The Subordination of Women and National Health Outcomes

A Bayesian Hierarchical Regression Model Approach

Kaylee Hodgson

Brigham Young University

March 20, 2019

Introduction

- Variation in inequality between men and women allows for empirical testing on its impacts
- Is the health and wellbeing of a nation impacted by the level that women are subordinated?

Purpose: identify whether higher levels of gender inequality in the household are significantly tied to poor health outcomes for women, children, and the general public.

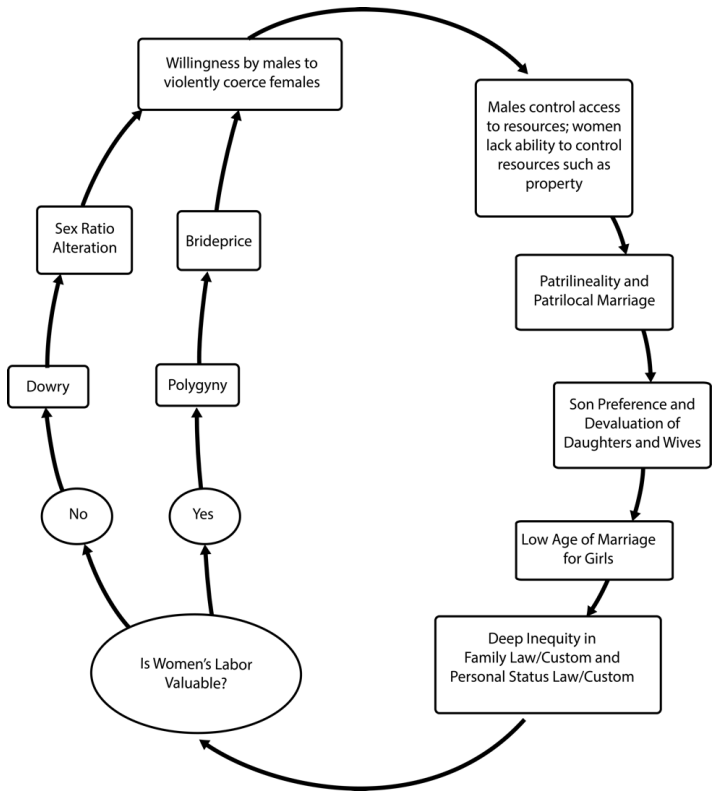
Background and Motivation

- Other studies have found direct link between specific aspects of treatment of women and specific health outcomes
- Women are often the main caregivers in their families and communities
- **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 them suffer negative health consequences as well**

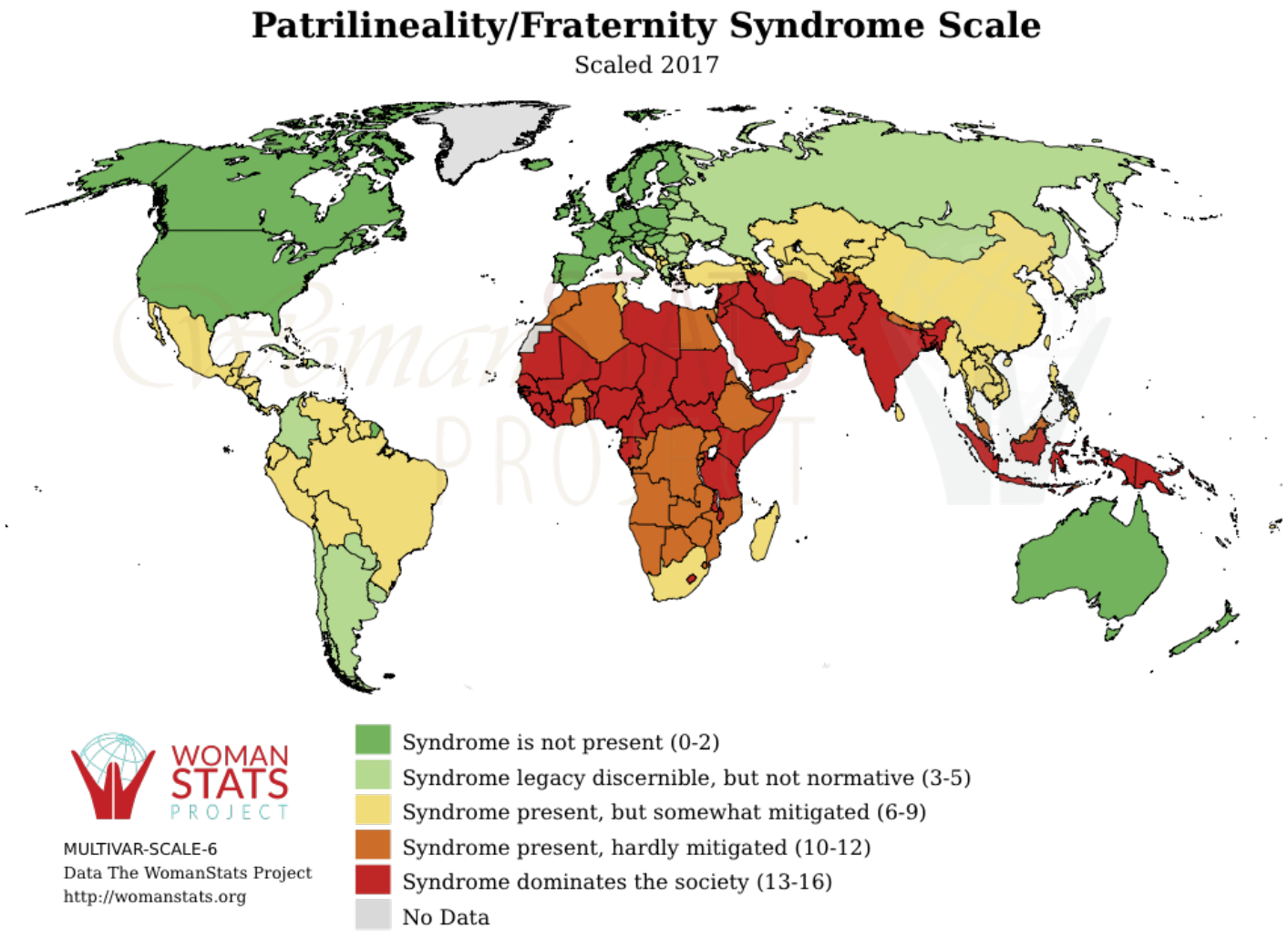
The Syndrome Scale

Combines 11 indicators of the subordination of women in the home:

- prevalence of patrilocal marriage
- prevalence of brideprice or dowry
- prevalence and legality of polygyny
- presence of counsin marriage
- age of marriage for girls
- laws and practices surrounding women's property rights
- presence of son preferences or sex ratio alteration
- presence of inequity in family law/custom that favors males
- overall level of violence against women in society
- presence of societal sanction for femicide
- whether there is legal exoneration for rapists who offer to marry their victims



The Syndrome Scale



Control Variables

- percentage of the population that lives in urban areas (The World Bank)
- aggregated civilization identification (based on Samuel Huntington's civilizational)
- colonial heritage - dichotomous, indicates whether a country was colonized (from Valerie Hudson and Donna Lee Bowen's scale in their forthcoming book)
- percentage of land that is arable (The World Bank)
- number of unique land neighbors (Wikipedia)
- level of ethnic fractionalization (Alesina et al., 2003)
- level of religious fractionalization (Alesina et al., 2003).

Health Outcomes: Women's Health

1. Life Expectancy at Birth for Females (World Health Organization)
2. Percent of Pregnant Women Receiving Prenatal Care (The World Bank)
3. Percent of Birth Attended by Skilled Staff (The World Bank)
4. Births per 1000 Women Ages 15-19 (The World Bank)
5. Lifetime Risk of Maternal Death (The World Bank)
6. Death by Communicable Diseases, Maternal, Prenatal and Nutrition Conditions as a percentage of Total Population (The World Bank)
7. Prevalence of HIV Among Women Ages 15+ (The World Bank)

Health Outcomes: Children's Health

1. % Under 5 Who are Stunted (The World Bank)
2. Infant Mortality Rate (The World Bank)
3. Prevalence of Wasting - % Under 5 (The World Bank)
4. % Under 5 who are Underweight (The World Bank)
5. Deaths due to Diarrhea of Children Under 5 (World Health Organization)
6. % Children Ages 12-23 Months Immunized Against Measles (The World Bank)

Health Outcomes: Overall Societal Health

1. % Total Population Using Open Defecation (WHO/UNICEF)
2. % Population Using Open Defecation in Urban Areas (WHO/UNICEF)
3. Life Expectancy (World Health Organization)
4. Health Expenditure as % of GDP (World Health Organization)
5. Health Expenditure Per Capita (World Health Organization)
6. Incidence of Tuberculosis per 100,000 People (The World Bank)
7. % of Population Between 15-49 with HIV (The World Bank)
8. Percentage of Adults Ages 15-49 with HIV/AIDS (CIA World Factbook)
9. % of Population that is Undernourished (2015), (The World Bank)
10. Total Alcohol Consumption Per Capita (The World Bank)
11. Cigarette Consumption, number of cigarettes smoked per person per year ages $>$ 15 (The Tobacco Atlas)
12. Access to Improved Water Sources (Social Progress Index)
13. Access to Improved Sanitary Facilities (Social Progress Index)
14. Global Hunger Index (Global Hunger Index)
15. Sustainable Society Index Human Wellbeing (Sustainable Society Index)
16. Average Dietary Energy Supply Adequacy (Food and Agriculture Organization of the UN)

Hypothesis

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

Preliminary Approaches: Multivariate Regression

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Xi}, \tag{1}$$

- Treats each row of $\mathbf{Y}$ as a single observation
- Observation only considered complete if the estimation for each response variable is given
- An option to deal with missing values: Imputation

Preliminary Model Approaches: Multivariate Regression

A Couple of Imputation Options

Expectation-Maximization and Multiple Imputation

- Treat missing components of $\mathbf{Y}$ as responses in regression model:

$$\hat{\mathbf{y}}_i^{(m)} = \hat{\boldsymbol{\mu}}_i^{(m)} + \mathbf{B}(\mathbf{x}_i^{(c)} - \hat{\boldsymbol{\mu}}_i^{(c)}), \tag{2}$$

- At each iteration, use predicted values of $\mathbf{y}_i^{(m)}$ to estimate $\mathbf{B} = \hat{\boldsymbol{\Sigma}}_{(m)(c)}\hat{\boldsymbol{\Sigma}}_{(c)(c)}^{-1}$ and $\hat{\boldsymbol{\mu}}$, until convergence.
- Take the final estimates, $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$, and use those to find the missing values of $\mathbf{Y}$, with an error term in the calculation:

$$\mathbf{y}_{i,[m]}^{(m)} = \boldsymbol{\mu}_i^{(m)} + \mathbf{B}(\mathbf{x}_i^{(c)} - \boldsymbol{\mu}_i^{(c)}) + \mathbf{e}_{i,[m]}^{(m)}, \tag{3}$$

$$\mathbf{e}_{i,[m]}^{(m)} \sim \text{MVNorm}(\mathbf{0}, \boldsymbol{\Sigma}_{(m)(m)} - \boldsymbol{\Sigma}_{(m)(c)}\boldsymbol{\Sigma}_{(c)(c)}^{-1}\boldsymbol{\Sigma}_{(c)(m)}) \tag{4}$$

Mean Hot-Deck Imputation

- Take the average of each column of $\mathbf{Y}$
- Impute that average value for every missing value

Preliminary Model Approaches: Multivariate Regression

Imputation Assumptions

- These imputation methods require that the data be missing at random
- Health dataset appears to have non-random missing data:
  1. Different databases have varying lists of countries, so missingness at least partially relies on the data source
  2. Health outcome data more likely to be reported by countries that have better health outcomes

Preliminary Model Approaches: Simple Linear Regression

$$\mathbf{Y}_j = \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \tag{5}$$

where $j = 1, ..., 29$ corresponds with each dependent variable.

- This approach treats each response variable as independent of all other response variables

Preliminary Model Approaches: Advantages and Disadvantages

1. Multivariate Regression
   - Advantage: Takes into account the relationship between the response variables
   - Disadvantage: Deletes too many observations, imputation requires data be missing at random
2. Simple Regression
   - Advantage: Keeps more observations
   - Disadvantage: Treats each response variable as independent from the others

Is there an alternative model that has the advantages of each of these?

Bayesian hierarchical approach to linear regression

- Allows analysis of each health indicator separately in a linear regression model - less need for imputation
- The relationship between the response variables are 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.

Standardizing Independent Variables for Analysis

Standardization of independent variables following the methods put forward in Gelman et al. (2014):

- Scale all to have mean=0, standard deviation=0.5
- Dichotomous/Categorical variable:

1. If $m > 2$, create $m - 1$ indicator variables
2. Find proportion of 0's (π_0) and 1's (π_1)
3. Redefine values: $0 = \pi_1$ and $1 = -\pi_0$

Standardizing Dependent Variables for Analysis

- First verify the normality of all response variables
- Transform variables that are heavily skewed using box cox transformation
- Standardize all response variables:

$$Y_{ij}^{s} = (Y_{ij} - \bar{Y}_j)/\text{sd}(Y_j) \tag{6}$$

- Make directionality consistent

Note: Scaling variables may be not be appropriate for analysis where we need specific interpretations of the coefficients.

Bayesian Hierarchical Linear Regression Model

The three-level hierarchical model is structured as follows:

$$\begin{aligned}
&\text{level 1: } y_{ij}|\boldsymbol{\beta}_j, \sigma_{ij} = \beta_{j0} + \sum_{k=1}^{10} x_{ijk}\beta_{jk} + \epsilon_i, \epsilon_i \sim \text{ Normal}(0, \sigma^2) \\
&\text{level 2: } \beta_{j0} \sim \text{ Normal}(0, 1),\ \beta_{jk} \sim \text{ Normal}(\mu_k, \phi_k),\ \sigma^2 \sim \text{ Uniform}(0, 1) \\
&\text{level 3: } \mu_k \sim \text{ Normal}(0, 1),\ \phi_k^2 \sim \text{ Uniform}(0, 2),
\end{aligned} \tag{7}$$

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.

Simulation Study

Multivariate data with values missing non-randomly (matching the health data missingness patterns)

1. Percentage missing different for each response variable
2. Lowest values for each variable deleted

Compare the following models' performance with this data:

1. Multivariate Linear Regression (removing all rows with missing values)
2. Multivariate Linear Regression (with mean hot-deck imputation)
3. Multivariate Linear Regression (with expectation-maximization and multiple imputation)
4. Simple Linear Regression (where each model is fitted independently)
5. Bayesian Hierarchical Linear Regression

Simulation Study

- $\mathbf{X}$ values are drawn independently and randomly from a standard normal distribution ($x_{ij} \overset{i.i.d.}{\sim} \text{Normal}(0, 1)$)
- $\boldsymbol{\beta}$ 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)

Simulation Study

At each iteration...

- Induce correlation between the response variables: $\boldsymbol{\epsilon}$ randomly generated such that $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\epsilon_i \sim$ MVNorm$(\mathbf{0}, \boldsymbol{\Sigma})$.
- Induce non-random missingness: percentage missing (0-30%) randomly chosen for each Y_j, and the missing values are non-randomly chosen.
- Use different methods to estimate the $\boldsymbol{\beta}$ values.
- Estimate the mean absolute error (MAE) and root mean squared error (RMSE) to compare the $\hat{\boldsymbol{\beta}}$ values to $\boldsymbol{\beta}$.

$$\text{RMSE} = \sqrt{\frac{1}{k+1}\sum_{r=0}^{k}(\beta_r - \hat{\beta}_r)^2} \quad (8) \qquad \text{MAE} = \frac{1}{k+1}\sum_{r=0}^{k}|\beta_r - \hat{\beta}_r| \quad (9)$$

Simulation Study Results

Table: 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	0.0970
	Model 2	0.3285	0.1846	0.1455	0.1563	0.1306	0.0908
	Model 3	0.2096	0.1756	0.0976	0.0868	0.0710	0.0637
	Model 4	**0.1823**	0.1588	**0.0910**	0.0868	0.0624	0.0587
	Model 5	0.1847	**0.1579**	0.0913	**0.0851**	**0.0623**	**0.0585**
MAE	Model 1	0.2764	0.2077	0.0953	0.1263	0.0702	0.0786
	Model 2	0.2612	0.1476	0.1131	0.1128	0.0942	0.0683
	Model 3	0.1781	0.1373	0.0796	0.0704	0.0570	0.0506
	Model 4	**0.1502**	0.1297	**0.0744**	0.0706	0.0510	0.0473
	Model 5	0.1524	**0.1290**	0.0746	**0.0695**	**0.0509**	**0.0471**

Simulation Study Results

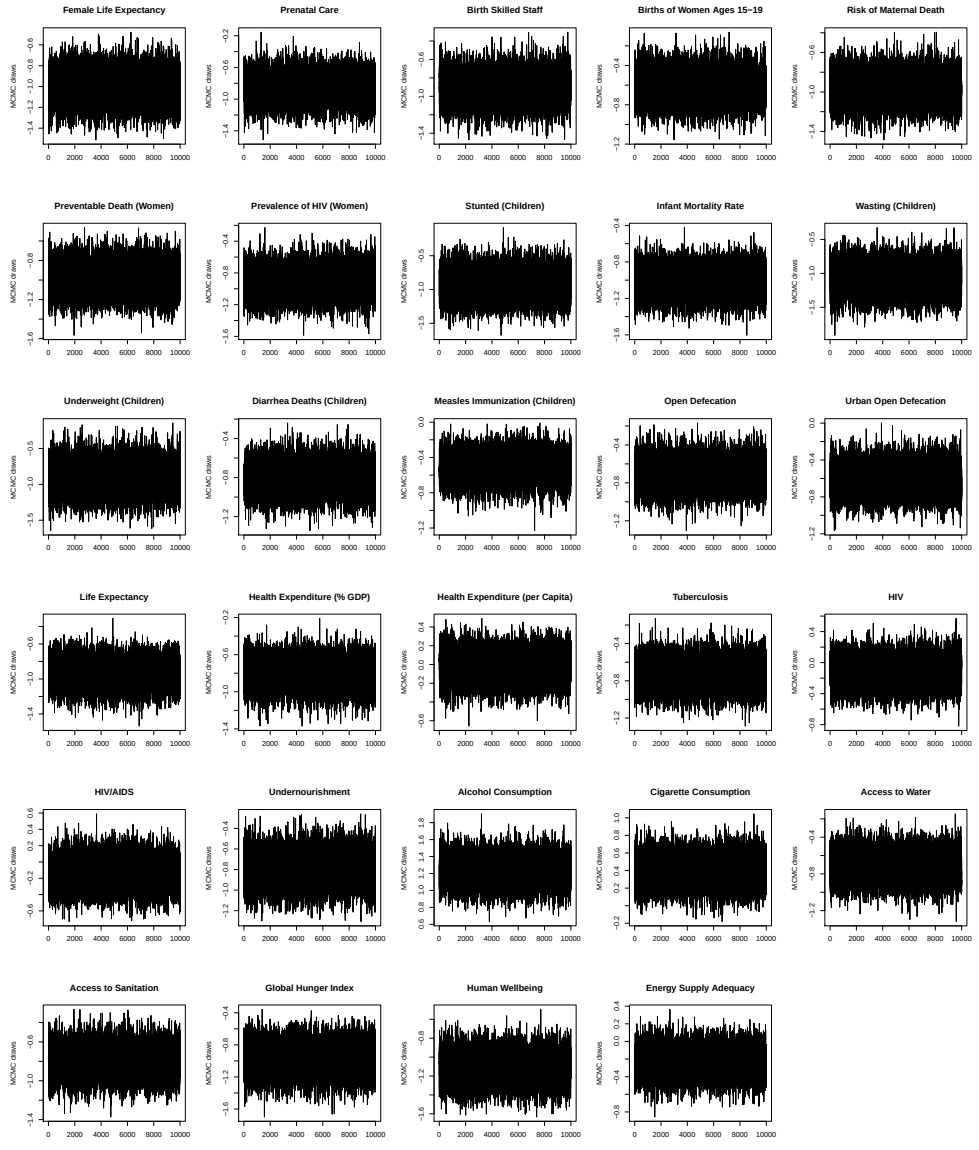
Table: Simulation Study Model Comparisons - Credible/Confidence Intervals

	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**

Bayesian Hierarchical Analysis of Health Data

- Analysis run in JAGS using Markov Chain Monte Carlo (MCMC)
- Chains=5, Iterations=12,000, Burn-in=2,000, Thinning=5 (to reduce autocorrelation)

Model Diagnostics



Model Diagnostics

Diagnostics for the Syndrome Coefficient

- Lowest Effective Sample Size in 29 Models: 8,618 (out of 10,000 posterior draws remaining after burn-in and thinning)
- $\hat{R}$ in 29 Models: ranged from 0.958 to 1.04

Model Results

	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

Findings and Conclusion

- The Syndrome variable significantly linked to all health outcomes for both women and children, and in 10/16 overall health models
- Higher subordination on average corresponds to worse health outcomes
- If women are disempowered, harmed, overly burdened, or silenced, their positive health practices and impacts are stifled.
- Political and social leaders interested in improving health should place empowerment of women at the center of their calculus for efforts to improve health outcomes