

Lancet Commission Metrics Working Group: Draft Report on Large-N Analyses Extensions

Oskar Timo Thoms*
Consultant

July 24, 2021

*This report builds and expands on the January 2021 report co-authored together with Dane Rowlands and Valerie Percival. The first two sections are edited and expanded versions of the first two sections in the previous report authored by Thoms, Rowlands and Percival, and one of the appendices reproduces the previous report's section on the cross-sectional regression analyses written by Rowlands.

Contents

1	Introduction	3
2	Findings of the previous report	4
3	* Extensions in this report *	7
4	* Data *	9
5	Methods	13
6	* Analyses of health and gender outcomes *	14
6.1	Gender or health variables	15
6.2	Classifications	15
6.3	Internal conflict	16
6.4	State repression	18
6.5	Non-state conflict	20
6.6	Homicides	21
7	* Analyses of violence outcomes *	23
7.1	Internal Conflict	23
7.2	State repression	24
7.3	Non-state conflict	26
7.4	Homicides	26
8	* Sequencing analysis: work-in-progress *	27
A	Summary tables for analyses extensions	29
B	Cross-sectional analyses in the previous report	34
B.1	Methods	34
B.2	Health and Gender Outcomes	35
B.3	Violence Outcomes	39
C	Panel analyses in the previous report	40
C.1	Health and gender outcomes	40
C.2	Violence outcomes	44
C.3	Discussion	44

* Starred * section titles include new materials not discussed in the previous report.

1 Introduction

Building on earlier work, this draft report summarizes extensions of large-N analyses conducted for the Lancet-SIGHT Commission, whose objective is to examine if and how gender equality and health equity can contribute to more peaceful societies. The central research question which guides this work 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 (V&V) 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 likely many other important 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. We see this as a first useful step in a larger research agenda exploring relationships between health, gender, and peaceful societies.

In the previous report, 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 statistical estimates. We believe that both approaches are valuable in examining the research questions.

In this report, I re-implement and extend the panel analyses to address some questions that emerged in the previous report with respect to the operationalization of V&V cycles, the robustness of results, and some differences in findings between the cross-sectional and panel analyses. Section 2 reviews the findings of the previous report, and Section 3 outlines what extensions are pursued and why. Sections 4 and 5 discuss the data and methods used, largely repeating the discussion of the previous report and explaining how the classifications are used in the extensions. Sections 6 and 7 summarize and discuss the results of the new analyses. Finally, Section 8 briefly introduces the exploratory work I have done on categorizing the sequencing of change in gender and health outcomes. This work is to be expanded into a pathways typology in the final report. This draft report is written in such a way that no knowledge of the previous report is required. The appendix provides tables summarizing the findings of the main models examined for this report, and reproduces the cross-sectional and panel analyses of the first report for reference.

2 Findings of the previous report

The present report focuses on extensions of the panel analyses, which are in part motivated by some differences in findings between the cross-sectional and panel analyses. In our prior report, the cross-sectional analyses yielded 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 are 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 would be a useful extension of this work.
- This cross-sectional 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) and 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.

For the interested reader not familiar with the previous report, the cross-sectional analyses and their discussion are reproduced in Appendix B.

The panel models presented in the previous report led 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 little 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 V&V 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 V&V 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. The extensions in this report examine whether variation in the middle group is consistent with the notion of virtuous cycles. Further 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.

The discussion and summary of the panel analyses in the previous report are reproduced in Appendix C. We concluded that the findings from the panel analyses suggested more support for vicious than virtuous cycles, particularly between health and gender and in their effects on violence. The evidence varied by measures used and there were many null findings. The effects of violence on gender and health were less clear and often even contradicted the theoretical framework of V&V cycles. Moreover, the analyses suggested that different types of violence matter in different ways, pointing to the need for further research into such differences.

This research represents useful ground-clearing for further analyses investigating V&V cycles between health, gender and violence outcomes. These findings warrant further 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. Such research should account for other possible factors, including, for instance, quality of institutions, inequality, functioning of economies, etc.

We also noted that our analyses were 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 scholarly 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 V&V cycles hypotheses. However, analyses using a richer classification (including the other groups resulting from our classification procedure) and other interactions may uncover other associations pertaining to the V&V cycles framework. 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.

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 is also useful to further explore what accounts for differences in findings between the cross-sectional and panel analyses. While we noted that the lack of evidence for virtuous 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.

3 * Extensions in this report *

In this draft report, I discuss findings from new panel data analyses that address some, though not all, of the limitations noted in the previous section. These analyses extend the previous ones in four different ways.

First, the V&V cycles hypothesis can be conceptualized and operationalized in different ways. As noted above, while the cross-sectional analyses examined correlations between the individual gender, health and violence measures, the panel analyses also employ classifications based on health and gender outcomes (see Section 4). To date, however, we only examined the lowest and highest classifications. The extensions in this report explore variation in the middle by also distinguishing countries which comparatively perform better on gender than on health and vice versa. The objective of this extension is to examine whether V&V cycles may be at work across the full range of health and gender outcomes, including in the middle classifications that were previously left out of the analyses. While this makes the analyses more complicated, it also provides interesting new insights.

While the panel analyses overall still find more evidence of vicious than of virtuous cycles, they do find more support for the latter than the panel analyses in the previous report, and this is largely due to examining the middle classifications. Another objective of examining the middle classifications was to explore whether particular configurations are more likely to result in virtuous cycles: for instance, are those countries that have comparatively better gender outcomes than health outcomes more likely experience virtuous cycles? The evidence from the extensions is not clear on this question. In some cases, the models do show such differences, but these findings are not

consistent across outcome and violence variables, making it difficult to generalize.

Second, in the previous report, the panel and cross-national analyses come to some different conclusions, as outlined in Section 2. These differences are largely about the strength and robustness of findings, specifically on virtuous associations, rather than contradictory evidence. Further, some members of the Lancet Commission have raised questions about the timeframes required to realize different outcomes. For example, gender equality and virtuous outcomes may take longer to realize change, and thus may only show clear statistical patterns in the cross-sectional analyses and not in the panel analyses because they are structured differently. The extensions address this by considering different timelines in the panel analyses to examine whether this changes the results to more closely match the findings of the cross-sectional analyses. In one set of alternative models, the sample of observations is limited to the period 1995-2015 to match the timeframe of the cross-sectional analyses, and in another set of alternative models, I maintain the timeframe of the panel analyses (1970-2018), but increase the time lag between the independent and dependent variables from five to fifteen years.¹ In some cases, considering these different periods or lags leads to more evidence for vicious or virtuous cycles, but in others it actually weakens the evidence. It is likely that the different findings of the cross-sectional and panel analyses in the previous report result from the combination of the slow-moving nature of change in health and gender outcomes (compared to more short-term variation in violence), the different data structures of the two modelling approaches, and the different conceptualizations and operationalizations of V&V cycles.

Third, the analyses of the previous report find suggestive evidence for vicious cycles between health (measured by life expectancy and infant mortality) and gender (measured by the mean years of schooling ratio of females to males and by the adolescent fertility rate), and with some types of violence. In robustness checks, I examine whether such statistical associations hold for other outcomes, by substituting other health (under-five mortality rate, UFM, and disability-adjusted life years, DALYs) and gender measures (the labour participation ratio of females to males and the Gender Inequality Index, GII) as the dependent variables of the regression analyses. These new measures are chosen based on data availability and to expand the conceptual range of the gender and health dimensions. In preview, the findings of these robustness checks are mixed. Unsurprisingly, the results for the UFM are very similar to those for the infant mortality rate (IMR) and largely supportive. The models of the labour participation ratio find no evidence of vicious or virtuous cycles, and find some contradictory evidence; therefore, I do not include this variable in the discussion of the findings. The results for DALYs and the GII provide a mix of supportive and contradictory evidence. Overall, these robustness checks do not substantially strengthen the evidence base, but they also do not weaken it.

Fourth, the initial analyses strongly suggest that different types of violence – internal conflict, state repression and societal violence – may matter in different ways, but these analyses only considered different types of violence separately. Therefore, the extensions include models which also account for other types when examining the role of specific types of violence, to further examine the robustness of findings on V&V cycles.² While the inclusion of these violence variables

¹I did not increase this to twenty years – the lag used in the cross-sectional analyses – because the longer the lag, the smaller the sample of observations, affecting statistical precision.

²Other possible approaches are to include aggregate measures, such as deaths from several types of violence, or interactions of different types of violence. Instead, I choose simple control variables, as aggregate measures are difficult to interpret because they conflate distinct phenomena, and additional interactions would add further complexity to the models.

sometimes weakens the evidence, in many cases it actually leads to clearer findings on vicious or virtuous cycles. Given the substantial evidence in the analyses that different types of violence, and even different measures of the same types of violence, lead to different conclusions, accounting for these different types increases the confidence in findings on V&V cycles.

The remaining sections discuss the data, methods, and findings of the new analyses. Full details of the data preparation and the analyses, including all regression details, are available at timothoms.github.io/LSC-MWG/.

4 * Data *

This report uses the same data and panel regression methods as the previous report which it builds on, but altering the research design and models to extend the analyses.

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.³ Many health and gender measures in particular are subject to many missing data points across countries and over time, and this missingness is unlikely to be random and thus presents a potential threat to inferences. Therefore, 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 IMR) and gender (the mean years of schooling ratio of females to males and the age-specific fertility rate for adolescents, AFR) dimensions. To measure violence, we chose several indicators to capture different types, including state-based internal armed conflict, repression and societal violence, and the total and civilian battle-related death rates (per population) resulting from varying types of conflict and homicide rates. These variables are listed in Table 1. The analyses for the previous report included measures of internal war and state torture. This report does not discuss any analyses using these two measures because the results are very similar to those for internal conflict and extra-judicial killings, respectively.

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. The affected variables are indicated by ***bold italics*** in Table 1. Readers familiar with these measures should keep in mind that this alters how one would usually interpret them. As data availability and quality generally improves after 1989 and more violence variables exist for this period, the primary period of the analyses is 1990-2018. 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.

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. For political factors, the models reported here only include a measure of participatory democracy, as this was the

³A detailed codebook is available at this [link](#).

Table 1: Measures used in the large-N analyses

Category	Variable	Source
health	life expectancy	WPP
health	<i>infant mortality rate</i> (IMR)	WPP
health	<i>under-five mortality rate</i> (UFMR)	WPP
health	<i>disability-adjusted life years</i> (DALYs)	IHME
gender	mean years of schooling ratio (MYS)	HDR
gender	<i>adolescent fertility rate</i> (AFR)	WPP
gender	labour participation ratio	ILO
gender	Gender Inequality Index (GII)	HDR
state-based conflict	internal conflict incidence	UCDP
state-based conflict	internal war incidence	UCDP
state-based conflict	internal conflict deaths (rate per population)	UCDP
state-based conflict	internal conflict civilian deaths (rate per population)	UCDP
state repression	one-sided violence (OSV) incidence	UCDP
state repression	one-sided violence (OSV) deaths (rate per population)	UCDP
state repression	<i>latent physical integrity measure</i>	Fariss
state repression	<i>state torture</i>	V-Dem
state repression	<i>extra-judicial killings</i>	V-Dem
societal violence	non-state conflict (NSC) incidence	UCDP
societal violence	non-state conflict (NSC) deaths (rate per population)	UCDP
societal violence	non-state conflict (NSC) civilian deaths (rate per population)	UCDP
societal violence	homicides (rate per population)	ODC/WHO

only measure of political structures that consistently showed associations in the initial analyses.

To help operationalize the V&V cycles hypotheses, we developed a classification of countries based on the four health and gender outcomes, creating groups of countries for comparison. 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. In the analyses for the previous report, we selected the bottom four cells as our low classification and the top four as our high classification. We examined the statistical effects of the lowest and highest 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. While this expectation was borne out in some models, it also became clear that the large middle category lumped together much potentially interesting variation. Therefore, one of the motivations for the extensions in this report is to explore the possibility that V&V cycles may also be at work in the middle. In the analyses in this report, five classifications are used based on the matrix in Table 2: low, mid, upp, and those that score comparatively high on health but low on gender (H>G), and those that score comparatively high on gender but low on health (G>H).

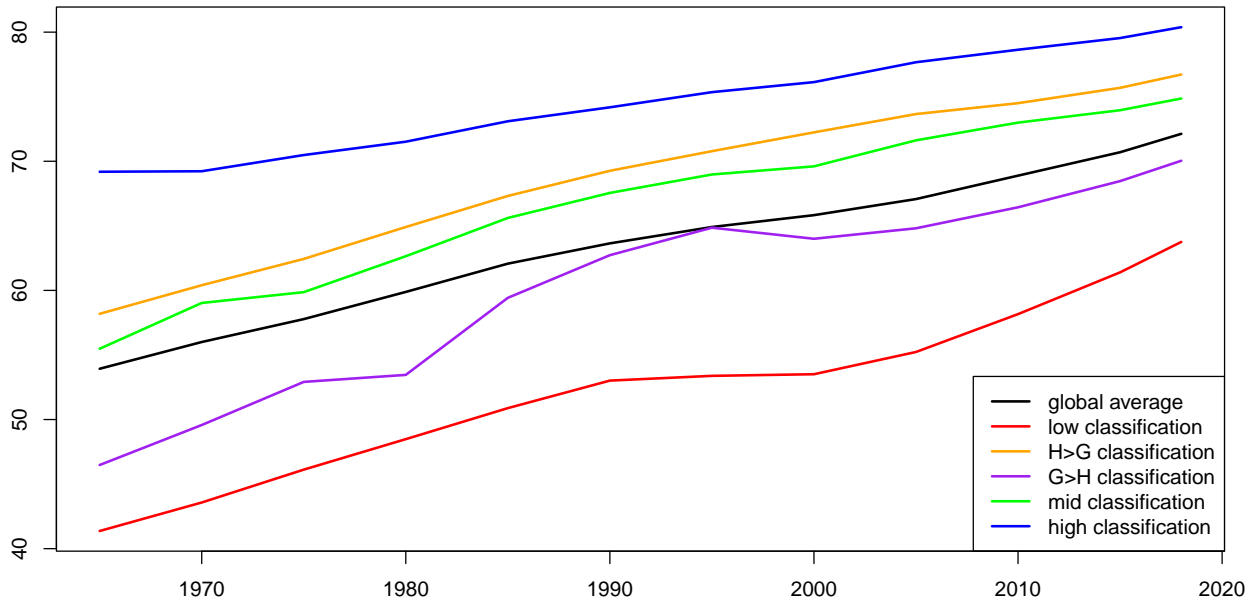
Plotting the averages of the health and gender variables at the global level and within the classifications show clear trends over time (see Figures 1-4).⁴ First, there is considerable global

⁴Note that here the original measures are shown as five-year averages, not the inverted versions used in the statistical analyses (see above). Also, the mean years of schooling ratio is available for a shorter time period, since 1990.

Table 2: Classification groups used in the analyses extensions

Health	Gender				
	1st quintile	2nd quintile	3rd quintile	4th quintile	5th quintile
1st quintile	low	low	G>H	G>H	G>H
2nd quintile	low	low	G>H	G>H	G>H
3rd quintile	H>G	H>G	mid	G>H	G>H
4th quintile	H>G	H>G	H>G	upp	upp
5th quintile	H>G	H>G	H>G	upp	upp

Figure 1: Life expectancy trends



improvement in the health and gender measures, regardless of how countries are classified. Second, overall improvements are greatest in the low classifications, with some interesting differences. Most improvement in the IMR occurs in the G>H classification, and most improvement in the AFR occurs in the H>G classification, followed by the low classification. Improvements tend to be much less pronounced in the high classification. This suggests possible ceiling effects for better performers that need to be accounted for in the analyses. Other graphical analyses not presented here also indicate that low performers on the health and gender dimensions tend to have the most violence on average over time (across different measures and different types of violence), and high performers experience much less violence. In a partial exception to this generalization, average homicide rates are highest in intermediate rather than the low classifications, but the high classification clearly has the lowest average homicide rates.

These classifications are used in the analyses below to examine the possibility of V&V cycles in two different ways, and these have different implications for how to conceptualize V&V cycles. The simplest tests are whether the individual gender and health variables are positively correlated with each other or with violence variables (controlling for other factors), but in this approach, it is not

Figure 2: Infant mortality trends

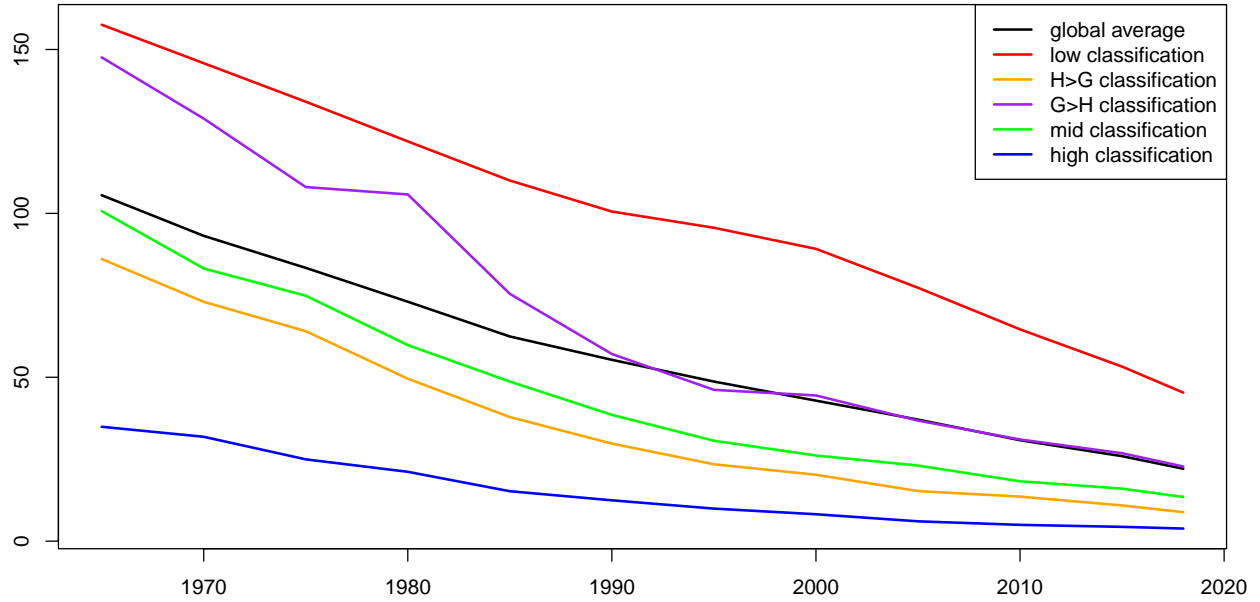
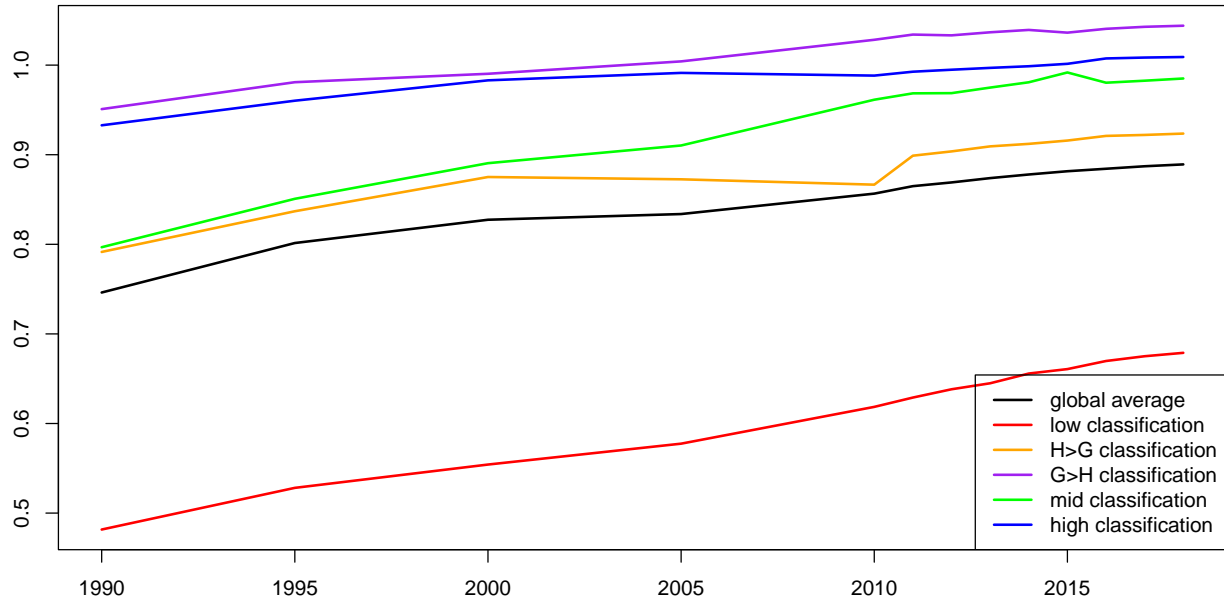


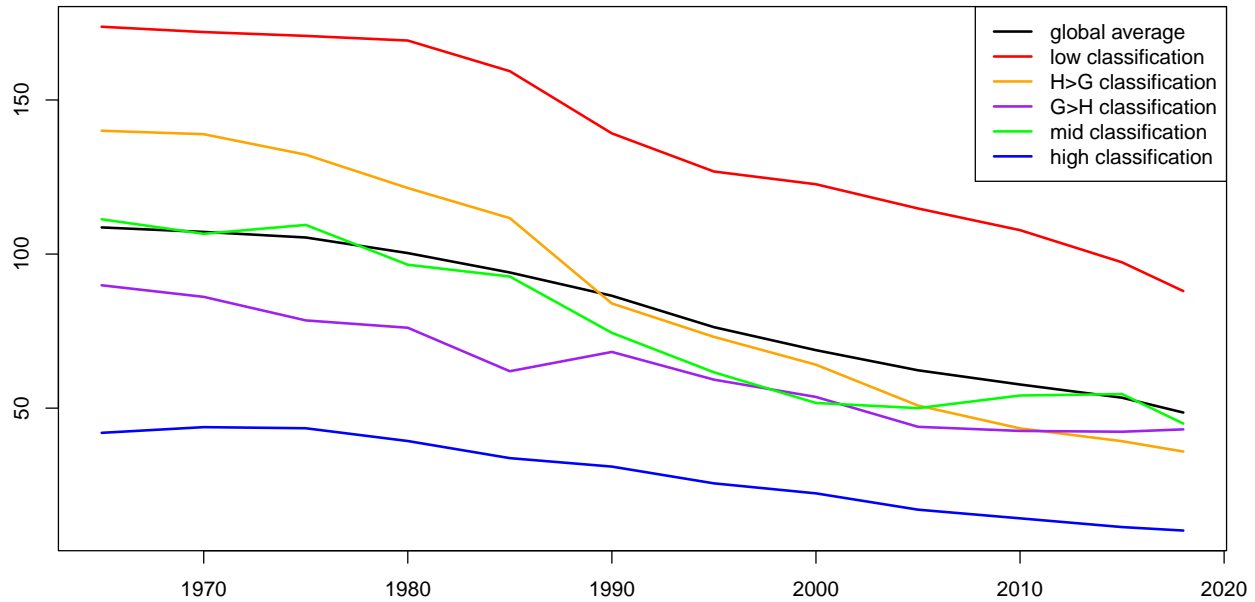
Figure 3: Mean years of schooling ratio trends



possible to determine whether the gender, health and violence variables actually mutually reinforce each other in cycles. Moreover, such simple correlations imply that V&V cycles are simply two sides of the same coin; one implies the other. I use the classifications for more direct tests, and to allow V&V cycles to operate separately as distinct processes.

First, to examine whether particular configurations of health and gender outcomes (i.e. classifications) are associated with improvements or worsening in gender, health and violence variables, I include binary (“dummy”) variables indicating the groups based on Table 2. This can be interpreted

Figure 4: Adolescent fertility trends



as the effect of past health and gender configurations on current health and gender outcomes or on violence outcomes. Here, associations of lower classifications with worse outcomes and of higher classifications with better outcomes would provide evidence that is consistent with “two-way” vicious or virtuous cycles, respectively.

Second, modelling interactions of these binary classification indicators with violence variables lets me calculate the statistical effects of past violence within each classification on current health, gender or violence outcomes. This can be interpreted as a “three-way” cycle. We can test whether past violence in particular classifications is associated with worse or better current health or gender outcomes or with more or less violence. For instance, if we find an association of past violence in lower classifications with more current violence or worse health or gender outcomes, this evidence would support vicious cycles because violence begets more violence where countries perform comparatively poorly on health and gender. Or, if we find past violence in higher classifications groups is associated with better health or gender outcomes or less violence, this would support virtuous cycles, because countries in the higher classifications were able to improve conditions despite their past violence. Moreover, the evidence does not need to be this clear-cut to be supportive of the V&V cycles framework. For instance, past violence may be associated with more current violence across several or all classifications, but more so in lower classifications relative to higher ones; such evidence would also be supportive of vicious cycles.

5 Methods

In contrast to the cross-sectional models in the previous report (and reproduced in Appendix B), which examine the growth of the outcome variables between the periods 1991-1995 and 2011-2015, 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 (usually between 6 and 11 periods). The dependent variables of the analyses are the individual health, gender, and violence variables. I test the V&V cycles hypotheses by examining the lagged gender or health variables, or the classification indicators, and the lagged violence variables or their interactions with the classification indicators, 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.⁵ I further include the lagged dependent variable in each model in order to account for trends, autocorrelation and ceiling effects.

When modelling the interactions, I also include additional models with controls for different types of violence. I proxy the presence and magnitude of other types of violence with deaths rates for internal conflict (combatants and civilians), one-sided violence (civilians), and non-state conflict (combatants and civilians). For this purpose, I do not use the homicide rate for societal violence, because these data have much more missing data points than the UCDP data. To avoid introducing further endogeneity, I do not include deaths rates related to the violence variable of interest in a given model. For instance, the models examining the role of internal conflict incidence do not include the closely related internal conflict death rate. Moreover, the models examining the role of repression or extra-judicial killings do not include the one-sided violence death rate because these measure overlapping phenomena. And models examining the role of homicides do not include the one-sided violence and non-state conflict death rates because these measures plausibly include at least some of the same deaths.

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 in cross-sectional data, because of the lack of independence of observations within countries over time, leading to issues of correlated unit (country) or time 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 sometimes leads to different and even contradictory results. Based on the results of these tests across models, we consider the fixed effects (FE, or "within") ordinary least squares estimator (OLS run on demeaned within-country data) with cluster-robust standard errors, and often preferably, the fixed effects general (or unrestricted) feasible generalized least squares (FGLS) estimator. Generally, the tests suggest that almost all models are subject to unit and/or time effects, and according to Hausman-test, random effects models are almost never justified. In many models, the tests further suggest persistent serial correlation, often of AR(1) type, leading to the preference for the FGLS over the FE estimator, but the discussion below and the summary tables in the appendix consider both.

6 * Analyses of health and gender outcomes *

This section and the next summarize the findings of the statistical analyses with examples. Details of all regression models and figures of all interaction effects are shown on the project website.⁶ Moreover, Appendix A provides summary tables of the findings of the main analyses across all

⁵In the initial analyses, 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.

⁶Available at timothoms.github.io/LSC-MWG/panel.html.

outcome variables (but not for the analyses using longer lags or shortening the time period of the sample). I report only associations that are statistically significant at the 0.05 alpha level or lower. The regression figures on the project website and the figures of interactions effects show effects with 95% confidence intervals. The discussion is organized by key independent variables to summarize their statistical associations across outcome variables, and focuses on the main models which include the classifications, explained in Section 4, and the control variables for types of violence explained in Section 5. I only note the results for alternative models where they diverge from the main findings.

6.1 Gender or health variables

With respect to health outcomes, the gender variables are positively associated with improvements in life expectancy, and the MYS ratio is also positively associated with improvements in the IMR, UFM, and DALYs. However, the AFR is negatively associated with these latter health outcomes, contradicting expectations. When limiting the sample to the 1995-2015 period or using a lag of 15 years between the dependent and independent variables, the evidence is weaker for life expectancy. Moreover, the AFR is positively associated with the IMR when limiting the sample to the shorter period, reversing the finding in the full sample, but using longer lags provides no such support.

Regarding gender outcomes, the health variables (life expectancy and IMR) have no statistically significant effects on the MYS ratio and the labour participation ratio, but they are both positively associated with the AFR, which is consistent with expectations. When limiting the study period to 1995-2015, the IMR is now also positively associated with the MYS ratio and the results for the AFR remain the same. When using the longer lags, there are some contradictory statistical effects of life expectancy. Finally, the IMR is negatively associated with the GII. Overall, these results show some associations that are broadly consistent with V&V cycles, but more interesting evidence is provided by the analyses using the classifications and interactions below.

6.2 Classifications

The effects of the classification indicators support V&V cycles in models of life expectancy; the low and G>H classifications have decreased life expectancy compared to the highest classification, while the G>H, mid and highest classifications have increased life expectancy compared to the low classification. However, when using the alternate time period or longer lags, the models find associations that actually contradict V&V cycles.

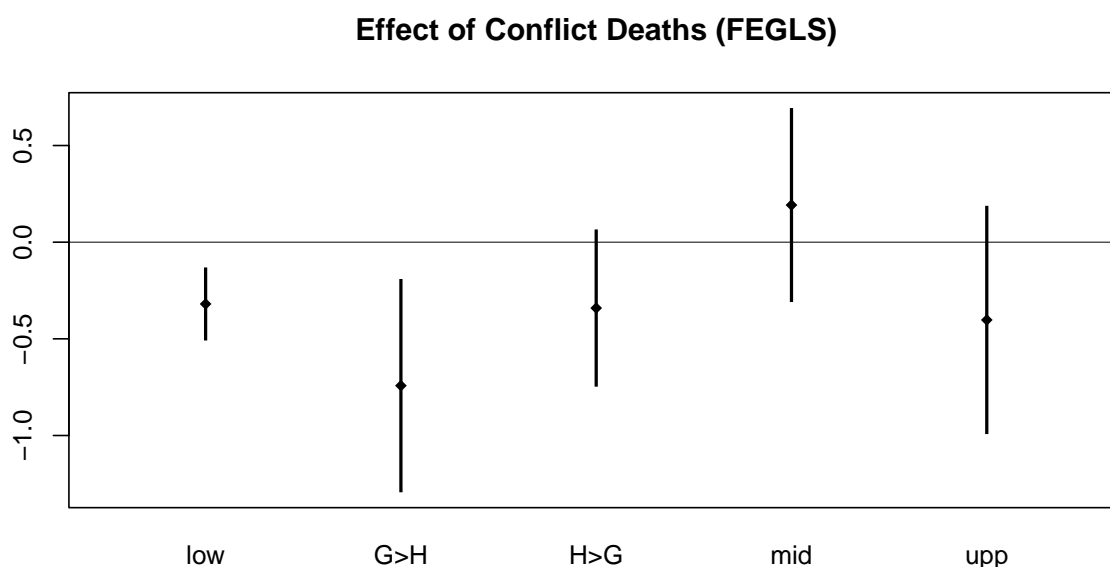
For the IMR and the UFM, the results are very similar, which is unsurprising since they are closely related, so I only discuss the former. The findings are mixed and weak in the main models: the G>H classification has worsened infant mortality compared to both the low and the highest classification, but only in the fixed effects models, while none of the other groups have statistically significant associations. When limiting the analysis to the shorter period, the evidence for vicious cycles is stronger because the low group has worse outcomes compared to other groups, but the H>G, G>H and mid classifications all have better outcomes than the highest classification. When using the longer lags, the findings with respect to the classifications strongly contradict V&V cycles, because the low group is associated with improvements compared to all others. The models of DALYs find either no support or contradictory evidence based on the classification indicators.

With respect to the main gender outcomes – the MYS ratio and the AFR – the classifications have associations that support V&V cycles. The H>G, mid, and highest classifications all have higher

MYS ratios than the low classification, which is consistent with the presence of a vicious cycle, and the statistically significant improvements in two of the middle classifications suggest the possibility of virtuous cycles in these groups. Similarly, the results of the AFR models suggest vicious cycles in the low and H>G classifications (compared to the highest classification) and virtuous cycles not just in the highest classification but also in the G>H and mid classifications. For the other gender outcome variables – the labour participation ratio and the GII – the statistical effects of the classification indicators do not provide any support for vicious or virtuous cycles.

6.3 Internal conflict

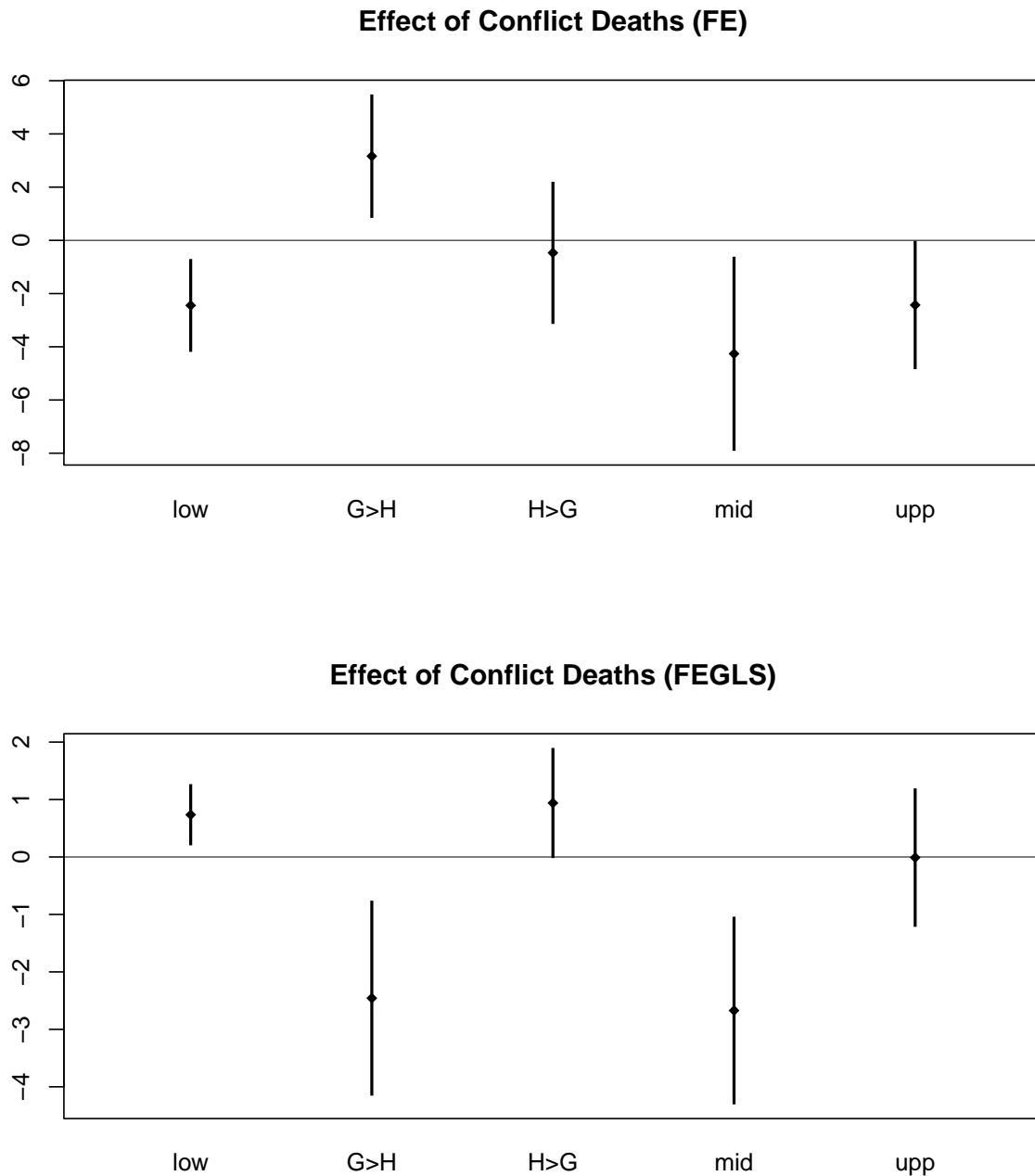
Figure 5: Effects of internal conflict deaths on life expectancy



The analyses consider the role of internal conflict with three variables: incidence (the proportion of years with a conflict coded in the UCDP data over the five-year period); total battle-related deaths; and civilian battle-related deaths. When the outcome is life expectancy, the interaction between conflict incidence and the classifications do not support V&V cycles, but they do suggest a vicious cycle involving deaths rates. Figure 5 shows the interaction effects from the FEGLS estimator with controls for other types of violence; it shows that past conflict deaths are associated with decreased life expectancy only in the low and G>H classifications. The results for civilian deaths are weaker but still suggestive of a vicious cycle for the G>H classification. When the sample is limited to the shorter time period or longer lags are used, the evidence becomes weaker and more mixed. Moreover, the analyses of the IMR and UFMR outcome variables find similar support for a vicious cycle involving internal conflict deaths or civilian deaths in the G>H classification. The results of the analyses of DALYs, however, possibly contradict the notion of vicious cycles, since some

of the internal conflict variables are associated with higher life expectancy in lower and higher classifications, and these results are not robust across models specifications.

Figure 6: Effects of internal conflict deaths on AFR



When examining internal conflict in models of gender outcomes, there is little support for V&V cycles. The models of the MYS ratio show no associations, except in one model which finds a negative effect in the mid classification. Decreasing the time period of the sample does not alter

this finding but when using the longer lags internal conflict incidence is associated with decreased MYS ratios in the low classification and conflict deaths are associated with higher MYS ratios in the G>H classification in fixed effects models. The latter raises the possibility of virtuous cycles even at a lower level but where countries perform comparatively better on gender than health, but these findings are not robust to using the FEGLS estimator.

In models of the AFR, the evidence is mixed but mostly contradicts V&V cycles, and the two estimators often lead to opposite findings. For instance, Figure 6 show some associations that are supportive but more that contradict, and importantly are clearly not robust. The results are also not supportive when decreasing the time period or using longer lags. The results of the GII models are supportive of virtuous cycles in the H>G and mid classifications, and possibly vicious cycles in the G>H classification, but the evidence is weak as the results are not robust to different specifications.

6.4 State repression

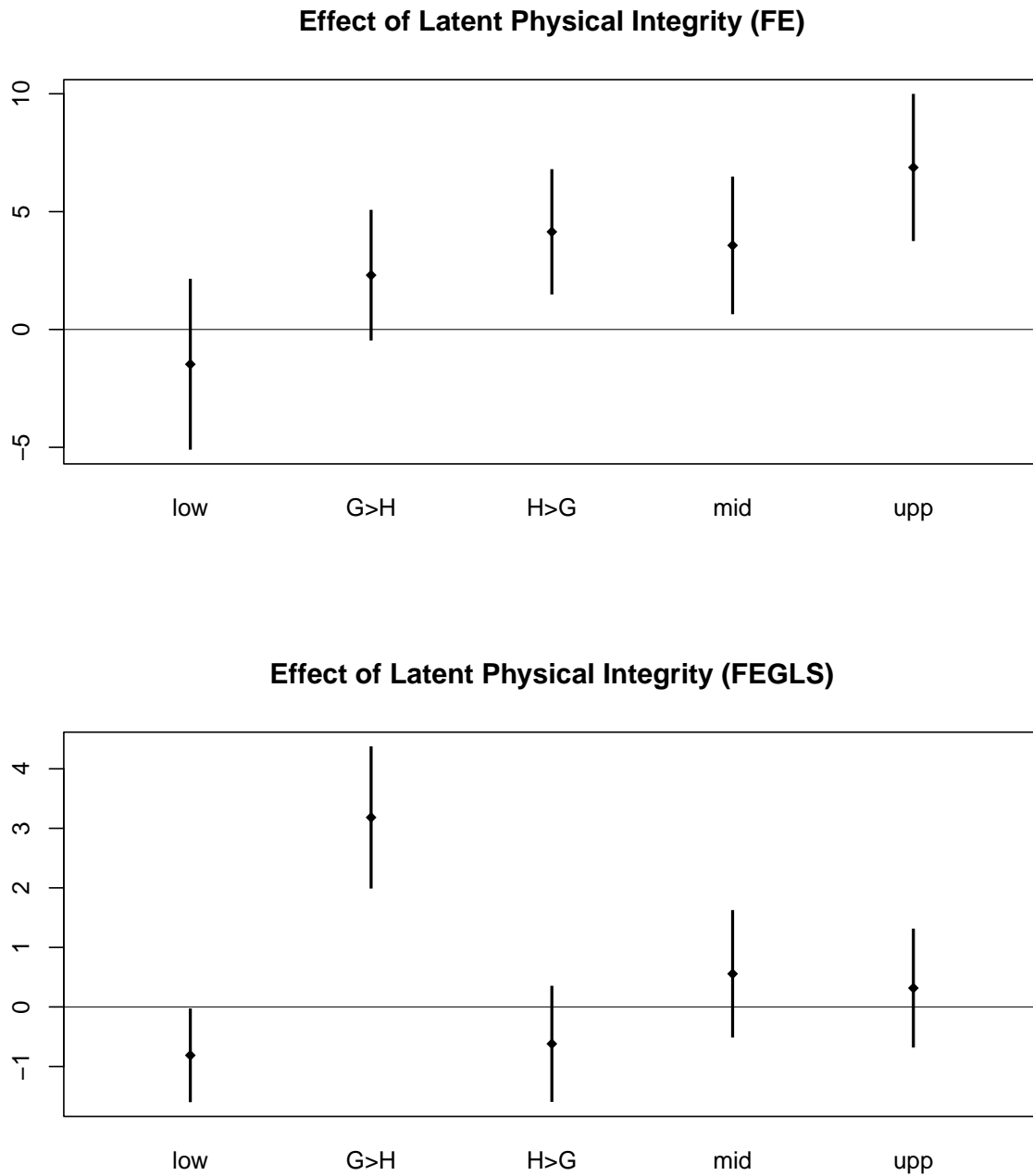
The analyses use four measures of state repression: UCDP's one-sided violence incidence and deaths; the latent physical integrity (LPI) measure; and a measure of extra-judicial killings. One-sided violence (OSV) is associated with higher life expectancy in some lower and higher classifications and this varies by model specifications. When using the shorter time period, the evidence remains mixed, but when using longer lags, they suggest vicious cycles in the G>H and H>G classifications. One-sided violence is also associated in several models with higher IMR and UFMR in the low classification; the evidence is much weaker for the analyses using the shorter time period and longer lags. There is no support for V&V cycles involving one-sided violence in the analyses of DALYs.

In the cases of physical integrity violations and extra-judicial killings, there is no evidence of V&V cycles with life expectancy. However, when using the shorter time period, these variables are associated with decreased life expectancy in the low classification. For the other health outcome variables, the interactions with the LPI and extra-judicial killings suggest some support. For instance, Figure 7 show the effects of the LPI on the IMR and how they differ by the estimator used. In the top panel, physical integrity violations are associated with improved infant mortality in the H>G, mid, and high classifications, but not in the low and G>H classifications. This suggest improvements in health outcomes despite violent repression only in some classifications. In the bottom panel, physical integrity violations are associated with worsening IMR in the low classification but improvements in the G>H group. While the FEGLS estimator is somewhat preferred here based on econometric tests, the FE estimator with cluster robust standard errors is also defensible. Thus, these models provide support for both V&V cycles. Models of the UFMR and of DALYs and those substituting extra-judicial killings for the LPI provide similarly supportive evidence.

With respect to gender outcomes, the models of the MYS ratio and of the AFR with interactions of one-sided violence variables provide no support and often contradict V&V cycles, regardless of the time period of the sample or the length of the lags. However, OSV incidence is strongly associated with increased GII in the H>G classification, while OSV deaths are strongly associated with declines in the GII in the G>H classification.

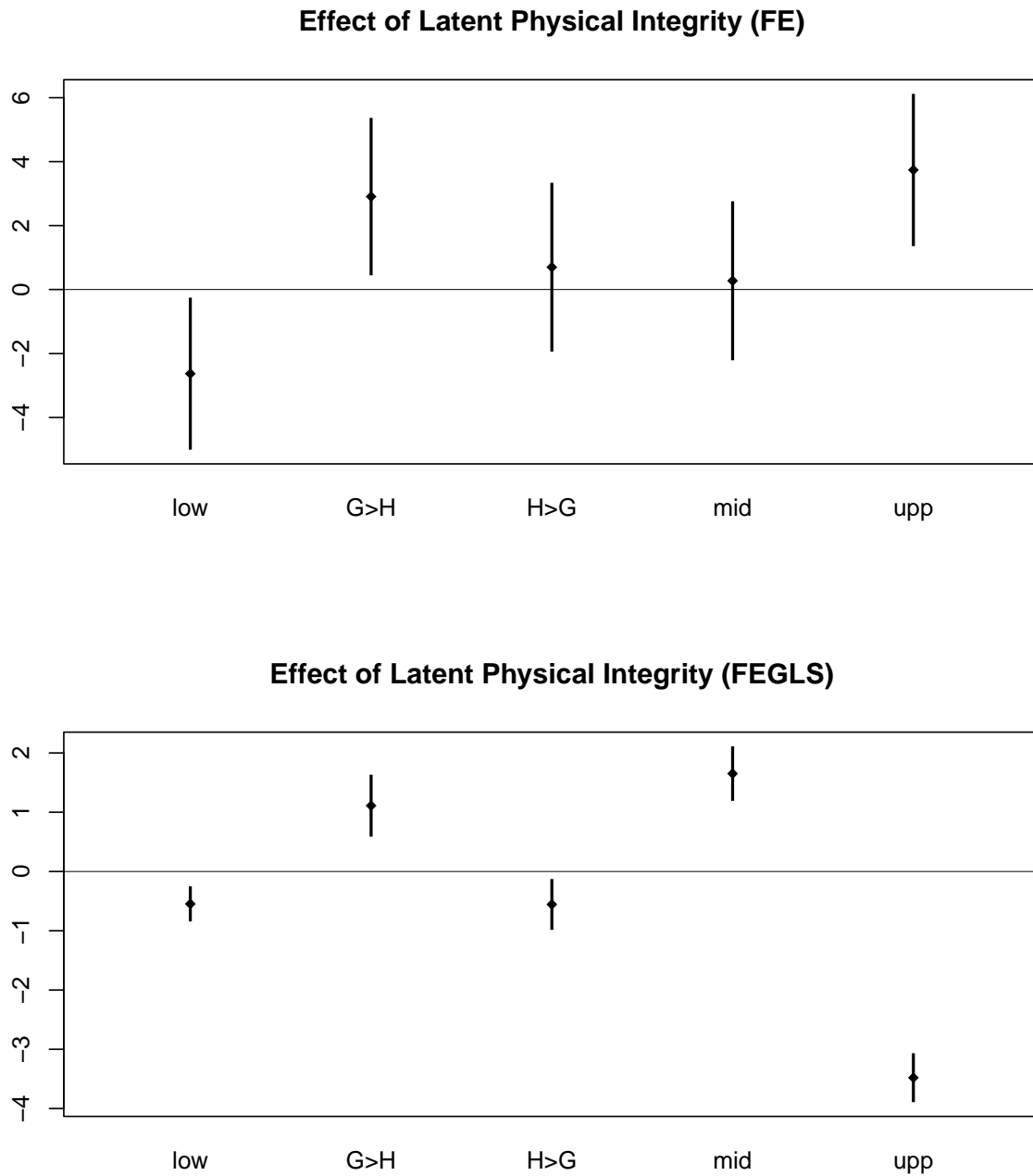
Physical integrity violations are associated with lower MYS ratios in the G>H, H>G and mid classifications. When using the shorter time period or the longer lags, these associations disappear. The models of the AFR have interesting results, shown in Figure 8. Regardless of the estimator, physical integrity violations are associated with worsened adolescent fertility rates in the low classification, but improved in the G>H classifications, suggesting both a vicious cycle, and a

Figure 7: Effects of physical integrity violations on IMR



virtuous cycle when countries perform better on gender than on health. Moreover, the vicious cycle also may apply to the H>G classification. Yet, the two estimators also lead to opposite findings regarding the effect in the highest classification. When shortening the time period or lengthening the lags, the evidence becomes weaker and more contradictory. Finally, the GII models are weakly supportive of virtuous cycles in higher classifications.

Figure 8: Effects of physical integrity violations on AFR



6.5 Non-state conflict

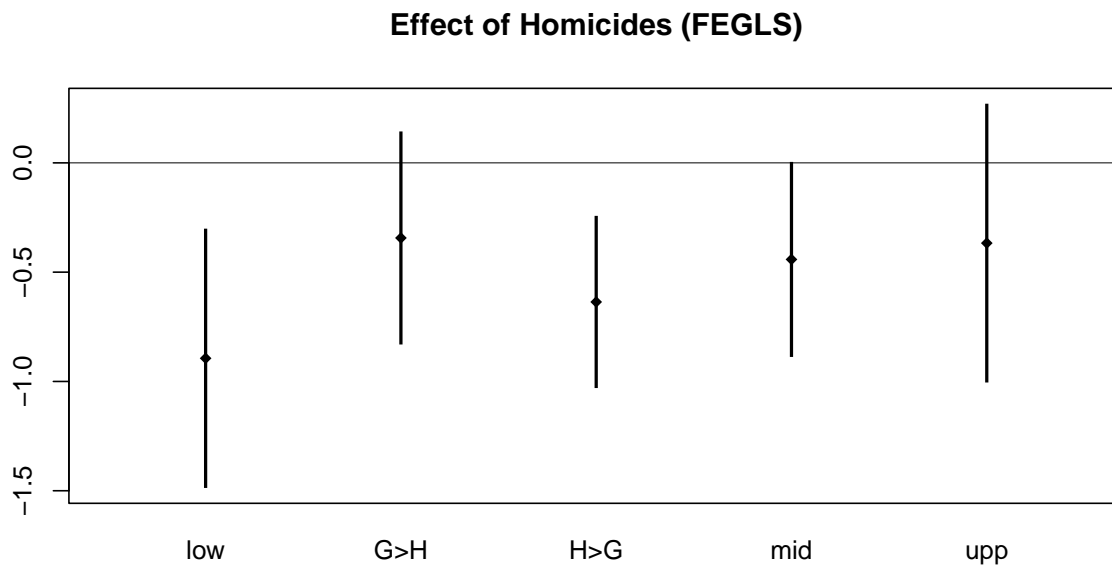
Non-state conflict is organized violence between societal actors, measured by three variables: incidence; battle-related deaths; and battle related civilian deaths. The models of life expectancy do not clearly support V&V cycles involving non-state violence, and some associations contradict

such cycles. However, when the shorter time period or the longer lags are used, some of these contradictions disappear, though the evidence remains weak. The models of the IMR and UFMR provide some support for V&V cycles. Non-state conflict deaths are associated with worse infant mortality in the H>G classification and better in the highest classification, but also with better IMR in the low classification in some models. Using the shorter time period or the longer lags does not strengthen this evidence, and the main models of the UFMR are quite similar to those of the IMR. The models of DALYs provide only limited support for vicious or virtuous cycles involving non-state conflict civilian deaths, and here the two estimators show contradictory associations.

The models of the MYS ratio provide very little support for V&V cycles involving non-state conflict, and the non-state conflict variables are associated with improved adolescent fertility rates even in the low, H>G, and G>H classifications in some models, contradicting theoretical expectations. In models of the GII, Non-state conflict deaths are strongly associated with lower GII in the G>H classification, which is consistent with a vicious cycle. Overall, non-state conflict appears to have less systematic effects on gender and health outcomes than some other types of violence.

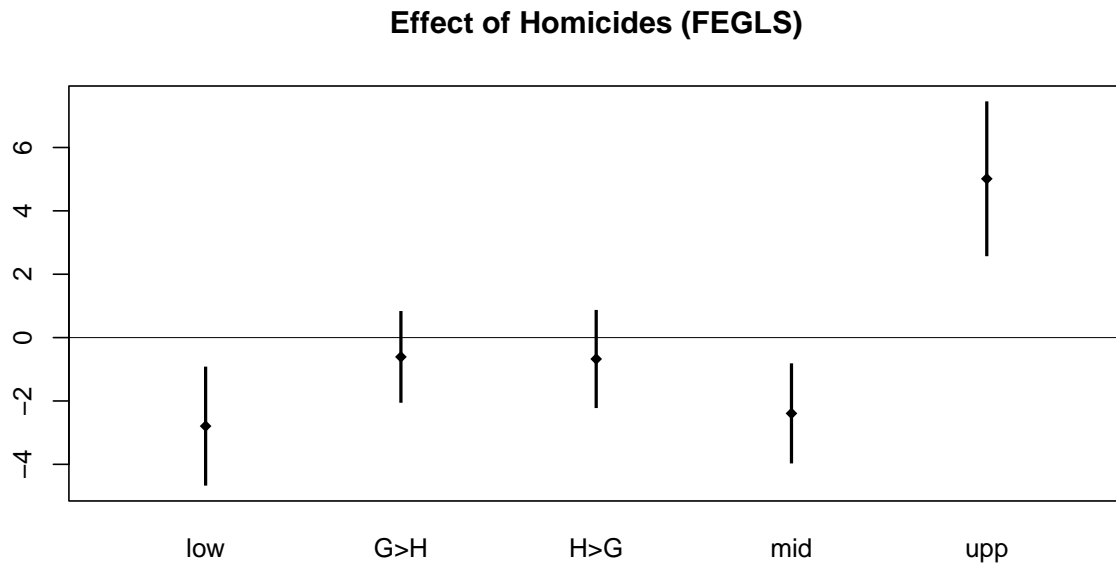
6.6 Homicides

Figure 9: Effects of homicides on life expectancy



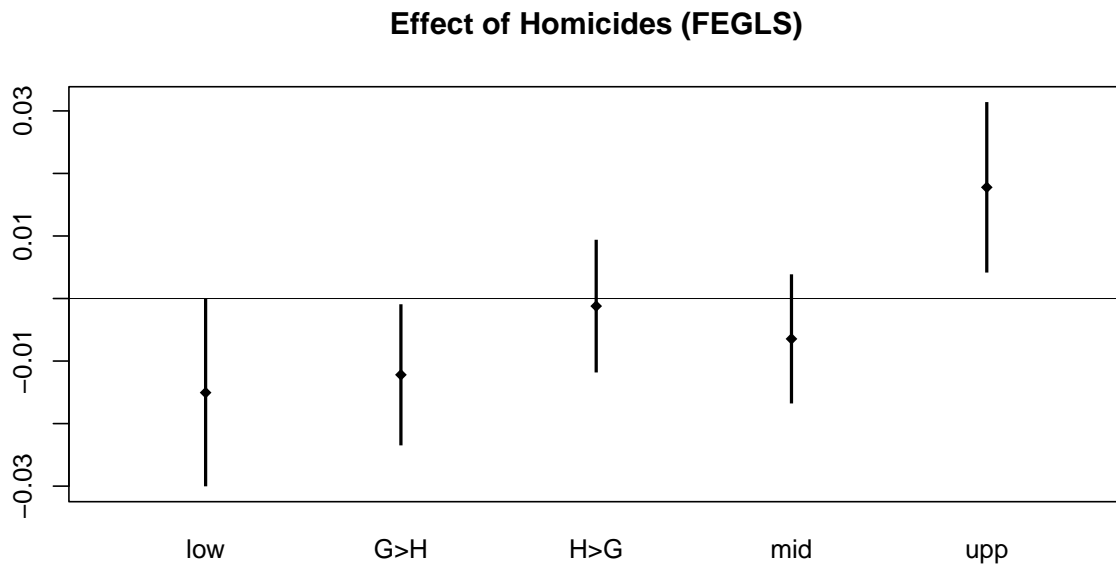
Homicides also represent societal violence but, unlike non-state conflict, not necessarily between organized groups. The models of life expectancy, IMR, UFMR, and DALYs all suggest vicious cycles involving homicides, and possibly some virtuous cycles. Figure 9 shows that homicides are associated with decreased life expectancy in the low, H>G, and mid classifications. In Figure 10, they have similar effects on worsening infant mortality in the low and mid classifications, but are

Figure 10: Effects of homicides on IMR



also associated with improved infant mortality in the highest classification. The results are similar for the UFMR, and models of DALYs also support vicious cycles in the lower classifications.

Figure 11: Effects of homicides on GII



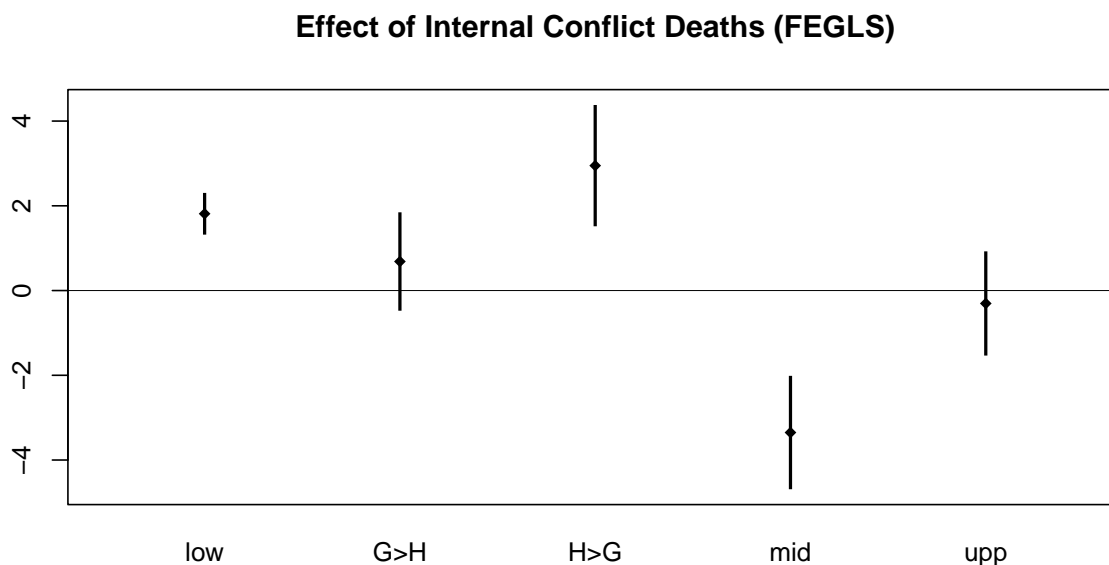
With respect to gender outcomes, there is no support of V&V cycles with homicides in models of the MYS ratio, and there is contradictory evidence when using the shorter lags. The models of AFR are supportive of vicious cycles in the low classification, and possibly of virtuous cycles in the highest classification, but this finding is weaker. There is clear support for V&V cycles in one of the models of GII, as shown in Figure 11, but this is not robust to the other estimator.

7 * Analyses of violence outcomes *

This section summarizes the findings of the models of violence outcomes. These are structured analogously to the models of health and gender outcomes. They examine the statistical effects of individual gender and health variables, of the classifications, and of their interactions with the lagged violence variables. As discussed in Section 4, the former set of models test the effects of gender and health outcomes on violence, while the interactions test the effects of past violence in particular health and gender contexts on current violence, providing evidence on three-way relationships.

7.1 Internal Conflict

Figure 12: Effects of past internal conflict deaths on current deaths

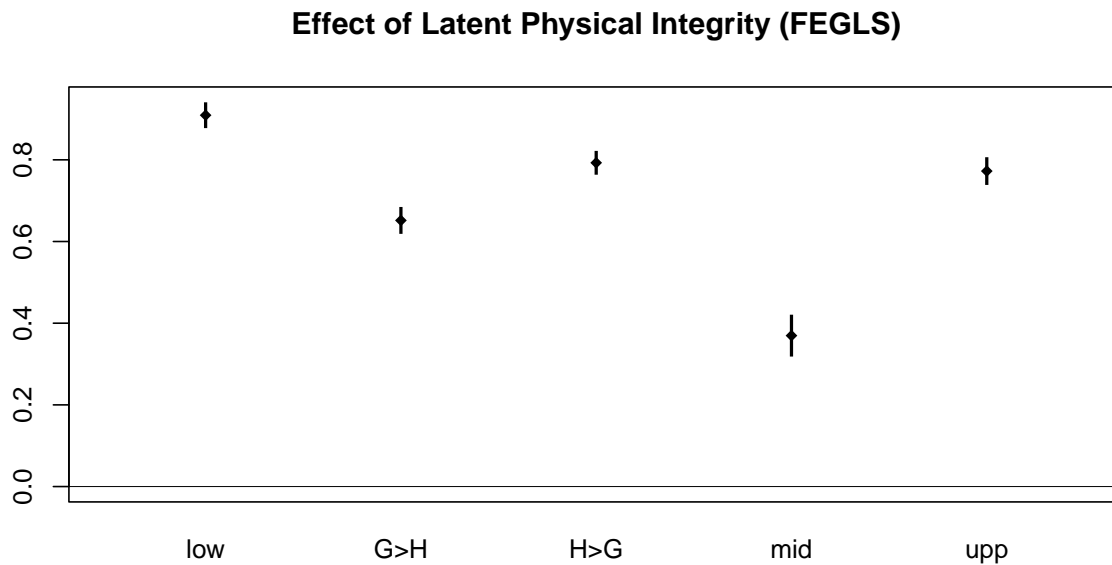


The models of internal conflict provide quite consistent evidence of V&V cycles. Improved adolescent fertility and infant mortality are associated with decreased internal conflict incidence. When the classification indicators are included, the H>G, mid, and highest classifications are associated with decreased internal conflict deaths, compared to the low classification. Regardless of the internal conflict measure used, the the models with interactions often suggest vicious cycles,

particularly in the low classification. For instance, Figure 12, shows that past deaths are associated with increased current deaths in the low and H>G classifications, but not in the G>H classification, and with decreased deaths in the mid classification. This shows that violence begets more violence in the former but that countries in the middle classification often can break out of such violent cycles. The models using longer lags also find support.

7.2 State repression

Figure 13: Effects of past physical integrity violations on current violations



There is some support for vicious cycles in the models of one-sided violence. Improvements on both the health variables are associated with less one-sided violence incidence, and higher MYS ratios are associated with decreased one-sided deaths. In models with the interactions, past incidence is associated with current incidence only in the low and mid classifications, and these associations are consistent across model specifications and regardless of estimator.

When analysing the other repression measures, the evidence is quite mixed. In several models of latent physical integrity, the individual gender and health variables have associations that contradict the expectations, and these results are not robust across estimators. Similarly, the effects of the classification indicators also contradict the expectations, and show no statistically distinguishable associations in the models of extra-judicial killings. The results of the models with interactions are again interesting. Figure 13 shows that, while past physical integrity violations are associated with increased current violations across all classifications, the magnitude and confidence intervals are such that the largest increase is found in the low classification and the smallest in the mid classification. This seems broadly consistent with the V&V framework. However, Figure 14 shows that, once measures of other types of violence are included, these differences disappear.

Figure 14: Effects of past physical integrity violations on current violations, with violence controls

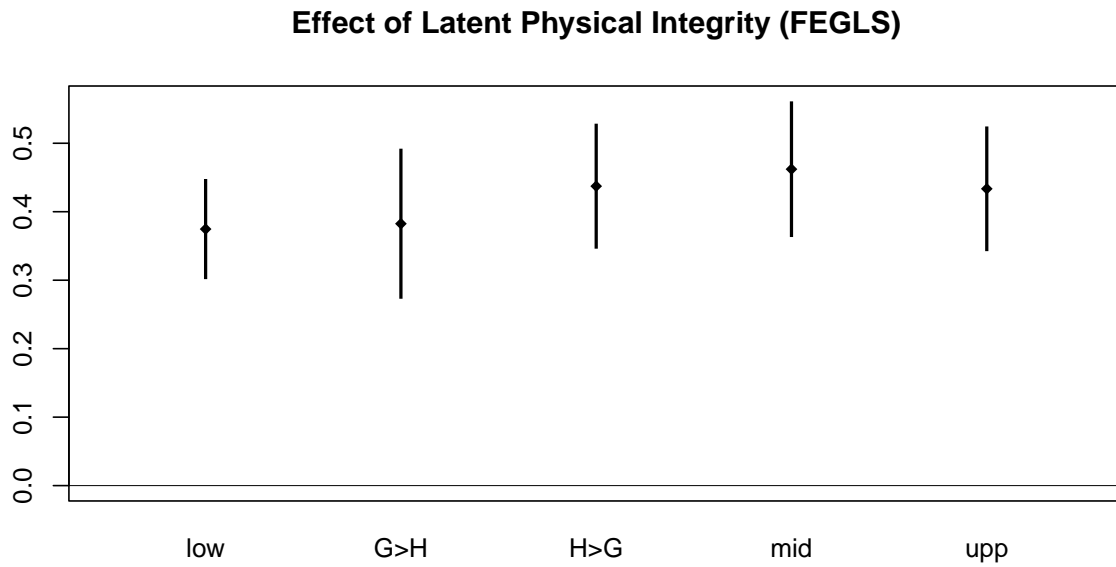
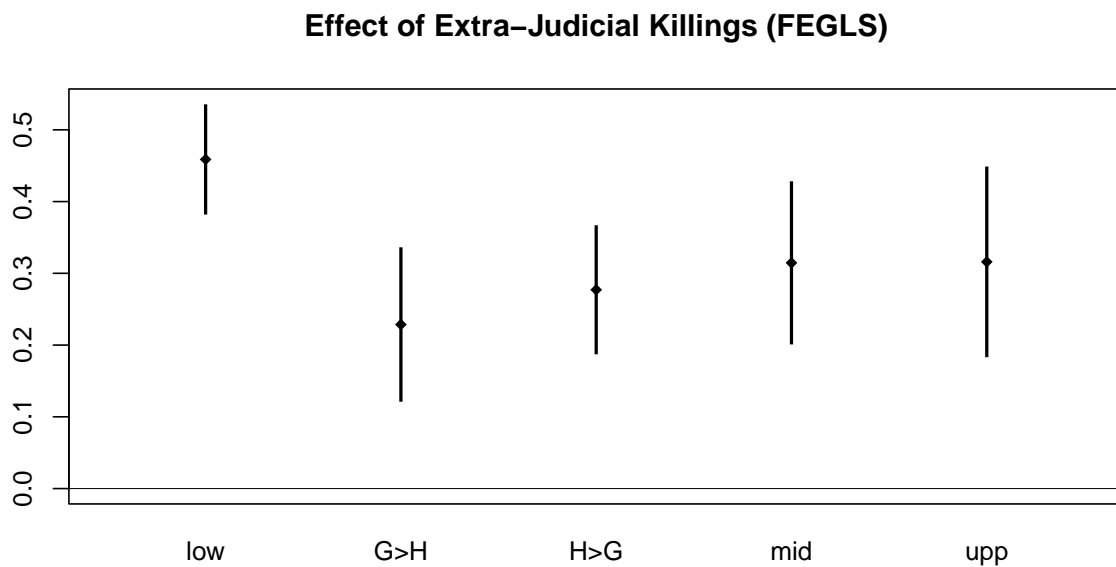


Figure 15: Effects of past extra-judicial killings on current killings

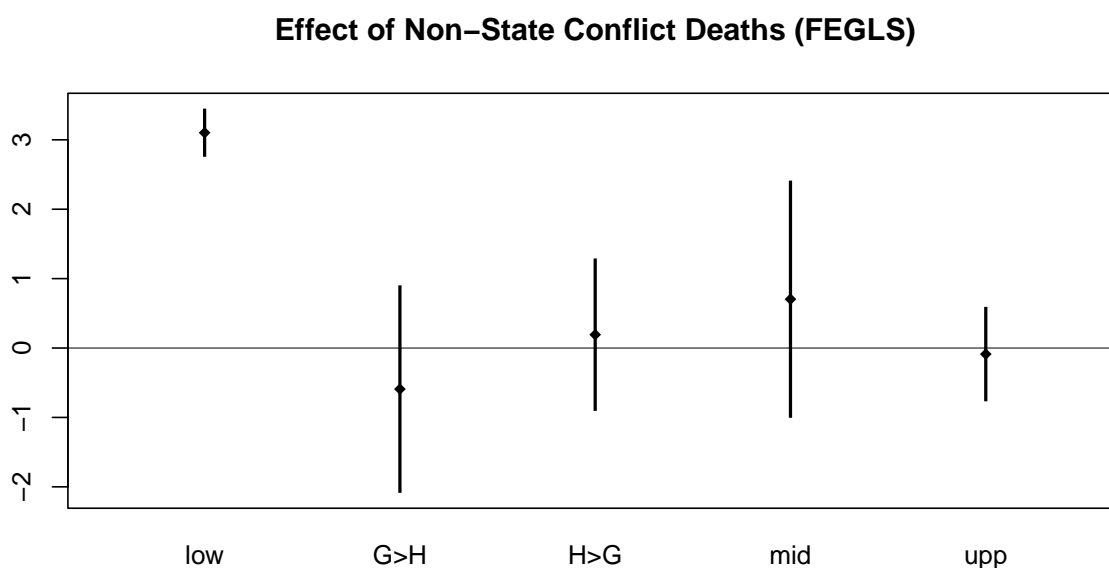


In the case of extra-judicial killings however, as shown in Figure 15, when controls of other types of violence are included, past extra-judicial killings are associated with more current killings

in the low classification than in the $G>H$ and $H>G$ classifications. This suggests that better health and gender outcomes are associated with less even at low levels.

7.3 Non-state conflict

Figure 16: Effects of past non-state conflict deaths on current deaths

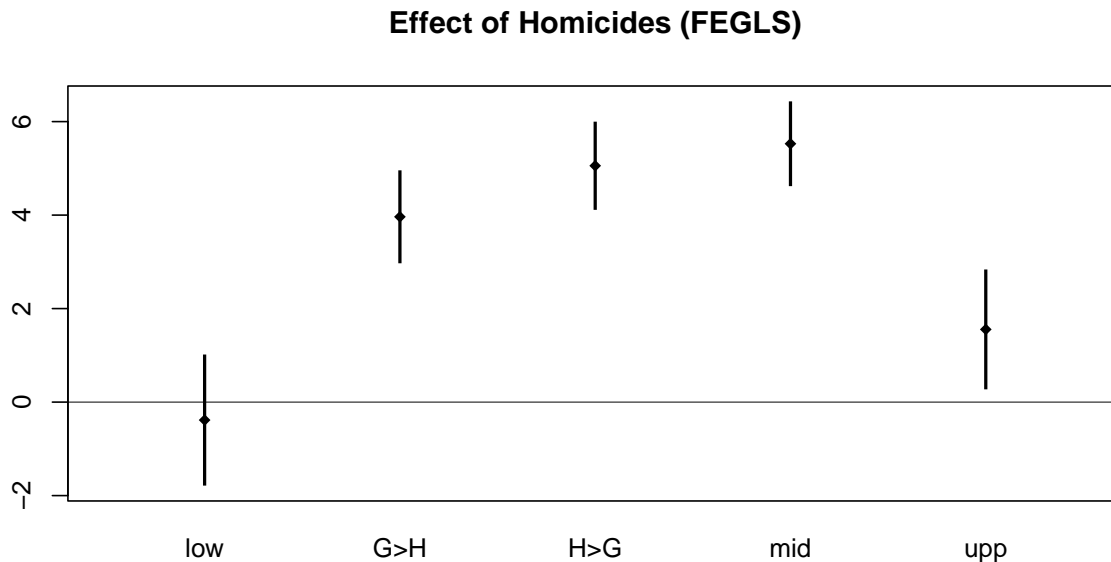


In the models of non-state conflict, none of the individual health and gender variables have clear associations. The low classification is associated with increased non-state deaths compared to all other classifications, suggesting a vicious cycle. The models with interactions show that past non-state conflict incidence is associated with more current incidence in the $H>G$ classification. Finally, in Figure 16, past deaths are strongly associated with more current deaths in the low classification, providing support for a vicious cycles, and this holds regardless of the estimator used.

7.4 Homicides

In the models of homicides, neither the individual health and gender variables, nor the classification indicators show statistically significant associations. However, the findings of the interactions are quite interesting. As shown in Figure 17, past homicides are associated with increased current homicides in all but the low classifications. This does not seem consistent with V&V cycles, but it is important to note that the magnitude of the associations for the $G>H$, $H>G$ and mid classifications are larger than and statistically distinguishable from the association for the highest classification. This does suggest a vicious cycle, just not for the low classification.

Figure 17: Effects of past homicides on current homicides



8 * Sequencing analysis: work-in-progress *

As noted in Section 1, the Lancet Commission is interested not only in whether V&V cycles are at work, but also if and how policy interventions to support gender equality and health equity can nudge societies into virtuous cycles. Toward this end, it would be useful to identify pathways of gender and health change to identify entry points and policy levers. A key question is: are there particular sequences of improvements in gender and health outcomes that are more likely to break countries out of vicious cycles? Therefore, another research stream is to map different pathways of change over time, subject to data availability, to determine which pathways are more common or likely to lead to improvements.

Initially, I mapped trajectories based on the classifications explained in Section 4, but I have concluded that this is not a good representation of change over time, because for any given five-year period the classifications are relative to all other countries. In the panel regression analyses, this is useful because it account for the global improvements across the gender and health measures, but it misrepresents change over time in trend analyses, because an underlying change from one period to the next may not be reflected in the classifications, or a classification may change without a significant change in the underlying health and gender measures.

Therefore, I have mapped the trajectories based on two of the underlying measures: the infant mortality rate (IMR) and the adolescent fertility rate (AFR). I consider these two variables, because the IMR is strongly correlated with life expectancy – and thus they include very similar information about trajectories – and the other gender variable – the mean years of schooling ratio of female to male – is only available for a shorter period of time, whereas the AFR is available for the same countries and years as the IMR. This is a promising approach. Figures of the trajectories for all

countries in the dataset are shown on the project website.⁷ The next step is to determine whether a useful sequencing typology can be developed from this representation.

⁷Available at timothoms.github.io/LSC-MWG/sequencing.html. Note that starting points are indicated in red, while endpoints are shown in green. As in the statistical analyses, the IMR and AFR are inverted such that higher values indicate “better” outcomes.

Appendix

A Summary tables for analyses extensions

Rather than presenting the many detailed regression tables here – these details are available on the project website – this appendix shows summary tables of statistically significant findings (at the 0.05 alpha level) of the main analyses. These show the outcome variables (DVs) in columns and key independent variables (IVs) in the rows, based on all models, with the following symbols indicating their statistical effects and specific notes; each cell in the table represents one or more separate models.

Table 3: Legend

Symbol	Note
+	statistically significant positive effect
–	statistically significant negative effect
()	statistical effect not found in the preferred model (FEGGLS)
na	variable not included in the model
DV	variable is the DV
ref	variable is the reference category
[l]	effect is relative to the low classification
[u]	effect is relative to the upper classification
[v]	effect is present when accounting for other types of violence

Table 4: Health & Gender Outcomes: base models

	Life Expectancy	IMR	UFMR	DALYs	MYS	AFR	Labour Participation Ratio	GII
lagged DV	+	+	+	+	+	+	+	+
Life Expectancy	DV	na	na	na		+		
Infant Mortality	na	DV	na	na		+		+
MYS Ratio	+	+	(+)	+	DV	na	na	na
Adolescent Fertility Rate	+	–	–	–	na	DV	na	na
low classification	–[u]				–[u]	–		
H>G classification	(+)[l]			(–)[l]	+ [l]	–[u]		
G>H classification	(–)[u] / +[l]	(–)	(–)[u]			+ [l]	–[l]	
mid classification	+ [l]			(–)[l]	+ [l]	+ [l]		
upp classification	+ [l]				+ [l]	+ [l]		

Table 5: Health & Gender Outcomes: Findings on Internal conflict

	Life Expectancy	IMR	UFMR	DALYs	MYS	AFR	Labour Participation Ratio	GII
Conflict	+			+			−[v]	+
— low interaction	+							(+)[v]
— H>G interaction	(+)	+				−		+ [v]
— G>H interaction		−		(+)[v]			(−)[v]	
— mid interaction			(+)			(−)[v] / +[v]		+ [v]
— upp interaction							−[v]	
Conflict Deaths		−	−	+		+		
— low interaction	−[v]	−	(−)	+		(−)[v] / +[v]		
— H>G interaction						+ [v]		
— G>H interaction	−[v]	−[v]	(−)[v]			(+)[v] / −[v]		−[v]
— mid interaction				(+)[v]	−[v]	−[v]		(+)[v]
— upp interaction				+		(−)[v]		
Conflict Civilian Deaths	+			+		+		
— low interaction	+			+ [v]		+		−[v]
— H>G interaction								
— G>H interaction	−[v]	−[v]	−[v]			−[v]		
— mid interaction		(−)[v]		(+)[v]		−[v]		
— upp interaction		+ [v]		+				

Table 6: Health & Gender Outcomes: Findings on State Repression

	Life Expectancy	IMR	UFMR	DALYs	MYS	AFR	Labour Participation Ratio	GII
OSV		(-)	(-)					
— low interaction		-[v]	- / (-)[v]		+[v]	+[v]		
— H>G interaction								+[v]
— G>H interaction	+[v]							
— mid interaction	+ / (+)[v]				-[v]	(-)[v]		
— upp interaction				(+)				
OSV Deaths	+	(-)	(-)	+		+		
— low interaction	+[v]	(-)(low)	(-)	+[v]		+	(+)	
— H>G interaction								
— G>H interaction		(-)	(-)		+[v]			-[v]
— mid interaction	+[v]				-[v]	(-)[v]		
— upp interaction	(+)[v]			(+)[v]			-	
LPI		+	+	+		(+)	(-)	+
— low interaction		-[v]	-[v]			+ / (-)[v] / -[v]		
— H>G interaction		+ / (+)[v] / -[v]	+ / (+)[v]	(+)[v]	-[v]	(+) / -[v]		+[v]
— G>H interaction	+[v]	+[v]	+	+	-[v]	+[v]		
— mid interaction		+ / (+)[v]	+ / (+)[v]	(+)[v]	-[v]			
— upp interaction	(-) / -[v]	+ / (+)[v]	+ / (+)[v]	(+)[v]		(+)[v] / -[v]	(-)[v]	+[v]
Killings								
— low interaction		(-) / -[v]	(-)	-[v]		-[v]		
— H>G interaction		+	+ / (+)[v]	(+)[v]		-[v]		
— G>H interaction	(+)			(+)[v]				
— mid interaction		+ / (+)[v]	+	(+)[v]				(+)[v]
— upp interaction		(+)[v]	(+)			(+)		

Table 7: Health & Gender Outcomes: Findings on Societal Violence

	Life Expectancy	IMR	UFMR	DALYs	MYS	AFR	Labour Participation Ratio	GII
NSC				+		+		
— low interaction				+		+[v]		-[v]
— H>G interaction		(-)	(-)			+[v]		-
— G>H interaction								
— mid interaction							(-)[v]	
— upp interaction					(+)[v]	+[v]		
NSC Deaths	+			+		+		
— low interaction	+	+[v]	+[v]	+	+[v]	+[v]		
— H>G interaction		(-)[v]	(-)[v]			+[v]		
— G>H interaction	(-)	(-)[v]				(+)[v]		-[v]
— mid interaction	(+)[v]						(-)	
— upp interaction		(+)[v]	(+)[v]	(+)[v]		(+)[v]		
NSC Civilian Deaths	+			+		+	(+)	
— low interaction	+	+[v]		+		+[v]	(+)[v]	
— H>G interaction		(-) / -[v]	(-)[v]	(-)[v]			(-)[v]	
— G>H interaction	+[v]	+	+[v]	(-)[v] / +[v]			(+)[v]	
— mid interaction								
— upp interaction	(+)[v]	(+)[v]	(+)[v]	(+)[v]	(-)	+[v]	(+)[v]	(+)
Homicides	-	+	+	(-)		+		
— low interaction	-[v]	-[v]	+ / -[v]	-[v]		-[v]	+[v]	-[v]
— H>G interaction	(-)[v]		+			-[v]		
— G>H interaction	-[v]	+		-[v]		-[v]	+[v]	-[v]
— mid interaction	(-)[v]	-[v]	(-)[v] / +	(-)[v]			+[v]	
— upp interaction	(-)[v]	+[v]	+[v]			+ / (+)[v]	-[v]	+[v]

Table 8: Violence Outcomes

	Conflict	Conflict Deaths	Civilian Deaths	OSV	OSV Deaths	LPI	Killings	NSC	NSC Deaths	NSC Civilian Deaths	Homicides
Life Expectancy			+	–		+					
Infant Mortality	–			(–)		(–)					
MYS Ratio					–	–					
Adolescent Fertility Rate	–					+	(+)				+
low classification	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
H>G classification		–				+		+	–		
G>H classification						+		+	–		
mid classification		–				+		+	–		
upp classification		–	–			+		+	–		
lagged DV	+	+	+	+		+	+	(+)	+	+	+
— low interaction	+ [v]	+ [v]	+ [v]	+ [v]		+ [v]	+ [v]		+ [v]	+ [v]	+ / – [v]
— H>G interaction	+	+ [v]	(–) [v]	(+)		+ [v]	+ [v]	+ [v]	(+) [v]		+ [v]
— G>H interaction	+ / (+) [v]	(+) [v]				+ [v]	+ [v]				+
— mid interaction	+	– [v]		+ [v]		+ [v]	+ [v]		(+) [v]		+ [v]
— upp interaction	+ / (+) [v]			+		+ [v]	+ [v]	(+) [v]			+ / – [v]

B Cross-sectional analyses in the previous report

This section reproduces the discussion of the cross-sectional analyses in the previous report, authored by Dane Rowlands.

B.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.

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.

B.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 9 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.⁸

⁸Recall 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 9: 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) ^a	0.0318 † (0.093) ^c			0.000532 † (0.083) ^c	-0.000365 † (0.079) ^c		
Conflict measure 1996-2015	-0.193 * (-0.049) ^{b,r}				-0.012 *** (0.003) ^{c,r}			
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)		-16.92 † (0.052)
Future Polyarchy 1996-2015	4.80 *** (0.002)				-0.0619 * (0.041)		-16.37 ** (0.024)	
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 ²	0.492	0.483	0.824	0.817	0.521	0.510	0.344	0.344
F-test p-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; ^a internal conflict indicator; ^b civilian death rate from internal conflict; ^c civilian death rate from one-sided-violence; ^r robust to variations in the other conflict measure.

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 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 9 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.

B.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.

Table 10: 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.

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 10.

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 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.

C Panel analyses in the previous report

C.1 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). The summary tables in this section 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 11 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

Table 11: Summary of Previous 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				

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 V&V 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 V&V cycles are possible, however, including considering V&V cycles separately. We explore an alternative conceptualization by examining the statistical effects of our low and high classifications, which operationalize the V&V 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 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 V&V 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 V&V 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 V&V cycles in violence outcomes, and here the evidence is clearer and more consistent.

Table 12: Summary of Previous 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										[+]		[+]	[(+)]

C.2 Violence outcomes

Table 12 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 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.

C.3 Discussion

The panel analyses find more support for vicious than virtuous cycles, and that these are more applicable to violence than gender and health outcomes. There is some evidence that health and gender outcomes reinforce each other, but the evidence is stronger for health than gender outcomes. 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, but only MYS is 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. 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. For the gender outcomes, life expectancy and the IMR are associated with an improved AFR, but they have no statistical effects on

MYS. When considering our classification, which operationalizes the V&V cycles hypotheses with worst and best combinations of health and gender outcomes, 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 support for vicious cycles between health and gender but not virtuous cycles.

Few of the violence variables have statistical effects on the gender and health outcomes that are consistent with the V&V cycles hypotheses. In fact, many models suggest that violence is associated with better health and gender outcomes, as indicated by + in the table, and contradicting our theoretical framework. (The findings for the AFR are particularly noteworthy in this respect.) A few results are consistent with the V&V cycles hypotheses, 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, overall, the health and gender models only support a vicious cycle between health and gender, but do not provide much or consistent evidence for a three-way vicious cycle. Moreover, many models turn up null results.

The analyses of various violence outcomes more clearly support the theoretical framework, though there are again several contradictory results with respect to the relationship between individual health and gender variables and violence variables, and many null results. Supporting the theoretical framework, higher life expectancy is associated with less one-sided violence episodes (but 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. Like the analyses of health and gender outcomes, these results are not strongly supportive.

However, the evidence is much more consistent and supportive of vicious cycles when using our classifications. 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 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. By contrast, positive interaction effects of the high classification contradict the notion of virtuous cycles.