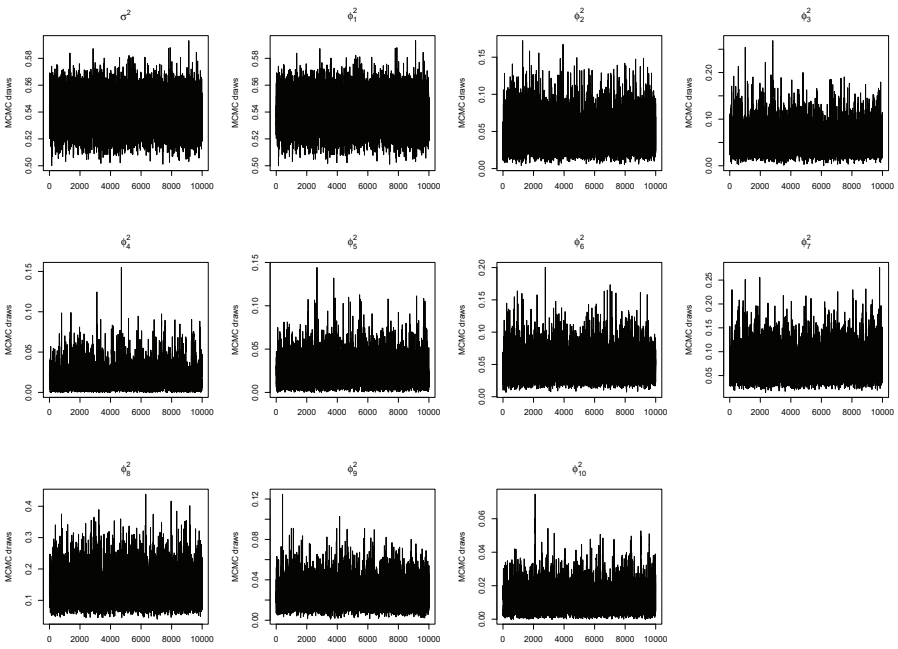

Figure 3.2: Convergence plots for variance terms.

I determine that the posterior draws have converged, the quantile estimates are accurate for the number of iterations used, and that the variance terms' prior distributions are appropriate, so I proceed to reporting the estimates and results of the models.

3.2 Model Results

Table 3.1 provides the Syndrome estimates for each of the 29 models, as well as the 95% credible intervals. Because for all of the standardized versions of the health indicators, lower scores indicate worse health outcomes, I expect from my hypothesis that all of the coefficients corresponding with the Syndrome, or the subordination of women, should be negative. A negative coefficient indicates that countries with worse subordination of women's interests to men's interests perform worse, on average, on that health indicator.

The credible intervals for all of the Women's Health and the Children's Health indicators (shown in Table 3.1) contain only negative values. This gives substantial evidence to support the claim that higher levels of subordination of women in the home are related

to worse health outcomes for women and children. The results for the Overall Societal indicators are more mixed. While most of the coefficient estimates are negative, some of the credible intervals also include positive values. However, 10 out of the 16 models that measure overall health outcomes indicate that the subordination of women is a significant indicator of worse health outcomes. Notably, subordination seems to have an opposite effect on cigarette and alcohol consumption - countries with worse subordination of women tend to have lower alcohol and cigarette consumption.

The last row of Table 3.1 gives the estimate and interval for $\mu_{Syndrome}$, which is the mean in the prior distribution for the $\beta_{Syndrome}$ in each model. This mean parameter has a hyperprior distribution in the model of Normal$(0, 1)$. The $\mu_{Syndrome}$ estimate then becomes the overall mean for all of the Syndrome coefficients in the model, and therefore of interest in this study. The estimate indicates that the overall average effect of the subordination of women on the health outcomes in this study is -0.56. Both ends of the credible interval are negative, indicating that there is a significant average overall effect of the Syndrome on these health outcomes.

Overall, the results seem to corroborate my hypothesis: that countries with worse treatment of women tend to perform worse on health outcomes. These results are more straightforward for the health outcomes for both women and children, where the models for all of those sub-indicators support the hypothesis. And, although the results for the models with overall health outcomes were not as consistent, I still find that in 10 out of those 16 models, treatment of women is a significant negative determinant of health outcomes.

The pattern in the results is perhaps unsurprising given the theory behind this statistical model. I proposed that the treatment of women in the household directly affects the women and children involved in the poor treatment, but also impacts others in the family and community who rely on these women for care. The impact of poor treatment of these women is more clear, however, on the health of the women and their children in the home, while the affect on the community has a less clear or direct tie. Treatment of women in

Table 3.1: Coefficient Estimate and Credible Interval Results for the Syndrome Predictor

	Health Variable	**Estimate**	**2.5%**	**97.5%**
Women's Health	Female Life Expectancy at Birth	-1.00	-1.29	-0.72
	Pregnant Women Receiving Prenatal Care	-0.86	-1.19	-0.53
	Births Attended by Skilled Staff	-0.91	-1.21	-0.60
	Births of Women Ages 15-19	-0.61	-0.89	-0.32
	Lifetime Risk of Maternal Death	-0.96	-1.25	-0.67
	Death by Preventable Conditions	-0.98	-1.27	-0.69
	Prevalence of HIV for Women Ages 15+	-0.91	-1.25	-0.58
Children's Health	Children Under 5 Who are Stunted	-0.91	-1.31	-0.52
	Infant Mortality Rate (IMR)	-1.02	-1.31	-0.73
	Children Under 5 Who are Wasting	-1.08	-1.48	-0.69
	Children Under 5 who are Underweight	-0.91	-1.31	-0.51
	Children Under 5 Deaths due to Diarrhea	-0.80	-1.08	-0.52
	Children Ages 12-23 Immunized Against Measles	-0.51	-0.80	-0.23
Overall Societal Health	Open Defecation	-0.70	-0.99	-0.42
	Open Defecation in Urban Areas	-0.61	-0.90	-0.30
	Life Expectancy	-0.94	-1.23	-0.65
	Health Expenditure as % of GDP	-0.81	-1.10	-0.52
	Health Expenditure per Capita	-0.03	-0.32	0.27
	Incidents of Tuberculosis	-0.73	-1.01	-0.44
	HIV Ages 15-49	-0.12	-0.46	0.22
	HIV/AIDS Ages 15-49	-0.14	-0.48	0.20
	Undernourishment	-0.80	-1.09	-0.50
	Alcohol Consumption per Capita	1.23	0.92	1.53
	Cigarette Consumption	0.40	0.09	0.70
	Access to Improved Water Sources	-0.72	-1.01	-0.42
	Access to Improved Sanitary Facilities	-0.80	-1.09	-0.51
	Global Hunger Index	-0.98	-1.33	-0.62
	Sustainable Society Index Human Wellbeing	-1.11	-1.41	-0.81
	Average Dietary Energy Supply Adequacy	-0.23	-0.53	0.06
Average Overall Effect	$\mu_{Syndrome}$	-0.56	-0.77	-0.36

the household is significant to all of the health indicators on women and children, giving strong evidence for this direct tie proposed in the theory. The affect of women's status in the household is not as consistently tied to overall health, but the 10 out of 16 significant results indicate that the impact of the treatment of women reaches outside of the household into their communities.

3.3 Discussion and Conclusion

I theorized in the introduction of this work that poor measures of national health are significantly associated with the dynamic between men and women that has consistently left women disadvantaged. I specifically hypothesize that countries with higher levels of sys-

tematic subordination of women in the household perform worse on health indicators for 1) women, 2) children, and 3) the overall public. The Bayesian hierarchical analysis results strongly support this hypothesis. I find consistently that the Syndrome variable is significantly linked to health outcomes for both women and children, and specifically that higher subordination on average corresponds to worse health outcomes. In 10 out of the 16 overall health models, I find the same results. Additionally, I estimate the overall average for all of the Syndrome coefficients, and find that there is an overall significant and negative average impact of poor treatment of women on the health outcomes included in this study.

While causal inference cannot be made regarding the statistical results, I propose, as is beautifully theorized in Hudson et al. (forthcoming), that the sexual relations are the first political order of a society, and that these sex dynamics determine how decisions are made and where priority is placed within the family as well as outside, including health decisions. Women are disproportionately the caregivers in their societies and communities, and therefore have high impacts on their own health and the health of their families and communities, if they are empowered to do so. If women are disempowered, harmed, overly burdened, or silenced, their positive health practices and impacts are stifled. I therefore submit that political and social leaders interested in improving health should begin to place the empowerment of women at the center of their calculus for efforts to improve health outcomes.

BIBLIOGRAPHY

Afghan Women's Writing Project: Oral Stories Project (2014), "Whatever Men Want is What Happens," http://awwproject.org/2014/07/whatever-men-want-is-what-happens/.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003), "Fractionalization," *Journal of Economic growth*, 8, 155–194.

American Cancer Society, Inc. and Vital Strategies (2018), "The Tobacco Atlas," https://tobaccoatlas.org.

Blimes, R. J. (2006), "The indirect effect of ethnic heterogeneity on the likelihood of civil war onset," *Journal of Conflict Resolution*, 50, 536–547.

Box, G. E., and Cox, D. R. (1964), "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, 26, 211–243.

Central Intelligence Agency (2017), "CIA WorldFactbook," https://www.cia.gov/library/publications/the-world-factbook/.

Das Gupta, M. (2009), *Family systems, political systems, and Asia's' missing girls': the construction of son preference and its unraveling*, The World Bank.

Emerson, S. A. (2011), "Desert insurgency: Lessons from the third Tuareg rebellion," *Small Wars & Insurgencies*, 22, 669–687.

Fearon, J. D., and Laitin, D. D. (2003), "Ethnicity, insurgency, and civil war," *American political science review*, 97, 75–90.

Food and Agriculture Organization of the United Nations (2017), http://www.fao.org/statistics/databases/en/.

Fox, J. (2004), "The rise of religious nationalism and conflict: Ethnic conflict and revolutionary wars, 1945-2001," *Journal of peace Research*, 41, 715–731.

Fukuyama, F. (2014), *Political order and political decay: From the industrial revolution to the globalization of democracy*, Macmillan.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian data analysis*, Chapman and Hall/CRC.

Harari, Y. N. (2011), *A brief history of humankind*, New York: HarperCollins.

Homer-Dixon, T. F. (1991), "On the threshold: environmental changes as causes of acute conflict," *International security*, 16, 76–116.

Hudson, V. M., Bowen, D. L., and Nielsen, P. L. (forthcoming), *The First Political Order: Sex, Governance, and National Security*, Columbia University Press.

Huntington, S. (1996), "The clash of civilizations and the making of a new world order," *New York*.

International Food Policy Research Institute (2016), "Global Hunger Index," http://www.ifpri.org/topic/global-hunger-index.

Juergensmeyer, M. (2017), *Terror in the mind of God: The global rise of religious violence* (Vol. 13), Univ of California Press.

McDermott, R. (2018), *The Evils of Polygyny: Evidence of Its Harm to Women, Men, and Society*, Cornell University Press.

McDermott, R., and Cowden, J. (2014), "Polygyny and violence against women," *Emory LJ*, 64, 1767.

McFarlane, J., Campbell, J. C., Sharps, P., and Watson, K. (2002), "Abuse during pregnancy and femicide: urgent implications for women's health," *Obstetrics & Gynecology*, 100, 27–36.

Mearsheimer, J. J., et al. (2001), *The tragedy of great power politics*, WW Norton & Company.

Nour, N. M. (2006), "Health consequences of child marriage in Africa," *Emerging infectious diseases*, 12, 1644.

—— (2009), "Child marriage: a silent health and human rights issue," *Reviews in obstetrics and gynecology*, 2, 51.

Pickering, S. (2011), "Determinism in the mountains: The ongoing belief in the bellicosity of mountain people'," *The Economics of Peace and Security Journal*, 6.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11.

Raftery, A. E., and Lewis, S. (1991), "How many iterations in the Gibbs sampler?" Tech. rep., WASHINGTON UNIV SEATTLE DEPT OF STATISTICS.

Raj, A., and Boehmer, U. (2013), "Girl child marriage and its association with national rates of HIV, maternal health, and infant mortality across 97 countries," *Violence against women*, 19, 536–551.

Rencher, A. C. (2003), *Methods of multivariate analysis* (Vol. 492), John Wiley & Sons.

Rizvi, N., Khan, K. S., and Shaikh, B. T. (2014), "Gender: shaping personality, lives and health of women in Pakistan," *BMC women's health*, 14, 53.

Starr, H., and Most, B. A. (1983), "Contagion and border effects on contemporary African conflict," *Comparative Political Studies*, 16, 92–117.

Starr, H., and Thomas, G. D. (2005), "The nature of borders and international conflict: Revisiting hypotheses on territory," *International Studies Quarterly*, 49, 123–139.

Su, Y.-S., and Yajima, M. (2012), "R2jags: A Package for Running jags from R," *R package version 0.03-08, URL http://CRAN. R-project. org/package= R2jags*.

Supreme Court of British Columbia (2011), "Reference re: Section 293 of the Criminal Code of Canada, 2011 BCSC 1588 (CanLII)," https://www.canlii.org/en/bc/bcsc/doc/2011/2011bcsc1588/2011bcsc1588.html#.

Sustainable Society Foundation (2017), "Sustainable Society Index," http://www.ssfindex.com/data-all-countries/.

The Social Progress Imperative (2016), "Social Progress Index," https://www.socialprogress.org.

The WomanStats Project (2018), "WomanStats Project Database," http://www.womanstats.org.

The World Bank (2014), "World Development Indicators," http://data.worldbank.org/data-catalog/world-development-indicators.

—— (2015), "World Development Indicators," http://data.worldbank.org/data-catalog/world-development-indicators.

—— (2018), "World Bank Open Data," https://data.worldbank.org.

United Nations Children's Fund (2005), "Early marriage: a harmful traditional practice. A statistical exploration," .

United Nations International Children's Emergency Fund (2017), "UNICEF Data," https://data.unicef.org.

Wikipedia (2018), "List of countries and territories by land borders," https://en.wikipedia.org/wiki/List_of_countries_and_territories_by_land_borders.

Wodon, Q. (2015), "Child marriage, family law, and religion: An introduction to the fall 2015 issue," *The Review of Faith & International Affairs*, 13, 1–5.

World Health Organization (2015), "SDG Health and Health-related Target Indicators," http://www.who.int/gho/en/.