Democracy and Human Development:
Issues of Conceptualization and Measurement

Appendices

Contents

APPENDIX A: Data Collection 3
Recruitment 4
Coding 5
Measurement 6
APPENDIX B: Variable Descriptions and Descriptive Statistics 8
Table B1: Variable Descriptions 9
Table B2: Descriptive Statistics 14
APPENDIX C: Robustness 16
Table C1: MPI and Mortality 18
Table C2: Restricted Sample Tests 24
Table C3: Tests of Functional Form 25
Table C4: Possible Non-linearities in MPI stock 26
Table C5: Instrumental Variable Tests 27
Table C6: Possible Non-linearities in GDP 28
Table C7: Additional Tests of MPI level 29
Table C8: Tests of MPI using Driscoll-Kraay Standard Errors 30
Table C9: Tests of MPI and GDPpc (ln) stock (10%) 31
Table C10: Tests of MPI using Annual Data 32
Table C11: Tests of MPI and Political Durability 33
Table C12: Tests of MPI not including Measurement Error 34
Table C13: Democracy Indices using Data from 1960-present 36
Table C14: Additional Tests of BMR 37
Table C15: Additional Tests of BNR 38
APPENDIX D: The Multiplicative Polyarchy Index (MPI) 39
Table D1: MPI level by Country in 2000 40
Table D2: Correlation Matrix of MPI Components 41
Disaggregation 42
Table D3: Excluding Components of MPI 42

Table D4: Disaggregating MPI 43
Table D5: Additional Tests of Clean Elections 44
Figure D1: Histograms of MPI stock (10%) and MPI level in 2000 45
Figure D2: MPI level and Polity2 in 1900 46
Figure D3: MPI level and Polity2 in 2000 47
Table D6: Horse-race Tests 48
Table D7: Varying Aggregation Rules 49
Figure D4: Change in Predictive Margins based on Aggregation Rules 50
APPENDIX E: Mediation Analysis 51
Table E1: Mediation Analysis using Annual Data 54
Table E2: Mediation Analysis using 10-year Intervals 55
APPENDIX REFERENCES 56

APPENDIX A: Data Collection

The MPI index is based on indicators drawn from the Varieties of Democracy (V-Dem) project, as are several of the other measures tested in Tables 1 and 2. It is important, therefore, to understand a bit about how the data was collected and aggregated across coders. (For further information see Coppedge et al. (2015) and Pemstein et al. (2015).)

Each indicator in the V-Dem dataset that is not factual in nature is coded by multiple Country Experts, generally about five (5). Most experts do not possess the requisite expertise to code the entire V-Dem questionnaire, which means that a single country will generally be coded by a dozen or more experts, each working on different facets of the questionnaire. To date, V-Dem has engaged in collaboration with over 2,500 Country Experts.

Recruitment

The following procedure is used to recruit Country Experts. First, we identify a list of potential coders for a country (typically 100-200 names per country). This bulk of names on the list are provided by Regional Managers (members of the V-Dem project located in universities and think-tanks throughout the world) in consultation with other members of the V-Dem team. Assistant Researchers (located at V-Dem Institute, University of Gothenburg) also contribute to this list, using information about potential country experts gathered from the web. Other members of the project team provide additional names if they have country-specific expertise. At present, V-Dem has accrued a roster of 15,000+ potential Country Experts.

For each potential Country Expert on the resulting list, we compile basic information – country of origin, current location, highest educational degree, current position, and area of expertise in terms of the surveys the expert could code as evidenced by a short biographical sketch and/or list of publications, website information and the like. We also take note of any possible biases that might affect their ability to code questions in a dispassionate manner.

In selecting whom to recruit from this list five criteria come into play:

The most important selection criterion, naturally, is expertise in the country(ies) and the section of the survey they are assigned to code. This is usually signified by an advanced degree in the social sciences, law, or history; a record of publications; and positions in civil society that establish their expertise in the chosen area (e.g. a well-known and respected journalist). Naturally, potential coders are drawn to areas of the survey that they are most familiar with, and are unlikely to agree to code topics they know little about. So, self-selection also works to achieve our primary goal of matching questions in the survey with country-specific expertise.

The second criterion is origin in the country to be coded. V-Dem's goal is that a minimum of three out of five (60%) Country Experts should be nationals or permanent residents of the country they code (preferably both). Exceptions are required for a few countries where it is difficult to find in-country coders who are both qualified and independent of the governing regime. This criterion should help avoid potential Western/Northern biases in the coding.

The third criterion is the prospective coder's seriousness of purpose. By this, we mean a person's willingness to devote time to the project, to deliberate carefully over the questions asked in the survey, and to report their honest judgment. Sometimes, personal acquaintanceship is enough to convince a Regional Manager that a person is fit, or unfit, for the job. Sometimes, this feature becomes apparent in communications with Project Coordinators that precede the offer to work on V-Dem.

The fourth criterion is impartiality. V-Dem aims to recruit coders who will answer survey questions in an impartial manner. This means avoiding those who might be beholden to powerful actors – by reason of coercive threats or material incentives – or who serve as spokespersons for a

political party or ideological tendency (in some instances, such as North Korea, this may entail avoiding all in-country coders). Where this is difficult, or where the reality is difficult to determine, we aim to include a variety of coders who, collectively, represent an array of views and political perspectives on the country in question.

The final criterion is obtaining diversity in professional background among the coders chosen for a particular country. For certain areas (e.g the media, judiciary, and civil society surveys) this entails a mixture of highly recognized professionals from the sector along with academics who study these topics. Generally, it also means finding experts who are located at a variety of institutions, universities and research institutes.

After weighing these five criteria, the 100-200 potential experts on the list are given a rank from "1" to "3" indicating order of priority.

The two Project Coordinators at the V-Dem Institute, University of Gothenburg, then handle the enrolment of Country Experts from the list of potential country experts. In handling the recruitment, the continuously review the resulting mix of actual country experts in light of the five criteria to ensure that V-Dem ends up with a set of experts for each country that fulfill our standards.

If the quota of five Country Experts per section of the survey for each country is not met, we work down the list of potential Country Experts until the quota is obtained. Others, following the same procedure, replace those who fail to complete the survey in a reasonable time. Coders receive a modest honorarium for their work that is proportional to the number of surveys they have completed.

A number of steps are taken to assure informed consent and confidentiality among participants. The on-line survey provides full information about the project (including this document) and the use of the data, so that coders are fully informed. It also requires that prospective coders certify that they accept the terms of the agreement. They can access the surveys only with a randomized username that we assign and a secret password that they create themselves. The data they supply is stored on a firewall-protected server. Any data released to the public excludes information that might be used to identify coders. All personal identifying information is kept in a separate database in order to ensure the protected identities of coders.

In order to ensure that we are able to recruit widely among potential experts, and in order to minimize confusion due to unfamiliarity with English, questions are translated from English into five additional languages: Arabic, French, Portuguese, Russian, and Spanish. Approximately 15 percent of the experts code in a non-English version of the questionnaire.

About 35 percent of the Country Experts are women, and over 80 percent have PhDs or MAs and are affiliated with research institutions, think tanks, or similar organizations.

Coding

Coding is carried out using the V-Dem online survey tool. The web-based coding interfaces are directly connected with a postgres database where the original coder-level data is kept, maintaining coder confidentiality.

In addition to country-specific ratings, Country Experts are requested to code several additional countries that they are familiar with for a shorter time-slice. This «bridge» or «lateral» coding assures cross-country equivalence by forcing coders to make explicit comparisons across countries, and provides critical information for the measurement model (described below).

For each question, and for each country-year, experts are required to report a self-assessed level of certainty. This is an indicator of their subjective level of uncertainty for the data point they

provide. This is scored on a scale from 0 to 100 with substantive anchor points for each 10-percent interval.

Measurement

Having discussed the process of data collection, we proceed to the task of measurement. Under this rubric, we include (a) the questionnaire, (b) our measurement model, (c) methods of identifying error in measurement, (d) studies of measurement error, and (e) methods of correcting error. In principle, the discussions are relevant for different types of data (A, B, and C in the V-Dem scheme) but most if not all of them are much more acute when it comes to expert-based coding of evaluative, non-factual yet critical indicators. Hence, most of the following is focused on the C-type indicators.

The most important feature of a survey is the construction of the questionnaire itself. In crafting indicators we have sought to construct questions whose meaning is clear and specific and not open to a wide variety of interpretations. They should mean the same thing (more or less) in each context and not suffer from temporal or spatial non-equivalence. Our methodology involves enlisting some of the leading scholars in the world on different aspects of democracy and democratization – known as Project Managers.

Each Project Manager was enrolled because of his/her specific and evidenced expertise in a particular area (e.g. legislatures, executives, elections, civil society, and so on) and with a view to generate a group that also had substantive experiences and expertise on all regions of the world. Starting in 2009, Project Managers designed survey-questions in their area to measure democraticness in relation to the different traditions of democratic theory. All suggestions were reviewed and refined collectively over the course of two years. The V-Dem pilot test carried out in 2011 served as an initial test of our questionnaire, prompting quite a few revisions in the next round of surveys. Another round of collective deliberation followed that also involved a number of consultations with scholars outside of the project team. The revised questions for C-coding thus went through several rounds of review with the Project Managers and outside experts over the course of two years before emerging in their final form, depicted in the Codebook.

Even with careful question design, a project of this nature cannot help but encounter error. This may be the product of linguistic misunderstandings (recall that most of our coders do not speak English as their first language and some take the survey in a translated form), misunderstandings about the way in which a question applies to a particular context, factual errors, errors due to the scarcity or ambiguity of the historical record, differing interpretations about the reality of a situation, variation in standards, coder inattention, errors introduced by the coder interface or the handling of data once it has been entered into the database, or random mistakes.

Some of these errors are stochastic in the sense of affecting the precision of our estimates but not their validity. Other errors are systematic, potentially introducing bias into the estimates that we produce.

Having five coders for each question is immensely useful, as it allows us to identify wayward coders as well as to conduct inter-coder reliability tests. These sorts of tests – standard in most social science studies – are rarely if ever employed in extant democracy indices.

While we select experts carefully, they clearly exhibit varying levels of reliability and bias, and may not interpret questions consistently. In such circumstances, the literature recommends that researchers use measurement models to aggregate diverse measures where possible, incorporating information characterized by a wide variety of perspectives, biases, and levels of reliability (Bollen & Paxton 2000, Clinton & Lapinski 2006, Clinton & Lewis 2008, Jackman 2004, Treier & Jackman

2008, Pemstein, Meserve & Melton 2010). To combine expert ratings for a particular country/indicator/year to generate a single "best estimate" for each question, we employ methods inspired by the psychometric and educational testing literature (see e.g. Lord & Novick 1968, Jonson & Albert 1999, Junker 1999, Patz & Junker 1999).

The underpinnings of these measurement models are straightforward: they use patterns of cross-rater (dis)agreement to estimate variations in reliability and systematic bias. In turn, these techniques make use of the bias and reliability estimates to adjust estimates of the latent—that is, only indirectly observed—concept (e.g. executive respect for the constitution, judicial independence, or property rights) in question. These statistical tools allow us to leverage our multi-coder approach to both identify and correct for measurement error, and to quantify confidence in the reliability of our estimates. Variation in these confidence estimates reflect situations where experts disagree, or where little information is available because few raters have coded a case. These confidence estimates are tremendously useful. Indeed, the tendency of most researchers to treat the quality of measures of complex, unobservable concepts as equal across space and time, ignoring dramatic differences in ease of access and measurement across cases, is fundamentally misguided, and constitutes a key threat to inference.

The majority of expert-coded questions are ordinal: they require raters to rank cases on a discrete scale, generally with four or five response categories. To achieve scale consistency, we fit ordinal IRT models to each question (see Johnson & Albert 1999 for a technical description of these models). These models achieve three goals. First, they work by treating coders' ordinal ratings as imperfect reflections of interval-level latent concepts. Therefore, while an IRT model takes ordinal values as input, its output is an interval-level estimate of the given latent trait (e.g. election violence). Interval-valued estimates are valuable for a variety of reasons; in particular, they are especially amenable to statistical analysis. Second, IRT models allow for the possibility that coders have different thresholds for their ratings (e.g. one coder's somewhat might fall above another coder's almost on the latent scale), estimate those thresholds from patterns in the data, and adjust latent trait estimates accordingly. Therefore, they allow us to correct for this potentially serious source of bias. This is very important in a multi-rater project like V-Dem, where coders from different geographic or cultural backgrounds may apply differing standards to their ratings. Finally, IRT models assume that coder reliability varies, produce estimates of rater precision, and use these estimates—in combination with the amount of available data and the extent to which coders agree—to quantify confidence in reported scores.

With lateral and bridge coding we are able to mitigate the incomparability of coders' thresholds and the problem of cross-national estimates' calibration. While helpful in this regard, our tests indicate that given the sparsity of our data, even this extensive bridge-coding is not sufficient in solving cross-national comparability issues. We therefore also employ a data-collapsing procedure. At its core, this procedure relies on the assumption that as long as none of the experts change their ratings for a given time period, we can treat the country-years in this period as one year. The results of our statistical models indicate that this technique is extremely helpful in increasing the weight given to lateral/bridge coders, and thus further mitigates cross-national comparability problems.

APPENDIX B: Variable Descriptions and Descriptive Statistics

Table B1: Variable Descriptions

Outcomes

Infant mortality rate (IMR). Babies who die prior to their first birthday as a share of 1,000 live births, transformed by the natural logarithm (unless otherwise noted). Missing data within a time series is interpolated using a linear model. *Source*: Gapminder (gapminder.org), drawing on the UN Inter-agency Group for Child Mortality Estimation (www.childmortality.org), the Human Mortality Database (www.mortality.org), UNICEF (www.childinfo.org), and B.R. Mitchell (1998a, 1998b, 1998c) – which, in turn, draw on a wide variety of sources. Data sources are listed for each observation in a lengthy Excel table, downloadable at: https://www.gapminder.org/data/documentation/gd002/ *Tag:* imr_gapminder_ipo_ln

Infant mortality rate (IMR), WDI. Babies who die prior to their first birthday as a share of 1,000 live births, transformed by the natural logarithm (unless otherwise noted). *Source:* WDI (World Bank 2013). *Tag:* wdi_mort_ln

Child mortality rate (CMR). Children who die prior to their fifth birthday as a share of 1,000 live births, transformed by the natural logarithm (unless otherwise noted). Missing data within a time series is interpolated using a linear model. *Source*: Gapminder (gapminder.org). *Tag:* cme_gapminder_ipo_ln

Life expectancy (LE). Expected longevity at birth based on current age-specific mortality rates. The variable is transformed by subtracting LE from the maximum value in our sample (85) and then taking the logarithm of that number. A low number signals a lower mortality rate. Missing data within a time series is interpolated using a linear model. *Source*: Gapminder (gapminder.org). *Tag*: le_gapminder_ipo_85mln

Components of MPI

Clean elections. Clean elections connote an absence of registration fraud, systematic irregularities, government intimidation of the opposition, vote buying, and election violence. The index is formed by taking the point estimates from a Bayesian factor analysis model of the indicators for EMB autonomy (v2elembaut), EMB capacity (v2elembcap), election voter registry (v2elrgstry), election vote buying (v2elvotbuy), election other voting irregularities (v2elirreg), election government intimidation (v2elintim), election other electoral violence (v2elpeace), and election free and fair (v2elfrfair). Since the bulk of these indicators are only observed in election years, the index scores have then been repeated within election regime periods (as defined by v2x_elecreg). *Source:* V-Dem. *Tag:* v2xel_frefair

Elected executive. This index attempts to measure whether the chief executive is elected, either directly elected through popular elections or indirectly through a popularly elected legislature that then appoints the chief executive. There are six different chains of appointment/selection to take into account in constructing this index, all of which are scaled to vary from 0 to 1. First, whether the head of state is directly elected (a=1) or not (a=0). Second, the extent to which the legislature is popularly elected (b), measured as the proportion of legislators elected (if legislature is unicameral), or the weighted average of the proportion elected for each house, with the weight defined by which house is dominant (if legislature is bicameral). Third, whether the head of state is appointed by the legislature, or the approval of the legislature is necessary for the appointment of the head of state (c1=1, otherwise 0).

Fourth, whether the head of government is appointed by the legislature, or the approval of the legislature is necessary for the appointment of the head of government (c2=1, otherwise 0). Fifth, whether the head of government is appointed by the head of state (d=1) or not (d=0). Sixth, whether the head of government is directly elected (e=1) or not (e=0). Define hosw as the weight for the head of state. If the head of state is also head of government (v2exhoshog==1), hosw=1. If the head of state has more power than the head of government over the appointment and dismissal of cabinet ministers, then hosw=1; if the reverse is true, hosw=0. If they share equal power, hosw=.5. Define the weight for the head of government as hogw=1-hosw. *Source:* V-Dem. *Tag: v2x_elecoff*

Free association. This index attempts to measure the extent to which parties, including opposition parties, are allowed to form and to participate in elections, and the extent to which civil society organizations are able to form and to operate freely. The index is formed by taking the point estimates from a Bayesian factor analysis model of the indicators for party ban (v2psparban), barriers to parties (v2psbars), opposition parties autonomy (v2psoppaut), elections multiparty (v2elmulpar), CSO entry and exit (v2cseeorgs) and CSO repression (v2csreprss). Since the multiparty elections indicator is only observed in election years, its values have first been repeated within election regime periods (as defined by v2x_elecreg). *Source:* V-Dem. *Tag:* v2x_frassoc_thick

Free expression. This index attempts to measure the extent to which the government respects press and media freedom, the freedom of ordinary people to discuss political matters at home and in the public sphere, as well as the freedom of academic and cultural expression. The index is formed by taking the point estimates from a Bayesian factor analysis model of the indicators for print/broadcast censorship effort (v2mecenefm), internet censorship effort (v2mecenefi), harassment of journalists (v2meharjrn), media bias (v2mebias), media self-censorship (v2meslfcen), print/broadcast media critical (v2mecrit), and print/broadcast media perspectives (v2merange), freedom of discussion for men/women (v2cldiscm, v2cldiscw) and freedom of academic and cultural expression (v2clacfree). *Source:* V-Dem. *Tag:* v2x_freexp_thick

Suffrage. The share of adult citizens (as defined by statute) that has the legal right to vote in national elections. This measure covers legal (de jure) restrictions, not restrictions that may be operative in practice (de facto). The scores reflect de jure provisions of suffrage extension in percentage of the adult population as of January 1 in a particular year. The adult population (as defined by statute) is defined by citizens in the case of independent countries or the people living in the territorial entity in the case of colonies. Universal suffrage is coded as 100%. Universal male suffrage only is coded as 50%. Years before electoral provisions are introduced are scored 0%. The scores do not reflect whether an electoral regime was interrupted or not. Only if new constitutions, electoral laws, or the like explicitly introduce new regulations of suffrage, the scores were adjusted accordingly if the changes suggested doing so. If qualifying criteria other than gender apply (such as property, tax payments, income, literacy, region, race, ethnicity, religion, and/or 'economic independence'), estimates have been calculated by combining information on the restrictions with different kinds of statistical information (on population size, age distribution, wealth distribution, literacy rates, size of ethnic groups, etc.), secondary country-specific sources, and – in the case of very poor information – the conditions in similar countries or colonies. *Source:* V-Dem. *Tag:* v2x_suffr

Other Measures of Democracy

Polity2. Measures the extent to which democratic or autocratic "authority patterns" are institutionalized in a given country. It takes into account how the executive is selected, the degree of checks on executive power, and the form of political competition. *Source:* Marshall et al. (2014). *Tag:* e_polity2.

UDS. A democracy index comprised of multiple indicators and aggregated through a Bayesian IRT measurement model. *Source:* Pemstein et al. (2010). *Tag:* e_uds_mean.

Contestation. Defined as the "extent and fairness of electoral competition between parties and distinct interests," including "the existence of independent political parties, the freedom of electoral competition, the extent of intra-governmental constraints, legislative membership by opposition parties and the closeness of national votes," as measured by a variety of extant indicators. *Source:* Miller (2015). *Tag:* contdim.

Inclusiveness. Defined as "the extent of popular electoral involvement across the citizenry," understood as including suffrage and turnout, and measured with a variety of extant indicators. *Source:* Miller (2015). *Tag:* partdim.

Participation. The participatory principle of democracy emphasizes active participation by citizens in all political processes, electoral and non-electoral. It is motivated by uneasiness about a bedrock practice of electoral democracy: delegating authority to representatives. Thus, direct rule by citizens is preferred, wherever practicable. This model of democracy thus takes suffrage for granted, emphasizing engagement in civil society organizations, direct democracy, and subnational elected bodies. This index is formed by averaging the following indices: civil society participation (v2x_cspart), direct popular vote (v2xdd_dd), elected local government power (v2xel_locelec), and elected regional government power(v2xel_regelec). *Source:* V-Dem. *Tag:* v2x_partip.

Deliberation. The deliberative principle of democracy focuses on the process by which decisions are reached in a polity. A deliberative process is one in which public reasoning focused on the common good motivates political decisions—as contrasted with emotional appeals, solidary attachments, parochial interests, or coercion. According to this principle, democracy requires more than an aggregation of existing preferences. There should also be respectful dialogue at all levels—from preference formation to final decision—among informed and competent participants who are open to persuasion. To measure these features of a polity we try to determine the extent to which political elites give public justifications for their positions on matters of public policy, justify their positions in terms of the public good, acknowledge and respect counter-arguments; and how wide the range of consultation is at elite levels. The index is formed by point estimates drawn from a Bayesian factor analysis model including the following indicators: reasoned justification (v2dlreason), common good justification (v2dlcommon), respect for counterarguments (v2dlcountr), range of consultation (v2dlconslt), and engaged society (v2dlengage). *Source:* V-Dem. *Tag:* v2xdl_delib.

Female power. Political empowerment is understood to include open discussion of political issues, participation in civil society organizations, freedom of movement, the right to private property, access to justice, freedom from forced labor, representation in the ranks of journalists, and an equal share in the overall distribution of power. The index is formed by taking the point estimates from a Bayesian factor analysis model of the indicators for CSO women's participation (v2csgender), female journalists (v2mefemjrn), freedom of domestic movement for women (v2cldmovew), freedom of discussion for women (v2cldiscw), freedom

from forced labor for women (v2clslavef), property rights for women (v2clprptyw), access to justice for women (v2clacjstw), and power distributed by gender (v2pepwrgen). *Source:* V-Dem. *Tag:* v2x_gender.

Civil society. The sphere of civil society lies in the public space between the private sphere and the state. Here, citizens organize in groups to pursue their collective interests and ideals. We call these groups civil society organizations (CSOs). CSOs include, but are by no means limited to, interest groups, labor unions, spiritual organizations (if they are engaged in civic or political activities), social movements, professional associations, charities, and other non-governmental organizations. The core civil society index (CCSI) is designed to provide a measure of a robust civil society, understood as one that enjoys autonomy from the state and in which citizens freely and actively pursue their political and civic goals, however conceived. The index is formed by taking the point estimates from a Bayesian factor analysis model of the indicators for candidate selection – national/local (v2pscnslnl), CSO consultation (v2cscnsult), CSO participatory environment (v2csprtcpt), and CSO women's participation (v2csgender). *Source:* V-Dem. *Tag:* v2x_cspart.

Individual liberty. To what extent are laws transparent and rigorously enforced and public administration impartial, and to what extent do citizens enjoy access to justice, secure property rights, freedom from forced labor, freedom of movement, physical integrity rights, and freedom of religion? The index is formed by taking the point estimates from a Bayesian factor analysis model of the indicators for rigorous and impartial public administration (v2clrspct), transparent laws with predicTable Dnforcement (v2cltrnslw), access to justice for men/women (v2clacjstm, v2clacjstw), property rights for men/women (v2clprptym, v2clprptyw), freedom from torture (v2cltort), freedom from political killings (v2clkill), from forced labor for men/women (v2clslavem v2clslavef), freedom of religion (v2clrelig), freedom of foreign movement (v2clfmove), and freedom of domestic movement for men/women (v2cldmovem, v2cldmovew). *Source:* V-Dem. *Tag:* v2xcl_rol.

BMR. Dichotomous democracy measure based on contestation and participation. Countries coded democratic have (1) political leaders that are chosen through free and fair elections and (2) a minimal level of suffrage. *Source:* Boix, Miller & Rosato (2013). *Tag:* e_boix_regime.

BNR. Following Dahl (1971), a country is defined as democratic if there is a high level of contestation and at least 50% of the adult population is allowed to vote. *Source:* Bernhard, Nordstrom & Reenock (2011). *Tag:* e_bnr_dem.

Covariates

GDP per capita (ln). Gross domestic product per capita, transformed by the natural logarithm. *Source:* Maddison Project (Bolt & van Zanden 2014). *Tag:* e_migdppc_ln.

Urbanization. Ratio of urban population to total population. *Source:* V-Dem, constructed from data from CLIO Infra (clio-infra.eu). *Tag:* e_miurbani.

Fertility (ln). Fertility rate, transformed by the natural logarithm. The fertility rate (aka total fertility rate, period total fertility rate, total period fertility rate) of a population is the mean number of children that would be born to a woman over her lifetime if (a) she were to experience the current age-specific fertility rates through her lifetime, and (b) she were to survive through the end of her reproductive life. It is obtained by adding single-year age-specific rates at a given time. *Source:* WDI (World Bank 2013). *Tag:* e_miferrat_ln.

Growth. Annual growth rate of GDP per capita. *Source:* Maddison Project (Bolt & van Zanden 2014). *Tag:* e_migdpgro.

Internal Conflict. Coded 1 if the country suffered in an internal armed conflict in a given year, 0 otherwise. The original source codebook (Brecke 2001) states that no war is coded as 0 and war is coded as 1. However, the data contains only 1's along with missing data (no 0's). Following the authors' instructions (personal communication), we re-code missing observations as non-conflict (0) for countries where at least one year in the original times series (which runs from 1500 until present) was coded as 1. *Sources:* Clio Infra (clio-infra.eu), drawing on Brecke (2001), compiled by V-Dem. *Tag:* e_miinterc.

External conflict. Coded 1 if the country participated in an international armed conflict in a given year, 0 otherwise. The original source codebook (Brecke 2001) states that no war is coded as 0 and war is coded as 1. However, the data contains only 1's along with missing data (no 0's). Following the authors' instructions (personal communication), we re-code missing observations as non-conflict (0) for countries where at least one year in the original times series (which runs from 1500 until present) was coded as 1. *Sources:* Clio Infra (clio-infra.eu), drawing on Brecke (2001), compiled by V-Dem. *Tag:* e_miinteco.

Corruption stock (10%). Includes indicators of corruption in the executive, the legislature, the judiciary, and the public sector at-large, aggregated with Bayesian factor analysis and then constructed as a historical stock with a 10% annual depreciation rate. *Source:* V-Dem. *Tag:* v2x_corr_stock_10.

Instruments

WAVE. This instrument for democracy codes whether the last regime change in a country, as registered by Polity IV and its regime duration variable, occurred within or outside on of Huntington's (1991) "reverse waves of democratization". *Sources:* Marshall et al. (2014), drawing on the coding rules described in Knutsen (2011). *Tag:* inreversewave.

Regional/global means of MPI. These instruments are calculated for MPI level, for each year, and the country in question (for which we instrument MPI stock/level scores) is always exempted when we calculate the regional and global averages. Thus the instruments represent the average MPI scores for all other countries in the region/globally that have data in that particular year. *Sources:* V-Dem. *Tags:* regionEDmt_minus; globalEDmt_minus

Table B2: Descriptive Statistics

	Obs.	*mean*	*SD*	*min*	*max*
Infant mortality rate (ln), Gapminder	1,246	3.901	1.057	0.642	5.761
Infant mortality rate (ln), WDI	719	3.537	1.095	0.588	5.334
Child mortality rate (ln)	1,043	4.099	1.202	0.956	6.129
Life expectancy (ln)	1,640	3.247	0.603	0.646	4.147
MPI	1,745	0.199	0.302	0.000	0.969
MPI stock (1%)	1,745	0.089	0.180	0.000	0.983
MPI stock (10%)	1,745	0.171	0.280	0.000	0.997
Clean elections (V-Dem)	1,792	0.352	0.343	0.000	0.989
Elected executive (V-Dem)	1,796	0.534	0.458	0.000	1.000
Free association (V-Dem)	1,785	0.450	0.320	0.022	0.959
Free expression (V-Dem)	1,784	0.466	0.309	0.012	0.986
Suffrage (V-Dem)	1,777	0.660	0.421	0.000	1.000
GDP per capita (ln)	1,256	7.848	1.059	5.414	10.439
Urbanization	1,732	0.365	0.251	0.008	1.000
Fertility (ln)	1,785	1.460	0.507	0.067	2.206
Growth	1,193	2.051	3.267	-23.261	32.393
Internal Conflict	1,274	0.099	0.242	0.000	1.000
External conflict	1,539	0.077	0.198	0.000	1.000
Corruption (10%)	1,784	3.958	2.553	0.093	9.350
Polity2 (Marshall)	1,310	0.537	0.349	0.000	1.000
Polity2 stock (1%)	1,310	0.235	0.228	0.000	0.988
Polity2 stock (10%)	1,310	0.441	0.319	0.000	1.000
UDS (Pemstein)	1,094	0.479	0.215	0.028	0.982
UDS stock (1%)	1,094	0.256	0.210	0.002	0.991
UDS stock (10%)	1,094	0.401	0.234	0.008	0.998
Contestation (Miller)	1,111	0.483	0.339	0.004	0.999
Contestation stock (1%)	1,111	0.220	0.229	0.001	0.990
Contestation stock (10%)	1,111	0.390	0.306	0.007	0.998
Inclusiveness (Miller)	1,113	0.529	0.239	0.000	0.904
Inclusiveness stock (1%)	1,113	0.276	0.226	0.000	0.987
Inclusiveness stock (10%)	1,113	0.482	0.266	0.000	0.997
Contestation * Inclusiveness	1,110	0.318	0.281	0.000	0.920
Contestation * Inclusiveness (1%)	1,110	0.164	0.210	0.000	0.987
Contestation * Inclusiveness (10%)	1,110	0.285	0.282	0.000	0.994
Participation (V-Dem)	1,798	0.349	0.214	0.000	0.823
Participation stock (1%)	1,798	0.231	0.208	0.000	0.988
Participation stock (10%)	1,798	0.374	0.254	0.000	0.992
Deliberation (V-Dem)	1,798	0.428	0.306	0.000	1.000
Deliberation stock (1%)	1,798	0.207	0.203	0.000	0.986
Deliberation stock (10%)	1,798	0.371	0.284	0.001	1.000
Female power (V-Dem)	1,451	0.505	0.259	0.002	0.997
Female power stock (1%)	1,451	0.232	0.202	0.001	0.981
Female power stock (10%)	1,451	0.415	0.258	0.005	0.999

Civil society (V-Dem)	1,777	0.438	0.302	0.011	1.000
Civil society stock (1%)	1,777	0.221	0.207	0.001	0.985
Civil society stock (10%)	1,777	0.383	0.279	0.005	0.999
Individual liberty (V-Dem)	1,798	0.500	0.305	0.005	0.997
Individual liberty stock (1%)	1,798	0.258	0.226	0.001	0.986
Individual liberty stock (10%)	1,798	0.436	0.290	0.003	1.000
BMR (Boix)	1,342	0.404	0.468	0.000	1.000
BMR stock (1%)	1,342	0.170	0.253	0.000	1.000
BMR stock (10%)	1,342	0.331	0.399	0.000	1.000
BNR (Berhard)	891	0.371	0.460	0.000	1.000
BNR stock (1%)	891	0.152	0.245	0.000	0.969
BNR stock (10%)	891	0.269	0.364	0.000	0.999
MPI global mean	1,683	1.493	1.052	0.100	3.847
MPI regional mean	1,679	1.478	2.006	0.000	8.330
WAVE	1,299	0.397	0.461	0.000	1.000

APPENDIX C: Robustness

In the paper, we show that the MPI is robust to several specifications and functional forms. In this appendix, we include a larger battery of robustness tests, along with a full discussion of Table 2 (reproduced as Table C1).

Model 1 in Table C1 includes the benchmark model from Table 1. Here, IMR (logged) is regressed against MPI along with per capita GDP, country fixed-effects, and year fixed-effects in an OLS model with standard errors clustered by country. Right-side variables are again lagged one time-period (one year, or a decade for our 10-year panel specifications) behind the outcome.

A key feature of all of the tests in Table C1 is the incorporation of measurement error drawn from the V-Dem measurement model, where multiple ratings are combined into a single point estimate along with a confidence interval for each country-year-indicator, as described briefly in Appendix A and more extensively in Pemstein et al. (2015). Measurement error associated with democracy and other macro-level indices, while often informally acknowledged, is rarely incorporated into empirical tests. We do so by running the specified model on 900 draws of the posterior distribution estimated for MPI, based on an aggregation of the posteriors for each component of the index (Pemstein, Meserve, & Melton 2010).

In Model 1, the benchmark, MPI is measured as a stock variable with a ten percent annual depreciation rate. This modest depreciation rate is intended to capture both short- and long-term effects of democracy on human development. (This model replicates one of the models from Column 13, Table 1, while incorporating the above-discussed measurement errors.) The estimated relationship is highly significant ($p<0.01$), and – as discussed in the paper – substantially large.

Model 2 offers the conventional "level" measure of MPI (no adjustment for historical stock). Results are robust. However, additional tests show that the level measure is less robust than the stock measure. Table C7 offers ten such models (re-running all tests from Table C1), scrutinizing the thesis that MPI's effect on infant mortality might be proximal rather than distal. Results suggest that the proximal relationship is quite strong (clearing the 90% confidence test in all but one model), but not as robust as the distal relationship. This corroborates our hypothesis that the full impact of political institutions on human development takes time to materialize. This is also borne out when comparing the substantive effects of the various models. The predicted effect on IMR from Model 2 for our hypothetical country (introduced in the paper for Model 13, Table 1) with $1000 GDP per capita (and mean on the country- and year-fixed effects) of going from 0 on MPI in t-1 to 1 on MPI in t (using 10-year panels) is a 29% drop in IMR, from 73 to 52. This compares to a 32% drop following such a democratization experience using the stock measure (Model 1).

Stock has an important theoretical meaning for institutional variables such as democracy, but not for variables whose impact on human development is expected to be proximal. However, to ensure that democracy stock is not registering a proxy effect, we calculate stock for our all-purpose control variable, per capita GDP, and then replicate the set of tests contained in Table C1. Results from this exercise, shown in Table C9, reveal that a stock measure of economic development bears no empirical relationship to IMR and that the inclusion of this factor does not disturb the relationship of MPI to IMR.

Table C1: MPI and Mortality

Outcome	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	OLS	OLS	FD, RE	OLS	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	Imputed	Full	10-yr panel	1960-	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
MPI		-0.314***											
		(0.090)											
MPI stock (10%)	-0.537***		-0.539***	-0.107*	-0.267***	-0.804***	-0.394***	-0.964***		-0.286*	-0.480***	-0.465***	-0.197***
	(0.125)		(0.120)	(0.060)	(0.082)	(0.136)	(0.122)	(0.127)		(0.161)	(0.150)	(0.146)	(0.073)
MPI stock (10%), FD					-0.071***								
					(0.023)								
MPI stock (10%), *T*-30									-0.316**				
									(0.125)				
GDPpc (ln)	-0.369***	-0.397***	-0.357***	-0.047	-0.141***		-0.350***	-0.456***	-0.373***	-0.217	-0.312***	-0.378***	-0.199***
	(0.078)	(0.079)	(0.075)	(0.030)	(0.040)		(0.055)	(0.043)	(0.084)	(0.149)	(0.063)	(0.082)	(0.041)
Urbanization							0.235						
							(0.311)						
Fertility (ln)							0.500***						
							(0.087)						
Growth							0.002						
							(0.004)						
Internal conflict							0.057						
							(0.040)						
External conflict							0.022						
							(0.060)						
Corruption stock (10%)							0.057***						
							(0.019)						
IMR^{t-1}				0.805***						0.690***			
				(0.029)						(0.164)			
Year cubed			✓										
Decade FE	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓
Time trend										✓			
Countries	154	154	154	154	153	168	108	170	133	154	154	155	154
Decades	11	11	11	11	10	11	11	11	9	11	5	9	11
Obs	993	993	993	923	839	1132	751	1578	885	800	659	841	1090
R2 (within)	(0.907)	(0.904)	(.905)	(0.965)	(0.215)	(0.885)	(0.942)		(0.894)		(0.837)	(0.878)	(0.887)
MPI, as above	-0.571***	-0.233***	-0.582***	-0.015**	-0.195***	-0.730***	-0.384***	-1.657***	-0.338***	-0.080**	-0.456***	-0.460***	-0.184***
(annual data)	(0.127)	(0.072)	(0.124)	(0.007)	(0.049)	(0.129)	(0.135)	(0.110)	(0.124)	(0.040)	(0.136)	(0.131)	(0.066)

Outcomes: IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. *Estimators:* OLS (ordinary least squares), FD (first-difference), RE (random effects). All models incorporate measurement error for MPI based on posteriors produced by the V-Dem measurement model. Robust standard errors clustered by country. ***p<.01 **p<.05 *p<.10

In Model 3, we return to the ten percent depreciation rate, this time using linear, squared and cubic trend variables to control for temporal effects. This offers a different way of modeling time effects, one that is more efficient and allows for a more natural interpretation of the coefficient for the variable of theoretical interest – which is almost unchanged from the benchmark.

In Model 4, we return to the ten percent depreciation rate, this time including a lagged dependent variable as an additional regressor. This model estimates how MPI stock at *t*-1 affects changes in infant mortality from *t*-1 to *t*. While the high coefficient (0.81) of the lagged dependent variable confirms the highly trended nature of IMR, higher MPI stock is still associated with reductions in mortality, and the coefficient is significant at 10%.

In Model 5, we adopt a first-difference specification with a random effects estimator in order to isolate the short- and long-term effects of MPI. Here, the dependent variable is measured as a change from *t*-1 to *t*. On the righthand side, lagged MPI stock (capturing distal effects) is included alongside the first-difference of MPI stock (capturing proximate effects) and GDP per capita. The relationship between MPI and IMR is highly significant in both cases, indicating the presence of both short- and long-term effects.

In Model 6, we remove per capita GDP from the model, leaving a bivariate specification in which IMR is regressed on MPI along with year and country fixed effects. The estimated impact of a change in MPI is enhanced, relative to the benchmark. If one is willing to believe that electoral democracy has a (positive) effect on GDP per capita growth (e.g., Acemoglu et al., 2014; Gerring et al., 2005), this model may be regarded as providing an estimate of the *total* effect of MPI on IMR. Because this is a contentious claim taking us beyond the scope of the present study, and because income may simultaneously affect regime type, we revert to the standard specification in subsequent tests.

In Model 7, we add several covariates to the benchmark that might be expected to affect infant mortality and perhaps MPI. These potential confounders include urbanization, fertility, GDP per capita growth, internal conflict, external conflict, and corruption. Although the sample is diminished, the coefficient estimate for MPI is comparable to the benchmark. The inclusion of indices measuring conflict and corruption is especially noteworthy, as it suggests that MPI is not serving as a proxy for state capacity. (The inclusion of other measures of state capacity from V-Dem confirms this result, as none of these covariates mitigates the estimated effect of MPI on IMR.) Our results are also robust when other plausible covariates such as regime duration, are added (see Table C11). We disregard these covariates in other models because they depress the sample and, more importantly, risk introducing post-treatment bias.

Sample bias is a potential problem when units are not chosen randomly from a known universe, when that universe cannot be represented in its entirety, and when missing data is not missing at random (see Ross, 2006). In particular, we must be wary of the possibility that data for democracy and IMR might be missing for poorly performing countries, low-income countries, and non-democracies (Halperin et al., 2005). To alleviate this concern, we impute missing data using the Amelia II software, which models the cross-section time-series structure of our data (Honaker, King, & Blackwell, 2011). The resulting datasets include 170 countries observed across 11 decades – or fewer, if the country was not independent throughout the entire 1900-2016 period – producing over 1,500 observations. Model 8 shows results of our benchmark model, with coefficients and standard errors accounting for variability across twenty imputed datasets following rules by Rubin (1987). The estimated coefficient for MPI increases in size relative to the unimputed benchmark, suggesting that our benchmark model may produce conservative estimates.

Another problem of causal identification concerns possible endogeneity between MPI and IMR. One approach to this problem utilizes time to "exogenize" the regressor of interest. In our benchmark, right-side variables are lagged one period behind the outcome. In Model 9, we take this

approach to an extreme, lagging MPI by three decades (*t*-30), which should offer more assurance against *X*/*Y* circularity and simultaneity (an unmeasured confounder that affects both *X* and *Y*). The estimated coefficient is diminished relative to the benchmark, as one might expect, but remains sizeable and significant.

In Model 10, we enlist a dynamic panel model, system generalized method of moments (GMM), developed explicitly for studying sluggish variables (Blundell & Bond, 1998). In order to comply with the standard recommendation (fewer instruments than cross-section units) we restrict the number of lags used for instrumentation to three (third to fifth lag). The model treats MPI as endogenous, and, in contrast with several alternative specifications (e.g., also modelling GDP as exogenous, or using 1-year panels) that we tried, it performs well on all relevant specification tests. The Hansen J-test p-value is .20 and the Ar(2)-test p-value is .16. This suggests that Model 11 yields a consistent estimate of the relationship between MPI and IMR. The GMM model corroborates our main result, as MPI is significant at 10% (t=1.8). While the coefficient on MPI suggests that the short-term effect is rather modest in size, the estimated long-term impact is sizeable. The predicted long-term impact of MPI on IMR – calculated as $\beta_{MPI}/(1-\beta_{lagged\ DV})$ – is roughly -0.92, an estimate that is even larger than that obtained from the benchmark. We remark that an alternative specification including also a second lag of the dependent variable performs even better, both in terms of the p-value of the MPI coefficient (p=0.05) as well as on the AR(X) specification tests (AR(2) p-value of 0.50 and AR(3) p-value of 0.39, whereas the Hansen J p-value is slightly lower at 0.15). Hence, also in this specification with two lags of ln IMR included as regressor, and which further assuages concerns that residual autocorrelation is driving results, we find a fairly clear relationship between MPI stock and ln IMR.

The final models in Table C1 focus on alternate mortality-based outcome measures, introduced in Section II. Model 11 employs a measure of IMR drawn from the World Development Indicators (WDI) (World Bank 2013), an alternate data source that is highly correlated with the Gapminder dataset but more limited in temporal coverage. This model, restricted to the post-1960 period, also serves to mitigate concerns about poor data quality earlier in the twentieth century. Model 12 adopts the child mortality rate ("CMR") as an outcome, transformed by the natural logarithm. Model 13 adopts Life expectancy ("LE"), a summary measure of mortality rates across the lifespan, with a transformed index (as described above). MPI is consistently associated with lower mortality.

To mitigate concerns about autocorrelation influencing standard errors, we have used decade-years rather than country-years as the principal units of analysis. (However, results are robust to using country-years, see Table C10.) However, one might still worry about problems of serial correlation caused by sluggish variables and artificially deflated standard errors (which we have attempted to further counteract with clustered errors). This is borne out by tests of first-order autocorrelation for the benchmark model (and even the OLS Model 5 including a lagged dependent variable as covariate), using the xtserial package in STATA, leading us to reject the hypothesis of no autocorrelation. Hence, the presence of autocorrelation is still a caveat to note when interpreting the results from our benchmark specification, despite the result being robust in the first-difference (Model 5) and GMM (Model 10) specifications, which should be even better suited than the benchmark for addressing this issue. Nonetheless, we note that another form of dependence that could influence results is un-modelled spatial correlation, e.g., due to spill-over effects produced by the diffusion of disease or life-saving technologies. Reassuringly, the results are robust when employing Driscoll-Kraay errors that account for cross-sectional dependence across units (Table C8).

Several additional robustness tests are presented in Appendix D. In Table C2, we conduct restricted-sample tests in order to gauge sensitivity to the exclusion of particular regions of the

world. The relationship between MPI and IMR is robust, though estimates vary as the sample changes, as one might expect.

In Table C3, we conduct tests of functional form. Recall that IMR is transformed by the natural logarithm, reflecting a theoretical expectation that mortality is more elastic at higher rates. This is in keeping with (a) the general downward trend in mortality throughout the modern world, (b) the left-bounded nature of IMR, and (c) a well-established tradition by which right-skewed variables (and IMR in particular) are logged. However, other transformations are possible. Models 1-2 adopt a square root transformation, which deals with the skewed distribution of IMR but is difficult to interpret. Another approach retains the linear (raw) format of IMR while adopting the Tobit estimator designed to handle skewed distributions (Long & Freese, 2014). Results, shown in Model 3, are robust. (Yet, one must be somewhat skeptical of these results given that we have dispensed with country fixed-effects.) In Model 4, we examine the independent variable of interest, MPI (discounted at 10% annually). The multiplicative aggregation rule, when combined with components that recognize a zero score, truncates this index at zero, generating a right-skewed distribution. Conceivably, results for our benchmark model may reflect a binary distinction – between country-years coded 0 and country-years receiving a positive score. Accordingly, in Model 4, we exclude observations for which MPI is equal to zero. All these tests of functional form corroborate previously reported findings.

In Table C4, we explore possible non-linearities in the relationship between MPI and human development, replicating models in Table D3 with an additional quadratic term – MPI stock $(10\%)^2$. The squared term is often (though not always) statistically significant. However, the sign of the coefficient varies. In sum, there is minimal support for non-monotonic effects.

In Table C5, we explore an instrumental-variable approach. Following recent work on the economic effects of democracy (e.g., Acemoglu et al., 2014), we look to processes of institutional diffusion for instruments that may affect the assignment of the treatment (electoral democracy) but not the outcome (IMR) – except as mediated by the theoretical variable of interest (MPI) – thus satisfying the exclusion restriction. That is, we want to capture exogenous variation in electoral democracy stemming from foreign institutional variation as transmitted through various regional and global spill-over effects. Simultaneously, we want to ensure that results are not contaminated by any direct link from the institutions of foreign countries to domestic human development outcomes. Thus, we employ the WAVE instrument of democracy from Knutsen (2011), as well as regional and global averages of MPI, to tap exogenous variation in domestic MPI stemming from international or regional trends. These are tested separately, and together, in Models 1-5, all of which employ MPI stock (10%). In Models 6 and 7, we apply the same technique – all instruments combined – using MPI level. In Models 4 and 6, we implement an additional safeguard to ensure that the exclusion restriction holds, controlling for regional and global averages of human development. These controls should block out other channels through which the instruments (democracy abroad) are linked with the outcome (human development domestically), such as the potential causal channel: *Democracy abroad → Human development abroad →Human development domestically.* (This channel could arise from spillover effects on mortality related to pandemics spreading across borders.)

Tests with a two-stage least squares estimator, shown in Table C5, follow our benchmark model, with country and year fixed-effects and clustered standard errors. With the exception of analysis only using WAVE as instrument, these tests yield highly significant and substantively large effects. Weak-instrument test F-values show that the instruments are very strong, and, for the most part, Hansen J-test p-values suggest that the exclusion restriction cannot be rejected. While not entirely robust, the models in Table C5 thus provide fairly strong corroborating evidence that improvements in MPI (both level and stock) reduce infant mortality rates.

In Table C6, we explore possible non-linearities in per capita GDP (logged). Model 1 reproduces the benchmark model as a baseline for comparison. Model 2 introduces a polynomial – $GDPpc^2$ – to capture a quadratic relationship, whereas Model 3 introduces another polynomial – $GDPpc^3$ – to capture a cubed relationship. MPI stock (10%), though somewhat attenuated in the latter specifications, remains robust.

Table C2: Restricted Sample Tests

Excluded region	**E. Europe, Post-Soviet**		**Latin America**		**MENA**		**Sub-Saharan Africa**		**W. Europe, No. America**		**East Asia**		**Southeast Asia**		**South Asia**		**Caribbean**	
Sample	1900-	1940-	1900-	1940-	1900-	1940-	1900-	1940-	1900-	1940-	1900-	1940-	1900-	1940-	1900-	1940-	1900-	1940-
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
MPI stock (10%)	-0.546 (0.142) ***	-0.439 (0.141) ***	-0.532 (0.133) ***	-0.430 (0.150) ***	-0.656 (0.119) ***	-0.640 (0.120) ***	-0.441 (0.123) ***	-0.312 (0.124) **	-0.271 (0.189)	-0.320 (0.150) **	-0.511 (0.123) ***	-0.459 (0.132) ***	-0.497 (0.121) ***	-0.461 (0.131) ***	-0.585 (0.114) ***	-0.491 (0.126) ***	-0.552 (0.119) ***	-0.498 (0.122) ***
GDPpc (ln)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Decade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Countries	126	126	134	134	137	137	107	107	134	134	148	148	146	146	148	148	152	152
Decades	11	7	11	7	11	7	11	7	11	7	11	7	11	7	11	7	11	7
Obs	878	773	837	730	894	769	733	608	782	728	953	831	940	819	951	831	976	855
R2 within	0.901	0.867	0.906	0.864	0.922	0.893	0.936	0.925	0.859	0.842	0.907	0.872	0.909	0.870	0.912	0.874	0.908	0.874

Outcome: infant mortality rate (ln). *Units of analysis:* country-decades. Right-side variables measured at T-1. *Estimator:* ordinary least squares regression, standard errors clustered by country. *** p<.01 **p<.05 *p<.10 Each model excludes a region of the world, as noted.

Table C3: Tests of Functional Form

Outcome	$\sqrt{Y}$	$\sqrt{Y}$	*Y*	*Y(ln)*
Sample	Full	Full	Full	MPI>0
Estimator	OLS	FD, RE	Tobit	OLS
Model	**1**	**2**	**3**	**4**
MPI stock (10%)	-0.202***	-0.090***	-15.136***	-0.486***
	(0.037)	(0.025)	(5.806)	(0.112)
GDPpc (ln)	✓	✓	✓	✓
Decade FE	✓	✓	✓	✓
Country FE	✓			✓
Countries	154	153	154	146
Decades	11	10	11	11
Obs	993	839	1005	839
R2 within	0.890	0.351	0.105	0.929

Outcome (Y): infant mortality rate. *Units of analysis:* country-years. Right-side variables measured at T-1. *Estimators:* OLS (ordinary least squares), FD (first-difference), RE (random effects). Standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C4: Possible Non-linearities in MPI stock

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	FD, RE	OLS	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Imputed	Full	10-yr panel	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
MPI stock (10%)	0.071	-0.354***	-0.275**	0.146	0.369	-0.296		-0.467**	0.014	-0.292	0.152
	(0.245)	(0.113)	(0.117)	(0.248)	(0.244)	(0.249)		(0.201)	(0.233)	(0.224)	(0.144)
MPI stock (10%)²	-0.719***	0.314***	-0.084	-1.102***	-0.925***	-0.739***		0.339	-0.645**	-0.214	-0.408**
	(0.241)	(0.117)	(0.134)	(0.240)	(0.247)	(0.277)		(0.393)	(0.262)	(0.258)	(0.160)
MPI stock (10%), *T*-30							-0.154				
							(0.365)				
MPI stock (10%)², *T*-30							-0.201				
							(0.368)				
GDPpc (ln)	-0.341***	-0.048	-0.150***		-0.029	-0.451***	-0.374***	-0.239	-0.295***	-0.367***	-0.185***
	(0.076)	(0.030)	(0.040)		(0.315)	(0.048)	(0.081)	(0.172)	(0.067)	(0.087)	(0.042)
Urbanization					0.584***						
					(0.082)						
Fertility (ln)					0.001						
					(0.004)						
Growth					0.046						
					(0.036)						
Internal conflict					0.012						
					(0.048)						
External conflict					0.049***						
					(0.017)						
Corruption (10%)					-0.029						
					(0.315)						
Y		0.829***						0.709***			
		(0.030)						(0.150)			
Decade FE	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓		✓	✓	✓	✓		✓	✓	✓
Time trend								✓			
Countries	154	154	153	168	108	170	133	154	154	155	154
Decades	11	11	10	11	11	11	9	10	5	9	11
Obs	993	923	839	1132	751	1666	879	800	659	841	1088
R2 (within)	(0.910)	(0.965)	0.206	(0.891)	(0.946)	(0.915)	(0.893)		(0.840)	(0.878)	(0.889)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C5: Instrumental Variable Tests

Instruments	WAVE	Regional mean	WAVE Regional mean	WAVE Regional mean Global mean	WAVE Regional mean Global mean	WAVE Regional mean Global mean	WAVE Regional mean Global mean
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**
MPI						-0.347***	-0.460***
						(0.098)	(0.111)
MPI stock (10%)	-0.638	-1.223***	-0.651***	-0.273***	-0.579***		
	(0.454)	(0.355)	(0.205)	(0.103)	(0.121)		
GDPpc (ln)	✓	✓	✓	✓	✓	✓	✓
Regional avg. dep. var.				✓		✓	
Global avg. dep. var.				✓		✓	
Year fixed effects	✓	✓	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Countries	149	150	146	146	146	146	146
Decades	11	11	11	11	11	11	11
Obs	909	962	880	858	880	858	880
Kleibergen-Paap F	13.15	76.31	71.01	131.9	121.1	91.56	74.85
Hansen J-test p-value	-	-	0.821	0.0503	0.862	0.0968	0.960

Two-stage least squares regression analyses with various instruments, as explained in the text. Second-stage results only (first-stage results available upon request). *Outcome:* infant mortality rate. *Units of analysis:* country-years. Right-side variables measured at T-1. Robust standard errors clustered by country. Instruments: WAVE is drawn from Knutsen (2011), and codes whether the last regime change in a country, as registered by Polity IV and its regime duration variable, occurred within or outside on of Huntington's (1991) "reverse waves of democratization". The regional and global mean instruments are calculated for MPI level, for each year, and the country in question (for which we instrument MPI stock/level scores) is always exempted when we calculate the regional and global averages. In other words, we calculate the average MPI scores for all other countries in the region/globally that have data in that particular year. *** p<.01 **p<.05 *p<.10

Table C6: Possible Non-linearities in GDP

Model	**1**	**2**	**3**
MPI stock (10%)	-0.530***	-0.407***	-0.408***
	(0.119)	(0.136)	(0.124)
GDPpc (ln)	-0.365***	0.543	11.916***
	(0.077)	(0.448)	(4.113)
GDPpc (ln)^2		-0.056*	-1.472***
		(0.030)	(0.506)
GDPpc (ln)^3			0.058***
			(0.021)
Decade FE	✓	✓	✓
Country FE	✓	✓	✓
Countries	154	154	154
Decades	11	11	11
Obs	993	993	993
R2 within	0.907	0.910	0.915

Outcome: infant mortality rate (ln). *Units of analysis:* country-years. Right-side variables measured at T-1. *Estimator:* ordinary least squares. Standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C7: Additional Tests of MPI level

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	FD, RE	OLS	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Imputed	Full	10-yr panel	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
MPI	-0.320***	-0.128***	-0.129***	-0.505***	-0.148*	-0.625***		-0.045	-0.237**	-0.277***	-0.094*
	(0.089)	(0.049)	(0.044)	(0.103)	(0.087)	(0.096)		(0.103)	(0.097)	(0.093)	(0.050)
MPI, *T*-30							-0.313***				
							(0.114)				
GDPpc (ln)	-0.394***	-0.045	-0.162***		-0.374***	-0.505***	-0.374***	-0.216**	-0.333***	-0.400***	-0.211***
	(0.077)	(0.030)	(0.042)		(0.056)	(0.045)	(0.079)	(0.084)	(0.066)	(0.078)	(0.042)
Urbanization					0.343						
					(0.317)						
Fertility (ln)					0.490***						
					(0.091)						
Growth					0.003						
					(0.004)						
Internal conflict					0.066						
					(0.041)						
External conflict					0.024						
					(0.058)						
Corruption stock (10%)					0.068***						
					(0.019)						
Y^{t-1}		0.808***						0.870***			
		(0.026)						(0.164)			
Decade FE	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓		✓	✓	✓	✓		✓	✓	✓
Time trend								✓			
Countries	154	154	153	168	108	170	133	154	154	155	154
Decades	11	11	10	11	11	11	9	10	5	9	11
Obs	993	923	839	1132	751	1666	879	800	659	841	1088
R2 (within)	(0.904)	(0.965)	0.186	(0.879)	(0.940)	(0.912)	(0.894)		(0.832)	(0.876)	(0.885)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-years. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C8: Tests of MPI using Driscoll-Kraay Standard Errors

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	LE
Estimator	OLS	OLS	OLS	OLS	FD, pooled OLS	OLS	OLS	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
MPI			-0.320***							
			(0.058)							
MPI stock (1%)		-0.741***								
		(0.144)								
MPI stock (10%)	-0.530***			-0.106*	-0.361***	-0.778***	-0.384***		-0.467***	-0.194***
	(0.070)			(0.052)	(0.054)	(0.059)	(0.060)		(0.035)	(0.048)
MPI stock (10%), *T*-30								-0.323***		
								(0.083)		
GDPpc (ln)	-0.365***	-0.369***	-0.394***	-0.046***	-0.186***		-0.351***	-0.376***	-0.315***	-0.198***
	(0.026)	(0.028)	(0.027)	(0.010)	(0.055)		(0.031)	(0.030)	(0.042)	(0.021)
Urbanization							0.212			
							(0.219)			
Fertility (ln)							0.495***			
							(0.046)			
Growth							0.002			
							(0.003)			
Internal conflict							0.059			
							(0.044)			
External conflict							0.022			
							(0.049)			
Corruption stock (10%)							0.056***			
							(0.011)			
Y^{t-1}				0.806***						
				(0.042)						
Decade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓		✓	✓	✓	✓	✓
Time trend										
Countries	154	154	154	154	153	168	108	133	154	154
Decades	11	11	11	11	10	11	11	9	5	11
Obs	993	993	993	923	839	1132	751	879	659	1088
R2 (within)	(0.907)	(0.906)	(0.904)	(0.965)	0.209	(0.885)	(0.942)	(0.893)	(0.837)	(0.886)

Outcomes (*Y*): IMR (infant mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-years. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference). The benchmark using Driscoll-Kraay standard errors and CMR as dependent variable is not reported, since the variance matrix is reported by STATA to be "nonsymmetric or highly singular". Due to the incompatibility of Random Effects and this error specification, we employ pooled OLS in the first-difference specification (Model 5). *** p<.01 **p<.05 *p<.10

Table C9: Tests of MPI and GDPpc (ln) stock (10%)

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	OLS	OLS	FD, RE	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	10-yr panel	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
MPI			-0.480***								
			(0.107)								
MPI stock (1%)		-1.120***									
		(0.217)									
MPI stock (10%)	-0.774***			-0.120**	-0.367***	-0.507***		-0.129	-0.618***	-0.725***	-0.325***
	(0.136)			(0.058)	(0.076)	(0.131)		(0.200)	(0.156)	(0.153)	(0.081)
MPI stock (10%), *T*-30							-0.520***				
							(0.143)				
GDPpc (ln) stock (10%)	0.003	0.002	0.003	0.002*	-0.001	-0.002	0.007***	-0.014**	-0.000	0.005**	0.002*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.007)	(0.002)	(0.002)	(0.001)
Urbanization						0.223					
						(0.346)					
Fertility (ln)						0.707***					
						(0.089)					
Growth						0.001					
						(0.004)					
Internal Conflict						0.090**					
						(0.042)					
External Conflict						0.071					
						(0.072)					
Corruption stock (10%)						0.074***					
						(0.023)					
Y^{t-1}				0.825***				0.784***			
				(0.025)				(0.116)			
Decade FE	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓	✓	✓		✓	✓		✓	✓	✓
Time trend								✓			
Countries	154	154	154	154	153	108	133	154	154	155	154
Years	11	11	11	11	10	11	9	10	5	9	11
Obs	993	993	993	923	839	751	879	800	659	841	1088
R2 (within)	(0.890)	(0.889)	(0.884)	(0.965)	0.163	(0.932)	(0.875)		(0.808)	(0.855)	(0.872)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C10: Tests of MPI using Annual Data

Outcome	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	OLS	OLS	FD, RE	OLS	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	Imputed	Full	5-yr panel	1960-	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
MPI		-0.233***											
		(0.072)											
MPI stock (10%)	-0.571***		-0.582***	-0.015**	-0.010***	-0.730***	-0.384***	-1.657***		-0.080**	-0.456***	-0.460***	-0.184***
	(0.127)		(0.124)	(0.007)	(0.003)	(0.129)	(0.135)	(0.110)		(0.040)	(0.136)	(0.131)	(0.066)
MPI stock (10%), FD					-0.195***								
					(0.049)								
MPI stock (10%), *T*-30									-0.338***				
									(0.124)				
GDPpc (ln)	-0.356***	-0.390***	-0.354***	-0.004	-0.057***		-0.364***	-0.387***	-0.382***	-0.059***	-0.299***	-0.341***	-0.179***
	(0.066)	(0.066)	(0.067)	(0.003)	(0.015)		(0.057)	(0.028)	(0.076)	(0.021)	(0.063)	(0.074)	(0.034)
Urbanization							0.297						
							(0.310)						
Fertility (ln)							0.421***						
							(0.085)						
Growth							0.002***						
							(0.001)						
Internal Conflict							0.024						
							(0.027)						
External Conflict							0.018						
							(0.032)						
Corruption stock (10%)							0.056***						
							(0.020)						
IMR^{t-1}				0.977***						0.925***			
				(0.004)						(0.026)			
Year cubed			✓										
Year FE	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓
Time trend										✓			
Countries	154	154	154	154	154	168	108	178	133	154	154	155	154
Years	111	111	111	111	110	112	110	116	81	22	51	81	111
Obs	9094	9094	9094	9004	8936	10335	6872	16771	7639	1743	6370	7382	10459
R2 (within)	(0.907)	(0.902)	(.904)	(0.996)	0.113	(0.890)	(0.936)		(0.887)		(0.842)	(0.858)	(0.896)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-years. Right-side variables measured at *T*-1 unless otherwise noted. All models incorporate measurement error for MPI based on posteriors produced by the V-Dem measurement model. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C11: Tests of MPI and Political Durability

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**
MPI			-0.222***			-0.291***
			(0.076)			(0.084)
MPI stock (1%)		-0.813***			-1.051***	
		(0.195)			(0.215)	
MPI stock (10%)	-0.401***			-0.503***		
	(0.110)			(0.124)		
GDPpc (ln)	-0.312***	-0.297***	-0.328***			
	(0.054)	(0.054)	(0.057)			
GDPpc (ln) stock (10%)				-0.003	-0.004*	-0.004*
				(0.002)	(0.002)	(0.002)
Urbanization	0.435	0.245	0.551*	0.424	0.187	0.574*
	(0.277)	(0.285)	(0.283)	(0.311)	(0.322)	(0.317)
Fertility (ln)	0.457***	0.527***	0.448***	0.630***	0.713***	0.630***
	(0.083)	(0.077)	(0.087)	(0.087)	(0.081)	(0.092)
Growth	0.001	0.000	0.002	-0.001	-0.001	0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Internal Conflict	0.007	0.012	0.010	0.023	0.028	0.027
	(0.036)	(0.036)	(0.038)	(0.038)	(0.039)	(0.039)
External Conflict	0.014	0.015	0.011	0.067	0.065	0.067
	(0.050)	(0.052)	(0.052)	(0.059)	(0.060)	(0.062)
Corruption stock (10%)	0.045***	0.045***	0.054***	0.059***	0.059***	0.071***
	(0.017)	(0.016)	(0.016)	(0.019)	(0.018)	(0.018)
Political Durability	-0.000	0.001	-0.001	-0.002*	0.000	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Decade FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓		✓
Countries	108	108	108	108	108	108
Decades	11	11	11	11	11	11
Obs	711	711	711	711	711	711
R2	0.951	0.952	0.950	0.944	0.946	0.942

Outcome (*Y*): IMR (infant mortality rate, logged). Units of analysis: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. Estimator: OLS (ordinary least squares). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10 Durability drawn from the Polity IV dataset (Marshall, Gurr & Jaggers 2014).

Table C12: Tests of MPI not including Measurement Error

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	OLS	OLS	OLS	FD, RE	OLS	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	Full	Imputed	Full	10-yr panel	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
MPI		-0.320***												
		(0.089)												
MPI stock (10%)	-0.530***		-0.531***	-0.106*	-0.128**	-0.067***	-0.778***	-0.384***	-0.890***		-0.301**	-0.467***	-0.457***	-0.194***
	(0.119)		(0.118)	(0.056)	(0.063)	(0.022)	(0.130)	(0.120)	(0.116)		(0.154)	(0.146)	(0.143)	(0.068)
MPI stock (10%), FD						-0.265***								
						(0.078)								
MPI stock (10%), *T*-30										-0.323***				
										(0.121)				
GDPpc (ln)	-0.365***	-0.394***	-0.357***	-0.046	-0.047	-0.143***		-0.351***	-0.468***	-0.376***	-0.211	-0.315***	-0.376***	-0.198***
	(0.077)	(0.077)	(0.073)	(0.030)	(0.039)	(0.040)		(0.055)	(0.045)	(0.080)	(0.150)	(0.065)	(0.081)	(0.041)
Urbanization								0.212						
								(0.306)						
Fertility (ln)								0.495***						
								(0.087)						
Growth								0.002						
								(0.004)						
Internal Conflict								0.059						
								(0.039)						
External Conflict								0.022						
								(0.056)						
Corruption stock (10%)								0.056***						
								(0.019)						
Y^{t-1}				0.806***	0.868***						0.691***			
				(0.028)	(0.054)						(0.163)			
Y^{t-2}					-0.194**									
					(0.067)									
Y^{t-3}					-0.051									
					(0.067)									
Year cubed			✓											
Decade FE	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓
Time trend											✓			
Countries	154	154	154	154	152	153	168	108	170	133	154	154	155	154
Decades	11	11	11	11	9	10	11	11	11	9	10	5	9	11
Obs	993	993	993	923	667	839	1132	751	1666	879	800	659	841	1088
R2 (within)	(0.907)	(0.904)	(0.905)	(0.965)	(0.965)	0.216	(0.885)	(0.942)	(0.914)	(0.893)		(0.837)	(0.878)	(0.886)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C13: Democracy Indices using Data from 1960-present

Aggregation	***Composite***				***Empowerment***					***Binary***		***Multiplicative***	
Index	Polity2 (Marshall)	UDS (Pemstein)	Contes -tation (Miller)	Inclusive -ness (Miller)	Partici -pation (V-Dem)	Deliber -ation (V-Dem)	Female Power (V-Dem)	Civil Society (V-Dem)	Individual Liberty (V-Dem)	BMR (Boix)	BNR (Bernhard)	Contest.* Inclusive (Miller)	MPI (authors)
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1. Level	-0.164**	-0.465***	-0.164*	-0.140*	-0.101	-0.025	0.022	0.090	-0.062	-0.095*	-0.134**	-0.259**	-0.255***
	(0.078)	(0.163)	(0.086)	(0.083)	(0.180)	(0.094)	(0.213)	(0.114)	(0.100)	(0.055)	(0.051)	(0.112)	(0.095)
2. Stock (10%)	-0.118	-0.103	-0.184	0.265**	-0.098	-0.063	-0.041	-0.002	-0.128	-0.176*	-0.263***	-0.340**	-0.480***
	(0.141)	(0.239)	(0.149)	(0.121)	(0.215)	(0.154)	(0.195)	(0.174)	(0.145)	(0.095)	(0.088)	(0.172)	(0.138)
3. Stock (10%), Y^{t-1}	-0.155**	-0.157	-0.176**	0.087	-0.180*	-0.121	-0.100	-0.175**	-0.171**	-0.131***	-0.131***	-0.212**	-0.189**
	(0.068)	(0.117)	(0.072)	(0.069)	(0.102)	(0.074)	(0.102)	(0.081)	(0.075)	(0.049)	(0.046)	(0.088)	(0.078)
4. Stock (1%)	-0.278	-0.700*	-0.856*	0.590	-0.343	-0.549	-0.880	-0.200	-0.927**	-0.657**	-0.521***	-1.060**	-1.105***
	(0.476)	(0.368)	(0.459)	(0.537)	(0.610)	(0.542)	(0.613)	(0.522)	(0.445)	(0.305)	(0.174)	(0.413)	(0.289)
GDPpc (ln)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Decade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Countries	152	154	150	150	154	154	154	154	154	154	152	150	154
Decades	5	5	5	5	5	5	5	5	5	5	5	5	5
Obs (approx.)	669	680	664	663	695	695	667	695	695	677	669	663	689

Outcome: Infant mortality rate (ln). For each index, we conduct four tests: (1) level, (2) stock (10% annual depreciation rate), (3) stock (10% annual depreciation rate) with a lagged dependent variable, (4) stock (1% annual depreciation rate). *Units of analysis:* country-years. *FE:* fixed effects. All right-side variables measured at t-1. *Estimator:* ordinary least squares, standard errors clustered by country. ***p<.01 **p<.05 *p<.10

Table C14: Additional Tests of BMR

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	OLS	OLS	FD, RE	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	Full	10-yr panel	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
BMR		-0.100*										
		(0.054)										
BMR stock (10%)	-0.178**		-0.151*	-0.103***	-0.126**	-0.233**	-0.052		-0.086	-0.182*	-0.211**	-0.064
	(0.083)		(0.079)	(0.037)	(0.049)	(0.090)	(0.088)		(0.078)	(0.100)	(0.092)	(0.052)
BMR stock (10%), *T*-30								-0.075				
								(0.078)				
GDPpc (ln)	-0.404***	-0.411***	-0.399***	-0.068**	-0.186***		-0.363***	-0.342***	-0.213**	-0.349***	-0.428***	-0.209***
	(0.058)	(0.059)	(0.055)	(0.033)	(0.041)		(0.053)	(0.058)	(0.093)	(0.067)	(0.064)	(0.041)
Urbanization							0.460					
							(0.282)					
Fertility (ln)							0.414***					
							(0.090)					
Growth							0.003					
							(0.004)					
Internal Conflict							0.038					
							(0.043)					
External Conflict							0.022					
							(0.052)					
Corruption stock (10%)							0.071***					
							(0.017)					
Y^{t-1}				0.756***					0.622***			
				(0.030)					(0.122)			
Year cubed			✓									
Decade FE	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓
Time trend									✓			
Countries	154	154	154	154	151	168	108	132	154	154	154	154
Decades	11	11	11	11	10	11	11	9	10	5	8	11
Obs	927	927	927	875	772	1005	719	712	786	659	789	986
R2 (within)	(0.923)	(0.922)	(0.920)	(0.967)	0.190	(0.900)	(0.948)	(0.931)		(0.832)	(0.892)	(0.886)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Table C15: Additional Tests of BNR

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	OLS	OLS	FD, RE	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	Full	10-yr panel	Full	Full	Full
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
BNR		-0.117**										
		(0.049)										
BNR stock (10%)	-0.222***		-0.217***	-0.095**	-0.180***	-0.272***	-0.098		-0.098	-0.261***	-0.212**	-0.063
	(0.075)		(0.075)	(0.042)	(0.048)	(0.084)	(0.078)		(0.158)	(0.091)	(0.083)	(0.047)
BNR stock (10%), *T*-30								-0.068				
								(0.087)				
GDPpc (ln)	-0.380***	-0.392***	-0.391***	-0.052	-0.228***		-0.325***	-0.217***	-0.519*	-0.333***	-0.395***	-0.206***
	(0.062)	(0.061)	(0.060)	(0.034)	(0.048)		(0.054)	(0.059)	(0.270)	(0.066)	(0.066)	(0.044)
Urbanization							0.329					
							(0.324)					
Fertility (ln)							0.509***					
							(0.086)					
Growth							0.003					
							(0.005)					
Internal Conflict							-0.034					
							(0.042)					
External Conflict							-0.038					
							(0.066)					
Corruption stock (10%)							0.060***					
							(0.018)					
Y_{t-1}				0.737***					0.585*			
				(0.034)					(0.346)			
Year cubed			✓									
Decade FE	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓
Time trend									✓			
Countries	152	152	152	152	151	166	108	130	152	152	152	152
Decades	7	7	7	7	6	7	7	5	7	5	7	7
Obs	795	795	795	756	643	866	613	519	696	658	759	809
R2 (within)	(0.886)	(0.885)	(0.885)	(0.947)	0.193	(0.854)	(0.925)	(0.876)		(0.834)	(0.883)	(0.806)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

APPENDIX D: The Multiplicative Polyarchy Index (MPI)

Table D1: MPI level by Country in 2000

Country	MPI	Country	MPI	Country	MPI	Country	MPI
Sweden	0.950	Croatia	0.643	Kenya	0.151	Dem. Rep. of Congo	0.000
Norway	0.943	Senegal	0.637	Seychelles	0.147	Dem. Rep. of Vietnam	0.000
Denmark	0.940	Barbados	0.637	Malawi	0.145	Eritrea	0.000
Uruguay	0.940	Jamaica	0.629	Liberia	0.140	Fiji	0.000
France	0.939	Dominican Republic	0.618	Lebanon	0.132	Jordan	0.000
Switzerland	0.938	Botswana	0.603	Georgia	0.122	Kosovo	0.000
Germany	0.936	Mexico	0.588	Ivory Coast	0.114	Laos	0.000
Costa Rica	0.932	Indonesia	0.582	Cambodia	0.110	Lesotho	0.000
Czech Republic	0.919	Mongolia	0.568	Algeria	0.108	Libya	0.000
Estonia	0.915	São Tomé och Príncipe	0.560	Nepal	0.103	Morocco	0.000
Spain	0.915	Romania	0.526	Cen. African Rep.	0.088	North Korea	0.000
Ireland	0.909	Benin	0.522	Uganda	0.079	Oman	0.000
Greece	0.905	Namibia	0.506	Gambia	0.077	Pakistan	0.000
Australia	0.905	Philippines	0.495	Gabon	0.068	Qatar	0.000
Netherlands	0.901	Turkey	0.474	Mauritania	0.063	Republic of the Congo	0.000
Canada	0.898	Nicaragua	0.467	Zanzibar	0.052	Rwanda	0.000
Poland	0.898	Moldova	0.458	Togo	0.048	Saudi Arabia	0.000
United Kingdom	0.898	Venezuela	0.457	Chad	0.048	Sierra Leone	0.000
Belgium	0.898	Vanuatu	0.448	Yemen	0.046	Somalia	0.000
New Zealand	0.898	Honduras	0.430	Kazakhstan	0.044	Somaliland	0.000
Finland	0.893	Guyana	0.417	Cameroon	0.043	Swaziland	0.000
Iceland	0.892	Niger	0.401	Malaysia	0.043	Timor-Leste	0.000
Portugal	0.883	Paraguay	0.388	Djibouti	0.036		
Brazil	0.869	Guatemala	0.381	Belarus	0.036		
Slovakia	0.866	Colombia	0.372	Zimbabwe	0.036		
Austria	0.865	Burkina Faso	0.369	Kyrgyzstan	0.035		
Chile	0.865	Bangladesh	0.358	Guinea	0.034		
Italy	0.854	El Salvador	0.353	Iran	0.033		
Slovenia	0.850	Mali	0.346	Azerbaijan	0.028		
Latvia	0.846	Palestine/West Bank	0.344	Ethiopia	0.026		
Argentina	0.839	Sri Lanka	0.339	Tajikistan	0.021		
Japan	0.832	Madagascar	0.335	Kuwait	0.017		
South Korea	0.816	Thailand	0.311	Equatorial Guinea	0.008		
Lithuania	0.814	Zambia	0.298	Egypt	0.006		
United States	0.812	Solomon Islands	0.289	Tunisia	0.004		
Cape Verde	0.808	Macedonia	0.274	Sudan	0.002		
Mauritius	0.793	Serbia	0.263	Uzbekistan	0.001		
Hungary	0.781	Papua New Guinea	0.257	Maldives	0.001		
Suriname	0.772	Nigeria	0.251	Syria	0.001		
Cyprus	0.770	Russia	0.236	Iraq	0.000		
Israel	0.755	Guinea-Bissau	0.213	Turkmenistan	0.000		
Taiwan	0.751	Albania	0.207	Afghanistan	0.000		
Bolivia	0.711	Armenia	0.199	Angola	0.000		
South Africa	0.682	Mozambique	0.193	Bhutan	0.000		
Panama	0.669	Montenegro	0.192	Bosnia and Herz.	0.000		
Bulgaria	0.665	Haiti	0.176	Burma/Myanmar	0.000		
India	0.651	Tanzania	0.170	Burundi	0.000		
Trinidad and Tobago	0.650	Peru	0.163	China	0.000		
Ecuador	0.649	Singapore	0.161	Comoros	0.000		
Ghana	0.643	Ukraine	0.155	Cuba	0.000		

Table D2: Correlation Matrix of MPI Components

	Clean elections	Elected executive	Free association	Free expression	Suffrage
Clean elections	1				
Elected executive	0.653	1			
Free association	0.823	0.651	1		
Free expression	0.793	0.611	0.935	1	
Suffrage	0.537	0.643	0.470	0.432	1

N = 1,746

Disaggregation

To further probe our argument, we investigate alternative modes of disaggregation.

We begin with the components of MPI. Which elements of this composite index – Clean elections, Elected executive, Suffrage, Free association, or Free expression – is most important in driving the observed relationship between democracy and infant mortality?

One approach to this question is to exclude each component of the MPI, seriatim, and re-test the resulting index in the benchmark model. This exercise, displayed in Table D3, affirms several aspects of our argument. First, none of these exclusions compromise the core relationship between electoral democracy and infant mortality, which remains negative and significant ($p<.05$) in all models, offering assurance that our finding is not the product of a particular concatenation of components. Second, the estimated impact of the indices with excluded components is lower than the full index, conrroborating that the relationship between electoral democracy is enhanced when all components are present. Finally, a close look at the coefficients across all the models in Table D3 reveals that the exclusion of Clean Elections attenuates the performance of the index more than the exclusion of any other component, confirming our contention that – even within an index focused on electoral democracy – the most election-focused component has the strongest relationship to infant mortality.

Table D3: **Excluding Components of MPI**

	1	2	3	4	5	6
MPI stock (10%)	-0.530***					
	(0.119)					
…excluding Clean Elections		-0.241**				
		(0.116)				
…excluding Elected Executive			-0.520***			
			(0.126)			
...excluding Suffrage				-0.530***		
				(0.119)		
…excluding Free Association					-0.478***	
					(0.124)	
…excluding Free Expression						-0.435***
						(0.118)
GDPpc (ln)	-0.365***	-0.404***	-0.367***	-0.365***	-0.375***	-0.378***
	(0.077)	(0.076)	(0.077)	(0.077)	(0.077)	(0.076)
Decade FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Countries	154	154	154	154	154	154
Decades	11	11	11	11	11	11
Obs	993	993	993	993	993	993
R2 within	0.907	0.902	0.907	0.907	0.906	0.905

Outcome: Infant mortality rate (ln). *Units of analysis:* country-decades. Right-side variables measured at T-1. All democracy sub-indices measured as stock from 1900 (or first year for which data is available) with 10% annual depreciation. *Estimator:* ordinary least squares with country and year fixed effects, robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Another approach to this question is to construct stock variables, using our standard ten percent annual depreciation rate, for each ingredient of MPI.[1] We then regress IMR against each of

[1] The five components of the MPI are correlated with each other, though not as highly one might expect. Pearson's r correlations range from 0.39 to 0.93 (see Table C2). In a principal components factor analysis (not shown) the first

these sub-indices in our benchmark model, as shown in Table D4. Consistent with our priors, only the purely electoral component, Clean Elections, predicts lowered infant mortality. Other components of MPI, somewhat surprisingly, predict higher IMR (Elected Executive and Suffrage) or have no relationship to IMR (Free Association, Free Expression).

Table D4: **Disaggregating MPI**

	1	2	3	4	5
Clean Elections	-0.418***				
	(0.155)				
Elected Executive		0.015*			
		(0.008)			
Suffrage			0.038***		
			(0.010)		
Free Association				-0.010	
				(0.012)	
Free Expression					-0.002
					(0.013)
GDPpc (ln)	-0.408***	-0.454***	-0.406***	-0.430***	-0.444***
	(0.077)	(0.076)	(0.076)	(0.075)	(0.077)
Decade FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
Countries	154	154	154	154	154
Years	11	11	11	11	11
Obs	1003	1005	998	1001	1005
R2 within	0.904	0.902	0.906	0.901	0.901

Outcome: Infant mortality rate (ln). *Units of analysis:* country-years. Right-side variables measured at T-1. All democracy sub-indices measured as stock from 1900 (or first year for which data is available) with 10% annual depreciation. *Estimator:* ordinary least squares with country and year fixed effects, robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

One might be tempted to conclude that the impact of MPI on human development is entirely driven by one component of the index. To test this possibility, we replicate our battery of tests with Clean Elections as the variable of theoretical interest, as shown in Table D5. It turns out that the performance of Clean Elections is neither as strong nor as robust as the composite MPI index when subjected to the full set of empirical tests. Evidently, improvements in human development are not solely the product of clean elections. This corroborates our contention that the various ingredients of electoral democracy have mutually reinforcing effects. One cannot account for the relationship between electoral democracy and human development without an aggregation technique that acknowledges these interactive properties.

component explains roughly 70% of the variance.

Table D5: Additional Tests of Clean Elections

Outcome (Y)	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR	IMR(WDI)	CMR	LE
Estimator	OLS	OLS	OLS	OLS	FD, RE	OLS	OLS	OLS	Sys. GMM	OLS	OLS	OLS
Sample	Full	Full	Full	Full	Full	Full	Full	Full	5-yr panel	Full	Full	Full
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Clean Elections			-0.216**									
			(0.100)									
Clean Elections stock (1%)		-0.884***										
		(0.270)										
Clean Elections stock (10%)	-0.418***			-0.124*	-0.196***	-0.559***	-0.016		-0.146	-0.492***	-0.532***	-0.129*
	(0.155)			(0.064)	(0.076)	(0.142)	(0.161)		(0.099)	(0.147)	(0.150)	(0.072)
Clean Elections *stock* (10%), *T*-30								-0.177				
								(0.158)				
GDPpc (ln)	-0.408***	-0.394***	-0.427***	-0.054*	-0.150***		-0.389***	-0.424***	-0.217***	-0.323***	-0.403***	-0.214***
	(0.077)	(0.079)	(0.077)	(0.031)	(0.040)		(0.055)	(0.077)	(0.066)	(0.065)	(0.075)	(0.042)
Urbanization							0.265					
							(0.303)					
Fertility (ln)							0.474***					
							(0.093)					
Growth							0.004					
							(0.004)					
Internal Conflict							0.084*					
							(0.043)					
External Conflict							0.034					
							(0.056)					
Corruption stock (10%)							0.078***					
							(0.019)					
Y^{t-1}				0.807***					0.635***			
				(0.027)					(0.093)			
Decade FE	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Country FE	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓
Time trend									✓			
Countries	154	154	154	154	154	168	108	135	154	154	155	154
Decades	11	11	11	11	10	11	11	9	10	5	9	11
Obs	1003	1003	1003	931	849	1143	756	898	806	663	849	1102
R2 (within)	(0.904)	(0.905)	(0.902)	(0.965)	0.186	(0.874)	(0.940)	(0.892)		(0.839)	(0.881)	(0.886)

Outcomes (*Y*): IMR (infant mortality rate, logged), CMR (child mortality rate, logged), LE (life expectancy, reverse scale, logged). Units of analysis: country-decades. Right-side variables measured at *T*-1 unless otherwise noted. Estimators: OLS (ordinary least squares), FD (first-difference), RE (random effects). Robust standard errors clustered by country. *** p<.01 **p<.05 *p<.10

Figure D1: **Histograms of MPI stock (10%) and MPI level in 2000**

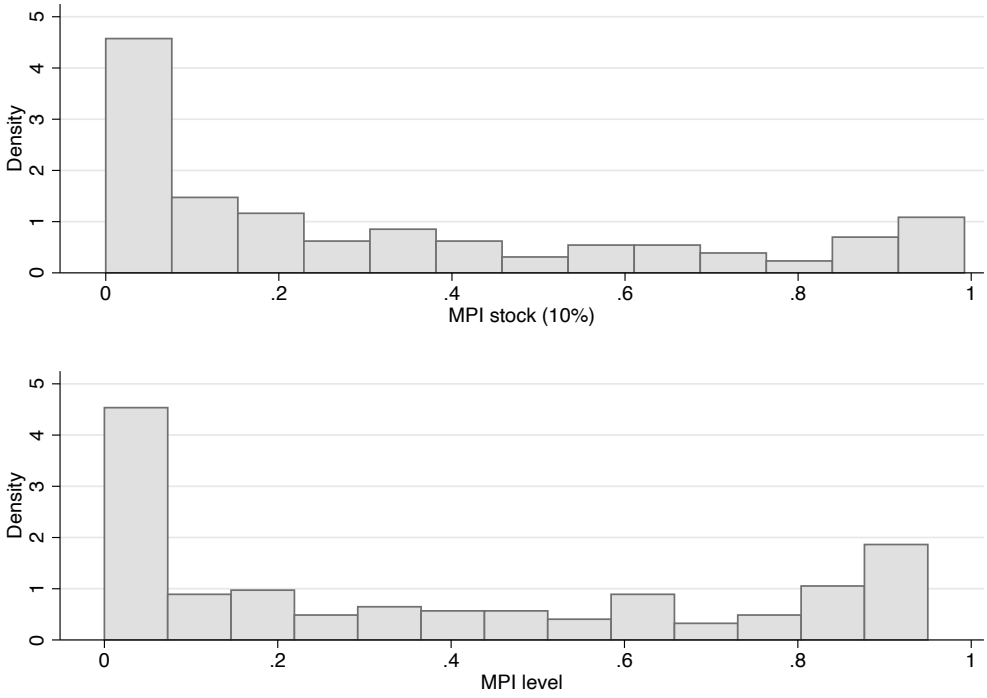
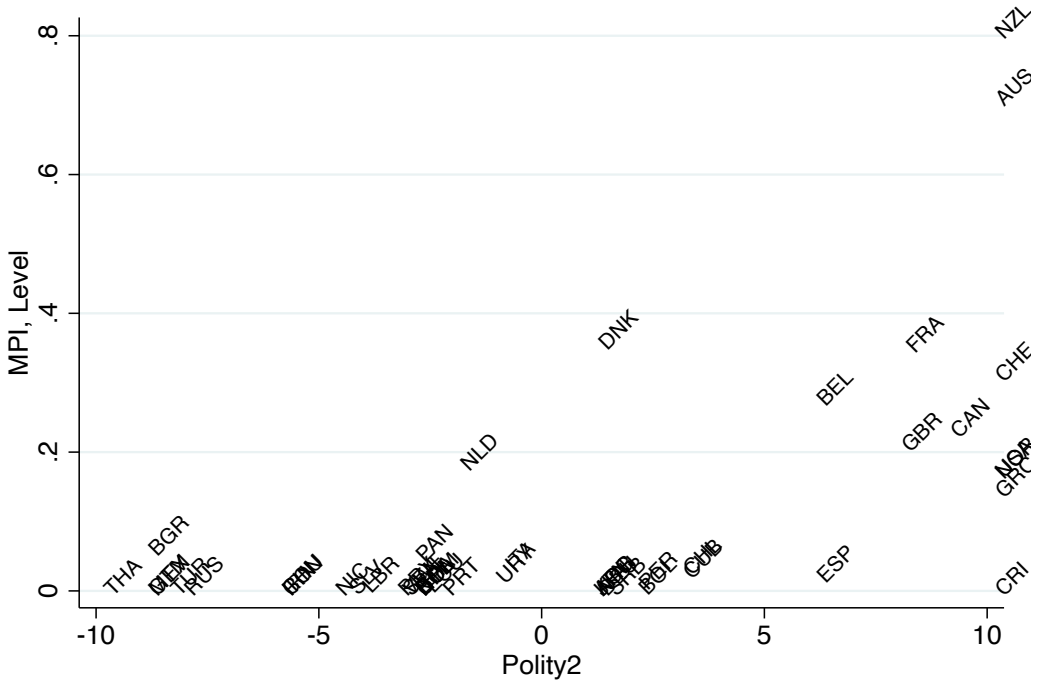
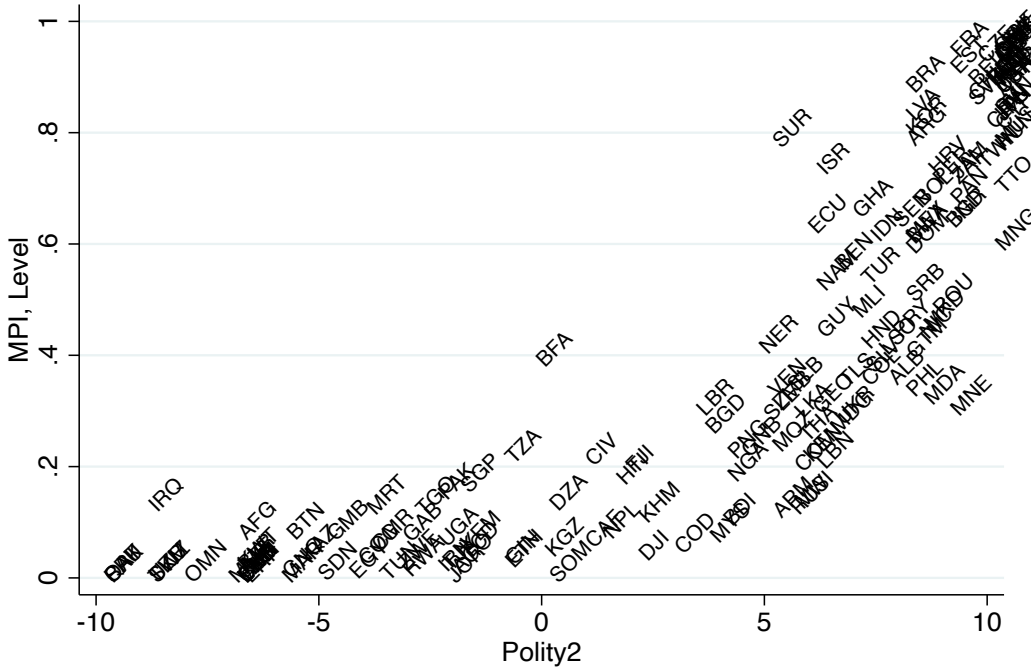


Figure D2: **MPI level and Polity2 in 1900**



Scatterplot of MPI (level) and Polity2 (Polity IV) in 1900. Pearson's r = 0.57. *N* = 45.

Figure D3: **MPI level and Polity2 in 2000**



Scatterplot of MPI (level) and Polity2 (Polity IV) in 2000. Pearson's r = 0.85. *N* = 157.

Table D6: Horse-race Tests

Alternate index	Polity2 (Marshall)	UDS (Pemstein)	Contes -tation (Miller)	Inclusive -ness (Miller)	Partici -pation (V-Dem)	Deliber -ation (V-Dem)	Female power (V-Dem)	Civil society (V-Dem)	Individual liberty (V-Dem)	BMR (Boix)	BNR (Bernhard)	Contest.* Inclusive (Miller)
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
MPI stock (10%)	-0.643*** (0.128)	-0.452*** (0.120)	-0.594*** (0.122)	-0.584*** (0.110)	-0.662*** (0.141)	-0.606*** (0.141)	-0.504*** (0.143)	-0.632*** (0.147)	-0.429*** (0.146)	-0.584*** (0.134)	-0.394*** (0.135)	-0.556*** (0.130)
Alternate index stock (1%)	0.353 (0.246)	-0.020 (0.231)	0.232 (0.234)	0.564* (0.303)	0.519 (0.322)	0.285 (0.329)	-0.157 (0.446)	0.354 (0.296)	-0.349 (0.327)	0.124 (0.195)	-0.096 (0.140)	0.103 (0.232)
GDPpc (ln)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Decade FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Countries	152	154	150	150	154	154	154	154	154	154	152	150
Decades	11	7	11	11	11	11	11	11	11	11	7	11
Obs	914	814	902	901	993	993	951	993	993	922	792	901
R2 within	0.929	0.893	0.929	0.930	0.908	0.907	0.917	0.907	0.907	0.928	0.889	0.929

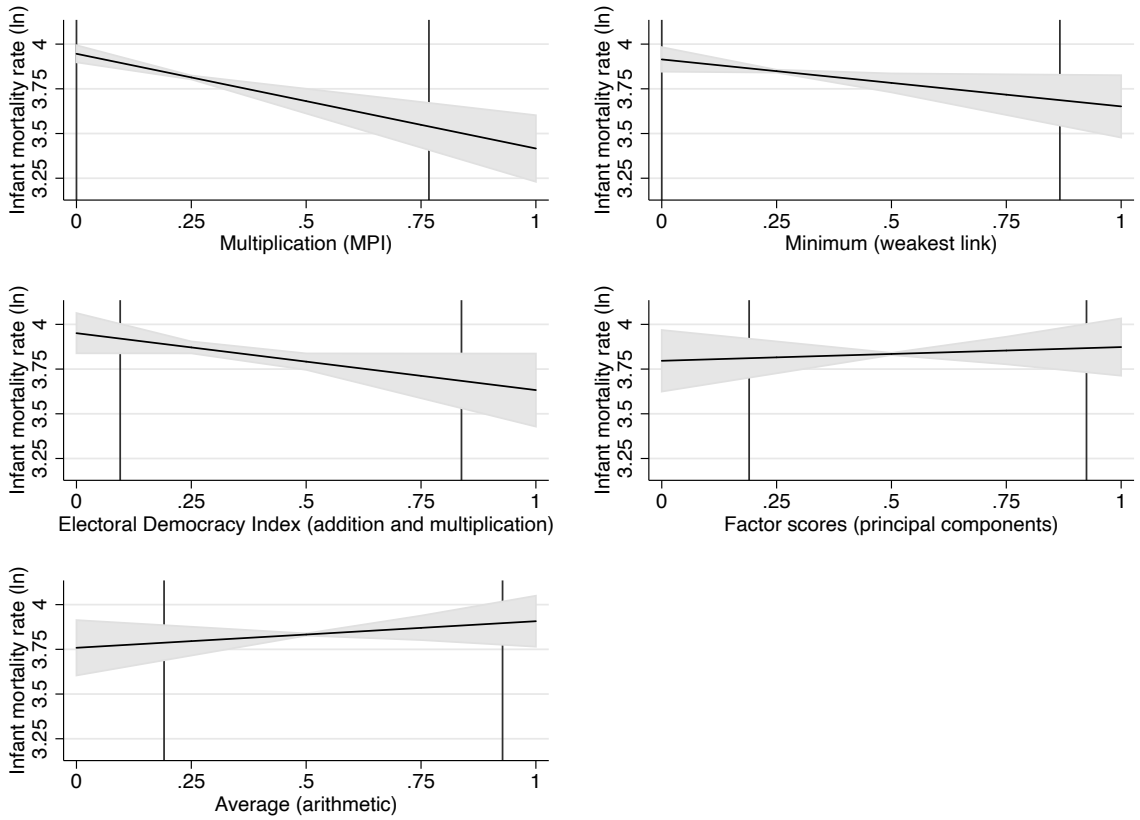
Outcome: infant mortality rate (ln). *Units:* country-decades. Right-side variables measured at T-1. *Estimator:* OLS. Standard errors clustered by country. ***p<.01 **p<.05 *p<.10

Table D7: Varying Aggregation Rules

Model	1	2	3	4	5
Multiplication	-0.530***				
(MPI)	(0.119)				
Minimum		-0.263**			
(weakest-link)		(0.124)			
Electoral Democracy Index			-0.318*		
(addition & multiplication)			(0.162)		
Factor scores				0.077	
(principal components)				(0.169)	
Average					0.148
(arithmetic)					(0.152)
GDPpc (ln)	-0.365***	-0.421***	-0.399***	-0.429***	-0.431***
	(0.077)	(0.077)	(0.077)	(0.076)	(0.076)
Decade FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
Countries	154	154	154	154	154
Decades	11	11	11	11	11
Obs	993	1005	993	994	994
R2 within	0.907	0.902	0.903	0.901	0.901

Outcome: Infant mortality rate (ln). *Units:* country-decades. Right-side variables measured at T-1. All democracy indices measured as stock from 1900 (or first year data is available) with 10% annual depreciation. *Estimator:* OLS with country and year fixed effects, standard errors clustered by country. ***p<.01 **p<.05 *p<.10

Figure D4: Change in Predictive Margins based on Aggregation Rules



Predictive margins from Models 1-5 in Table D7, with all other covariates at their means. Vertical lines denote 10th and 90th percentiles. A change in MPI from the 10th to 90th percentile results in a 10% decrease in IMR (ln). This compares to a 6% decrease when calculating the index using addition and multiplication (i.e., Electoral Democracy Index) or the weakest link formulation. Aggregating by factor analysis or arithmetic average actually leads to an increase in IMR (ln) of 1% or 3%, respectively.

APPENDIX E: Mediation Analysis

In the text, we focus on the hypothesized effect of electoral democracy on human development. Here, we turn to the question of mechanisms. Why might electoral democracy (measured by MPI) be robustly associated with improvements in quality of life? In Section I, we argued that public spending plays an important role, and that selection effects (into leadership positions) and post-selection incentives (via electoral accountability) induce politicians in democracies to increase redistributive and public goods spending.

Given our focus on mortality, the most relevant – and measurable – pathway is health care spending. Regrettably, it is not possible to obtain fine-grained measures of health care spending for a large number of countries, e.g., on money spent in rural versus urban areas, hospital expenditures versus expenditures targeted on primary care. Nonetheless, aggregate spending statistics are useful in this context as they generally reflect formal policies approved by top leaders. (Allocative decisions, by contrast, are sometimes the product of decisions made by unelected bureaucrats and are on this particular account less relevant for present purposes, since our theory focuses on those at the apex of the policymaking process.)

While public expenditures have been examined by prior studies, these analyses have focused either on regime type and spending (Fielding, Freytag & Munch 2014) or on spending and health (Filmer & Pritchett 1999) – or, occasionally, on both, analyzed separately (McGuire 2010). Our objective is to assess the role of health spending as a mechanism lying *in between* political institutions and health. To do so adequately we must combine all three elements – *X*, *M*, and *Y* – into the same model. Accordingly, we employ a series of mediation analyses designed to test different indicators of health spending, different specifications, and different lag structures, as shown in Table E1.

Three data sources for public health expenditures are utilized: Nooruddin & Rudra (2014), Jensen & Skaaning (2015), and World Development Indicators (World Bank 2013). Data from these sources are correlated, though not as highly as one might think – which suggests the utility of running robustness tests with all three sources. WDI offers the best country coverage, but the shortest time-series (16 years); consequently, we do not employ this data for models with country fixed effects. Public health expenditures may be measured as a share of (a) GDP, (b) total central government expenditures, or (c) population (per capita). Each seems relevant to our theory, though "effort" is probably best proxied by (a) or (b), so we regard these as providing more appropriate tests of our theory.

Mortality may be measured by infant mortality (IMR), child mortality (CMR), or life expectancy (LE). We have already argued for a stock approach to measuring MPI (with a 10% annual depreciation rate), but we also test the simpler level measure.

Specifications may include country or region fixed effects (dummies for each major region of the world). Although the latter departs from our benchmark model, our mediation tests are limited by data coverage to shorter panels, raising questions about a unit fixed-effect model with sluggish left and right side variables. All models include year dummies and per capita GDP (logged).

The lag structure among the key variables – *X*, *M*, and *Y* – may be modeled as X^{t}-M^{t}-Y^{t+1} or (perhaps more plausibly) as X^{t}-M^{t+1}-Y^{t+2}.

The first model in Table E1 follows our benchmark closely. IMR (logged) is regressed on MPI stock (10%), per capita GDP (logged), and country fixed effects, with public health expenditure (Jensen & Skaaning 2015) treated as the mediator in a 0-1-2 lag structure. Here, we find that the indirect effect via public expenditure accounts for about 19% of the total effect of MPI on IMR, and the indirect effect is highly significant.

Subsequent models introduce permutations of this benchmark, altering one or more of the features described above. In all but one of these tests the mediator is in the predicted direction, and it is significant at 1% in 16 of the 19 specifications. Thus, we find corroboration of our argument that the positive net effect of electoral democracy on health outcomes stems, in part, from the type

of policies that autocratic and democratic leaders choose to pursue. While "public health spending" can cover a variety of expenses – and we cannot distinguish these with our data – our theoretical argument suggests that politicians incentivized by competitive elections and extensive franchise spend more on public goods prized by the citizenry, and these expenditures play a role in lowering IMR.

One must bear in mind the strong assumptions required for mediation analysis (Imai et al 2011). In particular, estimates of an indirect effect are sensitive to the omission of other relevant mediators. In this instance, we have been able to identify only one potential (measurable) mediator. If there are others, and if these are correlated with health spending, our estimates for health spending are not consistent. Second, data coverage for health spending is limited. Even the longest panel is only 35 years, and the variables of theoretical interest tend to be sluggish, as noted. This feature of the data may introduce bias when including country-fixed effects (Nickell 1981) – hence, the importance of region fixed-effect models as supplemental tests. For these reasons, we have greater confidence in estimates of the net effect, as shown in previous tables, than in estimates of the indirect effect, as shown in Table E1.

Table E1: Mediation Analysis using Annual Data

	Y	*X*	Fixed effects		*M*	Lag structure	Total effect			Indirect effect				Sample		
		MPI	*Country*	*Region*	*Health care exp*	*X-M-Y*	*β*	*t stat*	*p*	*β*	*t stat*	*p*	*Mediated*	*Countries*	*Years*	*Obs*
1	IMR	Stock	✓		Public/GDP (J&S)	0-1-2	-0.199	-5.357	***	-0.0597	-5.181	***	30.07	100	29	1568
2	IMR	Stock	✓		Public/GDP (J&S)	0-0-1	-0.214	-5.851	***	-0.0651	-5.386	***	30.41	100	29	1570
3	IMR	Stock	✓		Public/GDP (N&R)	0-0-1	0.016	0.347		-0.0183	-1.827		-116.1	53	35	1295
4	IMR	Stock	✓		Public/GDP (N&R)	0-1-2	0.040	0.841		-0.0153	-1.531		-38.76	53	35	1295
5	IMR	Stock		✓	Public/GDP (WDI)	0-0-1	-0.647	-12.78	***	-0.306	-12.24	***	47.34	149	16	2262
6	IMR	Stock		✓	Public/GDP (N&R)	0-0-1	-0.470	-8.993	***	-0.194	-7.678	***	41.38	53	35	1295
7	IMR	Stock		✓	Public/GDP (J&S)	0-0-1	-0.517	-11.79	***	-0.0667	-5.539	***	12.90	100	29	1570
8	IMR	Stock		✓	Public/GDP (WDI)	0-1-2	-0.631	-12.28	***	-0.282	-11.53	***	44.68	149	16	2299
9	IMR	Stock		✓	Public/GDP (N&R)	0-1-2	-0.503	-9.485	***	-0.193	-7.647	***	38.40	53	35	1295
10	IMR	Stock		✓	Public/GDP (J&S)	0-1-2	-0.534	-12.14	***	-0.0653	-5.461	***	12.22	100	29	1568
11	IMR	Stock			Public/GDP (WDI)	0-1-2	-0.486	-12.05	***	-0.392	-15.61	***	80.66	149	16	2299
12	IMR	Level	✓		Public/GDP (J&S)	0-1-2	-0.125	-5.947	***	-0.0283	-4.971	***	22.72	100	29	1568
13	IMR	Level		✓	Public/GDP (J&S)	0-1-2	-0.306	-8.084	***	-0.0614	-5.706	***	20.05	100	29	1568
14	IMR	Stock		✓	Public/cap (WDI)	0-1-2	-0.632	-12.32	***	-0.0559	-5.138	***	8.842	149	17	2301
15	IMR	Stock	✓		Public/cap (N&R)	0-1-2	0.048	1.044		0.00453	0.505		9.372	52	33	1238
16	IMR	Stock		✓	Public/cap (N&R)	0-1-2	-0.463	-8.648	***	-0.106	-5.718	***	22.91	52	33	1238
17	CMR	Stock	✓		Public/GDP (J&S)	0-1-2	-0.171	-4.666	***	-0.0593	-5.182	***	34.62	99	29	1563
18	CMR	Stock		✓	Public/GDP (J&S)	0-1-2	-0.511	-11.12	***	-0.0808	-5.893	***	15.81	99	29	1563
19	LE	Stock	✓		Public/GDP (J&S)	0-1-2	-0.133	-5.889	***	-0.0225	-3.422	***	16.87	100	29	1568
20	LE	Stock		✓	Public/GDP (J&S)	0-1-2	-0.238	-10.14	***	-0.0328	-5.301	***	13.76	100	29	1568

Y: human development, measured as infant mortality (IMR), child mortality (CMR), or life expectancy (LE). *X:* MPI, measured as stock (10% depreciation rate) or level (untransformed). *M:* mediator, health care spending, measured as total spending or public spending and as share of GDP or per capita, using data from J&S (Jensen & Skaaning 2015), N&R (Nooruddin & Rudra 2014), or WDI (World Bank 2013). All models include year fixed effects (FE) and GDP per capita (ln). *Unit of analysis:* country-year. *Estimator:* ordinary least squares. *** p<.01 **p<.05 *p<.10

Table E2: Mediation Analysis using 10-year Intervals

	Y	*X*	Fixed effects		*M*	Lag structure	Total effect			Indirect effect				Sample		
		MPI	*Country*	*Region*	*Health care exp*	*X-M-Y*	*β*	*t stat*	*p*	*β*	*t stat*	*p*	*Mediated*	*Countries*	*Decades*	*Obs*
1	IMR	Stock	✓		Public/GDP (J&S)	0-1-2	-0.272	-2.620	***	-0.078	-2.108	**	28.64	100	3	220
2	IMR	Stock	✓		Public/GDP (J&S)	0-0-1	-0.285	-2.704	***	-0.108	-2.596	***	38.04	100	4	223
3	IMR	Stock	✓		Public/GDP (N&R)	0-0-1	0.013	0.103		-0.036	-0.946		-269	53	4	175
4	IMR	Stock	✓		Public/GDP (N&R)	0-1-2	0.052	0.397		-0.023	-0.674		-45.09	53	4	173
5	IMR	Stock		✓	Public/GDP (WDI)	0-0-1	-0.669	-5.342	***	-0.314	-5.191	***	46.94	151	3	407
6	IMR	Stock		✓	Public/GDP (N&R)	0-0-1	-0.478	-3.318	***	-0.217	-2.922	***	45.42	53	4	175
7	IMR	Stock		✓	Public/GDP (J&S)	0-0-1	-0.546	-4.525	***	-0.068	-2.024	**	12.37	100	4	223
8	IMR	Stock		✓	Public/GDP (WDI)	0-1-2	-0.579	-3.849	***	-0.285	-3.937	***	49.22	151	2	297
9	IMR	Stock		✓	Public/GDP (N&R)	0-1-2	-0.484	-3.312	***	-0.219	-2.858	***	45.31	53	4	173
10	IMR	Stock		✓	Public/GDP (J&S)	0-1-2	-0.551	-4.479	***	-0.065	-1.947	*	11.81	100	3	220
11	IMR	Stock			Public/GDP (WDI)	0-1-2	-0.462	-4.098	***	-0.391	-5.675	***	84.61	151	2	297
12	IMR	Level	✓		Public/GDP (J&S)	0-1-2	-0.166	-2.562	**	-0.049	-2.120	**	29.46	100	3	220
13	IMR	Level		✓	Public/GDP (J&S)	0-1-2	-0.417	-3.809	***	-0.075	-2.279	**	17.96	100	3	220
14	IMR	Stock		✓	Public/cap (WDI)	0-1-2	-0.544	-3.537	***	-0.055	-1.764	*	10.07	151	3	299
15	IMR	Stock	✓		Public/cap (N&R)	0-1-2	0.066	0.511		0.005	0.184		8.131	52	4	167
16	IMR	Stock		✓	Public/cap (N&R)	0-1-2	-0.402	-2.732	***	-0.127	-2.191	**	31.51	52	4	167
17	CMR	Stock	✓		Public/GDP (J&S)	0-1-2	-0.249	-2.363	**	-0.073	-1.979	**	29.45	99	3	219
18	CMR	Stock		✓	Public/GDP (J&S)	0-1-2	-0.496	-3.848	***	-0.082	-2.092	**	16.51	99	3	219
19	LE	Stock	✓		Public/GDP (J&S)	0-1-2	-0.181	-2.717	***	-0.019	-0.862		10.43	100	3	220
20	LE	Stock		✓	Public/GDP (J&S)	0-1-2	-0.256	-3.829	***	-0.029	-1.789	*	11.18	100	3	220

Y: human development, measured as infant mortality (IMR), child mortality (CMR), or life expectancy (LE). *X:* MPI, measured as stock (10% depreciation rate) or level (untransformed). *M:* mediator, health care spending, measured as total spending or public spending and as share of GDP or per capita, using data from J&S (Jensen & Skaaning 2015), N&R (Nooruddin & Rudra 2014), or WDI (World Bank 2013). All models include year fixed effects (FE) and GDP per capita (ln). *Unit of analysis:* country-decades. *Estimator:* ordinary least squares. *** p<.01 **p<.05 *p<.10

APPENDIX REFERENCES

Aboubarb, M.R., A.L. Kimball. 2007. "A New Dataset on Infant Mortality Rates, 1816-2002." *Journal of Peace Research*, 44: 743.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J.A. 2014. "Democracy does cause growth." NBER Working Paper No. 20004.

Bernhard, Michael, Timothy Nordstrom, Christopher Reenock. 2001. "Economic Performance, Institutional Intermediation, and Democratic Survival." *Journal of Politics* 63: 775–803.

Blundell, R., S. Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1), 115-43.

Boix, Carles, Michael K. Miller, Sebastian Rosato. 2013. "A Complete Data Set of Political Regimes, 1800-2007." *Comparative Political Studies* 46:12, 1523-54.

Bollen, Kenneth A., Pamela Paxton. 2000. "Subjective Measures of Liberal Democracy." *Comparative Political Studies* 33(1): 58–86.

Brecke, Peter. 2001. "The Long-Term Patterns of Violent Conflict in Different Regions of the World." Paper presented at the Uppsala Conflict Data Conference, Uppsala, Sweden.

Clinton, J. D., David Lewis. 2007. "Expert Opinion, Agency Characteristics, and Agency Preferences." Political Analysis 16(1): 3–20.

Clinton, J. D., J.S. Lapinski. 2006. "Measuring Legislative Accomplishment, 1877-1994." *American Journal of Political Science* 50(1): 232–249.

Fielding, David, Andreas Freytag, Angela Münch. 2014. "Public Spending on Health as Political Instrument?: Regime type dependency of public spending." Working paper, Department of Environmental and Business Economics, University of Southern Denmark.

Long, J. Scott, Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata, 3rd ed.* Stata Press.

Gandhi, J., E. Lust-Oskar. 2009. Elections under authoritarianism. *Annual Review of Political* Science, 12, 403-422.

Gapminder. Accessed 2015. http://www.gapminder.org/data/

Gerring, John, Philip Bond, William Barndt, Carola Moreno. 2005. "Democracy and growth: A historical perspective." *World Politics* 57(3), 323-364.

Halperin, M.H., J.T. Siegle, Weinstein. 2005. *The democracy advantage: How democracies promote prosperity and peace*. New York: Routledge.

Honaker, J., Gary King, Matthew M. Blackwell. 2011. "Amelia II: A program for missing data." *Journal of Statistical Software*, 45(7), 1-47.

Human Life-Table Database: Methodology note http://www.lifetable.de/methodology.pdf

Human Mortality Database. 2008. *The Human Mortality Database*. Berkeley: University of California and Rostock: Max Planck Institute for Demographic Research. Available at www.mortality.org or www.humanmortality.de

Jackman, Simon. 2004. "What Do We Learn from Graduate Admissions Committees? A Multiple Rater, Latent Variable Model, with Incomplete Discrete and Continuous Indicators." *Political*

Analysis 12 (4): 400–424.

Jensen, Carsten, Svend-Erik Skaaning. 2015. "Democracy, Ethnic Fractionalization, and the Politics of Social Spending: Disentangling a Conditional Relationship." *International Political Science Review* 36:4, 457-72.

Johnson, Valen E., James H Albert. 1999. *Ordinal Data Modeling*. New York: Springer.

Junker, B. 1999. "Some Statistical Models and Computational Methods that may be Useful for Cognitively-Relevant Assessment." Carnegie Mellon University.

Knutsen, C.H. 2011. "Democracy, dictatorship and protection of property rights." *Journal of Development Studies*, 47(1), 164-182.

Levitsky, S., L.A. Way. 2010. *Competitive authoritarianism: Hybrid regimes after the cold war*. Cambridge: Cambridge University Press.

Lord, F. M., M.R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Miller, Michael K. 2015. "Democratic Pieces: Autocratic Elections and Democratic Development since 1815." *British Journal of Political Science* 45:3, 501-30.

Mitchell, BR. 1998a. *International Historical Statistics. Africa, Asia & Oceania 1750-1993*. Third edition. Basingstoke: Macmillan.

Mitchell, BR. 1998b. *International Historical Statistics. The Americas 1750-1993*. Fourth edition. Basingstoke: Macmillan.

Mitchell, BR. 1998c. *International Historical Statistics. Europe 1750-1993*. Fourth edition. Basingstoke: Macmillan.

Montevideo-Oxford Latin America Economic History Database. Available at: http://moxlad.fcs.edu.uy/

Nooruddin, Irfan, Nita Rudra. 2014. "Are Developing Countries Really Defying the Embedded Liberalism Hypothesis?" *World Politics* 66:04 (October) 603-40.

Patz, Richard J., Brian W. Junker. 1999. "A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models." *Journal of Educational and Behavioral Statistics* 24: 146-178.

Pemstein, Daniel, Stephen A. Meserve, James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18:4, 426-449.

Pemstein, D., K.L. Marquardt, E. Tzelgov, Y Wang, F. Miri. 2015. "The V-Dem measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data." University of Gothenburg, Varieties of Democracy Institute: V-Dem Working Papers Series No. 20.

Riley, James C. 2005a. "Estimates of Regional and Global Life Expectancy, 1800–2001". *Population and development review* 31:3, 537–543.

Riley, James C. 2005b. "The Timing and Pace of Health Transitions around the World." *Population and development review* 31:4, 741–764.

Roodman, D. 2009. "A note on the theme of too many instruments." *Oxford Bulletin of Economics and Statistics*, 71(1), 135-158.

Ross, M. 2006. "Is democracy good for the poor?" *American Journal of Political Science* 50(4), 860–874.

Rubin, D.B. 1987. *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schedler, A. 2006. "The logic of electoral authoritarianism." In Schedler, A. (ed.), *Electoral*

authoritarianism: The dynamics of unfree competition. Boulder: Lynne Rienner Publishers.

Treier, Shawn, Simon Jackman. 2008. "Democracy as a Latent Variable." American Journal of Political Science 52(1): 201–17.

UNICEF. 2005. *Multiple Indicator Cluster Survey (MICS)*. New York: UNICEF. Available at: www.childinfo.org

United Nations Population Division. 2010. *World Population Prospects: The 2008 Revision. Life expectancy at birth, both sexes, estimates 1950-2009*. New York: United Nations. Available at: http://esa.un.org/unpp/.

Wang, Y, V. Mechkova, F. Andersson. 2018. "Does Democracy Enhance Health? New Empirical Evidence 1900-2012." *Political Research Quarterly* (forthcoming).

World Bank. 2013. *World Development Indicators 2013*. Washington, DC: World Bank.