

# Lancet Commission Metrics Working Group Large-N Analyses Summary

Oskar Timo Thoms, Dane Rowlands, and Valerie Percival

March 9, 2021

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Key Findings to Date</b>	<b>3</b>
<b>3</b>	<b>Data</b>	<b>5</b>
<b>4</b>	<b>Cross-sectional Analyses</b>	<b>7</b>
4.1	Methods . . . . .	7
4.2	Health and Gender Outcomes . . . . .	8
4.3	Violence Outcomes . . . . .	12
<b>5</b>	<b>Panel Analyses</b>	<b>13</b>
5.1	Methods . . . . .	13
5.2	Health and Gender Outcomes . . . . .	14
5.3	Violence Outcomes . . . . .	16
<b>6</b>	<b>Possible extensions</b>	<b>18</b>

## 1 Introduction

The objective of the Lancet-SIGHT Commission is to examine if and how gender equality and health equity can contribute to more peaceful societies. The central research question which guides the report is “how does variation in gender equality and health equity (improvements or declines) influence variation in violence and conflict.”

The Metrics Working Group explores this question through quantitative analysis of large-N cross-national data, as well as studies based on systematic reviews of the existing literature and the use of descriptive data. To understand the quality of the data for both the large-N and the case studies, we have undertaken a rigorous analysis of indicators used in this research.

Through the analysis of a large number of observations, large-N analysis establishes associations to determine if and how variables relate to each other and if these relationships have statistical

significance. Careful statistical analysis can partly account for endogeneity and lessen the chances of spurious associations. Without doing this large-N analysis, we risk making false causal claims about the generalizability of the relationships between gender equality, health equity and peace. In short, large-N analysis can provide greater confidence of the external validity of our analysis.

Inspired by an article by Suri et al (2010), which argues human development and economic growth operate in vicious and virtuous cycles, we decided to test this concept with three variables - health, gender and violence/conflict. Specifically we want to explore:

- if and how gender inequalities and health inequities and violence reinforce each other in vicious cycles;
- if and how gender equality, health equity, and peace reinforce each other in virtuous cycles;
- and if and how interventions to support gender equality and health equity, under the right conditions, have the ability to nudge communities and societies from vicious to virtuous cycles.

Our large-N analyses begin to build an evidence base for broad statistical relationships between gender, health, and violence outcomes. Based on the virtuous and vicious cycles framework proposing that health, gender, and violence variables interact with each other to produce mutually reinforcing negative (vicious) and/or positive (virtuous) patterns, these analyses are guided by the following questions:

- While accounting for the overall improvements that most countries have experienced in health and gender outcomes in recent decades, do those countries that score low in earlier periods improve less than the global average and/or high performers, or do they perhaps even decline?
- While accounting for possible ceiling effects, do those countries that score high in earlier periods a) maintain stable and high levels and b) avoid violence?
- Do countries that experience higher levels of violence a) improve less on gender equality and health outcomes than those with less violence, or b) do their gender and health outcomes perhaps even deteriorate?
- Are countries that score low on health and gender outcomes more prone to violence than those that score higher?

These questions imply considerable complexity, as they examine statistical effects between gender and health outcomes; from gender and health outcomes to violence; and from violence to gender and health outcomes. Our multivariate regression analyses test these questions in probabilistic terms, not in deterministic fashion. While there are many other factors influencing health, gender and violence outcomes, we do not develop full causal models of each outcome, but examine whether associations between the three categories of outcomes are consistent with the theoretical framework of virtuous and vicious cycles.

## 2 Key Findings to Date

We examined multivariate models of both cross-sectional and panel data. These two different approaches to modeling cross-national data have complementary advantages and disadvantages. First, the cross-sectional models allow us to examine long-term changes in outcome variables, over 20-year periods, while the panel models examine short-term variation, from one five-year period to the next, but over longer study periods than the cross-sectional analyses. Second, the two approaches have different methodological trade-offs. While the cross-sectional models make it somewhat easier to address problems of endogeneity and heteroscedasticity, they also have fewer degrees of freedom due to their small sample sizes (i.e. the number of countries included). Panel models are subject to more concerns about endogeneity, heteroscedasticity and serial correlation due to the structure of the data used (several observations of countries over time), but they also have much larger sample sizes, allowing for more precision in estimates. We believe that both approaches are valuable in examining our research questions.

The cross-sectional analyses yield the following conclusions:

- Over the period from 1995 to 2015 there was a general improvement in the four measures of health and gender used in this analysis.
- Analyzing performance is complicated due to the presence of ceiling effects. Once a country has obtained a high level of performance in their health and gender performance, the room for additional improvements is quite limited. How we measure improvements is important, as is the manner in which we take into account the limited scope for changes in performance. In the regressions it is generally the case that past good health performance is associated with smaller future improvements, and the same for gender performance measures.
- In general there is reasonable evidence consistent with the presence of a virtuous feedback loop whereby better past gender performance is associated with improved future health measures, and better past gender measures are associated with improved future health performance.
- The cross-section variation in health and gender measures is typically well explained by these fairly simple models. The percentage of variation explained by the models (the  $R^2$  value) ranges from a low of about 35% (for the adolescent fertility rate) to over 80% (for the infant mortality rate).
- For the most part, the equations also perform well when using only past values of key explanatory variables (no concurrent values). These prediction models are less subject to some of the simultaneity and endogeneity problems with the base equations, and generally still yield comparable  $R^2$  values.
- As is well known from the literature, there is a high degree of inertia in violence on conflict over time. In other words, once countries experience violent conflict, they likely to do so in the future.
- In the analysis prior conflict is often positively related to future health and gender improvements, which appears to be the consequence of a reduction in past performance and the possibility of recovery. By contrast, and as expected, concurrent (future) violence and conflict is generally negatively related to health and gender improvements.

- There are many different measures of conflict and violence, and their relationship amongst one another and with health and gender measures is varied and often nuanced. It is difficult to identify one “summary” and “best” measure of conflict and violence, or to ascribe one relationship that applies to all of these measures. Investigating the nuances of different violence and conflict measures might be a very useful extension of this work.
- This analysis focuses on three conflict and violence measures (number of years experiencing significant conflict over a period of time, civilian deaths associated with civil conflict, and civilian deaths associated with one-sided violence) all seem to exhibit a negative association with better health and gender measures. More years of female education relative to male appears to have the most consistent negative association with future death and conflict rates. This evidence is supportive of the virtuous/vicious circle proposition, in which better health and gender performance is associated with less violence, which in turn generally contributes to better future health and gender performance.

The panel models lead to the following conclusions:

- As in the cross-sectional analyses, gender and health outcomes are clearly positively correlated with each other. This is consistent with the notion that they reinforce each other.
- Unlike in the cross-sectional analyses, there is no suggestion of ceiling or convergence effects in these analyses. Better health and gender outcomes in one period are associated with improved outcomes in the next period. This difference between the cross-sectional and panel analyses is likely due to the fact that they are differently structured to examine long-term versus shorter-term associations.
- While in bivariate associations, poor health and gender outcomes are clearly associated with more violence on average, the role of violence variables in health and gender outcomes is complex, as violence in one period is sometimes associated with better health and gender outcomes in the next period. This result may seem at odds with the vicious and virtuous cycles, and points to the need for fuller statistical models based on deeper theoretical work. Moreover, the analyses suggests that different types of violence may matter in different ways.
- The panel analyses also consider an alternative conceptualization and operationalization of vicious and virtuous cycles, by examining the effects of low and high classifications based on health and gender outcomes. These analyses find clear and consistent support for vicious cycles between health and gender outcomes and in their effects on violence outcomes.
- While we do not find an analogous result for virtuous cycles, we caution that this may be due to our particular conceptualization, which focuses on the extreme ends of our classifications, i.e. the worst and best performers in comparison to all the countries in between, assuming that virtuous and vicious cycles would be most pronounced in these groups. An extension of this research should examine whether variation in the middle group is consistent with the notion of virtuous cycles. Such research may also advance understanding of pathways of change if it considers sequencing.

- In models of violence outcomes with interactions between lagged dependent variables with the low and high classifications, we also find strong evidence for three-way vicious cycles, whereby previous violence is associated with renewed (or continued) violence particularly in countries that have very poor health and gender outcomes.

This research represents useful ground-clearing for further analyses investigating vicious and virtuous cycles between health, gender and violence outcomes, and we provide suggestions for extensions below. The remainder of this memo discusses the data, methods, and findings leading us to these conclusions. Further details on data preparation and the analyses (including regression details) are available at [oskarthoms.net/lcwg/](https://oskarthoms.net/lcwg/).

### 3 Data

Guided by the indicator mapping carried out by the Metrics Working Group, we identified a wide range of health, gender and violence measures as candidates for analyses, and assembled a comprehensive cross-national time-series dataset with the best available data on relevant indicators and key covariates.<sup>1</sup> Many health and gender measures in particular are subject to many missing data points across countries and over time. Based on our evaluation of the quality and coverage of these data, and based on within-category correlations, we chose two representative and commonly accepted variables each for the health (life expectancy and the infant mortality rate, IMR) and gender (the ratio of female-to-male mean years of schooling and the age-specific fertility rate for adolescents) dimensions. To measure violence, we chose several indicators to capture different types, including state-based armed conflict (internal conflict/war), repression (one-sided violence [OSV], latent physical integrity, state torture and extra-judicial killings) and societal violence (non-state conflict [NSC] and homicide rates), and the total and civilian battle-related death rates (per population) resulting from varying types of conflict.<sup>2</sup> For control variables, we use per capita income to represent the level of a country's economic development, and measures of political structure representing the presence or absence of different facets of democracy. While not exhaustive, these variables exhibit useful features in terms of country and year coverage, collection and reporting reliability, and theoretical plausibility.

It is important to note that, for easier comparisons of results across variables within each category, we have inverted some variables for the statistical analyses such that higher values of the health and gender variables always indicate “better” outcomes, while higher values of the violence variables always mean higher levels of violence.

To help operationalize the vicious and virtuous cycles hypotheses, we developed a classification of countries based on the four health and gender outcomes, creating groups of countries for comparison.<sup>3</sup> In some of our models, we examine the statistical effects of the lowest and highest

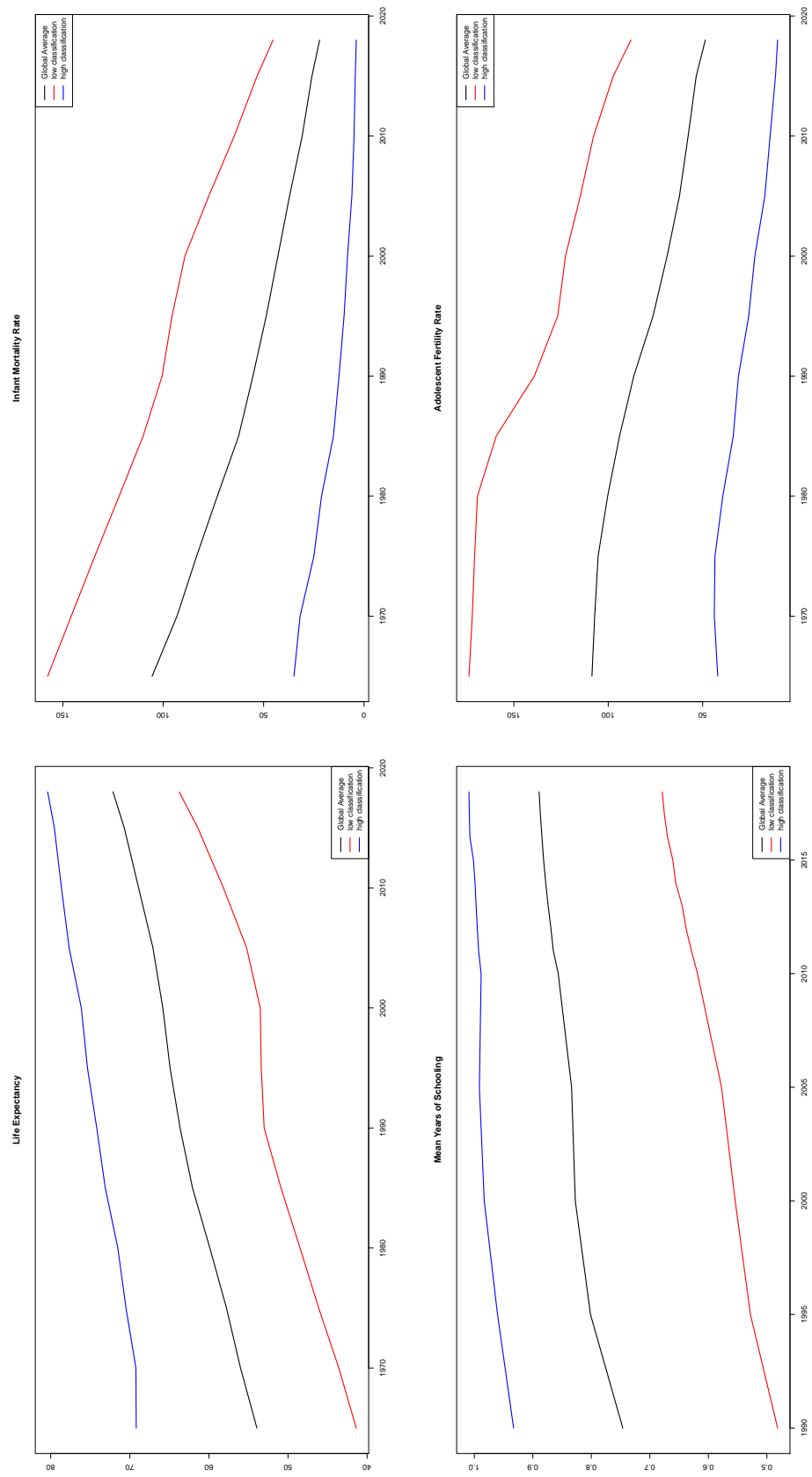
---

<sup>1</sup>A detailed codebook is available at this [link](#).

<sup>2</sup>As data availability and quality generally improves after 1990 and more violence variables exist for this period, the primary period of analyses is 1990-2015 or 2018. All variables are aggregated to five-year averages because three of the key health and gender variables are provided as such averages by the World Populations Prospects Database.

<sup>3</sup>We divided each of the four measures into quintiles at any given point in time and aggregated them along the gender and health dimensions, to produce a 5x5 matrix. We selected the bottom four cells as our low classification and the top four as our high classification. In the analyses presented here, we use only these groups, but the other groups resulting from this classification procedure could be analysed in extensions of this research.

Figure 1: Health and Gender Trends



classifications, making the assumption that if there are virtuous and vicious cycles present in the data, then they surely would be observed for these two groups.

Initial graphical analyses of the averages of the health and gender variables at the global level and within the two classification groups show clear improving trends over time (see Figure 1). First, there is considerable global improvement in the health and gender measures, regardless of how countries are classified. Second, improvements are greatest in the low classification, and tend to be less pronounced in the high classification. This suggests possible ceiling effects for better performers that need to be accounted for in the analyses; this is discussed further below. Other graphical analyses (not shown here) also indicate that low performers on the health and gender dimensions tend to have the most violence on average over time, and high performers experience much less violence.<sup>4</sup>

## 4 Cross-sectional Analyses

### 4.1 Methods

The cross-sectional regression analyses examine longer-term effects. The dependent variables are the improvements (increase from the 1996 start period to the 2015 end period) in country performance from 1996 to 2015 for each measure of health and gender performance, and a measure of cumulative violence or conflict over the same time-period. The key explanatory variables are the values of the different dependent variables representing conditions prior to, or in, 1995, as well as the control variables for the level and subsequent changes in per capita income and political structure, and for the subsequent presence of conflict. The time structure of the analysis means that the regressions are essentially testing for any consistent and significant statistical relationships indicating a possible direct association between future health, gender, and conflict performance over a long (twenty year) period, starting conditions of health, gender, conflict, economic and political conditions, and indirect connections through subsequent economic, political, and conflict developments. Using the cross-section analysis and the long time period means that we are examining longer-term trends rather than year-to-year variation. The structure also minimizes the likelihood of reverse causality and endogeneity, and versions of the regressions are essentially trying to predict future performance using only prior information.

These key regression equations are also supplemented by similar equations to examine the factors associated with economic improvements over the 1996-2015 period, and changes in political structure. This analysis tests for potential indirect associations between different performance measures. For example, a direct association would be suggested by the analysis if initial gender or health performance is related to subsequent improvements in those measures. An indirect relationship might be suggested if, instead, subsequent improvements in gender or health over the long term are only related to economic, political, or conflict patterns over the long run as well, but that these are in turn influenced by initial health and gender performance. So the causal route from better initial gender performance to improved health performance might not be directly from gender to health, but through initial gender conditions affecting subsequent economic performance, which in turn affects health performance.

---

<sup>4</sup>In a partial exception, average homicide rates are highest in intermediate rather than the lower classification groups.

Finally, the process of narrowing down the most informative regressions is extensive. In addition to standard diagnostics on heteroscedasticity and multicollinearity, several versions of the regressions are examined to gauge the sensitivity of the results to variations in the equation, included observations, and variable forms. Due to the presence of multicollinearity in some cases, and since the number of observations is occasionally fairly small, variables with statistically insignificant coefficient estimates are removed from the equations to test for the sensitivity of results. This step-wise process is carefully scrutinized at each step to avoid cherry-picking results and to ensure that the reported results are robust to most changes. We also drop some variables for theoretical reasons, and this process is often quite informative in identifying potentially important associations. We clearly identify the inferences associated with these changes.

## **4.2 Health and Gender Outcomes**

There are two important observations on the data that need to be recognized. First, on average most of the key indicators of health and gender exhibit steady improvement over the 1995-2015 period both in aggregate and for almost all countries in the sample. As a result, in the cross-section analysis virtuous circles and vicious circles are essentially opposite sides of the same coin as the performance of countries are essentially measured relative to one another. The “virtuous” group performing relatively well in terms of future performance (virtuous circle) implies that the “vicious” group is performing relatively poorly (vicious circle). One line of inquiry that might be worth pursuing is to impose an absolute measure, for example sustained declining performance, and ask which country groups exhibit absolute decline. In the cross section data it is rare that well-off countries end up having sustained reduction in health and gender performance, while poorly-off countries exhibit a much broader range of future performance, ranging from sustained declines to dramatic improvements.

In addition, the pace of improvement tends to be significantly faster for those countries that have poorer initial performance in these categories. This pattern suggests the presence of a “ceiling effect”: those countries that have already achieved high performance in health and gender equality have very little room for additional improvement, while those with weak initial performance have ample scope for improvement. The presence of a ceiling effect complicates the analysis considerably, since the data will tend to suggest the opposite of virtuous and vicious circles simply because initially poor performers will find it far easier to outperform initially strong ones. An alternative interpretation of this robust result is the convergence hypothesis whereby poorer performers generally tend to catch up with the top performers. Differentiating between the two interpretations is difficult, since both are plausible. We conduct a preliminary investigation of this problem by doing supplementary analysis using improvement measures that are adjusted for initial conditions. In effect, this supplementary analysis asks how countries with better gender or health measures perform relative to comparable countries in terms of income per capita.

The use of per capita income to adjusted performance expectations is linked to the second key observation: countries with higher per capita income also tend to have better initial measures of for health, gender, conflict and violence, and democracy. These strong cross-variable correlations mean that there is often considerable multicollinearity in the model, especially between health and gender variables, as well as with lower levels of violence. This static correlation is indeed suggestive of virtuous and vicious circles, but it does not imply a direct causal and dynamic relationship. Instead, it may simply be the case that a single common variable, such as economic performance, causally



affects these measures in the same way. The regression analysis is structured to address this problem, though imperfectly.

Table 1 presents the coefficient estimates from the regressions examining improvements in the health and gender outcomes. The first two columns show the results for changes in life expectancy; the reported results are representative of the analysis of life expectancy, and illustrative of many of the findings of other performance measures as well.<sup>5</sup>

The base model (column 1) provides the results for the best, most representative stepwise reduced estimating equation. The negative coefficient on both the life expectancy and infant mortality variables is indicative of the common “ceiling effect” observed in almost all of the regressions. Essentially, if a country is already performing well in terms of health performance, that country is likely to have little room for significantly better performance. For each one-year increase in life expectancy in 1995, the increase in life expectancy over the next 20 years declines is 0.3 years lower; the effect of a decline of one infant death per 1000 is to lower future life expectancy improvement by 0.064 years over the next 20 years.

A result that appeared quite robustly in the different regressions was the positive association between improvements (numerical declines) in the adolescent fertility rate of a country, and subsequent improvements in life expectancy. This association indicates that if the number of children born per 1000 adolescent girls fell by one, future life expectancy would improve by .05 years. The association could be indicative of a broad virtuous relationship between better overall rights for women (to the extent that these are primary factors in decreasing teenage pregnancy), or it could be more directly causal in the sense that reduced physical and emotional stress from adolescent pregnancy and childbirth may increase future life expectancy for a girl as she ages. The robust association, however, is encouraging.

The conflict and violence variables have the expected relationship with life expectancy improvements but are complex. A sensible and fairly robust pattern emerged from the regression analysis: previously high rates of conflict or violence tended to be associated with future improvement in life expectancy, which is possible if previous conflict had artificially reduced life expectancy levels below what would otherwise have been the case. Subsequent life expectancy improvements, therefore, might simply reflect the reversion to expected levels after a conflict ends. By contrast, measures of conflict or violence over the period for which life expectancy improvements are being measured are associated with declines in that improvement, which is again sensible. In some regressions, removing the future conflict and violence measure often led to a reversal of signs for the previous conflict and violence measure, which we think is due to the tendency for violence to have a high degree of inertia, so that the decline in life expectancy is reflecting the likely presence of future conflict, not the presence of past violence.

The economic and political control variables all have plausibly signed, coefficient estimates. Future economic growth and future high levels of democratic performance are both associated with improvements in life expectancy over the same period. Finally, the equation as a whole performs remarkably well, explaining almost half of the cross-sectional variance in life expectancy improvement for a sample of 156 countries.

When the life expectancy measure in 1995 is removed from the estimating model (not shown) to

---

<sup>5</sup>Recall that some variables have been rescaled so that an increase in any health or gender measure represents an improvement in performance; for example, infant mortality has been inverted so that when the measure of infant mortality increases, it represents a decline in infant mortality rates. The only variables where an increase is “bad” are the conflict and violence measures.

Table 1: Health and Gender Equality Improvements (change during 1995-2015)

	Life Expectancy base	Life Expectancy prediction	IMR base	IMR prediction	MYS Ratio base	MYS Ratio prediction	AFR base	AFR prediction
Life expectancy 1995	-0.302 † (0.074)	-0.259 † (0.080)						
Infant mortality rate 1995	-0.0643 ** (0.012)	-0.0606 ** (0.016)	-0.628 *** (0.000)	-0.583 *** (0.000)			0.136 ** (0.017)	0.151 ** (0.013)
Adolescent fertility rate 1995	0.0506 ** (0.013)	0.0449 ** (0.011)	0.123 *** (0.009)	0.0962 * (0.032)	0.000588 ** (0.023)	0.00062 * (0.017)	-0.265 *** (0.000)	-0.271 *** (0.000)
Gender schooling ratio 1995					-0.368 *** (0.000)	-0.354 *** (0.000)		
Conflict measure 1989-1995	0.0601 † (0.091) <sup>a</sup>	0.0318 † (0.093) <sup>c</sup>			0.000532 † (0.083) <sup>c</sup>	-0.000365 † (0.079) <sup>c</sup>		
Conflict measure 1996-2015	-0.193 * (-0.049) <sup>b,r</sup>				-0.012 *** (0.003) <sup>c,r</sup>			
GDPpc 1995			-0.2361 *** (0.009)					
GDPpc growth 1996-2015	0.0630 *** (0.008)		0.137 * (0.027)					
Past Polyarchy 1991-1995		4.22 *** (0.003)	8.97 ** (0.02)			-0.0558 † (0.071)	-16.37 ** (0.024)	-16.92 † (0.052)
Future Polyarchy 1996-2015	4.80 *** (0.002)				-0.0619 * (0.041)			
Constant	22.8 † (0.072)	21.1 † (0.060)	-1.30 (0.676)	1.824 * (0.05)	0.470 *** (0.000)	0.452 *** (0.000)	16.4 *** (0.000)	16.13 *** (0.002)
Sample size	156	165	156	180	135	135	165	165
R <sup>2</sup>	0.492	0.483	0.824	0.817	0.521	0.510	0.344	0.344
F-test <i>p</i> -value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

P-values in parentheses, derived from robust standard errors; \*\*\*, \*\*, \*, † indicate statistical significance at the 0.01, 0.025, 0.05 and .10 levels respectively; <sup>a</sup> internal conflict indicator; <sup>b</sup> civilian death rate from internal conflict; <sup>c</sup> civilian death rate from one-sided-violence; <sup>r</sup> robust to variations in the other conflict measure.

ensure no spurious connection between the dependent variable and its lag, the ceiling effect is still present as indicated in the negative association with the remaining health measure (infant mortality), since these health measures are highly correlated (0.96). The other results are fairly comparable except that the polyarchy measure that retains its statistical significance switches from the future values to the prior value. While this equation instability suggests that some caution is warranted in making inferences about the relationship between political structure and life expectancy, it is the case that there is a high degree of continuity in the polyarchy measure, with the pre- and post-1995 measures having a correlation of 0.9. Finally, there is only a small loss in explanatory power as represented by the  $R^2$  value.

The model variant in the second column uses only pre-1996 variables to generate a prediction of post-1995 life expectancy improvement. The results are comparable to those of the previous, and while the sample increases slightly, the proportion of variance explained in the model remains very high (0.48). Finally, another model (not shown) examined the improvement in life expectancy adjusted for a country's initial life expectancy level and initial per capita income. While this form of the dependent variable yields a very different measure (the correlation between the actual improvement and the adjusted measure is only 0.08), the results are remarkably similar in terms of the explanatory variables with statistically significant coefficient estimates, and the sign and magnitude of those estimates. The main difference is that lower prior per capita income is associated with better subsequent improvement, which is consistent with a recovery in life expectancy from previously lower than expected levels.

Overall, the analysis of life expectancy improvement suggests the following important relationships. First, there appears to be a strong ceiling effect in which previously good performance on health measures limits the scope for future improvement. Second, lower levels of adolescent fertility rates (AFR) in the past are consistently linked with better future life expectancy improvements. While AFR may reflect both health and gender balance conditions, the relationship is consistent with the virtuous circle concept, and possibly the importance of gender balance as a factor in future health improvements. Past conflict is associated with subsequent life expectancy improvement, which is presumably because of its effect on suppressing the starting point for measuring future improvement. Future conflict, however, clearly suppresses future life expectancy. There is also evidence that democracy is associated with improved life expectancy performance, as is economic growth.

The analysis for infant mortality rate is generally consistent with the findings for life expectancy. The third and fourth columns in Table 1 present the key results. As with life expectancy, in both the base model and the predicting model there is a strong ceiling effect, but also a strong positive connection to previously lower rates of adolescent fertility. There is no strong and consistent link to prior or future conflict. Democracy is positively associated with improvements in infant mortality. There is weak evidence that poor prior economic performance, and better future economic performance, are linked to infant mortality rate improvements. These two variables are themselves correlated (0.56) though the evidence of strong multicollinearity in the equation is weak, and when one is removed from the equation the statistical significance of the coefficient estimate for the other drops below acceptable levels, while the other coefficient estimates remain qualitatively unchanged. So, there is a need to treat the relationship between income per capita measures and infant mortality improvements with some caution. Both the base equation and the predictive equation have very high rates of explanatory power, with the  $R^2$  being about 0.82 in both cases.

The estimations for improvements in gender schooling years ratios and adolescent fertility rates

are presented in the remaining columns. The results all point to the presence of a ceiling effect (i.e. a consistent negative relationship between the previous level and the future improvement rate). There is no evidence that either of the health measures affects the schooling measure, but solid evidence of a positive link between infant mortality rates and future AFR improvements. Conflict affects schooling improvements in the same way as life expectancy: past conflict is associated with future improvements (attributed to catch up) and future violence is associated with lower improvements in female-to-male ratios of schooling years. The one anomaly is in the prediction model for gender balance in schooling, where current conflict is negatively associated with improvements. The likely reason for this reversal is the strong inertia in conflict measures, with past violence being highly correlated with future conflict (correlation = 0.69), so when future conflict is not included in the equation its effect is transferred to the associated past measure. Neither gender-related variable is associated with economic performance. The only unexpected effect is the frequent, though often statistically marginal, negative relationship between democracy and future improvements in both gender measures. This result requires future investigation.

### 4.3 Violence Outcomes

The analysis so far has focused on the possible direct effects of prior gender and health performance on their future improvements, controlling for key characteristics of conflict, the economy, and political structure. These last three variables may also provide indirect channels for affecting the future. Health and gender levels could amplify their effects through these key controls. To investigate these potential indirect connections we use regression analysis to estimate the association between past health and gender measures and future conflict, economic growth, and political structure.

A key set of relationships for our work involves health, gender, and violent conflict. We investigated several measures of violence in our analysis, including in the previous regressions on health and gender improvements. Here, the primary violence indicators we use are a count measure for the number of years a country experiences violent conflict, a measure of the number of civilian deaths from internal conflict, and the number of civilian deaths associated with one-sided violence. All the future cumulative values of these three variables for the 1996-2015 period are positively correlated with each other and negatively correlated with the past measures of the health and gender variables in the analysis. The formal regression analysis uses these future violence variables as the dependent variable, while the explanatory variables are past health and gender measures, the lagged dependent variable, and measures of economic performance and political structure. Some illustrative results are presented in Table 2.

In the prediction of future violence the results were consistent. In all cases the best step-wise reduced equation includes one of the health and gender measures, often along with the lagged version of the dependent variable. Past education balance is significantly associated with all three future violent conflict measures, and all four health and gender measures are associated significantly with future death rates from one-sided violence. Both health measures are also statistically significantly associated with the variable measuring the number of conflict years, but not civilian deaths from internal conflict. Therefore, there is considerable evidence that better performance in health and gender generally is associated with fewer future deaths, which in turn is often associated with greater improvements in health and gender in the future.

The analyses of future economic growth and future levels of democracy suggest that these indirect channels have very little role to play (not shown). Current health and gender measures do

Table 2: Predicting future conflict/violence phenomena (1996-2015)

	Future Internal Conflict Years	Future Internal Conflict Years	Future Internal Conflict Years	Future Civilian Deaths (Conflict)	Future Civilian Deaths (OSV)
Life expectancy 1995	-0.103 *** (0.000)				
Infant mortality rate 1995		-0.0265 *** (0.000)			
Mean years schooling 1995			-5.00 *** (-0.010)	-2.20 * (0.030)	-1.422 *** (0.010)
Lagged dependent variable	0.407 *** (0.000)	0.404 *** (0.000)	0.374 *** (0.000)		0.0683 *** (0.000)
Constant	7.71 *** (0.000)	-0.248 (0.498)	5.27 *** (0.003)	0.549 * (0.034)	1.37 *** (0.008)
Sample size	180	180	143	143	143
$R^2$	0.471	0.473	0.413	0.013	0.54
F-test $p$ -value	0.000	0.000	0.000	0.030	0.000

P-values in parentheses, derived from robust standard errors; \*\*\*, \*\*, \*, † indicate statistical significance at the 0.01, 0.025, 0.05 and .10 levels respectively.

not appear to be significantly associated with either the economy or political structure, with the exception of a weak relationship between life expectancy and future economic growth.

## 5 Panel Analyses

### 5.1 Methods

In contrast to the cross-sectional models, the panel models examine short-term changes from one five-year period to the next, over as many periods as data availability allows; they use so-called "wide" panels of many countries but only a small number of time periods (between 6 and 11 periods). The dependent variables of the analyses are the individual health, gender, and violence variables. We test the vicious and virtuous cycles hypotheses by examining the lagged gender or health variables, or the low and high classification, and the lagged violence variables, as explanatory variables. All the panel models include covariates for logged GDP per capita, logged population size, and participatory democracy, in order to account for socio-economic and institutional conditions.<sup>6</sup> We further include the lagged dependent variable in each model in order to account for trends, autocorrelation and ceiling effects.

Analysis of observational data always requires careful attention to omitted variable biases and endogeneity of various kinds. These issues become even more pressing concerns in panel data than

<sup>6</sup>We also substituted other political regime variables, but only included participatory democracy in the final models because the others did not show statistical associations with the outcome variables.

in cross-sectional data, because of the lack of independence of observations within countries over time, leading to issues of correlated unit (country) effects, heteroscedasticity and serial correlation. These are problems in our models and data, as evidenced by a variety of econometric tests developed for panel data, which we carried out for each regression equation to decide which estimators are most defensible. This is important as the choice of estimator often leads to very different results. Based on the results of these tests across models, we chose the fixed effects (or "within") ordinary least squares estimator (OLS run on demeaned within-country data), and more preferably, the fixed effects general (or unrestricted) feasible generalized least squares (GGLS) estimator.

## 5.2 Health and Gender Outcomes

Rather than presenting the many detailed regression tables of these analyses, we show summary tables of statistically significant findings (at the 0.05 alpha level).<sup>7</sup> The summary tables below show the outcome variables (DVs) in columns and explanatory variables and covariates (IVs) in the rows, such that each cell represents a separate statistical model. Statistically positive and negative effects are indicated by "+" and "-", respectively, and "[ ]" indicates associations that contradict our expectations. (Effects found in FE OLS estimations but not in the preferred FE GGLS estimations are indicated by "()"; we consider this weaker evidence.) Table 3 shows the findings of the models with health or gender dependent variables.

As in the cross-sectional analyses, there is clear evidence that health and gender outcomes reinforce each other, but the evidence is somewhat stronger for health than gender outcomes, with two caveats. Both the mean years of schooling (MYS) ratio in favor of females and the adolescent fertility rate (AFR) are positively associated with increased life expectancy, and MYS is also positively associated with improvements in the infant mortality rate (IMR). Recall here that both the IMR and AFR were inverted for the purpose of the analyses. This means that the higher the MYS ratio, the better the IMR. In a contradictory and surprising finding, the AFR, which was also inverted, is negatively associated with the inverted IMR: this means that improvement in the AFR is associated with worsening in the IMR. This result requires further investigation. For the gender outcomes, life expectancy and the IMR are associated with an improved AFR, but they have no statistical effects on the MYS ratio. The null findings regarding associations between the health variables and the MYS ratio may well be due to the substantially lower sample size, as this measure is available for fewer time periods. Finally, across the four outcome variables, the lagged dependent variable always has a positive association, meaning that higher values in the previous period are associated with better outcomes in the next period. Thus, unlike in the cross-sectional analyses, the panel analyses do not suggest possible ceiling or convergence effects.

As discussed in the cross-sectional analyses, when conceptualizing vicious and virtuous cycles as two sides of the same coin, positive correlations between our outcome measures provide evidence that is consistent with the theoretical framework. Other conceptualizations of vicious and virtuous cycles are possible, however, including considering vicious and virtuous cycles separately. We explore an alternative conceptualization by examining the statistical effects of our low and high classifications, which operationalize the vicious and virtuous cycles hypotheses with the worst and best combinations of health and gender outcomes, and compare these to all countries in-between these groups. (Other conceptualizations are still possible and present promising avenues for extensions

---

<sup>7</sup>The details of individual regressions are provided at [oskarthoms.net/lcwg/panel.html](http://oskarthoms.net/lcwg/panel.html).

Table 3: Summary of Findings on Health and Gender Outcomes

	Life Expectancy	IMR	MYS Ratio	AFR
Life Expectancy	+	na		+
Infant Mortality	na	+		+
Mean Years of Schooling ratio	+	+	+	na
Adolescent Fertility Rate	+	[-]	na	+
low classification	-	-	-	-
high classification				
Conflict	[+]			
Conflict: low interaction				
Conflict: high interaction		+		
War				-
War: low interaction				
War: high interaction				
Conflict Deaths		-		[+]
Conflict Deaths: low interaction				[+]
Conflict Deaths: high interaction				
Conflict Civilian Deaths	[+]			[+]
Conflict Civilian Deaths: low interaction	[+]			[+]
Conflict Civilian Deaths: high interaction				
OSV		(-)		
OSV: low interaction		(-)	[+]	[+]
OSV: high interaction				
OSV Deaths	[+]	(-)		[+]
OSV Deaths: low interaction				
OSV Deaths: high interaction				
NSC				[+]
NSC: low interaction				
NSC: high interaction				
NSC Deaths	[+]			[+]
NSC Deaths: low interaction				
NSC Deaths: high interaction		[(+)]		
NSC Civilian Deaths	[+]			[+]
NSC Civilian Deaths: low interaction				
NSC Civilian Deaths: high interaction			[(-)]	
LPI		[+]	-	[(+)]
LPI: low interaction				
LPI: high interaction		(+)		
Torture				
Torture: low interaction				
Torture: high interaction		(+)		
Killings				
Killings: low interaction	-	-		
Killings: high interaction		(+)		
Homicides	-	[+]		[+]
Homicides: low interaction				
Homicides: high interaction				



of these analyses.) When considering the low and high classification, the findings are clear and consistent: the low classification is associated with worse health and gender outcomes across all four variables, but the high classification is never statistically significant. This provides strong support for vicious cycles between health and gender but is not consistent with virtuous cycles among countries with the best health and gender outcomes. It is important to note here that this null finding regarding virtuous cycles applies to the comparison between high performer and those in the middle group, and thus does not speak to possible virtuous cycles in the latter. This, again, suggests a useful extension for further analyses.

Thus far, these findings speak to associations between health and gender outcomes, but the vicious and virtuous cycles framework also proposes associations with violence and peace, respectively. We examine the role of violence by including the various violence measures and their interactions with the low and high classifications in separate models. Few of the violence variables have statistical effects on the gender and health outcomes that are consistent with this framework. In fact, many models suggest that violence is associated with better average health and gender outcomes, as indicated by “[+]” in the table. (The findings for the AFR are particularly noteworthy in this respect.) A few results are consistent with the vicious and virtuous cycles framework, such as the associations of internal war with a worse AFR; internal conflict deaths and one-sided violence with worse infant mortality; the latent physical integrity measure with lower MYS ratios; and, unsurprisingly, homicides and extra-judicial killings with lower life expectancy. However, many models with violence variables turn up null results, or even contradictory ones. These results suggest that the role of violence in health and gender outcomes is more complicated and deserves deeper investigations. In the next section, we explore vicious and virtuous cycles in violence outcomes, and here the evidence is clearer and more consistent.

### 5.3 Violence Outcomes

Table 4 summarizes the findings of the violence models, with thirteen different measures of violence outcomes. These analyses also partially support the theoretical framework, though there are again some contradictory results with respect to the relationship between individual health and gender variables and violence variables, and many null results. In supporting evidence, higher life expectancy is associated with less one-sided violence episodes (though, surprisingly, more deaths from such violence) and possibly with less state torture; improved infant mortality is associated with less internal conflict and possibly less one-sided violence and physical integrity violations; higher MYS ratios are associated with less internal conflict, one-sided violence deaths and physical integrity violations. While an improved AFR is associated with less internal conflict, it is also associated with more state violence and homicides. These results are mostly supportive, but are not consistent across health and gender variables and different types of violence. Useful extensions would consider further health and gender measures where data availability allows it, and build fuller models of these violence outcomes.

When considering our low and high classifications, there is clear and consistent evidence for vicious cycles. The low classification is associated with more internal war, internal conflict deaths, non-state conflict deaths and homicides. The high classification (indicating strong performance on both health and gender) is associated with less internal conflict civilian deaths and homicides, but also, surprisingly, with more violent state repression. Importantly, a three-way vicious cycle between health, gender and violence is strongly supported in many models by the interaction of the



Table 4: Summary of Findings on Violence Outcomes

	Conflict	War	Conflict Deaths	Civilian Conflict Deaths	OSV	OSV Deaths	NSC	NSC Deaths	NSC Civilian Deaths	LPI	Torture	Killings	Homicides
Life Expectancy				[+]	-	[(+)]				[+]	(-)		
IMR	-				(-)	-				(-)			
MYS ratio		-								-		[(+)]	
AFR	-									[+]/(-)			[+]
low classification		+	+				[-]	+		[+]		[+]	+
high classification				-					+	+	+	+	-
lagged DV	+	+	+	+	+		+	+		+	+	+	+
low interaction		+	+	+				+		+		+	+
high interaction		+	+							[+]		+	[(+)]

lagged violence variables with the low classification; these models examine the effect of previous levels of violence in countries that perform poorly on health and gender on subsequent levels of violence. These interactions are associated with more internal war, internal conflict deaths and civilian deaths, non-state conflict deaths, physical integrity violations, extra-judicial killings and homicides, compared to countries not included in the low and high classifications. These findings strongly suggest that violence begets more violence where countries perform very poorly on both health and gender, compared to better performing countries. While the high classification does not turn up consistent results and its interactions with the lagged violence variables sometimes has unexpected positive effects, it is important to note again that this should not be regarded as evidence against the presence of virtuous cycles, since it may be an artifact of our conceptualization, which does not explore variation in the middle groups between the low and high classifications.

## 6 Possible extensions

Throughout the discussion, we have noted possible avenues for extensions of this research. Deeper theoretical work and fuller statistical models are needed to address questions arising from these analyses and explore potential pathways from vicious to virtuous cycles.<sup>8</sup>

Our analyses are limited in several ways. First, they examine broad associations and include the same covariates across models. However, each of the outcome variables may be better modeled with a different set of covariates, based on different literatures. Such work may shed light, for instance, on the unexpected association between violence and improved adolescent fertility rates. Second, the analyses do not examine interactions between different types of violence, or even control for other types of violence in a given model. Third, the analyses use only the low and high classification to operationalize the vicious and virtuous cycles hypotheses. However, as noted above, analyses using a richer classification (including the other groups resulting from the classification procedure) and other interactions may uncover other associations pertaining to vicious and virtuous framework. For instance, what are the relative statistical contributions of gender and health outcomes to levels of violence? Moreover, examining more variation in the classifications could also help with exploring pathways of change. For instance, are there particular sequences of improvements in gender and health outcomes that are more likely to break countries out of vicious cycles? Another way to explore pathways may be mapping sequences of change in health and gender over decades to determine which pathways are most common or likely to lead to improvements.<sup>9</sup>

Despite these limitations, the strong support in particular for vicious cycles between health and gender, and with violence, is convincing and novel. This result could be further strengthened with robustness checks using other health and gender measures. It would also be useful to further explore what accounts for differences in findings between the cross-sectional and panel analyses. While we have noted that the evidence for vicious cycles in the panel analyses may be an artifact of our conceptualization, it is plausible that this result is also due to the different time frames of the analyses. It is possible that virtuous cycles are slow-moving, and thus do not show clear patterns in the panel analyses, while vicious cycles may operate faster.

---

<sup>8</sup>Such research should account for other possible factors, including, for instance, quality of institutions, inequality, functioning of economies, etc.

<sup>9</sup>In addition, case study analyses of outliers in our models, i.e. countries that defy the general prediction of the models, could be helpful in uncovering pathways of change.