**Title:**

# Assessing the impact of health care expenditures on mortality using cross-country data

**Authors*:**

Farasat Bokhari (f.bokhari@uea.ac.uk) [a]

Karl Claxton (karl.claxton@york.ac.uk) [b, c]

James Lomas (james.lomas@york.ac.uk) [c]

Rodrigo Moreno-Serra (rodrigo.morenoserra@york.ac.uk) [c]

Ryota Nakamura (ryota.nakamura@york.ac.uk) [d]

Marc Suhrcke (marc.suhrcke@york.ac.uk) [a]

[a] School of Economics, University of East Anglia, Norwich, Norfolk, NR4 7TJ, United Kingdom

[b] Department of Economics and Related Studies, University of York, York, YO10 5DD, United Kingdom

[c] Centre for Health Economics, University of York, York, YO10 5DD, United Kingdom

[d] Hitotsubashi Institiute for Advanced Study, Hitotsubashi University, 2-1 Naka Kunitachi Tokyo, Japan

* Authors listed in alphabetical order. Lead author is Nakamura.

**Acknowledgements**

This study is financially supported by the International Decision Support Initiative (iDSI). The authors would like to thank Mead Over, Sebastian Bauhoff, Amanda Glassman, Jessica Ochalek, Mark Sculpher, Peter Smith, Paul Revill, Yot Teerawattananon, Stephen Martin, Alex Rollinger and all participants of the iDSI-workshops in York and London for their very helpful support, comments and suggestions. We also thank the authors of Bokhari et al (2007) and Moreno Serra & Smith (2015) for openly sharing their data and codes.

**Abstract**

Prior literature reports significant variation in the impact of public health expenditure on mortality. We re-examine the literature that identifies this effect using cross-country data. Our analysis builds on the two instrumental variables (IV) approaches embodied by key publications in this field – Bokhari et al. (2007) and Moreno-Serra and Smith (2015). We start by successfully replicating their findings. However, further analyses using updated data and a hybrid specification reveal sensitive results. In particular, the relevance of the instruments is compromised in the updated data, leading to imprecise estimations of the relationship. While our results should not be taken to imply that there is no impact of public health care expenditures on mortality, the findings do call for further methodological work, for instance in terms of identifying more suitable IVs or by applying other estimation strategies.

**Keywords**

Mortality; Public healthcare expenditure; Cross-country data; Instrumental variables

## 1. Introduction

The question of whether, and if so, by how much, public health care expenditures impact on mortality outcomes has occupied and challenged researchers for some decades (Gravelle and Backhouse, 1987). While it is hard to imagine that – if the true relationship were in fact measurable – public health care expenditures would not 'buy' mortality reductions, all else equal, in reality there may well be 'weak links in the chain' from spending to outcomes (Filmer et al., 2000). Out of the significant body of empirical studies that have attempted to estimate the relationship, some previous studies confirm the expected sign and significance of the relationship, yet the magnitude of the impact varies considerably between studies, and several studies show statistically insignificant effects. For instance, an influential study by Filmer and Pritchett (1999), using global cross-country panel data, found at best a small impact of public health care expenditures on under-five-mortality rates, with a 1% increase in spending leading to a – statistically insignificant – reduction in under-5 mortality by 0.19%, and in infant mortality rates by 0.078%. By contrast, Bokhari et al. (2007) – using a cross-section of data from the year 2000 and an instrumental variable strategy that broadly follows Filmer and Pritchett (1999) – find that a 1% increase in public health care expenditure significantly (both in statistical and economic terms) reduces under-5 mortality by between 0.25 and 0.42%, and maternal mortality by 0.42 to 0.52%. Those findings have been reinforced very recently by Moreno-Serra and Smith (2015), using again a large cross-country panel analysis with a estimation strategy that allows them to explicitly model the reverse causality from mortality onto spending. Some studies have also pointed to potential nuances in the relationship, e.g. in the sense that public spending matters most for the health of poor people in lower income countries (Bidani and Ravallion, 1997), or that mortality is only reduced in response to public spending when the overall quality of governance in a country exceeds a certain threshold level (Wagstaff and Claeson, 2004; Rajkumar and

Swaroop, 2008). However, while a number of papers do claim to have established a causal relationship between expenditure and mortality outcomes, the magnitude of the impact varies considerably between studies.

In light of these ambiguities, and in order to reflect potential changes in the relationship between public expenditure on health and mortality over time, we revisit the literature that identifies this effect using cross-country data.

Re-estimating this relationship is important for several reasons. To the extent that we are able to derive a robust empirical relationship, this would allow us to capture the marginal productivity of the publicly funded health care system, i.e. the opportunity costs of public health care spending, which in turn would be a crucial element to derive more useful cost-effectiveness threshold estimates than have hitherto been promoted, for instance in the widely used form of the 1-3 times per capita GDP rule-of-thumb. This rule was initially proposed by the WHO Commission on Macroeconomics and Health (WHO 2001), and has subsequently been taken up by the World Health Organization (WHO 2015). From there its use has spread massively, not least for lack of obvious other alternatives. However, the limitations – and potential harms – of this approach have increasingly been recognised (Revill et al., 2014; Newall et al., 2014; Marseille et al. 2015). Recent UK-focused research has shown how cost-effectiveness thresholds can empirically be estimated, starting from rigorous empirical estimates of the impact of spending on health outcomes (Claxton et al., 2015). Cross-country based analysis of the kind presented in the present paper has the potential to inform empirically derived thresholds in countries world-wide, including low- and middle-income countries (LMICs) where the prioritization of health care resources to maximise population health gains is particularly warranted (Ochalek et al., 2015).

In this paper we employ an empirical strategy rooted in the literature that has estimated the mortality effects of public health care expenditure using instrumental variable strategies – by far the most popular strategy to tackle endogeneity in this literature. While we cannot fully reflect the specific approach taken in every single paper in this literature (see Table A10 for an overview of a select set of studies, and Gallet and Doucouliagos (2015) for a meta-analysis of the broader literature), we first examine the robustness of two exemplary studies – by Bokhari et al., (2007) and Moreno-Serra and Smith (2015), henceforth referred to as BGG and MSS, respectively – to new data, whilst preserving the papers' original model specifications. In the second, major step of the analysis we then abstract from the two exact specifications to construct panel data models using instrumental variables and covariates taken from our review of the existing literature. In doing so we still preserve the spirit of the two approaches, by applying both a traditional approach to instrumentation (broadly following Bokhari et al (2007)) along with an application of Bruckner (2013), which is the approach followed by Moreno-Serra and Smith (2015) (discussed further in the methods section). As part of the second step we also undertake a series of robustness checks, not least to accommodate some other estimation approaches used in the literature (e.g. dynamic panel analysis). Taken together we argue that following the empirical approaches proposed in these two studies – first literally and then more generally – allows us to capture the currently most promising approaches that have been used in this literature.

We argue that Bokhari et al. (2007), apart from having become a highly cited paper in this literature, broadly represents the IV approach adopted in several such papers, starting from Filmer & Pritchett (1999) and Wagstaff and Claeson (2004). Moreno-Serra and Smith (2015), by contrast, offer a different (though also IV based) approach that, for the first time in this literature, allows for an explicit assessment of the extent of reverse causality, running from mortality to health spending.

By following the above broad strategy we comprehensively re-examine the robustness of the empirical relationship between public health care spending and mortality outcomes using cross-country data, as it was presented in the main approaches used by the previous literature. Across the two replicated studies, we find that in the majority of replication scenarios the effect of public health care expenditure on mortality is negative, although qualitative conclusions based on statistical significance vary and the magnitude of the impact is sensitive to revisions and updates in the data used. New results from our hybrid approach, using a traditional IV approach, are unreliable on account of the chosen instrument (per capita military expenditure of neighbouring countries) being found to not be statistically relevant. For another hybrid model, using a identification strategy based on Bruckner (2013) and Moreno-Serra and Smith (2015), at least in some cases the estimated effects are broadly consistent with the previous literature, finding that public spending reduced mortality, although at generally lower effect magnitudes and with mixed degrees of statistical significance. Our very mixed findings should not be taken to imply that the conclusions from the previous studies are necessarily invalid, and conclusions need to be seen in the light of the considerable methodological challenges (relating to data and econometric techniques) faced by this strand of research.

## 2. Methods

Our basic model is specified as follows. We use country-level panel data including low-, middle- and high-income countries. Our primary focus is to estimate the causal impact of public health expenditure on mortality outcomes. To this end, we consider the following simple linear model:

$$\text{Eq. (1):} \quad \mathrm{MO_{jt}} = \beta_j + \beta_1 \mathrm{EX_{jt}} + \mathbf{X}'_{jt}\boldsymbol{\beta} + \varepsilon_{jt}$$

MO is mortality and EX is public healthcare expenditure in country *j* at time *t*. The vector X represents covariates. Our main goal is to estimate $\beta_1$, which represents the expenditure elasticity when the natural logarithm of MO and EX are used in the regression model i.e. the effect of a 1% increase in expenditure on the % change in mortality. While health care expenditure may reduce mortality, policymakers may increase health expenditure in response to an increase in mortality – i.e. the idea of reverse causality. Also, mortality and health expenditure may be correlated with a third variable (which is often not observed in the data), which gives rise to omitted variable bias on the estimated coefficients. Given these concerns in this observational setting, we need an exogenous variation of expenditure EX to identify the causal impact of EX on MO (which is represented by the coefficient $\beta_1$).

Following both BGG and MSS, we use an instrumental variable method. BGG and MSS use distinct IV approaches which differ in their econometric specification and also in their strategy in identifying the causal impact of expenditure. Employing more than one approach allows us to assess the robustness and sensitivity of the estimated impacts to different econometric approaches using the same data. Hence we start our analyses by replicating the analyses of BGG and MSS respectively, using their original data as well as our updated data. In order to assess the extent to which their conclusions may be driven by their specific features of the models (e.g. choice of variables and data years), we also construct a hybrid version of econometric model with our own set of variables (based on a review of the literature), and augment the two distinct approaches to estimate the impact of health expenditure on mortality outcomes.

### 2.1 BGG model

BGG use a traditional IV approach that may be seen as broadly representative for a range of existing studies published before. However, this particular study uses cross sectional data from 2000 (as opposed to panel data from multiple years), and also employs a structurally motivated model that investigates the impact of health expenditure as well as GDP per capita on mortality.

The main outcome variables in the BGG model are log of under-5 mortality and log of maternal mortality. There are four key independent variables which are treated as endogenous. The endogenous variables include GDP per capita, public health care expenditure and interaction of public health expenditures with deviation in donor funding on basic health from historical mean and measure of paved roads per unit area in the country (as a proxy of economic development). The model includes the following covariates: log of proportion of non-literate (proxy of education); log of quantity of paved roads per unit area; log of donor funding; and log of proportion of population with access to improved sanitation.

BGG use a number of instrumental variables in their first stage regression with the four endogenous variables. The IVs include: 1) log of neighbours' military expenditure per capita, and its interaction with a) deviation in donor funding on basic health from historical mean and b) the quantity of paved roads per unit area; 2) indicators of quality of policies and institutions (based on World Bank's annual assessment of countries), and again its interaction with deviation in donor funding from historical mean and paved road rate; and finally 3) the log of the consumption-investment ratio.

### 2.2 MSS model

MSS use a more novel IV approach inspired by Bruckner (2013), which explicitly takes into account the reverse causality from mortality to health expenditure. In this approach, as the first step, we estimate the reverse causality – the impact of mortality on health care expenditure:

$$\text{Eq. (2):}\quad \mathrm{EX_{jt}} = \gamma_{\mathrm{j}} + \gamma_1 \mathrm{MO_{jt}} + \mathbf{X}'_{\mathrm{jt}}\boldsymbol{\gamma} + \epsilon_j$$

Note that we use the same set of covariates (vector X) in this reverse causality model. Once we estimate $\gamma_1$ by a panel fixed effect instrumental variable regression, we calculate the following residual:

$$\text{Eq. (3):}\quad \mathrm{EX^{*}_{jt}} = \mathrm{EX_{jt}} - \widehat{\gamma_1}\mathrm{MO_{jt}}$$

Assuming the instruments used in Eq. (2) are valid and relevant, this residual is now orthogonal to the variation of mortality. In the second step, we use this residual as an IV for health expenditure in Eq. (1) to estimate the impact of health expenditure on mortality.

The main outcome variables in the MSS study are: under-5 mortality, adult female mortality, and adult male mortality. MSS specify four endogenous indicators of health expenditure: a) public health care expenditure; b) private out-of-pocket health care expenditure; c) expenditure from private voluntary health insurance; and d) immunization coverage. We therefore estimate Eq. (2) and calculate the residual in Eq. (3) for each endogenous variable separately, but estimate the impact of these variables in one main regression Eq. (1) with four endogenous variables. Since the panel data are clustered at country level, the estimation of the standard errors takes this data structure into account through the use of cluster-robust standard errors.

In the MSS model the covariates include: GDP per capita, primary education enrolment rate, proportion of population under 14, and proportion of population over 65. Note that, in contrast with BGG, in the MSS analysis GDP per capita is treated as an exogenous control variable. In the streamlined model below we follow this simplification of the model because our main purpose is specifically to estimate the impact of health expenditure on mortality, which we pursue using a reduced form specification.

## 2.3 A Hybrid Model

As described above, BGG and MSS have specific features in terms of data, selection of variables, and strategy to estimate causal impacts. For instance, BGG use cross-sectional data whereas MSS use panel data[1]; they use different set of health expenditure variables (some of which are not our primary interest) and also different sets of covariates. These particular features might have led to different magnitudes of the main estimates or even their conclusions. In order to still follow the spirit of the two approaches represented by these studies without strictly adhering to their specific models, we construct our own specifications (which we refer to as a hybrid model) as described below.

[1] On the whole there is good reason to expect that *all else equal*, using panel data allows can be more reliable than cross-sectional data as the former allows the researcher to control for unobserved time invariant factors at country level directly via a fixed effects estimator. This can be important if such a source of unobserved heterogeneity is both correlated with health spending and mortality outcomes as it would then reduce the need for finding appropriate IVs for the health spending variable (assuming of course that there is no other source of endogeneity). Alternatively, if the omitted time invariant variable is not correlated with health spending, one can use random effects models instead of fixed effects for a gain in efficiency. Generally, both types of panel estimators are an improvement over simple cross-sections. The problem, and hence the trade-off in panel vs cross-section data comes from not being able to obtain reliable time series on all variables of interest. This can happen when some of the variables of interest have been provided by data dissemination agencies by interpolating for missing values.

#### *2.3.1 Data structure*

We use country level panel data from 1995 to 2011. This means that in applying a traditional IV approach we complement the above-described, cross-sectional BGG analysis by incorporating further, more recent data points, and by allowing for country-level fixed effects in the model and estimate a within-country impact of health expenditure. This will additionally take into account any unobserved, time invariant country-level heterogeneity. Furthermore, the country level fixed effect is also likely to account for some of the (unobserved) heterogeneity from factors that vary only slowly over time (and particularly in shorter time series), such as cultural factors and institutional quality. MSS's original study uses data from 1995 to 2008. We therefore include more recent years (up to 2011) in our Bruckner IV approach analysis (Bruckner, 2013). Descriptive statistics are given in Table 3.

#### *2.3.2 Outcome variables*

We follow MSS in choosing our outcome variables. We use: under-5 mortality, adult female mortality, and adult male mortality. However, while MSS use under-5 mortality estimated by the Institute of Health Metrics and Evaluation, we use estimates produced by the United Nations (as in BGG). This is because more data are available in the UN mortality data (up to 2011) than in the IHME data (up to 2010).

#### *2.3.3 Health care expenditure*

While BGG and MSS specified a number of endogenous variables, in our streamlined model we focus only on public health care expenditure following a reduced form approach. In particular, in line with MSS, in our main model we treat GDP per capita as an exogenous control variable to keep the model parsimonious.[2] Moreover, although BGG include two interaction terms involving the rate of paved road and the amount of donor funding on basic health, we drop these interaction terms from our hybrid model because some of the data are not available, and also because these factors are not of primary interest to us.[3] Similarly, we exclude three of MSS's key explanatory variables, i.e. out-of-pocket expenditure, private insurance expenditure and immunization coverage, for the same reason.[4]

### *2.3.4 Instrumental variables*

Based on our review of the existing literature, we use instrumental variables, which have been shown to be relevant in the previous literature, including BGG and MSS. However, since we estimate within-country impact by using fixed effect models, our instrumental variables should vary over time. Therefore we cannot use time-invariant IVs (i.e. certain dimensions of governance quality and their interactions) in the current specification.

As explained in the sections 2.1 and 2.2, in the traditional IV analysis we need instruments for public health care expenditure, whereas the Bruckner IV analysis requires instruments for mortality. In the traditional IV approach, we use neighbours' military expenditure (in % of GDP) and neighbours' public health expenditure (in % of GDP) as instrumental variables (Filmer and Prichett, 1999; Wagstaff and Claeson, 2004).[5 6]

---

[2] Appendix Table A9 presents the model where per capita GDP is treated as an endogenous variable. Following BGG, we used the consumption-investment ratio as the instrument for per capita GDP. The estimated effect of per capita GDP is negative and in some cases the effect is statistically significant, i.e. higher national income reduces mortality. The estimated impact of public health expenditure in this model is largely similar to the estimates presented in the main paper. Also, using BGG's original data and following their specification, we estimated the model with per capita GDP treated as exogenous variable. We find that the impacts of public expenditure on the mortality outcomes are largely maintained (the impact is -0.390 (SE=0.128) for under-5 mortality; and it is -0.607 (SE=0.190) for maternal mortality). These additional results imply that assuming per capita GDP as exogenous should not substantially bias the estimation of the impact of public healthcare expenditure on mortality outcomes in our data.

[3] Using BGG's original data, we estimated the BGG's model while excluding the interaction terms and also assuming per capita GDP to be exogenous. The impacts of public expenditure on mortality outcomes are qualitatively maintained (the impact is -0.419 (SE=0.096) for under-5 mortality; and it is -0.565 (SE=0.163) for maternal mortality).

[4] Dropping private expenditure might lead to misspecification of the model and therefore bias the estimated impact of public healthcare expenditure. However, as shown in Table 4, dropping the private expenditure from the original MSS specification does not affect the impact of public expenditure to a substantive degree. Also, in Appendix Table B1, we estimate the impact of *total* healthcare expenditure (which includes both public and private expenditure) on mortality, and find that the results are similar to the principal estimation results presented in the main text (Table 6). This might be because public healthcare expenditure accounts for a majority of healthcare resources in many (though by no means all) countries and thus the potential bias from omitting private expenditure may be negligible.

[5] In preliminary analyses we also employed a 'just identified' model with one instrumental variable, where we used per capita military expenditure of neighbouring countries for the single endogenous variable (i.e. public health care expenditure) in the

Our instruments in the Bruckner IV approach include per capita emission of $CO_2$ and number of deaths by conflict per 100,000 people, following MSS. As a sensitivity check we also use WHO's data on external causes of death as additional IVs predicting exogenous variations in mortality.[7] Test statistics for the relevance of the IVs are presented and discussed in the Results section.

### *2.3.5 Covariates*

Previous studies (including but not limited to BGG and MSS) use a variety of covariates, and there appears to be no common set of covariates that has been applied in all studies. As Gallet and Doucouliagos (2015) point out, the choice of covariates is indeed a significant contributor of the diversity in the estimated impact of expenditure on mortality. In this study we use the following principle-based selection of covariates. We conduct a literature review of the related, cross-country type studies. We identified and reviewed 11 studies and recorded covariates from each study (Afridi and Ventelou, 2013; Akinci et al. 2014; Anyanwu and Erhijakpor, 2009; Bokhari et al. 2007; Farag et al. 2013; Filmer and Pritchett, 1999; Gani, 2009; Hu and Mendoza 2013; Moreno-Serra and Smith, 2015; Rajkumar and Swaroop, 2008; and Wagstaff and Claeson, 2004).[8] From the full set of candidates we obtain a shortlist of covariates in the following way: first, we exclude variables if we decide they can be outcomes of health expenditure (e.g. number of hospital beds), as including those

main model. We found this single instrument was too weak in terms of the first stage F statistics (F is less than 1 in most cases).

[6] A rationale for the IV is that if neighbouring countries spend more on military equipment and services, the domestic government would increase or decrease military expenditure to maintain the military balance, which would affect the level of healthcare expenditure per capita. Note that this IV has been widely used in the previous literature - see Anyanwu and Erhijakpor (2009), Bokhari et al. (2007), Filmer and Prichett (1999), and Wagstaff and Claeson (2004). We follow this popular (though not un-contestable) approach in the existing cross-country literature.

[7] External cases of death include: transport accidents, smoke and fire, poisoning, falls, drowning, and assault, which largely randomly occur across the population.

[8] These studies were collected using a 'snowballing' strategy, in which we started by looking at the reference list of BGG and MSS studies for relevant studies, then once we found a related study (initially judged by title, and then by full text), we included the study in our review and also checked the reference list of the study for further references. We iterated this process for several times, thus converging towards the set of studies included in our review.

variables would cause selection bias in the regression (Angrist and Pischke, 2009). Second, we follow Hauck et al. (2015) in our selection of further covariates:

1) We drop a variable if, in the currently available data, more than 60% of observations between 1995 and 2011 are missing across years and countries;

2) In order to avoid multi-collinearity among covariates, we calculate the pair-wise correlation between the de-meaned candidate variables and drop one of the variables, if the correlation coefficient is above 0.6.[9]

Our final set of covariates is as follows: GDP per capita, primary education enrolment rate, proportion of population under 14, proportion of population over 65, rate of paved road, access to improved sanitation, urbanisation, and two measures of efficiency of governance (government efficiency and control of corruption). See Table 3 for more detail.

Finally, in the hybrid model we use log-scale variables (which is consistent with BGG but not MSS) in order to estimate elasticities, i.e. the percentage reduction/increase in mortality associated with a 1% increase in expenditure.

Following BGG and MSS (and other studies in the literature), all models are estimated via Generalized Method of Moments.[10] All analyses are conducted using Stata 13 MP.

[9] De-meaned variables (i.e. we calculate the mean of a given variable for each country and then subtract the country's mean over the observed time period) are used because we use fixed effect models, in which de-meaned variables are analysed in the regression.

[10] Appendix Table A10 provides the results of the main streamlined model analysis via Limited Information Maximum Likelihood (LIML) estimation, which could be more robust to weak instrumental variables than GMM or 2SLS. The results using LIML are largely similar to the ones using GMM, but they actually give slightly less precise estimation. This generally applies to the results presented in other tables in this paper as well as those in Appendix. The full replications of the whole paper using LIML are available upon request.

## 3. Data

For the original replication of the published results the authors of the original articles generously allowed us full access to the original data (and codes) that were used for the published results of those papers. When updating the data to include the most recently available ones, we also used, wherever accessible, the sources indicated in the two papers. It should be noted that for most of the variables even the older data are regularly revised in the publicly available data sources, and it is not always possible to obtain older versions of the database.

The majority of our data comes from three widely used sources: the World Bank's World Development Indicators (WDI), the World Health Organization's Global Health Observatory, and the Institute of Health Metrics and Evaluation's 'Global Health Data Exchange'. The sources, definitions, and downloaded data of all the variables we used in our empirical analysis study are given in Appendix Table A1. [11]

Tables 1 and 2 give the descriptive statistics of each of the composed datasets we use in our analysis: First, for the original and updated data analysis of the BGG paper (Table 1), and second for the MSS paper (Table 2). In these tables, there are some notable discrepancies between original data and updated data, which could be attributable to data updates by the data providers, or to differences in the exact metric used for a given variable.[12] In Table 1, the original data consist of 127 countries whereas the updated data consist of 238 countries.

[11] Note that data on mortality are from the World Bank's WDI database (i.e. for adult female mortality rate, adult male mortality rate, maternal mortality rate, and under-5 mortality rate). The under-5 mortality data available in the WDI database are based on United Nations' estimates. IHME have produced alternative estimate of under-5 mortality, and it does not appear obvious which one of the two is closer to the truth (Alkema and You, 2012).

[12] For instance, the updated data has GDP per capita based on purchasing power parity in 2011 constant dollars, but MSS used GDP per capita based on purchasing power parity in 2005 constant dollars, and BGG used GDP per capita in 2000 international dollars (also based on purchasing parity). It should be noted that Table 1 and 2 compare outcomes from different set of countries.

Similarly, in Table 2, the original data consist of 153 countries whereas the updated data consist of 236 countries. Appendix Table C1 and C2 present the descriptive statistics for the subset of updated data using the same set of countries as in the original data. The discrepancies between the original and updated data become smaller. In the results section we examine the impacts of using different versions of GDP data on the regression results (see Appendix Tables A2-A5 for detailed results).

### 3.1 Data imputation

Although we use the most up-to-date data that is currently available in the public domain, significant numbers of missing values do remain. The sample size in different studies (BGG, MSS and others) varies significantly (from around 100 to 150 countries) as a result, depending on the selection of variables and versions of the data source used in their analyses.

**Table 1: Descriptive statistics of BGG data (original data and updated data)**

| | Original data (2000) | | | Updated data (2000)* | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N | # missing | max N | Missing rate |
| Under 5 motality | 73.49 | 68.91 | 127 | 57.67 | 58.17 | 192 | 44 | 236 | 18.6% |
| Maternal mortality | 345.39 | 425.07 | 127 | 264.23 | 361.86 | 183 | 53 | 236 | 22.5% |
| GDP per capita† | 6962.60 | 7987.95 | 127 | 14396.85 | 18467.87 | 185 | 51 | 236 | 21.6% |
| Public health expenditure‡ | 311.21 | 509.70 | 127 | 542.53 | 866.97 | 186 | 50 | 236 | 21.2% |
| Neighbors' military expenditure per capita | 161.04 | 183.22 | 127 | 271.48 | 383.53 | 222 | 14 | 236 | 5.9% |
| World Bank CPI on Social inclusion and equity | 3.63 | 0.76 | 127 | . | . | . | . | . | . |
| World Bank CPI on Economic management | 3.96 | 1.05 | 127 | . | . | . | . | . | . |
| Investment-GDP ratio | 14.87 | 8.71 | 127 | 22.79 | 9.92 | 189 | 47 | 236 | 19.9% |
| Donor funding on basic health per capita | 2.96 | 6.35 | 127 | . | . | . | . | . | . |
| Nonliterate rate | 21.22 | 22.15 | 127 | 21.32 | 21.01 | 101 | 135 | 236 | 57.2% |
| Paved road rate | 31.14 | 66.63 | 127 | 49.37 | 32.86 | 169 | 67 | 236 | 28.4% |
| Sanitation rate | 73.98 | 26.67 | 127 | 68.51 | 31.15 | 198 | 38 | 236 | 16.1% |

Notes:

* In this column the analysis include as many countries as possible (max available number of countries is 236) from the updated data sources.

†In the original data, GDP per capita using international dollars as of 2000 is used. In the updated data, GDP per capita using constant US dollars 2011 (PPP adjusted) is used.

‡In the original data, government health expenditure per capita (international dollars as of 2000) is used. In the updated data, government health expenditure per capita in 2005 constant dollars is used.

**Table 2: Descriptive statistics of MSS data (original data and updated data)**

| | Original data (1995-2008) | | | Updated data (1995-2008)* | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | N (153 countries) | Mean | SD | N (236 countries) | # missing | max N | Missing rate |
| Under 5 motality | 45.68 | 48.87 | 1397 | 52.26 | 52.93 | 2618 | 686 | 3304 | 20.8% |
| Adult female mortality | 155.85 | 114.87 | 1225 | 173.97 | 129.77 | 2696 | 608 | 3304 | 18.4% |
| Adult male mortality | 226.39 | 116.92 | 1225 | 242.39 | 123.48 | 2696 | 608 | 3304 | 18.4% |
| public health expenditure (1/100)† | 5.99 | 8.43 | 1397 | 5.93 | 10.19 | 2602 | 702 | 3304 | 21.2% |
| Privately pooled health expenditure (1/100) | 0.59 | 2.10 | 1397 | 2.77 | 14.27 | 2602 | 702 | 3304 | 21.2% |
| Out-of-pocket health expenditure (1/100) | 1.88 | 2.04 | 1397 | 0.58 | 1.93 | 2233 | 1071 | 3304 | 32.4% |
| Immunization coverage (1/10) | 8.61 | 1.42 | 1397 | 8.43 | 1.66 | 2674 | 630 | 3304 | 19.1% |
| GDP per capita (1/100)‡ | 121.15 | 131.67 | 1397 | 148.27 | 183.98 | 2578 | 726 | 3304 | 22.0% |
| Primary education enrollment rate (1/10) | 8.61 | 1.59 | 1397 | 8.84 | 1.54 | 1477 | 1827 | 3304 | 55.3% |
| Proportion of population under 14 (1/10) | 3.07 | 1.05 | 1397 | 3.19 | 1.05 | 2716 | 588 | 3304 | 17.8% |
| Proportion of population over 65 (1/10) | 0.75 | 0.52 | 1397 | 0.70 | 0.47 | 2716 | 588 | 3304 | 17.8% |
| CO2 emission per capita | 5.16 | 6.51 | 1397 | 4.79 | 6.74 | 2728 | 576 | 3304 | 17.4% |
| Death by conflict | 1.45 | 25.76 | 1397 | 1.12 | 17.08 | 3304 | 0 | 3304 | 0.0% |

Notes: * In this column the analysis include as many countries as possible (max available number of countries is 236) from the updated data sources.

† In the original data, government health expenditure per capita (international dollars as of 2005) is used. In the updated data, government health expenditure per capita in 2005 constant dollars is used.

‡ In the original data, GDP per capita using international dollars as of 2005 is used. In the updated data, GDP per capita using constant US dollars 2011 (PPP adjusted) is used.

The variable 'Death by conflict' does not have any missing values. The original SIPRI data (see Appendix Table A1) lists the cases of conflicts and the number of deaths where they observed. Following Morreno-Serra and Smith (2015), we assume that there were not conflicts (and therefore no deaths) if the original data do not list any information for a given country.

**Table 3: Descriptive statistics of the streamline model data (1995-2011)**

| | Mean | SD | N (236 countries) | # missing | max N | Missing rate |
|---|---|---|---|---|---|---|
| *Mortality outcomes* | | | | | | |
| Under 5 motality (UN estimate) | 51.70 | 53.78 | 3264 | 748 | 4012 | 18.6% |
| Adult female mortality | 171.16 | 128.47 | 3239 | 773 | 4012 | 19.3% |
| Adult male mortality | 238.45 | 122.25 | 3239 | 773 | 4012 | 19.3% |
| *Health expenditure* | | | | | | |
| Total health expenditure | 1085.90 | 4025.17 | 3163 | 849 | 4012 | 21.2% |
| Public health expenditure | 631.03 | 1206.37 | 3163 | 849 | 4012 | 21.2% |
| *Covariates* | | | | | | |
| GDP per capita | 15234.72 | 18652.36 | 3147 | 865 | 4012 | 21.6% |
| Sanitation rate | 69.69 | 30.77 | 3299 | 713 | 4012 | 17.8% |
| Primary education enrollment rate | 89.06 | 14.66 | 1849 | 2163 | 4012 | 53.9% |
| Proportion of population under 14 | 31.38 | 10.62 | 3299 | 713 | 4012 | 17.8% |
| Proportion of population over 65 | 7.10 | 4.77 | 3299 | 713 | 4012 | 17.8% |
| Paved road rate | 53.93 | 33.04 | 1621 | 2391 | 4012 | 59.6% |
| Urban rate | 55.46 | 24.47 | 3587 | 425 | 4012 | 10.6% |
| Efficiency of governance | -0.01 | 1.00 | 2628 | 1384 | 4012 | 34.5% |
| Control of corruption | -0.01 | 1.00 | 2634 | 1378 | 4012 | 34.3% |
| *Instrumental variables* | | | | | | |
| CO2 emission per capita | 4.76 | 6.63 | 3383 | 629 | 4012 | 15.7% |
| Death by conflict (per 100,000) | 1.00 | 15.56 | 4012 | 0 | 4012 | 0.0% |
| Neighbors' military expenditure per capita | 296.13 | 442.08 | 3774 | 238 | 4012 | 5.9% |
| Neighbors' health expenditure per capita | 589.29 | 951.79 | 3774 | 238 | 4012 | 5.9% |
| Neighbors' military expenditure (% of GDP) | 0.03 | 0.02 | 3774 | 238 | 4012 | 5.9% |
| Neighbors' health expenditure (% of GDP) | 0.04 | 0.02 | 3774 | 238 | 4012 | 5.9% |
| *External cause of death* | | | | | | |
| Transport accident | 12.43 | 6.66 | 1563 | 2449 | 4012 | 61.0% |
| Smoke, fire, flames | 1.35 | 1.88 | 1563 | 2449 | 4012 | 61.0% |
| Falls | 5.55 | 5.82 | 1563 | 2449 | 4012 | 61.0% |
| Drowning and submersion | 3.23 | 3.30 | 1555 | 2457 | 4012 | 61.2% |
| Assault | 6.70 | 9.80 | 1563 | 2449 | 4012 | 61.0% |
| Poisoning | 3.46 | 6.79 | 1563 | 2449 | 4012 | 61.0% |
| All other external cause | 9.68 | 14.17 | 1557 | 2455 | 4012 | 61.2% |

Table 1 and 2 also show the number of missing values in each variable in the updated data, for the replication of the BGG model and the MSS model, respectively. For the BGG data, except for the World Bank in-house data and the data on donor funding on basic health which we failed to obtain[13] the missing values in the updated data are up to 57% (of 236 countries in the updated data), with the missingness in the illiteracy rate and the rate of paved roads being particularly high. This means that without imputation or without amending the modelling

[13] While in the original BGG data, per capita donor funding for health promotion from the Development Assistance Committee (DAC) of the Organization for Economic Co-operation and Development (OECD) in year 1998 (which we failed to update) is used. However, we obtained a qualitatively similar data, i.e. data on per capita amount of Development Assistance for Health (DAH), available at the Institute of Health Metrics and Evaluation database, which we use in the replication analysis of BGG using data from year 2010.

specification we cannot replicate the BGG analysis with a sufficient number of observations. On the other hand, for the MSS data the extent of missingness is relatively moderate: over 15% of observations are missing for all variables (except for `death by conflict'), and more than 50% of `primary education enrolment rate' are missing.

Given the large numbers of missing values in the data (for both BGG and MSS), we decide to implement a standard multiple imputation by chained equations (MICE) approach, following Hauck et al. (2015), and available as a STATA command ('mi') (StataCorp, 2013).[14] MICE is a regression based imputation method, in which variables with missing values are regressed on the other variables in the dataset, and then the predicted value of the missing part will be imputed. This procedure is sequentially applied to the other variables with missing values, one at a time. In particular, again following Hauck et al. (2015), we use a predictive mean matching technique (Landerman et al., 1997). According to this technique, a missing value is imputed using non-missing values of samples with similar predictive mean. In the present study we use three neighbours of the predictive mean for the imputation (White et al. 2011). Also, since the data are clustered at country and year level, we take this into account in the analysis by including year and country fixed effects in the imputation equations. Throughout the present paper, we replicate the procedure and generate 50 imputed datasets, and hence conduct 50 separate regression analyses at a time, before the estimates are merged into the final estimation result.[15] Finally, if all values of the variables in the regression models are missing for a country, we drop the country from the analysis.[16]

[14] See also Perkins and Neumayer (2014) for an application of MICE in combination with instrumental variables regressions. The programme package "mi" is a set of Stata's official commands.
[15] Once the number of replications is specified, this procedure is automatically conducted using the 'mi' command in Stata 13 MP. For more detail, see StataCorp (2013). The rule-of-thumb suggests that the number of replications should about correspond to the percentage of missing values in the data. Since less than 50% of the data are missing in our data (Table 1, 2 and 3), the number of replications is considered to be sufficiently large.
[16] In the present study we conduct regression analyses for multiple outcome variables (under-5 mortality and adult mortality) separately. However, in the imputation equation we include the other outcome variable, e.g. we use among other variables under-5 mortality to impute missing values of adult mortality. Also, in the regression analysis we use the imputed dependent

Although in principle all variables in the regression model should be included in the imputation equation, in the present study we do not impute our instrumental variables in the chained multiple imputation. This is because a key condition in an instrumental variable approach is that instruments should not be correlated directly with outcome variable (capturing the exclusion restriction) (Wooldridge 2010). Imputing the instrumental variables by (partly) using the outcome variable would make it difficult to justify the exclusion restriction. Instead, for the instrumental variables we impute missing values with the mean of the values from neighbouring observations from the same country, i.e. when the value in time t is missing, we impute the mean of the values in t-1 and t+1. If all values are missing in the country, we again do not impute the missing values, and therefore that country is dropped from the final estimation sample.

Finally, we assessed the performance of the multiple imputations. We randomly chose 1000 non-missing observations, and artificially dropped them from the working data. We imputed the dropped observations (as well as the originally missing observations) using the MICE technique, and then compared the imputed values with the original non-missing values. For all variables used in the streamlined models, the multiple imputations worked well – the mean difference between the original and imputed value was consistently less than one standard deviation point value of each variable.

## 4. Results

In what follows we first present the results of the replication exercise for each of the two approaches – BGG (section 4.1.1) and MSS (section 4.1.2). Subsequently, we present the results of the hybrid approach – in section 4.2.1 for BGG and in section 4.2.2 for MSS.

---

variables. We checked the main regression results by not using the imputed dependent variables but instead using the un-imputed dependent variable, and confirmed that our main conclusion is not affected (details not reported here).

## 4.1 Replication results

### *4.1.1 Replication of BGG*

Table 4 starts by replicating the results of the original approaches used in BGG. Column (1) provides the results as they were given in the original paper. As we obtained the original data (as well as the command files) from the authors of both papers, we then tried to replicate the results using exactly the same specification as in the original specification. The results of this replication correspond exactly to those in the original paper. Table 4 also shows additional test statistics for the first stage regression for multiple endogenous variables: in addition to the standard F statistic, we present an augmented F-statistic allowing for multiple endogenous variables (Sanderson and Windmeijer, 2013), as well as Kleibergen and Paap's Wald rk F statistic for under-identification (Kleibergen and Paap, 2006), showing that the instruments are reasonably 'strong' (the rule-of-thumb threshold for F is 10).

As the next step we replace the original data from the BGG paper by the revised data that is available in the relevant data sources as of August 2015, but for exactly the same year that was covered in the paper, i.e. 2000 (see Column (2) in Table 4). This exercise entailed a reduction in the sample size, down from 127 to 66 countries (due to a particularly large number of missing values in the literacy and paved road variables), reducing the comparability of those specific estimates, with respect to the results of the original BGG paper. The reduction in sample size also provides the case for boosting the sample size using imputation methods.[17] We use two different ways of imputation in Table 4 – first we fill the gaps by using BGG's original data directly, wherever the updated data from readily available sources does not have this information (see Column (3)); second, we use a statistical multiple

[17] BGG's original data are manually imputed using similar data sources (e.g. CIA World Factbook), and by using information from adjacent years, where data for the year of their analysis (2000) was missing. By contrast, here we use a statistical method (i.e. multiple imputations) to impute missing values.

imputation method (MICE) to fill the gaps in the currently available data (see Column (4)).[18] The results in columns (3) and (4) of Table 4 give results that are qualitatively comparable but quantitatively different from the original research – the size of the impacts of the public health expenditure is remarkably larger in the analyses using the updated data with imputation. For under-5 mortality, the original BGG study found that a 1% increase in expenditure will reduce mortality by 0.341% (column (1)). However, when applying MICE to the updated data, the magnitude of the impact is about twice as large as in the original study (the elasticity is -0.564). A similar, but relatively modest increase in the size of the impact is found for maternal mortality (from -0.519 to -0.641). It is important to note that the relevance of the same instrumental variables is now weaker in the updated data, implying the results from the original BGG analysis (1) are more credible than those found in (3) and (4).

[18] We note that the large number of missing variables comes in the readily available data sources, compared to the data used by BGG, is due to extra-efforts expended by BGG in filling the gaps manually by using information from other data sources (e.g. CIA fact book) and by interpolation, in the case of variables that are likely to not fluctuate much from year to year (e.g. paved roads).

## Table 4: Replication of BGG analysis

| | Under 5 mortality | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log public health expenditure | -0.341*** | -0.810*** | -0.502*** | -0.564*** | -1.030*** | -0.873** |
| | (0.127) | (0.245) | (0.178) | (0.213) | (0.376) | (0.358) |
| log public health expenditure X donor funding | 0.007** | 0.009 | 0.009** | 0.006*** | -0.009 | 0.006*** |
| | (0.003) | (0.007) | (0.004) | (0.002) | (0.009) | (0.002) |
| log public health expenditure X paved road | -0.000 | -0.002** | -0.001*** | -0.000 | 0.001 | 0.001 |
| | (0.000) | (0.001) | (0.000) | (0.001) | (0.001) | (0.001) |
| log GDP per capita | -0.404* | 0.174 | -0.167 | 0.182 | -0.591 | 0.242 |
| | (0.210) | (0.341) | (0.303) | (0.361) | (0.433) | (0.499) |
| F for log public health expenditure | 10.508 | 3.337 | 5.185 | 5.216 | 1.525 | 19.747 |
| F for log public health expenditure X donor funding | 135.137 | 117.490 | 34.163 | 23.509 | 33.040 | 178.037 |
| F for log public health expenditure X paved road | 1407.622 | 61.444 | 282.896 | 100.276 | 25.112 | 57.691 |
| F for log GDP per capita | 31.596 | 3.538 | 4.252 | 5.281 | 3.340 | 7.620 |
| SW-F for log public health expenditure | 14.411 | 4.762 | 3.538 | 4.871 | 3.477 | 3.094 |
| SW-F for log public health expenditure X donor fundin | 35.880 | 10.443 | 69.138 | 46.163 | 5.722 | 33.189 |
| SW-F for log public health expenditure X paved road | 69.637 | 118.709 | 52.156 | 64.920 | 11.280 | 28.286 |
| SW-F for log GDP per capita | 14.334 | 4.387 | 5.508 | 5.092 | 3.419 | 3.409 |
| Kleinbergen-Paap Wald statistics | 8.534 | 2.245 | 2.012 | 2.565 | 1.481 | 1.55 |
| N | 127 | 66 | 127 | 127 | 44 | 127 |

| | Maternal mortality | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log public health expenditure | -0.519*** | -1.274*** | -0.769*** | -0.641** | -1.729*** | -0.883** |
| | (0.178) | (0.373) | (0.274) | (0.275) | (0.595) | (0.429) |
| log public health expenditure X donor funding | 0.007 | 0.016* | 0.016*** | 0.010*** | 0.001 | 0.006** |
| | (0.005) | (0.008) | (0.005) | (0.003) | (0.013) | (0.003) |
| log public health expenditure X paved road | 0.000 | -0.002** | -0.003*** | -0.001 | -0.000 | 0.000 |
| | (0.000) | (0.001) | (0.001) | (0.001) | (0.002) | (0.001) |
| log GDP per capita | -0.440 | 0.507 | 0.099 | 0.048 | 0.046 | 0.236 |
| | (0.272) | (0.484) | (0.457) | (0.438) | (0.705) | (0.641) |
| F for log public health expenditure | 10.508 | 3.337 | 5.185 | 5.216 | 1.525 | 19.747 |
| F for log public health expenditure X donor funding | 135.137 | 117.490 | 34.163 | 23.509 | 33.040 | 178.037 |
| F for log public health expenditure X paved road | 1407.622 | 61.444 | 282.896 | 100.276 | 25.112 | 57.691 |
| F for log GDP per capita | 31.596 | 3.538 | 4.252 | 5.281 | 3.340 | 7.620 |
| SW-F for log public health expenditure | 14.411 | 4.762 | 3.538 | 4.871 | 3.477 | 3.094 |
| SW-F for log public health expenditure X donor fundin | 35.880 | 10.443 | 69.138 | 46.163 | 5.722 | 33.189 |
| SW-F for log public health expenditure X paved road | 69.637 | 118.709 | 52.156 | 64.920 | 11.280 | 28.286 |
| SW-F for log GDP per capita | 14.334 | 4.387 | 5.508 | 5.092 | 3.419 | 3.409 |
| Kleinbergen-Paap Wald statistics | 8.534 | 2.245 | 2.012 | 2.565 | 1.481 | 1.55 |
| N | 127 | 66 | 127 | 127 | 44 | 127 |

Notes: Model (1) shows the results as published and replicated; Model (2) shows the results from updated data (year 2000) without imputation; Model (3) shows the results from updated data (year 2000), imputed using the original data where missing; Model (4) shows the results from updated data with multiple imputation; Model (5) shows the results from updated data (year 2010) without imputation; Model (6) shows the results from updated data (year 2010) with multiple imputation.

In Model (5) and (6), since we do not have access to the World Bank in-house data on quality of governance, we impute the variable using the BGG's original data (as of 2000).

In Table 1, non-literate rate and paved road rate exhibit high rates of missing values in the updated data (81.9% and 47.9% , respectively). Since these variables are not likely to vary overtime, we impute the data for 2000 if there are any information available either in 1998, 1999, 2001 and 2002 (two years before and after 2000). The same imputation of the literacy variable is applied to the updated data from 2010.

The BGG original analysis and our replications use data from 2000. In columns (5) and (6) we analyse the updated data for 2010.[19] With and without multiple imputations of the data, the estimated impacts of public health expenditure are greater than the estimated impacts from the 2000 data (in column (6) with multiple imputations, the magnitude of the impact is about -0.9 for both under-5 and maternal mortality).

*Sensitivity checks*

By way of sensitivity checks of the above replication analysis using the 2000 data, starting from the original BGG data, we sequentially update only one variable at a time, in order to assess which variable is causes the larges difference. The results are summarised in Appendix Table A2 and A3. The size of the impact of public health expenditure displays sensitivity with the impact on under-5 mortality now ranging from -0.257 to -0.459 (Appendix Table A3).

#### *4.1.2 Replication of MSS*

When trying to replicate the results of MSS using the original data (column (1) in Table 5), the model failed to produce the results. By comparing the results to those obtained using an older version of Stata, for which we were able to exactly replicate the original results, it turned out that this significant discrepancy to the original results is the result of an updated version of the Stata code used in the IV fixed effect estimation - xtivreg2 (Schaffer, 2010; we

[19] The descriptive statistics of the 2010 data are presented in Appendix Table C3.

used the July 2015 version).[20] Therefore we modified the MSS approach in that we dropped three of the four endogenous variables, leaving only our key variable of interest – public health expenditure – as the one endogenous variable. Re-running the MSS model, using exactly the same data as in the original paper, we obtain largely similar coefficients on the public health expenditure variable as in MSS (see column (2) in Table 5).

As was done with the BGG analysis above, we then substitute the original data from the MSS paper by the revised data that is available in the relevant data sources as of August 2015, but for exactly the same years as those covered in the papers, i.e. 1995-2008 (see column (3) of Table 5). Following this procedure produces less robust results compared to those based on the original data: the coefficient of interest remains statistically significant but reverses sign, suggesting a counter-intuitive relationship between expenditures and mortality rates, i.e. health expenditure *increases* mortality. This is likely the result of the less than ideal instruments for mortality (per capita CO2 emission and number of deaths by conflict), as suggested by the very low relevance of the instruments in the first stage regression (F-statistics is less than 4 in all models), reducing the reliability of the estimates.

[20] The updated code rejects the model as it detects multi-collinearity between the four instrumental variables used in the first stage regression of the MSS model. We used different versions of - ivreg2 - and – ranktest – commands which are both used within –xtivreg2 – routine. In the analyses presented throughout the present study is based on - ivreg2 - version 4.1.08 26 July 2015 and – ranktest – version 1.3.05 22 Jan 2015. Using older version of the routine - ivreg2 - version 2.2.08 15 Oct 2007 and – ranktest – 1.1.02 15 Oct 2007, we successfully replicate the results published in the MSS study.

## Table 5: Replication of MSS analysis

| A: Under 5 mortality | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Public health expenditure | -13.193** | -13.779** | 6.759*** | -41.302 | -0.486 | -4.251*** | -0.013 |
| | (0.018) | (5.794) | (1.915) | (37.414) | (0.321) | (1.498) | (0.082) |
| Out-of-pocket health expenditure | -6.143 | . | . | . | . | . | . |
| | (0.507) | . | . | . | . | . | . |
| Privately pooled health expenditure | 2.685 | . | . | . | . | . | . |
| | (0.594) | . | . | . | . | . | . |
| Immunization coverage | -2.203* | . | . | . | . | . | . |
| | (0.073) | . | . | . | . | . | . |
| 1st stage F of reverse causality estimate | 3.16 | 3.070 | 0.150 | 0.057 | 0.770 | 0.937 | 1.355 |
| 1st stage F of main estimate | . | 7.402 | 21.979 | 1.285 | 2630.650 | 40.332 | 5.61E+07 |
| Country | 153 | 153 | 154 | 159 | 166 | 159 | 162 |
| N | 1397 | 1398 | 1366 | 1602 | 1569 | 2217 | 2583 |
| **B: Adult female mortality** | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Public health expenditure | -2.583** | -3.957** | 12.224** | 7.483 | 0.552 | -1.041** | -0.591* |
| | (0.050) | (1.552) | (5.908) | (4.639) | (0.468) | (0.479) | (0.310) |
| Out-of-pocket health expenditure | 5.153 | . | . | . | . | . | . |
| | (0.161) | . | . | . | . | . | . |
| Privately pooled health expenditure | -23.385** | . | . | . | . | . | . |
| | (0.040) | . | . | . | . | . | . |
| Immunization coverage | -9.841** | . | . | . | . | . | . |
| | (0.030) | . | . | . | . | . | . |
| 1st stage F of reverse causality estimate | 7.19 | 7.341 | 3.763 | 2.321 | 4.195 | 3.148 | 4.985 |
| 1st stage F of main estimate | . | 4857.69 | 36.830 | 34.091 | 71586589.1 | 226878.2 | 110959.4 |
| Country | 148 | 148 | 153 | 158 | 165 | 158 | 162 |
| N | 1222 | 1223 | 1359 | 1677 | 1562 | 2203 | 2699 |
| **C: Adult male mortality** | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Public health expenditure | -2.210** | -2.944*** | 8.854** | 6.148* | 0.646 | -0.620 | -0.542** |
| | (0.025) | (1.026) | (3.704) | (3.475) | (0.552) | (0.382) | (0.238) |
| Out-of-pocket health expenditure | 8.731 | . | . | . | . | . | . |
| | (0.172) | . | . | . | . | . | . |
| Privately pooled health expenditure | -15.545** | . | . | . | . | . | . |
| | (0.016) | . | . | . | . | . | . |
| Immunization coverage | -7.858** | . | . | . | . | . | . |
| | (0.020) | . | . | . | . | . | . |
| 1st stage F of reverse causality estimate | 11.60 | 12.021 | 2.270 | 1.487 | 5.173 | 1.859 | 2.615 |
| 1st stage F of main estimate | . | 13790.4 | 89.679 | 52.797 | 4.216e+08 | 499648.8 | 243220.4 |
| Country | 148 | 148 | 153 | 158 | 165 | 158 | 162 |
| N | 1222 | 1223 | 1359 | 1677 | 1562 | 2203 | 2699 |

Notes:

- Model (1) shows the results as published
- Model (2) shows the results (public health expenditure only) from original data using updated Stata 13 (authors' own estimation)
- Model (3) shows the results from updated data (1995-2008) without imputation
- Model (4) shows the results from updated data (1995-2011) without imputation
- Model (5) shows the results from updated data (1995-2008), imputed using the original data where missing

- Model (6) shows the results from updated data (1995-2008) with multiple imputation
- Model (7) shows the results from updated data (1995-2011) with multiple imputation

However, once we impute the data again, first by filling the (now very few) gaps in the data with values from the original MSS data (Column (5)) and second by using multiple imputation techniques (Column (6) and (7), now with considerably larger sample size), the sign of the coefficients reverse again to the expected direction, but their magnitude still varies considerably between the specifications, and the instruments for mortality continue to be 'weak', i.e. they produce only fairly low F-statistics, according to the benchmarks suggested by Stock and Yogo (2005) and Sanderson and Windmeijer (2013).[21]

*Sensitivity checks*

Similarly to section 4.1.1, we assess which variable is causing the discrepancy between the estimation results using the original and updated data, by replacing every variable in the original data with the corresponding variable in the updated data. Appendix Tables A4 and A5 show the results. For under-5 mortality, replacing CO2 emission and GDP per capita lead to counter-intuitive estimates, while for adult female/male mortality, replacing education and public health expenditure lead to counter-intuitive results.[22] Similar to the BGG analysis above, the magnitude of the impact varies considerably upon updating only one variable at a time: the effect estimates range from +1.105 to -14.774 (Appendix Table A5).

## 4.2 Hybrid Model Analysis

We use a hybrid approach in which we focus on estimating the impact of public health care expenditure using approaches that are broadly (though no longer literally) in line with those employed in the previous literature, including specifically BGG and MSS.

[21] In the present analysis we follow Sanderson and Windmeijer's F statistics for relevance of the instruments. If the endogenous variable(s) are just identified i.e. the number of instruments equals the number of endogenous variables, this augmented F statistic coincides with the standard F statistic.

[22] As explained in the Data section, the original MSS study use per capita GDP in 2005 constant US dollars (PPP adjusted), whereas in the updated data analysis we use per capita GDP in 2011 constant US dollars (PPP adjusted).

Table 6 presents the key results of this model, using different strategies to estimate the causal impacts of public health expenditure: fixed effect instrumental variable approach, pooled cross-section IV approach with additional (time invariant) instruments. Comparing the estimation results with and without the statistical multiple data imputation, we note again the considerable numbers of missing values in the raw data (see Table 3 for more details on missing values).[23]

#### *4.2.1 Hybrid model analysis BGG: the 'traditional IV approach'*

In the case of the traditional IV approach, using the hybrid model, the results do not confirm the large and significant impacts found in the BGG study. In the first stage regression, the relevance of the instrumental variable (log of per capita military expenditure of neighbouring countries) is too weak to identify the causal effect, and hence the large standard errors of the point estimate, making the model essentially invalid.

[23] The publicly available mortality data are frequently estimated rather than directly observed (Alkema and You, 2012). In order to (partially) avoid the potential problem of serial correlation due to the data estimation, we analysed the same econometric model with more sparse, five-year interval data (1995, 2000, 2005, and 2010). The result is presented in Appendix Table A8. The results are qualitatively similar to the results in Table 6.

**Table 6: Traditional IV and Bruckner IV approach using streamlined model**

| | Traditional IV | | | | Bruckner IV | | | |
|---|---|---|---|---|---|---|---|---|
| | Fixed effect | | Pooled cross section | | Fixed effect | | Pooled cross section | |
| **A: log under 5 mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log Public health expenditure | 0.159 | 0.682 | 0.683 | 0.053 | 0.110** | -0.224** | -0.185* | -0.063 |
| | (0.753) | (0.782) | (0.629) | (0.200) | (0.047) | (0.103) | (0.098) | (0.040) |
| First stage F of reverse equation | . | . | . | . | 1.262 | 3.707 | 3.584 | 2.335 |
| First stage F of main equation | 0.245 | 0.748 | 0.785 | 1.761 | 2323.615 | 57.13 | 2413.079 | 79102.720 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| Country | 100 | 153 | 115 | 148 | 101 | 155 | 105 | 155 |
| N | 661 | 2601 | 648 | 2516 | 666 | 2631 | 634 | 2631 |
| **B: log adult female mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log Public health expenditure | 0.203 | 0.293 | 0.025 | -0.033 | 1.836** | 0.100*** | -0.063 | -0.029 |
| | (1.102) | (0.532) | (0.377) | (0.213) | (0.792) | (0.031) | (0.060) | (0.050) |
| First stage F of reverse equation | . | . | . | . | 0.096 | 4.408 | 6.940 | 4.168 |
| First stage F of main equation | 0.108 | 0.613 | 0.803 | 1.728 | 6.040 | 1107.273 | 1424.858 | 1924.423 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| Country | 99 | 153 | 114 | 148 | 100 | 155 | 105 | 155 |
| N | 642 | 2560 | 631 | 2477 | 647 | 2590 | 618 | 2590 |
| **C: log adult male mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log Public health expenditure | 0.194 | 0.224 | 0.147 | -0.156 | 0.337*** | -0.144* | -0.037 | -0.008 |
| | (1.188) | (0.397) | (0.345) | (0.156) | (0.077) | (0.082) | (0.057) | (0.031) |
| First stage F of reverse equation | . | . | . | . | 1.293 | 3.526 | 8.135 | 4.827 |
| First stage F of main equation | 0.108 | 0.613 | 0.803 | 1.728 | 82.351 | 31.314 | 2050.833 | 11815.54 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| Country | 99 | 153 | 114 | 148 | 100 | 155 | 105 | 155 |
| N | 642 | 2560 | 631 | 2477 | 647 | 2590 | 618 | 2590 |

Note: The estimates presented in the columns labelled 'pooled cross section' include more time invariant instrumental variables in addition to original set of time varying instruments. For Traditional IV approach, additional instruments for public health expenditure include the legal origin of the country: British, French, Socialist, German and Scandinavian laws (Rajkumar and Swaroop, 2008; Hu and Mendoza, 2013). For Bruckner IV approach, additional instruments for mortality include indicators of climate zone based on Kopper's classification and malaria ecology index (Lorentzen et al. 2008 and Kizsewski et al. 2004)

*Sensitivity checks*

Since the results in Table 6 (columns under 'Traditional IV') highlight the limitations in the existing instruments with the updated data, we have sought to improve the quality of the predictive power of the instruments by experimenting with different functional forms of the original instruments and by using alternative ones, following suggestions from the literature. The outcome of this exercise is presented in Appendix Table A6 for the traditional IV approach. Despite these efforts, we fail to find results that would confirm our expectations.

Next, in the hybrid model, per capita GDP is treated as an exogenous variable, whereas in the original BGG specification it is endogenous, and there is hence a concern that this could have led to biased estimates due to model misspecification. To examine this possibility, we conduct an analysis that does treat per capita GDP as another endogenous variable. Following BGG, the consumption-investment ratio was used as the additional instrumental variable for per capita GDP. Appendix Table A9 shows the results. Although it is interesting to see that the models with the multiple imputations show that higher per capita GDP significantly reduces under-5 mortality, we fail to find that public health expenditure reduces mortality outcomes.

#### *4.2.2 Streamlined model analysis MSS: the 'Bruckner IV approach'*

As for the Bruckner IV approach, the results using multiple data imputation appear mixed as well, showing the expected negative and statistically significant impacts of public health expenditure in particular for under-5 mortality, though less so for adult mortality (see Table 6 under 'Bruckner IV'). The instrumental variables turn out less relevant in a statistical sense (informed by F statistics of the first stage regression). The magnitude of the elasticity is substantially smaller than what was implied in the original studies. For instance, where intuitive direction and statistical significance are achieved, the magnitude of the impact on under-5 mortality is -0.224, and that on adult male mortality is -0.144, i.e. a 1% increase in expenditure leads to a reduction of mortality by 0.224 or 0.144%.

*Sensitivity checks*

Testing a set of different IVs, in an effort to overcome the limitations in the existing instruments with the updated data (see Appendix Table A7), provides the same conclusions as in the 'traditional IV' case above: we obtain insignificant or counter-intuitive results, the relevance of the instruments remains low and the discrepancy to the results in the original studies is large.

### 4.3 Specification and robustness checks

In Appendix B1, we conduct further specification checks of the analysis based on the hybrid model. Notably, the main models presented above have examined the impact of a change in expenditure on a change in mortality in the same year. Also, in the main model we only address the impact of public health expenditure, ignoring other sources of health care spending. First, we investigated the lagged (t-1, t-2, or t-5) impacts of public health expenditure on current mortality outcomes, with only one lag-specification included at a time. Also, we examined the impact of total health expenditure (public and private expenditure). For more detailed description, see Appendix B.1. The results of the analyses are presented in Appendix Table B1. For the Traditional IV approach, again the estimation results are not reliable due to low F statistics. For the Bruckner IV approach, when multiple imputation is applied, we again find a small, negative and statistically significant impact for under-5 mortality – the size of elasticity is around -0.1 to -0.2 for t-1,t-2 and t-5 models, and the impact diminishes upon increasing the number of lags. For adult female and male mortality, the effect is even smaller and statistically indistinguishable from zero. Similarly, the impacts of total expenditure were not statistically significant or intuitive, except for the case of Bruckner's IV approach (see Block D, Column (4) of Table B1), where the elasticity of (current) total health expenditure on under-5 mortality is -0.299.

We then allow for lagged structure in the estimation of the impact of expenditure. We include current and lagged health expenditure (t-1, t-2 and t-3) *in the same econometric model* and estimated the conditional impacts of expenditure at each period on current mortality outcomes. The results are summarised in Appendix Table B2. The results from the Traditional IV approach show non-significant impacts of contemporaneous as well as any lagged public health expenditure in all models. For the Bruckner IV approach, when the multiple imputations are applied, the impact of contemporaneous public health expenditures (column (4)) is negative and statistically significant for under-5 mortality and adult male mortality. However, the impacts of lagged public health expenditure (conditional on contemporaneous expenditure) are small and largely statistically insignificant.

The final set of specification checks involves different assumptions for the identification of causal impacts. Although in the main analysis we use instruments that were demonstrated to be statistically relevant in the previous cross-country data studies, we fail to confirm this in the updated data and in the hybrid model. This could suggest that alternative analytic approaches are called for that do not rely on the identification of suitable external instruments. Following this idea, we pursue two approaches: first, we use lagged health expenditure to avoid reverse causality while still using fixed effect modelling to take into account time-invariant country heterogeneity[24]; second, we incorporate a dynamic structure into the model and construct a set of IVs from observations from previous years – a dynamic panel data approach (Blundell and Bond, 1998; Ssozi and Amlani, 2015). The results of the estimations from the two approaches, with and without the data imputation, are presented in Table Appendix B3. In the table, Block A, B and C use only fixed effect effects to estimate the

[24] We make a stronger assumption about exogeneity of lagged health expenditure here than the previous models presented in this section. In fact, lagged health expenditure is predetermined but not necessarily exogenous.

effect of contemporaneous and lagged health expenditure. In Block D, the estimates from Blundell-Bond estimation are presented. The estimates of the impact of health expenditure turn out rather mixed, though overall the magnitudes of the impacts, whether statistically significant or not, are much smaller than expectations based on the results from BGG and MSS.[25]

## 5. Discussion

In this paper we have sought to re-examine the robustness of the empirical relationship between public health care spending and mortality outcomes using cross-country data, as it was presented in the main approaches used by the previous literature. We have based our analysis on the data and approaches embodied by two key publications in the field – by Bokhari et al. (2007) and Moreno-Serra and Smith (2015). We started by quite successfully replicating the published results, using the exact same data used by the studies. Results were more nuanced once we made use of data for the same year(s) as in the original study but available from publicly accessible sources at the time of data retrieval (August 2015). In 'updating' the data in this way, while still following the specific estimation strategies from the original papers, we also employed different imputation strategies to fill major data gaps for some variables. On the whole we found the BGG-based results to still hold qualitatively, though with increased effect size and with weaker performance of the instruments. By contrast, the MSS-based results were substantively affected, and the relevance of the instruments did deteriorate, producing less resemblance to the original results. Across the two studies, in the majority of replication scenarios the effect of public health care expenditure on mortality was found to be negative, although qualitative conclusions based on

[25] The results of the model using the imputed data should be interpreted with caution. As P-values of AR tests indicate, there is no second-order serial correlation. However, the Hansen test indicates that the over-identifying restriction is not valid, meaning that some of the IVs are endogenous.

statistical significance vary and the magnitude of the impact is sensitive to revisions and updates in the data used. [26]

In a further step, we abstracted from the specific estimation strategies in the original papers, among others in order to follow the spirit of the approaches used, without the need for overly strict adherence to their idiosyncratic features, and in order to capture some commonalities with other previous studies in this field. We thus constructed our hybrid versions of the two approaches, complemented by a battery of robustness checks, as well as some changes to the specifications to accommodate certain other empirical approaches used in the literature (e.g. dynamic panel analysis). In sum, the results showed that in the traditional IV model our instrument for public health care expenditure (per capita military expenditure of neighbouring countries) is not statistically relevant and thus the IV estimation coefficients become unreliable. For the Bruckner IV streamlined model, at least in some cases the estimated effects were broadly consistent with the previous literature, finding that public spending reduced mortality, although at generally lower effect magnitudes and with mixed degrees of statistical significance.

Our very mixed findings should not be taken to imply that the conclusions from the previous studies are necessarily invalid. Countries covered in the published data and in our updated data do indeed differ and therefore it is to be expected that the analysis leads to different conclusions. Nevertheless, our findings do highlight the potential lack of generalisability of the results to different settings.

[26] Similar sensitivities of the results to updates in the underlying data have been encountered by Easterly et al. (2004), Roodman (2008), Clemens et al. (2012), and Roodman (2015) in the context of replicating the results from the literature on the impact of aid on economic growth.

Our conclusions need to be seen in the light of the considerable methodological challenges faced by this strand of research, many of which may form the basis of a potential research agenda in this area, as discussed below.

### 5.1 Data limitations

The data used for the replication of the two key papers have been obtained directly from the authors of the paper in the case of Bokhari et al (2007) and from data made available via the web appendix for the Moreno Serra & Smith (2015) paper. With very few exceptions, this data is in principle available from readily and freely accessible data sources, e.g. the WHO's Global Health Observatory and the World Bank's World Development Indicators. These (as well as several other publicly accessible databases) are also the data sources we have used when updating and expanding the data employed in our empirical estimates. While the data from these sources represent the most comprehensive ones available and have been (and most certainly will be) used in numerous well-published studies[27], they are not without quality concerns that may affect the reliability of our impact estimates. This is the case despite the considerable and laudable efforts that international organizations or research institutes have invested to try and establish common standards of how data should be defined and collected. The degree to which quality issues are present will vary by indicator and across countries. For instance, the data collection for one of our key variables – GDP per capita – should be approximately similar across countries given longstanding efforts to ensure consistent definition of this measure.

Measurement error is likely more severe in the case of our key independent variable (i.e. public health expenditures). As the WHO's methodological background note (World Health

[27] The literature in other development areas, e.g. on the role of aid on economic growth, has been shown to suffer from similar data challenges (Easterly et al. 2004; Roodman, 2008).

Organization, 2012) to their Global Health Expenditure database acknowledges, there are major gaps in the data delivered to WHO from countries (which, if it is delivered, is typically taken on at face value), with the gaps then being filled by various forms of interpolation[28], extrapolation, imputation and estimation.

Random measurement error in the health expenditure variable would normally bias the estimates of the impact on mortality towards the null, while random measurement error in our dependent variable (mortality) would produce higher standard errors of the impact estimates, reducing the statistical significance of the results (Wooldridge, 2010).[29] As health expenditure enters as a right hand side variable in our regression model, it is likely that a valid instrumental variable strategy could reduce some of the bias resulting from such measurement error. In our case though, for many of the empirical models the instruments used were weak – and it is this that has arguably been a far greater problem in our set of estimates, compared to data quality issues. Moreover, there are at least two sources of measurement error – one embedded in the numbers reported to the data collecting agencies (i.e. WHO), and an additional one as a result of the (not always fully replicable) imputation approaches adopted by such agencies. In this case it is at least conceivable (though very hard to tell for certain) that measurement error in the health expenditure variable may account for the weakness of the instrumentation in the first stage estimates.

Arguably, measurement error is far more severe in the case of our dependent variables of interest, i.e. the under-5 and gender-specific adult mortality-rates – the underlying problem being the often complete absence of functioning vital registration systems in particular in

[28] Given that some of the original variables are estimated using previous years' observations, in Appendix Table A8 we present the same analysis based on the streamlined model where we use quinquennial data.
[29] Note also that in case the measurement error is non-random but constant over time, then the fixed effects approach would resolve this problem.

many low income countries (Alkema and You, 2012). In response, an impressive research 'industry' has emerged, developing and applying sophisticated modelling methods to fill the mortality data gaps statistically. The fact that different research teams find considerable differences in, for instance, their under-5 mortality estimates for many LMICs does, however, underscore the measurement problem at hand.

Short of obtaining the 'perfect' data (which will never happen), there are, however, ways in which more purposefully selected, specific data sources would be worthwhile using, if there is reason to believe that they provide better quality data, and even at the cost of reducing the country coverage. Examples on the mortality side might include using data from Demographic and Health Surveys only or on the expenditure side to focus on information from detailed IMF Government Finance Statistics, the World Bank's Poverty Assessments and Public Expenditure Reviews (see e.g. Gupta et al., 2003).

### 5.2 Methodological limitations

There are at least three methodological issues to be borne in mind.

First, following the previous literature that predominantly accounts for country-level heterogeneity and for endogeneity of health expenditure and mortality, we have employed a panel fixed effect approach, combined with instrumental variables techniques. Incorporating country fixed effect in the model means that we compromise statistical power because the estimation relies only on within-country variations in health care expenditure and mortality. Because there are relatively small variations in these variables over time, and also because the analysis often relies on 'weak' instrumental variables, our methods (though being standard in the literature) may be of insufficient power to statistically detect the impact. This potential downside of the fixed-effect approach, however, has to be weighed up against the likely considerable benefits of controlling more adequately for endogeneity through the regression control over unobserved heterogeneity in time-invariant characteristics and also in factors that change slowly over time such as cultural factors or the quality of institution.

Second, our results using the multiple imputations (MI) should be interpreted with caution. Although MI has been widely adapted in the applied statistic literature, its applications in health econometric models (in particular as part of instrumental variables approaches) have been rare (Perkins and Neumayer, 2014). We carefully construct our imputation so that it does not violate the exclusion restriction assumption in the IV approach. Yet detailed statistical properties of the application of the MI approach should be expanded upon in future research. That said, and as we showed, the alternative of not imputing the data would mean that the econometric analysis would have to do without a considerable number of countries with any missing values in the model specifications, producing a highly unrepresentative sample and causing selection bias, the impacts of which in terms of main effects are a priori unknown (though likely significant). At the same time, it must be noted that the statistical imputation increases the likelihood of measurement error in the regressions, thereby augmenting the impact estimates as well as the standard errors.

Third, the instrumental variable technique has been the standard workhorse for causal analysis in the relevant literature. As in the present study, the validity and relevance of the instruments can generally be questioned – i.e. some instruments occasionally prove to be good in a statistical sense, but that could be sensitive to choice of data and other variables in the model. Furthermore, it should be noted that the instrumental variable approach does not necessarily provide average treatment effects (ATE), which would be the key quantity of interest, but instead it gives the local average treatment effect (LATE), in which the impact of health expenditure is identified only for the subset of countries for which expenditure is *actually* affected by the instruments (Angrist and Imbens, 1995; Angrist et al., 1996). In the literature it is common to assume homogeneity of the IV impact, rather than allowing for heterogeneity of the impacts across countries, so that the estimated IV impacts conveniently represents ATE, despite that in practice it is hard to believe that the impacts of expenditure will be common across countries.  For simplicity, the present study, as did previous studies in the literature, focuses on assumed homogeneous impacts and abstains from discussing likely differences between the ATE and LATE in this context.[30]

### 5.3 Policy implications

The primary policy question is whether or not greater public health care expenditures improve population health.  The two prominent studies – BGG and MSS – claim that health expenditures do reduce mortality, and that the magnitude of the effect is considerable, hence supporting greater publicly funded health coverage across countries.  However, the cross-country based work presented here (as well as some of previous research, e.g., Filmer and Pritchett 1999) shows that the magnitude and precision of this effect is sensitive to the choice of data and econometric methods.  Our results do not necessarily suggest that policies expanding public health expenditure do not save lives at all. Rather, our results confirm the common problems in cross-country data analysis – that the noise in the cross country data may be so large that the current econometric methods could sometimes fail to robustly detect the health gains at global level. The observation that impact estimates can vary greatly and may also be small in magnitude and/or insignificant is at least partly confirmed by the meta-

[30] In this context it is of note that Gallet and Doucouliagos (2015) report in their meta-regression analysis of the relevant empirical studies (which do include within-country studies) that the magnitudes of the elasticity of life expectancy found by studies using IV approaches are significantly larger than comparable 'baseline' non-IV estimates.

analysis in Gallet and Doucouliagos (2015), who find a considerable spread in the impact estimates of health expenditures on mortality, as well as a low mean elasticity (i.e. 0.079).[31]

Another important implication of the analysis would be the extent to which our estimates reliably inform marginal productivity of healthcare systems across countries – the key variable that would be needed to inform country-specific cost-effectiveness thresholds (Ochalek et al 2015). In light of the limited robustness to changes in model specification and in data of the empirical findings, coupled with imprecise estimation (due to weak IVs), our results again may not uncover the true (if any) productivity of healthcare systems across countries, and therefore further improvement in either data, empirical methods, or even research design are required.[32]

### 5.4 Implications for future research

Future research should involve the improvement in both the methods for the analysis of cross country data and in data quality. Global efforts to obtain finer data should improve the data quality over time. Some arguably 'better' data already exists, if at the cost of a considerably smaller sample of countries; for example, international household surveys (e.g. the World Bank's Living Standard Measurement Surveys) do provide finer data on health and household healthcare spending at more granular level (e.g. observations at household level), which could potentially overcome some of the key limitations in the cross country data.

[31] Using a meta-regression model of Gallet and Doucouliagos (2005) (using their "Specific" model of conditional mortality rate), the predicted elasticity of the BGG study is -0.283 for under-5-mortality, and -0.383 for maternal mortality. As MSS did not estimate elasticities their study is not included in the Gallet and Doucouliagos' meta-analysis.

[32] An online tool based on Ochalek et al. (2015) has been made available at: https://www.york.ac.uk/media/che/documents/Calculating%20cost%20per%20DALY%20averted%20thresholds%20for%20LMICs.xlsm, which enables users to enter new estimates of the elasticity of health outcomes with respect to public health care expenditure, or to carry out sensitivity analysis when using existing published estimates when estimating cost-effectiveness thresholds for LMICs.

Since the impact of public health expenditure on mortality could also be heterogeneous across countries, future research should explore such heterogeneity. The BGG model does allow for differential impacts, conditional on the level of economic development (proxied by the proportion of paved roads) and on the amount of donor funding for health. Using the updated data for the year 2000 in the BGG specification, Appendix Figure A1 summarises the distribution of the impacts of public health expenditure on mortality. We find that the impacts for under-5 mortality are considerably more homogenous than for maternal mortality. Similarly, Wagstaff and Claeson (2004) found that the mortality-reducing impact of public expenditure becomes evident only once the quality of governance exceeds a certain threshold level. It may hence be worthwhile to assess possible additional factors that mitigate or promote the impact of public health expenditure, including, for instance, baseline mortality. Quantile treatment effect models could be useful to test and quantify the impacts of public health expenditure across countries (Chernozhukov, and Hansen 2005).

Another possibility would be to use within-country data to investigate the domestic impact of health expenditure on health outcomes. Although there are is some such research in high income countries (for example, Martin et al., 2008; Card et al., 2009; Claxton et al., 2015), the evidence is still limited in LMICs, where policy issues relating to universal health coverage become more relevant and also healthcare resources are generally scarce, except for some existing studies in major countries such as India (Bhalotra, 2007; Farahani et al., 2010).

At the same time, our findings add further weight to the case for journals to require authors to submit their data along with their paper, so that the data becomes permanently available to future scholars, thereby allowing for the discrepancies resulting from retrospective corrections to be traced to changes in the data. The problem we encountered when trying to replicate the exact original MSS results using an updated, revised version of a Stata command (xtivreg2) also suggest that whenever authors use a public-domain user-written program, that the journal should require them to provide the program version information as well as the data and relevant do-files.

Likewise, the agencies producing and publishing the data would benefit research by making available each version of the data, rather than just the latest one that would supersede all previous ones. The producers of such data could also usefully increase the degree of transparency about the exact methods that have been used in imputing gaps in the underlying data, and indeed the extent to which imputation has been undertaken.

Finally, the current literature is largely based on an instrumental variable approach in estimating causal impacts of health expenditure, despite the fact that instruments could be sometimes questioned (even if statistically relevant occasionally, they are not always theoretically convincing). An alternative approach would be to use actual policy variations in healthcare expenditure to inform the impact estimation. Examples of sources of such variations include expansion of public insurance schemes (see Acharya et al., 2012 for a comprehensive review), funding decision rules in healthcare (Dykstra et al., 2015), and large scale field or natural experiments that involve the provision of additional healthcare resources/funding (Wagstaff, 2011; Olken et al., 2014). Carefully evaluating those policy and experimental shocks on expenditure could possibly provide more transparent and robust evidence.

**References**

Acharya A, Vellakkal S, Taylor F, Masset E, Satija A, Burke M, Ebrahim S. *The Impact of Health Insurance Schemes for the Informal Sector in Low-and Middle-Income Countries: A Systematic Review*. The World Bank Research Observer 2012; lks009.

Afridi MA, Ventelou B. Impact of health aid in developing countries: The public vs. the private channels. *Economic Modelling* 2013;31; 759-765.

Akinci F, Hamidi S, Suvankulov F, Akhmedjonov A. Examining the impact of health care expenditures on health outcomes in the Middle East and N. Africa. *Journal of Health Care Finance* 2014.

Alkema L, You D. Child Mortality Estimation: A comparison of UN IGME and IHME estimates of levels and trends in under-five mortality rates and deaths. *PLoS Medicine* 2012;9.

Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 1995;90; 431-442.

Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996;91; 444-455.

Angrist JD, Pischke J-S. *Mostly Harmless Econometrics: An Empiricist's Companions*. Princeton University Press; 2009.

Anyanwu JC, Erhijakpor AEO. *Health Expenditures and Health Outcomes in Africa*. African Development Review 2009;21; 400-433.

Bhalotra S. Spending to save? State health expenditure and infant mortality in India. *Health Economics* 2007;16; 911-928.

Bidani B, Ravallion M. Decomposing social indicators using distributional data. *Journal of Econometrics* 1997;77; 125-139.

Blundell R, Bond S. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 1998;87; 115-143.

Bokhari FAS, Gai Y, Gottret P. Government health expenditures and health outcomes. *Health Economics* 2007;16; 257-273.

Brückner M. On the simultaneity problem in the aid and growth debate. *Journal of Applied Econometrics* 2013;28; 126-150.

Card D, Dobkin C, Maestas N. Does Medicare save lives? *The Quarterly Journal of Economics* 2009;124; 597-636.

Chernozhukov V, Hansen C. An IV model of quantile treatment effects. *Econometrica* 2005;73; 245-261.

Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith P, Sculpher M. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *HealthTechnology Assessment* 2015;19; 1.

Clemens MA, Radelet S, Bhavnani RR, Bazzi S. Counting chickens when they hatch: timing and the effects of aid on growth. *The Economic Journal* 2012;122; 590-617.

Dykstra S, Glassman A, Kenny C, Sandefur J. *Are Vaccines Fungible? Regression Discontinuity Evidence from a Large Aid Program*. Center for Global Development Working Paper 394;2015.

Easterly W, Levine R, Roodman D. Aid, Policies, and Growth: Comment. *The American Economic Review* 2004;94; 774-780.

Farag M, Nandakumar AK, Wallack S, Hodgkin D, Gaumer G, Erbil C. Health expenditures, health outcomes and the role of good governance. *Int J Health Care Finance Econ* 2013;13; 33-52.

Farahani M, Subramanian SV, Canning D. Effects of state-level public spending on health on the mortality probability in India. *Health Economics* 2010;19; 1361-1376.

Filmer D, Hammer JS, Pritchett LH. Weak links in the chain: a diagnosis of health policy in poor countries. *The World Bank Research Observer* 2000;15; 199-224.

Filmer D, Pritchett L. The impact of public spending on health: does money matter? *Social Science & Medicine* 1999;49; 1309-1323.

Gani A. Health care financing and health outcomes in Pacific Island countries. *Health Policy and Planning* 2009;24; 72-81.

Gallet CA, Doucouliagos C. The impact of healthcare spending on health outcomes: a meta-regression analysis. Economics Series SWP 2015/11, 2015. Deakin University Australia.

Gravelle HSE, Backhouse ME. International cross-section analysis of the determination of mortality. *Social Science & Medicine* 1987;25; 427-441.

Gupta S, Verhoeven M, Tiongson ER. Public spending on health care and the poor. *Health Economics* 2003;12; 685-696.

Hauck K, Martin S, Smith P. Social Determinants of Health in Low Income Countries: Do the Data support the Rhetoric? Mimeo 2015.

Hu B, Mendoza RU. Public health spending, governance and child health outcomes: Revisiting the links. *Journal of Human Development and Capabilities* 2013;14; 285-311.

Kizsewski A, Malaney P, Mellinger A, Spielman A, Sachs SE, Sachs JD. A global index of the stability of malaria transmission based on the intrinsic properties of anopheline mosquito vectors. *American Journal of Tropical Medicine and Hygiene* 2004;70; 486-498.

Kleibergen F, Paap R. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 2006;133; 97-126.

Landerman LR, Land KC, Pieper CF. An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods & Research* 1997;26; 3-33.

Lorentzen P, McMillan J, Wacziarg R. Death and development. *Journal of Economic Growth* 2008;13; 81-124.

Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost–effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization* 2015;93; 118-124.

Martin S, Rice N, Smith PC. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of Health Economics* 2008;27; 826-842.

Moreno-Serra R, Smith PC. Broader health coverage is good for the nation's health: evidence from country level panel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2015;178; 101-124.

Newall A, Jit M, Hutubessy R. Are current cost-effectiveness thresholds for low-and middle-income countries useful? Examples from the world of vaccines. *Pharmacoeconomics* 2014;32; 525-531.

Ochalek J, Lomas J, Claxton K. 2015. Cost per DALY thresholds for low and middle income countries: evidence from cross country data. *Centre for Health Economics Reseach Paper 122. University of York: York;* 2015.

Olken BA, Onishi J, Wong S. Should aid reward performance? Evidence from a field experiment on health and education in Indonesia. *American Economic Journal: Applied Economics* 2014;6; 1-34.

Perkins R, Neumayer E. Adoption and compliance in second-hand smoking bans: a global econometric analysis. *International Journal of Public Health* 2014;59; 859-866.

Rajkumar AS, Swaroop V. Public spending and outcomes: Does governance matter? *Journal of Development Economics* 2008;86; 96-111.

Revill P, Walker S, Madan J, Ciaranello A, Mwase T, Gibb DM, Claxton K, Sculpher MJ. 2014. Using cost-effectiveness thresholds to determine value for money in low-and middle-income country healthcare systems: Are current international norms fit for purpose*? Centre for Health Economics Research Paper No 98. York*; 2014.

Roodman D. Through the looking glass, and what ols found there: on growth, foreign aid, and reverse causality. *Center for Global Development Working Paper* 2008.

Roodman D. A replication of ‘counting chickens when they hatch’ (Economic Journal 2012). *Public Finance Review 2015*;doi:10.1177/1091142114537895.

Sanderson E, Windmeijer F. 2013. A weak instrument F-test in linear IV models with multiple endogenous variables. *CEMMAP working paper, Centre for Microdata Methods and Practice;* 2013.

Schaffer ME. xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. 2010. Available at: http://ideas.repec.org/c/boc/bocode/s456501.html

Ssozi J, Amlani S. The effectiveness of health expenditure on the proximate and ultimate goals of healthcare in sub-Saharan Africa. *World Development* 2015; 76; 165-179.

StataCorp. *Stata Multiple-Imputation Reference Manual Release 13*. StataCorp LP: College Station TX; 2013.

Stock JH, Yogo M. Testing for weak instruments in linear IV regression. In andrews DWK. *Identification and Inference for Econometric Models.* Cambridge University Press, New York, 2005.

Wagstaff A. Fungibility and the impact of development assistance: Evidence from Vietnam's health sector. *Journal of Development Economics* 2011;94; 62-73.

Wagstaff A, Claeson M. Rising to the challenges: the millennium development goals for healths. *World Bank*; 2004.

White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* 2011;30; 377-399.

Wooldridge JM. Econometric analysis of cross section and panel datas. *MIT press*; 2010.

World Health Organization. *General Statistical Procedures Used to Construct WHO Health Expenditure Database.* Geneva: World Health Organization. 2015.

World Health Organization. *Cost Effectiveness and Strategic Planning (WHO-CHOICE).* 2015.
Available at: http://www.who.int/choice/costs/CER_levels/en/ [Accessed December, 2015].

World Health Organization Commission on Marcroeconomics and Health. 2001. *Macroeconomics and Health: Investing in Health for Economic Development*. World Health Organization: Geneva; 2001.

## Appendix – Not for publication

## Appendix A

- Table A 1: List of sources of updated data
- Table A 2: BGG analysis using the original data
- Table A 3: BGG analysis using the original data: updated data are imputed using original data where missing
- Table A 4: MSS analysis using the original data
- Table A 5: MSS analysis using the original data: updated data are imputed using original data where missing
- Table A 6: Traditional IV approach (with streamlined model) using different specifications of instrumental variables
- Table A 7: Bruckner IV approach (with streamlined model) using different specifications of instrumental variables
- Table A 8: Regression analysis based on data from 1995, 2000, 2005, and 2010 only (replication of the analysis in Table 6, first column with fixed effect IV model)
- Table A 9: Traditional IV approach (with streamlined model) in which logGDP per capita is treated as an endogenous variable
- Table A10: Traditional IV and Bruckner IV approach using streamlined model, estimated using Limited Information Maximum Likelihood (A replication of Table 6)
- Figure A 1: Heterogeneity of the impact of public health expenditure on mortality based on the BGG specification, using year 2000 updated data

## Table A 1: List of sources of updated data

| Variable | Description | Database | URL |
|---|---|---|---|
| Public health expenditure | General government expenditure on health in constant 2010 US$ per capita, PPP | World Health Organization, Global Health Expenditure Data Base | http://apps.who.int/nha/database/ViewData/Indicators/en |
| Total health expenditure | Total expenditure on health in constant 2010 US$ per capita, PPP | World Health Organization, Global Health Expenditure Data Base | http://apps.who.int/nha/database/ViewData/Indicators/en |
| Private health insurance | Private insurance in constant 2010 US$ per capita, PPP | World Health Organization, Global Health Expenditure Data Base | http://apps.who.int/nha/database/ViewData/Indicators/en |
| Out of pocket health expenditure | Out of pocket expenditure in constant (2005) US$ per capita, PPP | World Health Organization, Global Health Expenditure Data Base | http://apps.who.int/nha/database/ViewData/Indicators/en |
| Immunization: measles | Immunization, measles (% of children ages 12-23 months) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SH.IMM.MEAS/countries |
| Immunization coverage (other than measles) | WHO/UNICEF coverage estimates for 1980-2014, as of July 2015.; Reported official target population, number of doses administrered and official coverage | World Health Organization, Immunization, Vaccines and Biologicals Data | http://www.who.int/immunization/monitoring_surveillance/data/en/ |
| Adult female mortality rate | Mortality rate, adult, female (per 1,000 female adults) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SP.DYN.AMRT.FE |
| Adult male mortality rate | Mortality rate, adult, male (per 1,000 male adults) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SP.DYN.AMRT.MA |
| Under 5 mortality rate (UN estimate) | Mortality rate, under-5 (per 1,000 live births) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SH.DYN.MORT |
| Under 5 mortality rate (IHME estimate) | Mortality rate, under-5 (per 1,000 live births): Infant and Child Mortality Estimates by country | Institute of Health Metrics and Evaluation, Global Health Data Exchange | http://ghdx.healthdata.org/record/infant-and-child-mortality-estimates-country-1970-2010 |
| Maternal mortality rate | Maternal mortality ratio (modeled estimate, per 100,000 live births) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SH.STA.MMRT |
| Gross domestic product (GDP) | GDP per capita, constant 2011 US$, PPP | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/NY.GDP.PCAP.KD |
| Primary education enrollment rate | Adjusted net enrollment rate, primary (% of primary school age children) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SE.PRM.TENR?page=4&display=defaut3Fiframe3Dtrue |

## Table A 1: List of sources of updated data (continued)

| Variable | Description | Database | URL |
|---|---|---|---|
| Population aged 0-14 | Population ages 0-14 (% of total) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SP.POP.0014.TO.ZS |
| Population aged 65+ | Population ages 65 and above (% of total) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS |
| CO2 emission per capita | CO2 emissions (metric tons per capita) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/EN.ATM.CO2E.PC |
| Battle related deaths | Number of battle-related deaths by country | Uppsala University, UCDP Battle-related Deaths dataset | http://www.pcr.uu.se/research/ucdp/datasets/ucdp_battle-related_deaths_dataset/ |
| WHO external cause of death | 8 indicators of deaths which are externaly caused | World Health Organization, Global Health Expenditure Data Base | http://www.who.int/healthinfo/mortality_data/en/ |
| WDI worldwide governance indicators | governmental performance | World Bank, Worldwide Governance Indicators data | http://info.worldbank.org/governance/wgi/index.aspx#home |
| Total population | Population, total | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SP.POP.TOTL |
| SIPRI military expenditure | Military expenditure in 2011 constant USD | Stockholm International Peace Research Institute Military Expenditure Database | http://www.sipri.org/research/armaments/milex/milex_database |
| Legal origins of country | Legal origins of countries (Brithsh, French, Socialist, German, and Scandinavian) | World Bank Global Development Network Growth Database: Lost decades social indicaotors and fixed factors | http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20701055~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html |
| Urban population rate | Urban population (% of total) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS |
| Sanitation rate | Improved sanitation facilities (% of population with access) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SH.STA.ACSN |
| Paved road rate | % paved road | United Nations Database (mirroing World Development Indicators database) | http://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3AIS.ROD.PAVE.ZS |
| Literacy rate | literacy rate (% of people ages 15 and above) | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SE.ADT.LITR.ZS |
| Primary education completion rate | % of the total population of the theoretical entrance age to the last grade of primary education | World Bank, World Development Indicators database | http://data.worldbank.org/indicator/SE.PRM.CMPT.ZS |
| Donor funding on basic health | Development assistance for health (DAH) | Institute of Health Metrics and Evaluation, Global Health Data Exchange | http://ghdx.healthdata.org/record/development-assistance-health-database-1990-2014 |
| Consumption-investment ratio | Investment Share of PPP Converted GDP Per Capita at 2005 constant prices | University of Pennsilvania, Penn World Tables | https://pwt.sas.upenn.edu/php_site/pwt71/pwt71_form.php |

**Table A 2: BGG analysis using the original data (replacing every variable from updated data sequentially)**

| | Variable replaced | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Under 5 mortality | Maternal mortality | GDP per capita | Public health expenditure | Neighbours military expenditure per capita | Investment-GDP ratio | Nonliteracy rate | Paved road rate | Sanitation rate |
| **A. Under 5 Mortality** | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Public health expenditure | -0.459*** | -0.341** | -0.339*** | -0.289** | -0.268* | -0.338** | -0.464* | -0.323 | -0.342*** |
| | (0.130) | (0.132) | (0.127) | (0.132) | (0.147) | (0.135) | (0.253) | (0.208) | (0.129) |
| First stage F | 14.411 | 14.411 | 9.643 | 3.263 | 13.060 | 14.821 | 2.949 | 7.777 | 16.795 |
| N | 127 | 127 | 125 | 126 | 127 | 127 | 33 | 81 | 125 |
| **B. Adult female mortality** | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Public health expenditure | -0.519*** | -0.613*** | -0.517*** | -0.355* | -0.454** | -0.523*** | -0.750** | -0.731** | -0.508*** |
| | (0.184) | (0.185) | (0.169) | (0.210) | (0.179) | (0.177) | (0.353) | (0.285) | (0.172) |
| First stage F | 14.411 | 14.414 | 9.643 | 3.263 | 13.060 | 14.821 | 2.949 | 7.777 | 16.795 |
| N | 127 | 125 | 125 | 126 | 127 | 127 | 33 | 81 | 125 |

**Table A 3: BGG analysis using the original data (replacing every variable from updated data sequentially): updated data are imputed using original data where missing**

| | Variable replaced | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Under 5 mortality** | **Maternal mortality** | **GDP per capita** | **Public health expenditure** | **Neighbours military expenditure per capita** | **Investment-GDP ratio** | **Nonliteracy rate** | **Paved road rate** | **Sanitation rate** |
| **A. Under 5 Mortality** | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Public health expenditure | -0.459*** | -0.341** | -0.341*** | -0.257** | -0.268* | -0.338** | -0.415*** | -0.432*** | -0.339*** |
| | (0.130) | (0.132) | (0.126) | (0.129) | (0.147) | (0.135) | (0.143) | (0.137) | (0.129) |
| First stage F | 14.411 | 14.411 | 9.759 | 3.366 | 13.060 | 14.821 | 12.508 | 14.643 | 16.962 |
| N | 127 | 127 | 127 | 127 | 127 | 127 | 127 | 127 | 127 |
| **B. Adult female mortality** | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Public health expenditure | -0.519*** | -0.608*** | -0.500*** | -0.270 | -0.454** | -0.523*** | -0.629*** | -0.617*** | -0.517*** |
| | (0.184) | (0.182) | (0.167) | (0.211) | (0.179) | (0.177) | (0.222) | (0.186) | (0.174) |
| First stage F | 14.411 | 14.411 | 9.759 | 3.366 | 13.060 | 14.821 | 12.508 | 14.643 | 16.962 |
| N | 127 | 127 | 127 | 127 | 127 | 127 | 127 | 127 | 127 |

**Table A 4: MSS analysis using the original data (replacing every variable from updated data sequentially)**

| | Variable replaced | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CO2 emission | Conflict death | Education | GDP per capita | Public health expenditure | Adult female mortality | Adult male mortality | Under 5 mortality | Population under 14 | Population over 65 |
| **A. Under 5 Mortality** | | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Public health expenditure | 1.105*** | -17.663** | -9.323** | 2.496*** | -14.774** | -13.779** | -13.779** | -13.779** | -14.165** | -12.671** |
| | (0.325) | (8.050) | (3.755) | (0.615) | (6.342) | (5.794) | (5.794) | (5.794) | (6.029) | (5.187) |
| First stage F of reverse equation | 2.818 | 1.842 | 0.485 | 1.197 | 3.070 | 3.070 | 3.070 | 3.070 | 2.870 | 2.703 |
| First stage F of main equation | 16316.984 | 5.783 | 10.581 | 582.578 | 6.447 | 7.402 | 7.402 | 7.402 | 7.145 | 8.040 |
| N | 1404 | 1398 | 1346 | 1375 | 1398 | 1398 | 1398 | 1398 | 1398 | 1398 |
| **B. Adult female mortality** | | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Public health expenditure | -4.785** | -2.523*** | 5.007 | -5.778** | -2.818** | -27.641** | -3.957** | -3.957** | -1.158* | -2.667*** |
| | (1.935) | (0.943) | (3.297) | (2.601) | (1.130) | (12.696) | (1.552) | (1.552) | (0.689) | (1.000) |
| First stage F of reverse equation | 7.485 | 14.846 | 0.571 | 8.447 | 7.341 | 3.991 | 7.341 | 7.341 | 6.333 | 7.304 |
| First stage F of main equation | 1904.345 | 57128.000 | 236.640 | 533.398 | 7759.480 | 10.180 | 4857.691 | 4857.691 | 5751295.449 | 37410.162 |
| N | 1229 | 1223 | 1165 | 1204 | 1223 | 1398 | 1223 | 1223 | 1223 | 1223 |
| **C. Adult male mortality** | | | | | | | | | | |
| Public health expenditure | -4.954*** | -1.927*** | 2.010 | -4.766** | -14.774** | -2.944*** | -19.772** | -2.944*** | -0.794 | -1.745*** |
| | (1.841) | (0.670) | (1.369) | (1.890) | (6.342) | (1.026) | (7.745) | (1.026) | (0.532) | (0.655) |
| First stage F of reverse equation | 11.767 | 17.764 | 0.880 | 14.104 | 3.070 | 12.021 | 6.542 | 12.021 | 10.708 | 11.880 |
| First stage F of main equation | 1091.073 | 155334.965 | 3138.602 | 874.922 | 6.447 | 13790.401 | 19.354 | 13790.401 | 1095374.893 | 183319.494 |
| N | 1229 | 1223 | 1165 | 1204 | 1398 | 1223 | 1398 | 1223 | 1223 | 1223 |

**Table A 5: MSS analysis using the original data (replacing every variable from updated data sequentially): updated data are imputed using original data where missing**

| | Variable replaced | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CO2 emission | Conflict death | Education | GDP per capita | Public health expenditure | Adult female mortality | Adult male mortality | Under 5 mortality | Population under 14 | Population over 65 |
| **A. Under 5 Mortality** | | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Public health expenditure | 1.105*** | -17.663** | -4.335*** | 3.121*** | -14.774** | -13.779** | -13.779** | -13.779** | -14.165** | -12.671** |
| | (0.325) | (8.050) | (1.573) | (0.778) | (6.342) | (5.794) | (5.794) | (5.794) | (6.029) | (5.187) |
| First stage F of reverse equation | 2.818 | 1.842 | 4.249 | 1.558 | 3.070 | 3.070 | 3.070 | 3.070 | 2.870 | 2.703 |
| First stage F of main equation | 16316.984 | 5.783 | 36.995 | 217.160 | 6.447 | 7.402 | 7.402 | 7.402 | 7.145 | 8.040 |
| N | 1404 | 1398 | 1512 | 1398 | 1398 | 1398 | 1398 | 1398 | 1398 | 1398 |
| **B. Adult female mortality** | | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Public health expenditure | -4.785** | -2.523*** | -5.186** | -6.245** | -2.818** | -27.641** | -3.957** | -3.957** | -1.158* | -2.667*** |
| | (1.935) | (0.943) | (2.153) | (2.796) | (1.130) | (12.696) | (1.552) | (1.552) | (0.689) | (1.000) |
| First stage F of reverse equation | 7.485 | 14.846 | 6.796 | 7.842 | 7.341 | 3.991 | 7.341 | 7.341 | 6.333 | 7.304 |
| First stage F of main equation | 1904.345 | 57128.000 | 1018.047 | 412.293 | 7759.480 | 10.180 | 4857.691 | 4857.691 | 5751295.449 | 37410.162 |
| N | 1229 | 1223 | 1320 | 1223 | 1223 | 1398 | 1223 | 1223 | 1223 | 1223 |
| **C. Adult male mortality** | | | | | | | | | | |
| Public health expenditure | -4.954*** | -1.927*** | -4.887*** | -5.390** | -14.774** | -2.944*** | -19.772** | -2.944*** | -0.794 | -1.745*** |
| | (1.841) | (0.670) | (1.792) | (2.096) | (6.342) | (1.026) | (7.745) | (1.026) | (0.532) | (0.655) |
| First stage F of reverse equation | 11.767 | 17.764 | 11.404 | 12.536 | 3.070 | 12.021 | 6.542 | 12.021 | 10.708 | 11.880 |
| First stage F of main equation | 1091.073 | 155334.965 | 1073.583 | 620.003 | 6.447 | 13790.401 | 19.354 | 13790.401 | 1095374.893 | 183319.494 |
| N | 1229 | 1223 | 1320 | 1223 | 1398 | 1223 | 1398 | 1223 | 1223 | 1223 |

**Table A 6: Traditional IV approach (with streamlined model) using different specifications of instrumental variables**

| | Unimputed model | | | | Imputed model | | | |
|---|---|---|---|---|---|---|---|---|
| **A: Under 5 mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | -0.048 | -0.003 | 0.038 | 0.193 | 1.069 | 0.035 | -0.060 | 0.560 |
| | (0.310) | (0.291) | (0.266) | (1.801) | (1.413) | (0.265) | (0.196) | (1.912) |
| First stage F | 0.664 | 0.707 | 0.503 | 0.079 | 0.506 | 1.564 | 3.221 | 0.157 |
| N | 661 | 661 | 661 | 661 | 2601 | 2601 | 2601 | 2601 |
| **B: Adult female mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | -0.118 | -0.315 | -0.264 | -0.303 | 0.381 | 0.271 | 0.233 | 0.448 |
| | (0.420) | (0.439) | (0.410) | (1.470) | (0.526) | (0.245) | (0.189) | (1.373) |
| First stage F | 0.546 | 0.757 | 0.537 | 0.076 | 0.506 | 1.564 | 3.221 | 0.157 |
| N | 642 | 642 | 642 | 642 | 2601 | 2601 | 2601 | 2601 |
| **C: Adult male mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | -0.174 | -0.447 | -0.471 | -0.501 | 0.284 | 0.328 | 0.312* | 0.802 |
| | (0.354) | (0.344) | (0.347) | (1.507) | (0.410) | (0.218) | (0.169) | (1.790) |
| First stage F | 0.546 | 0.757 | 0.537 | 0.076 | 0.506 | 1.564 | 3.221 | 0.157 |
| N | 642 | 642 | 642 | 642 | 2601 | 2601 | 2601 | 2601 |

Notes:
IVs in the main model (Table 6 of the main paper) are log neighbours military expenditure as % of GDP, and log neighbours public health expenditure as % of GDP.

- Model (1) uses Raw IV (note that main model in Table 6 used logged IVs)
- Model (2) adds squared raw IV to model (1)
- Model (3) further adds cubic raw IV to model (2)
- Model (4) uses total neighbours military and public health expenditure (not per capita expenditure)

**Table A 7: Bruckner IV approach (with streamlined model) using different specifications of instrumental variables**

| | Unimputed model | | | | Imputed model | | | |
|---|---|---|---|---|---|---|---|---|
| **A: Under 5 mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | -0.765*** | -0.493*** | -0.227*** | 0.030 | -0.272** | -0.193** | -0.060 | -0.152*** |
| | (0.157) | (0.092) | (0.045) | (0.042) | (0.120) | (0.091) | (0.039) | (0.038) |
| First stage F of reverse equation | 1.679 | 0.845 | 1.322 | 1.167 | 1.836 | 1.324 | 1.597 | 2.277 |
| First stage F of main equation | 51.819 | 141.607 | 1355.738 | 1036090.6 | 36.317 | 85.065 | 3427.838 | 3910.690 |
| N | 669 | 669 | 669 | 402 | 2631 | 2631 | 2631 | 1230 |
| **B: Adult female mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | -0.262*** | -0.237*** | -0.305*** | 0.048 | -0.240** | -0.117* | -0.049 | 0.121*** |
| | (0.060) | (0.056) | (0.068) | (0.036) | (0.121) | (0.068) | (0.039) | (0.045) |
| First stage F of reverse equation | 2.432 | 1.686 | 2.356 | 3.780 | 1.975 | 5.044 | 3.792 | 1.252 |
| First stage F of main equation | 252.397 | 331.238 | 172.733 | 7018.4 | 19.599 | 89.859 | 693.625 | 802.905 |
| N | 650 | 650 | 650 | 386 | 2590 | 2590 | 2590 | 1198 |
| **C: Adult male mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | -0.346*** | -0.192*** | -0.263*** | -0.010 | -0.392*** | -0.301** | -0.086 | 0.077*** |
| | (0.070) | (0.048) | (0.058) | (0.032) | (0.145) | (0.125) | (0.064) | (0.027) |
| First stage F of reverse equation | 1.139 | 1.693 | 3.137 | 3.851 | 1.180 | 2.990 | 2.095 | 1.370 |
| First stage F of main equation | 102.915 | 360.303 | 172.024 | 1233884.9 | 14.540 | 15.566 | 67.351 | 2711.931 |
| N | 650 | 650 | 650 | 386 | 2590 | 2590 | 2590 | 1198 |

Notes:
IVs in the main model (Table 6 of the main paper) are log per capita CO2 emission; and log number of deaths by conflict

- Model (1) uses Raw IV (note that main model in Table 6 used logged IVs)
- Model (2) adds squared raw IV to model (1)
- Model (3) further adds cubic raw IV to model (2)
- Model (4) uses main IV (CO2 and Conflict deaths) plus log of external causes of deaths

**Table A 8: Regression analysis based on quinquennial data (replication of the analysis in Table 6, first column with fixed effect IV model)**

| | Traditional IV | | | | Bruckner IV | | | |
|---|---|---|---|---|---|---|---|---|
| **A: log under 5 mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | -0.709 | -0.522 | 0.365 | 0.498 | 0.188* | -0.408*** | -0.237*** | -0.135* |
| | (0.584) | (7.415) | (0.542) | (0.555) | (0.108) | (0.121) | (0.065) | (0.081) |
| First stage F of reverse equation | . | . | . | . | 4.782 | 2.234 | 1.657 | 3.749 |
| First stage F of main equation | 1.300 | 0.385 | 0.658 | 0.839 | 222.019 | 46.742 | 748.225 | 315.139 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 131 | 612 | 308 | 612 | 131 | 618 | 308 | 620 |
| **B: log adult female mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | 0.403 | -0.460 | 0.178 | 0.078 | 0.201* | 0.483*** | 0.689*** | 0.222** |
| | (0.869) | (8.255) | (0.553) | (0.268) | (0.110) | (0.159) | (0.182) | (0.086) |
| First stage F of reverse equation | . | . | . | . | 0.958 | 1.048 | 0.224 | 1.256 |
| First stage F of main equation | 0.581 | 0.385 | 0.392 | 0.839 | 153.770 | 18.868 | 31.989 | 64.353 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 119 | 612 | 298 | 612 | 119 | 603 | 298 | 607 |
| **C: log adult male mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log Public health expenditure | 0.188 | -0.291 | 0.346 | 0.120 | -0.298** | -0.259*** | -0.095** | 0.043 |
| | (0.684) | (6.252) | (0.514) | (0.229) | (0.112) | (0.090) | (0.047) | (0.032) |
| First stage F of reverse equation | . | . | . | . | 0.953 | 1.631 | 0.105 | 2.019 |
| First stage F of main equation | 0.581 | 0.385 | 0.392 | 0.839 | 64.963 | 35.783 | 895.420 | 12969.315 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 119 | 612 | 298 | 612 | 119 | 603 | 298 | 607 |

Notes:

- Model (1) and (2) show the results based on the data from 1995, 2000, 2005 and 2010 (5-year interval data), with and without data imputation, respectively.
- Model (3) and (4) show the results based on the data where five-year average of the variables (1995-1999; 2000-2004; 2005-2009; and 2010-2011) are calculated and used in the regression, with and without data imputation, respectively.

**Table A 9: Traditional IV approach (with streamlined model) in which log GDP per capita is treated as an endogenous variable**

| | Fixed effect | | Pooled cross section | |
|---|---|---|---|---|
| **A: log under 5 mortality** | | | | |
| | (1) | (2) | (3) | (4) |
| log Public health expenditure | 0.297 | 0.188 | 0.660 | 0.007 |
| | (0.519) | (0.296) | (0.600) | (0.222) |
| log GDP per capita | -0.055 | -0.834* | -1.262 | -0.402** |
| | (0.846) | (0.439) | (0.785) | (0.188) |
| F for log Public health expenditure | 1.263 | 2.573 | 0.849 | 1.666 |
| F for log GDP per capita | 0.728 | 2.964 | 0.860 | 3.060 |
| Data imputation | No | Yes | No | Yes |
| N | 617 | 2415 | 607 | 2335 |
| **B: log adult female mortality** | | | | |
| | (1) | (2) | (3) | (4) |
| log Public health expenditure | 0.267 | 0.001 | 0.081 | -0.014 |
| | (0.517) | (0.212) | (0.400) | (0.212) |
| log GDP per capita | 0.041 | -0.177 | -0.236 | -0.045 |
| | (0.952) | (0.302) | (0.442) | (0.174) |
| F for log Public health expenditure | 1.344 | 2.573 | 0.857 | 1.666 |
| F for log GDP per capita | 0.605 | 2.964 | 0.863 | 3.060 |
| Data imputation | No | Yes | No | Yes |
| N | 606 | 2415 | 597 | 2335 |
| **C: log adult male mortality** | | | | |
| | (1) | (2) | (3) | (4) |
| log Public health expenditure | 0.116 | -0.007 | 0.094 | -0.160 |
| | (0.362) | (0.166) | (0.401) | (0.183) |
| log GDP per capita | -0.003 | -0.164 | -0.410 | -0.072 |
| | (0.582) | (0.253) | (0.414) | (0.168) |
| F for log Public health expenditure | 1.344 | 2.573 | 0.857 | 1.666 |
| F for log GDP per capita | 0.605 | 2.964 | 0.863 | 3.060 |
| Data imputation | No | Yes | No | Yes |
| N | 606 | 2415 | 597 | 2335 |

Notes: Instrumental variables for log Public health expenditure are the same as the Table 6 (neighbours' military expenditure as % of GDP and neighbours' public health expenditure as % of GDP). Following Bokhari et al. (2007), the instrumental variable for log GDP per capita is the Consumption-Investment ratio of the country (available from Penn World Table, accessed in August 2015).

**Table A10: Traditional IV and Bruckner IV approach using streamlined model, estimated using Limited Information Maximum Likelihood (A replication of Table 6)**

| | Traditional IV | | | | Bruckner IV | | | |
|---|---|---|---|---|---|---|---|---|
| | Fixed effect | | Pooled cross section | | Fixed effect | | Pooled cross section | |
| **A: log under 5 mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log Public health expenditure | 0.108 | 0.793 | 2.019 | -7.974 | 0.027 | -0.291** | -0.331** | 0.138* |
| | (1.109) | (1.026) | (3.806) | 520266.507) | (0.039) | (0.127) | (0.132) | (0.070) |
| First stage F of reverse equation | . | . | . | . | 1.262 | 3.707 | 3.584 | 2.335 |
| First stage F of main equation | 0.245 | 0.748 | 0.785 | 1.761 | 22093.959 | 31.452 | 255.860 | 551.121 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| Country | 100 | 153 | 115 | 148 | 101 | 155 | 105 | 155 |
| N | 661 | 2601 | 648 | 2516 | 666 | 2631 | 634 | 2631 |
| **B: log adult female mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log Public health expenditure | 2.790 | 0.957 | 8.384 | 3.811 | 3.073 | -0.246** | -0.162* | -0.015 |
| | (158.770) | (66.655) | (157.401) | (65.693) | (1.949) | (0.124) | (0.092) | (0.044) |
| First stage F of reverse equation | . | . | . | . | 0.096 | 4.408 | 6.940 | 4.168 |
| First stage F of main equation | 0.108 | 0.613 | 0.803 | 1.728 | 2.607 | 18.856 | 299.146 | 3271.514 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| Country | 99 | 153 | 114 | 148 | 100 | 155 | 105 | 155 |
| N | 642 | 2560 | 631 | 2477 | 647 | 2590 | 618 | 2590 |
| **C: log adult male mortality** | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log Public health expenditure | 0.794 | 1.065 | 7.422 | 24.867 | 0.284*** | -0.136* | -0.036 | -0.003 |
| | (4.759) | (7966.741) | (109.382) | 1.542e+09) | (0.062) | (0.080) | (0.057) | (0.029) |
| First stage F of reverse equation | . | . | . | . | 1.293 | 3.526 | 8.135 | 4.827 |
| First stage F of main equation | 0.108 | 0.613 | 0.803 | 1.728 | 126.514 | 33.662 | 2119.952 | 15690.246 |
| Data imputation | No | Yes | No | Yes | No | Yes | No | Yes |
| Country | 99 | 153 | 114 | 148 | 100 | 155 | 105 | 155 |
| N | 642 | 2560 | 631 | 2477 | 647 | 2590 | 618 | 2590 |

See Notes of Table 6 of the main paper.

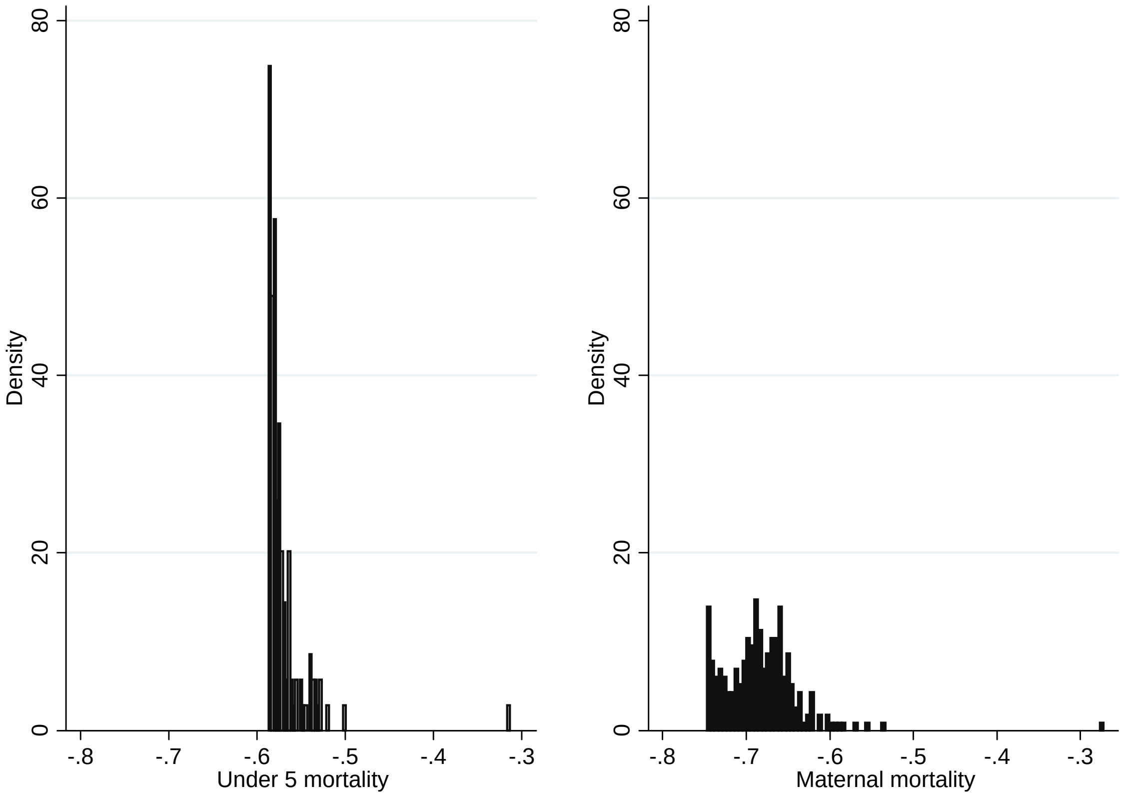


**Figure A 2: Heterogeneity of the impact of public health expenditure on mortality based on the BGG specification, using year 2000 updated data**

Note: Based on the BGG model (column (4) of Table 4), we estimated the impact of public health expenditure on mortality outcomes, allowing for variations by the level of economic development (proxied by the proportion of paved roads) and by donor funding on health. The histograms are based on 127 countries which are included in the regression analysis.

## Appendix B - Alternative estimation strategies explored

**B.1 Allowing for lagged impacts of public health expenditure**

A potential reason why we fail (as did some previous studies e.g. Filmer and Prichett, 1999) to confirm a negative impact of public health expenditure on health could be due to our model estimating the impact of current expenditure on current health outcome. Indeed there may well be lags between healthcare investment and its impact upon population health. To allow for this possibility, we now account for lagged impacts of public health expenditure in the streamlined model. Instead of contemporaneous health expenditure, we now use lagged expenditure (t-1, t-2 and t-5) in the equation as endogenous variables.

**B.2 Substitution between public and private expenditure**

Another potential reason why we do not find robust enough results to support the expectation that public health expenditures consistently and significantly reduce mortality rates may be that as public health care spending increases, individuals are likely to respond by reducing their previously private expenditures (from out-of-pocket or privately insured sources). As a result the overall resources available to the health care system may not increase as much, compared to a situation in which there would be no such substitution effect, and, hence, unless the marginal productivity of public spending vastly outperforms that of private spending, the overall mortality or health effect associated with a given increase in public health expenditures may be limited or non-existent. While ideally we would want to explicitly model the interrelationship between public and private health expenditures, this is beyond the scope of the present exercise. Instead, in order to circumvent the explicit modelling of the interrelationship it may be instructive to examine the impact of total health expenditure on mortality rates. At least, this would allow us to see whether we are in a position to detect *any* relationship between expenditures and outcomes, and if that was the case, this could be seen as indirect evidence of the importance of a substitution effect.

**B.3 Allowing for more lagged structure for the impacts of public health expenditure**

We further investigate the potential lagged structure by including lagged public health expenditure (t, t-1, t-2, and t-3) at the same time in the regression models. These models capture the impacts of previous years' expenditure on mortality in the current year, conditional on the impact of the expenditure in the current year. For both Traditional and Bruckner IV approach, the lagged values of the instrumental variables are used as IVs of the lagged public health expenditure, i.e. the IV of the lagged expenditure (t-1) is the lagged IV (t-1).

**B.4 Further estimation approaches**

In Block D of Appendix Table B3, we present the estimation results based on a dynamic panel approach - Blundell-Bond estimation. In this approach, a 1-year lagged mortality rate is included as an explanatory variable as well as public health expenditure and the same set of covariates (no lag were taken for these variables):

$$\text{Eq. (4):}\quad \mathrm{MO_{jt}} = \beta_j + \rho \mathrm{MO_{j,t-1}} + \beta_1 \mathrm{EX_{jt}} + \mathbf{X}'_{jt}\boldsymbol{\beta} + \varepsilon_{jt.}$$

Taking the first difference for all variables in Eq. (4) eliminates the country level fixed effect $\beta_j$ . The first-differenced error term $\varepsilon_{jt} - \varepsilon_{j,t-1}$ is uncorrelated with second or further lags of mortality, health expenditure and the other variables. We use these properties as the moment conditions in the estimation of the model by system Generalised Method of Moments. In this dynamic panel data approach, we are only interested in the impact of contemporaneous public health expenditure on contemporaneous mortality. The impact dMOt/dEXt is given by the single coefficient beta.

**Table B 1: Impact of lagged, and total public health expenditure on mortality (streamlined model using instrumental variable approach)**

**A: Impact of 1-year lagged health expenditure**

| | log under 5 mortality | | | | log adult female mortality | | | | log adult male mortality | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log health expenditure (t-1) | -0.262 | 0.720 | 0.036 | -0.199** | -0.408 | 0.208 | 1.238* | 0.019 | -0.344 | 0.131 | 0.148*** | -0.120 |
| | (0.462) | (0.766) | (0.040) | (0.093) | (0.671) | (0.318) | (0.722) | (0.012) | (0.532) | (0.252) | (0.043) | (0.075) |
| First stage F of expenditure (t-1) | 0.690 | 0.712 | 2804.160 | 61.301 | 0.458 | 0.712 | 3.871 | 10690547.8 | 0.458 | 0.712 | 80.659 | 31.121 |
| Imputation | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 661 | 2447 | 382 | 2476 | 642 | 2447 | 369 | 2451 | 642 | 2447 | 369 | 2451 |

**B: Impact of 2-year lagged health expenditure**

| | log under 5 mortality | | | | log adult female mortality | | | | log adult male mortality | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log health expenditure (t-2) | -0.218 | 0.669 | 0.061 | -0.177** | -0.118 | 0.081 | 1.664 | 0.015 | -0.132 | 0.053 | 0.155*** | -0.094 |
| | (0.252) | (0.741) | (0.046) | (0.083) | (0.246) | (0.252) | (1.479) | (0.012) | (0.214) | (0.217) | (0.051) | (0.066) |
| First stage F of expenditure (t-2) | 1.878 | 0.746 | 1516.7 | 71.363 | 1.426 | 0.746 | 1.382 | 9753700.8 | 1.426 | 0.746 | 54.508 | 32.963 |
| Imputation | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 603 | 2295 | 413 | 2321 | 585 | 2295 | 400 | 2308 | 585 | 2295 | 400 | 2308 |

**C: Impact of 5-year lagged health expenditure**

| | log under 5 mortality | | | | log adult female mortality | | | | log adult male mortality | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log health expenditure (t-5) | -0.036 | -0.281 | -0.033 | -0.113** | -0.157 | -0.308 | 1.299 | 0.011 | -0.014 | -0.394 | 0.080 | -0.036 |
| | (0.116) | (0.465) | (0.045) | (0.056) | (0.109) | (0.361) | (1.071) | (0.010) | (0.074) | (0.393) | (0.054) | (0.039) |
| First stage F of expenditure (t-5) | 2.892 | 0.921 | 889.299 | 88.489 | 2.991 | 0.921 | 1.728 | 8346576 | 2.991 | 0.921 | 28.507 | 38.357 |
| Imputation | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 556 | 1832 | 226 | 1856 | 538 | 1832 | 214 | 1844 | 538 | 1832 | 214 | 1844 |

**D: Impact of total health expenditure**

| | log under 5 mortality | | | | log adult female mortality | | | | log adult male mortality | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log total health expenditure | -0.197 | 0.606 | -0.093* | -0.299*** | -0.404 | -0.486 | 0.018 | 0.228*** | -0.232 | -0.219 | -0.331*** | 0.021 |
| | (1.174) | (2.009) | (0.048) | (0.045) | (0.864) | (0.894) | (0.037) | (0.052) | (0.624) | (0.774) | (0.083) | (0.031) |
| First stage F of reverse equation | . | . | 1.262 | 3.696 | . | . | 0.096 | 4.461 | . | . | 1.293 | 3.696 |
| First stage F of main equation | 0.393 | 0.244 | 36756.8 | 1776.2 | 0.491 | 0.244 | 373761.6 | 635.921 | 0.491 | 0.244 | 122.514 | 129224.7 |
| Imputation | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 661 | 2601 | 666 | 2631 | 642 | 2601 | 647 | 2590 | 642 | 2601 | 647 | 2590 |

Notes: In Block A, B and C, the first stage F of the reverse causality model in the Bruckner IV approach is not presented in the table. Following Moreno-Serra and Smith (2015), once the residual from the reverse causality model is estimated (see Eq.(3) in Method section), its lag is calculated and used as instrumental variable in the main regression model.

## Table B 2: Impact of present and lagged health expenditure on mortality (streamlined model using instrumental variable approach)

**A: Impact of contemporaneous and lagged (1 year only) health expenditure**

| | log under 5 mortality | | | | log adult female mortality | | | | log adult male mortality | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log health expenditure (t) | 0.267 | 0.610 | 0.064 | -0.153*** | 0.390 | 0.300 | 1.720* | 0.017** | 0.490 | 0.253 | 0.304*** | -0.104*** |
| | (0.487) | (0.592) | (0.042) | (0.051) | (0.721) | (0.223) | (1.021) | (0.008) | (0.709) | (0.180) | (0.096) | (0.035) |
| log health expenditure (t-1) | -0.367 | 0.042 | 0.008 | -0.088 | -0.400 | -0.085 | 0.648 | 0.007 | -0.393 | -0.096 | 0.010 | -0.048 |
| | (0.287) | (0.514) | (0.035) | (0.067) | (0.244) | (0.245) | (0.714) | (0.011) | (0.238) | (0.200) | (0.046) | (0.057) |
| First stage F of expenditure (t) | 0.576 | 1.712 | 7776.995 | 512.554 | 0.290 | 1.712 | 5.622 | 2.916e+08 | 0.290 | 1.712 | 58.792 | 1156.531 |
| First stage F of expenditure (t-1) | 2.225 | 1.797 | 41340.818 | 450.828 | 2.054 | 1.797 | 9.018 | 2.197e+08 | 2.054 | 1.797 | 79.947 | 750.556 |
| Imputation | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 661 | 2447 | 382 | 2476 | 642 | 2447 | 369 | 2434 | 642 | 2447 | 369 | 2434 |

**B: Impact of contemporaneous and lagged (1 and 2 years) health expenditure**

| | log under 5 mortality | | | | log adult female mortality | | | | log adult male mortality | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log health expenditure (t) | 0.376 | 0.322 | 0.096** | -0.166*** | 0.459 | 0.019 | 2.795 | 0.010 | 0.378 | 0.050 | 0.278*** | -0.113*** |
| | (0.488) | (0.341) | (0.045) | (0.047) | (0.734) | (0.136) | (4.173) | (0.009) | (0.665) | (0.123) | (0.097) | (0.034) |
| log health expenditure (t-1) | -0.075 | 0.193 | -0.022 | -0.009 | -0.424 | 0.107 | 1.708 | 0.010** | -0.445 | 0.066 | 0.041 | -0.008 |
| | (0.397) | (0.285) | (0.027) | (0.026) | (0.555) | (0.112) | (4.342) | (0.004) | (0.520) | (0.095) | (0.058) | (0.014) |
| log health expenditure (t-2) | -0.357 | -0.237 | -0.031 | -0.088 | -0.060 | -0.249 | 1.839 | 0.003 | -0.058 | -0.218 | -0.029 | -0.036 |
| | (0.240) | (0.395) | (0.039) | (0.057) | (0.293) | (0.161) | (4.033) | (0.010) | (0.269) | (0.141) | (0.055) | (0.049) |
| First stage F of expenditure (t) | 0.628 | 2.152 | 10571.0 | 2086.008 | 0.539 | 2.152 | 0.866 | 3.522e+08 | 0.539 | 2.152 | 52.780 | 1029.291 |
| First stage F of expenditure (t-1) | 0.987 | 1.736 | 43483.6 | 1592.700 | 0.801 | 1.736 | 0.421 | 6.542e+08 | 0.801 | 1.736 | 49.621 | 938.015 |
| First stage F of expenditure (t-2) | 2.423 | 1.835 | 29486.1 | 2415.063 | 2.243 | 1.835 | 0.814 | 4.166e+08 | 2.243 | 1.835 | 75.358 | 814.302 |
| Imputation | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 603 | 2295 | 304 | 2321 | 585 | 2295 | 292 | 2280 | 585 | 2295 | 292 | 2280 |

**C: Impact of contemporaneous and lagged (1, 2 and 3 years) health expenditure**

| | log under 5 mortality | | | | log adult female mortality | | | | log adult male mortality | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | | Traditional IV | | Bruckner IV | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log health expenditure (t) | 0.318 | -0.030 | 0.148*** | -0.136*** | 0.420 | -0.073 | 1.006** | 0.012 | 0.260 | -0.065 | 0.285*** | -0.084** |
| | (0.419) | (0.193) | (0.038) | (0.050) | (0.602) | (0.116) | (0.476) | (0.008) | (0.512) | (0.103) | (0.089) | (0.035) |
| log health expenditure (t-1) | -0.143 | 0.045 | -0.003 | -0.040* | -0.341 | -0.017 | 0.138 | 0.002 | -0.347 | -0.014 | -0.004 | -0.034*** |
| | (0.321) | (0.165) | (0.027) | (0.023) | (0.390) | (0.098) | (0.415) | (0.005) | (0.357) | (0.086) | (0.048) | (0.012) |
| log health expenditure (t-2) | -0.263 | 0.126 | 0.007 | -0.021 | -0.189 | 0.118 | 0.463 | 0.008* | -0.178 | 0.075 | -0.001 | -0.006 |
| | (0.226) | (0.205) | (0.023) | (0.015) | (0.213) | (0.105) | (0.577) | (0.004) | (0.165) | (0.093) | (0.052) | (0.011) |
| log health expenditure (t-3) | 0.034 | 0.105 | -0.029 | -0.072 | 0.045 | -0.254* | 0.441 | 0.001 | 0.078 | -0.219* | 0.022 | -0.026 |
| | (0.182) | (0.243) | (0.036) | (0.056) | (0.216) | (0.150) | (0.737) | (0.009) | (0.181) | (0.131) | (0.050) | (0.045) |
| First stage F of expenditure (t) | 0.733 | 3.473 | 10578.6 | 3619.021 | 0.691 | 3.473 | 4.380 | 3.906e+08 | 0.691 | 3.473 | 68.129 | 1378.982 |
| First stage F of expenditure (t-1) | 2.042 | 2.555 | 125656.0 | 3679.820 | 1.757 | 2.555 | 3.524 | 6.033e+08 | 1.757 | 2.555 | 70.385 | 1449.253 |
| First stage F of expenditure (t-2) | 1.782 | 1.762 | 40257.7 | 57381.659 | 1.710 | 1.762 | 3.648 | 6.635e+08 | 1.710 | 1.762 | 30.885 | 1922.683 |
| First stage F of expenditure (t-3) | 1.284 | 1.820 | 33858.4 | 27009.915 | 1.361 | 1.820 | 2.131 | 4.418e+08 | 1.361 | 1.820 | 73.709 | 1009.680 |
| Imputation | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 603 | 2142 | 238 | 2166 | 585 | 2142 | 226 | 2126 | 585 | 2142 | 226 | 2126 |

Notes: For Bruckner's IV approach, following Moreno-Serra and Smith (2015), once the residual from the reverse causality model is estimated (see Eq.(3) in Method section), its lag is calculated and used as instrumental variable of the corresponding expenditure variable (t, t-1, t-2 and t-3) in the main regression model.

**Table B 3: Alternative estimation approaches (fixed effect plus lagged expenditure; dynamic panel specification)**

| | log under 5 mortality | | log adult female mortality | | log adult male mortality | |
|---|---|---|---|---|---|---|
| **A: Impact of contemporaneous health expenditure: Fixed effect model** | | | | | | |
| | (1) | (2) | (1) | (2) | (1) | (2) |
| log health expenditure | -0.026 | -0.005 | -0.026 | 0.025** | -0.021 | 0.042** |
| | (0.035) | (0.020) | (0.029) | (0.011) | (0.026) | (0.018) |
| Imputation | No | Yes | No | Yes | No | Yes |
| N | 688 | 2635 | 669 | 2635 | 669 | 2635 |
| **B: Impact of 1-year lagged health expenditure: Fixed effect model** | | | | | | |
| | **log under 5 mortality** | | **log adult female mortality** | | **log adult male mortality** | |
| | (1) | (2) | (1) | (2) | (1) | (2) |
| log health expenditure (t-1) | -0.044 | -0.015 | -0.022 | 0.020* | -0.015 | 0.038* |
| | (0.029) | (0.021) | (0.025) | (0.012) | (0.022) | (0.020) |
| Imputation | No | Yes | No | Yes | No | Yes |
| N | 687 | 2479 | 668 | 2479 | 668 | 2479 |
| **C: Impact of 2-year lagged health expenditure: Fixed effect model** | | | | | | |
| | **log under 5 mortality** | | **log adult female mortality** | | **log adult male mortality** | |
| | (1) | (2) | (1) | (2) | (1) | (2) |
| log health expenditure (t-2) | -0.052* | -0.024 | -0.028 | 0.017 | -0.025 | 0.035* |
| | (0.030) | (0.023) | (0.026) | (0.011) | (0.022) | (0.020) |
| Imputation | No | Yes | No | Yes | No | Yes |
| N | 636 | 2325 | 618 | 2325 | 618 | 2325 |
| **D: Impact of contemporaneous health expenditure: Blundell-Bond dynamic panel estimation** | | | | | | |
| | **log under 5 mortality** | | **log adult female mortality** | | **log adult male mortality** | |
| | (1) | (2) | (1) | (2) | (1) | (2) |
| lagged mortality (t-1) | 1.058*** | 1.128*** | 1.063*** | 1.061*** | 1.058*** | 1.033*** |
| | (0.166) | (0.197) | (0.046) | (0.044) | (0.129) | (0.034) |
| log health expenditure | -0.011 | -0.021 | -0.039 | -0.019 | 0.065 | -0.024 |
| | (0.331) | (0.049) | (0.043) | (0.026) | (0.185) | (0.031) |
| Imputation | No | Yes | No | Yes | No | Yes |
| P value of Hansen test | 0.791 | 0.002 | 0.478 | 0.001 | 0.361 | 0.008 |
| P value of AR(1) test | 0.808 | 0.269 | 0.003 | 0.006 | 0.038 | 0.006 |
| P value of AR(2) test | 0.709 | 0.701 | 0.272 | 0.229 | 0.348 | 0.204 |
| Number of instruments | 28 | 31 | 28 | 31 | 28 | 31 |
| N | 688 | 2480 | 669 | 2480 | 669 | 2480 |

Notes:
The analysis based on the dynamic panel estimation was conducted using Stata command “xtabond2” (st0159_1), written by David Roodman. For reference see: Roodman, D. (2009) How to do xtabond2: an introduction to difference and system GMM in Stata. The Stata Journal 9(1) 86-136.

## Appendix C – Further descriptive statistics

### Table C 1: Descriptive statistics of BGG data (original data and updated data)

| | Original data (2000) | | | Updated data (2000) | | |
|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N |
| Under 5 motality | 73.49 | 68.91 | 127 | 65.83 | 59.82 | 127 |
| Maternal mortality | 345.39 | 425.07 | 127 | 290.93 | 364.00 | 125 |
| GDP per capita† | 6962.60 | 7987.95 | 127 | 10227.33 | 11868.58 | 125 |
| Public health expenditure‡ | 311.21 | 509.70 | 127 | 433.55 | 843.80 | 126 |
| Neighbors' military expenditure per capita | 161.04 | 183.22 | 127 | 162.13 | 263.19 | 127 |
| World Bank CPI on Social inclusion and equity | 3.63 | 0.76 | 127 | . | . | . |
| World Bank CPI on Economic management | 3.96 | 1.05 | 127 | . | . | . |
| Investment-GDP ratio | 14.87 | 8.71 | 127 | 21.63 | 9.20 | 127 |
| Donor funding on basic health per capita | 2.96 | 6.35 | 127 | . | . | . |
| Nonliterate rate | 21.22 | 22.15 | 127 | 22.61 | 22.25 | 76 |
| Paved road rate | 31.14 | 66.63 | 127 | 44.93 | 31.58 | 113 |
| Sanitation rate | 73.98 | 26.67 | 127 | 62.74 | 32.02 | 125 |

Notes:

* In this column the analysis include as many countries as possible (max available number of countries is 238) from the updated data sources.
†In the original data, GDP per capita using international dollars as of 2000 is used. In the updated data, GDP per capita using constant US dollars 2011 (PPP adjusted) is used.
‡In the original data, government health expenditure per capita (international dollars as of 2000) is used. In the updated data, government health expenditure per capita in 2005 constant dollars is used.

### Table C 2: Descriptive statistics of MSS data (original data and updated data)

| | Original data (1995-2008) | | | Updated data (1995-2008) | | |
|---|---|---|---|---|---|---|
| | Mean | SD | N (153 countries) | Mean | SD | N (153 countries) |
| Under 5 motality | 45.68 | 48.87 | 1397 | 45.68 | 48.87 | 1397 |
| Adult female mortality | 155.85 | 114.87 | 1225 | 161.41 | 122.46 | 1397 |
| Adult male mortality | 226.39 | 116.92 | 1225 | 230.83 | 118.42 | 1397 |
| public health expenditure (1/100)† | 5.99 | 8.43 | 1397 | 6.58 | 9.80 | 1397 |
| Privately pooled health expenditure (1/100) | 0.59 | 2.10 | 1397 | 2.60 | 13.42 | 1397 |
| Out-of-pocket health expenditure (1/100) | 1.88 | 2.04 | 1397 | 0.72 | 2.39 | 1264 |
| Immunization coverage (1/10) | 8.61 | 1.42 | 1397 | 8.63 | 1.44 | 1397 |
| GDP per capita (1/100)‡ | 121.15 | 131.67 | 1397 | 168.17 | 194.49 | 1374 |
| Primary education enrollment rate (1/10) | 8.61 | 1.59 | 1397 | 8.81 | 1.57 | 1235 |
| Proportion of population under 14 (1/10) | 3.07 | 1.05 | 1397 | 3.08 | 1.06 | 1397 |
| Proportion of population over 65 (1/10) | 0.75 | 0.52 | 1397 | 0.76 | 0.52 | 1397 |
| CO2 emission per capita | 5.16 | 6.51 | 1397 | 5.09 | 6.52 | 1397 |
| Death by conflict | 1.45 | 25.76 | 1397 | 1.29 | 23.88 | 1397 |

Notes:
* In this column the analysis include as many countries as possible (max available number of countries is 229) from the updated data sources.
† In the original data, government health expenditure per capita (international dollars as of 2005) is used. In the updated data, government health expenditure per capita in 2005 constant dollars is used.
‡ In the original data, GDP per capita using international dollars as of 2005 is used. In the updated data, GDP per capita using constant US dollars 2011 (PPP adjusted) is used.

**Table C 3: Descriptive statistics of the updated BGG data 2010**

| | Original data (2000) | | | Updated data (2010) | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N | # missing | max N | Missing rate |
| Under 5 motality | 73.49 | 68.91 | 127 | 39.36 | 41.44 | 192 | 44 | 236 | 18.6% |
| Maternal mortality | 345.39 | 425.07 | 127 | 182.55 | 242.17 | 183 | 53 | 236 | 22.5% |
| GDP per capita† | 6962.60 | 7987.95 | 127 | 17010.99 | 19625.66 | 189 | 47 | 236 | 19.9% |
| Public health expenditure‡ | 311.21 | 509.70 | 127 | 799.50 | 1792.02 | 187 | 49 | 236 | 20.8% |
| Neighbors' military expenditure per capita | 161.04 | 183.22 | 127 | 341.46 | 529.26 | 222 | 14 | 236 | 5.9% |
| World Bank CPI on Social inclusion and equity | 3.63 | 0.76 | 127 | . | . | . | . | . | . |
| World Bank CPI on Economic management | 3.96 | 1.05 | 127 | . | . | . | . | . | . |
| Investment-GDP ratio | 14.87 | 8.71 | 127 | 24.53 | 11.50 | 189 | 47 | 236 | 19.9% |
| Donor funding on basic health per capita* | 2.96 | 6.35 | 127 | 8.17 | 13.78 | 193 | 43 | 236 | 18.2% |
| Nonliterate rate | 21.22 | 22.15 | 127 | 15.91 | 18.96 | 133 | 103 | 236 | 43.6% |
| Paved road rate | 31.14 | 66.63 | 127 | 59.02 | 32.05 | 94 | 142 | 236 | 60.2% |
| Sanitation rate | 73.98 | 26.67 | 127 | 72.27 | 29.97 | 186 | 50 | 236 | 21.2% |

Notes:
*While in the original BGG data, per capita donor funding for health promotion from the Development Assistance Committee (DAC) of the Organization for Economic Co-operation and Development (OECD) in year 1998 is used. In the updated 2010 data, we instead use the data on per capita amount of Development Assistance for Health (DAH), available at the Institute of Health Metrics and Evaluation database.
†In the original data, GDP per capita using international dollars as of 2000 is used. In the updated data, GDP per capita using constant US dollars 2011 (PPP adjusted) is used.
‡In the original data, government health expenditure per capita (international dollars as of 2000) is used. In the updated data, government health expenditure per capita in 2005 constant dollars is used.

## Appendix D - Review of related studies using cross-country data

### Table D 1: Review of related studies using cross-country data

| Author | Publication year | Region | Data source | Data years | Outcome | Expenditure | Main result (Not necessarily significant) | Model | IV | Covariates |
|---|---|---|---|---|---|---|---|---|---|---|
| Afridi & Ventelou | 2013 | All, 113 countries | IHME, WHO | 1995-2006 | Adult mortality | Aid (public and private); Govt health expedinture | Aid reduces adult mortality | Dynamic panel system GMM | Lagged aid | per capita health aid by government; per capita health aid by private channel; per capita non-health aid; per capita government health expenditure; per capita private health expenditure; per capita GDP |
| Akinci et al | 2014 | MENA, 19 countries | WHO, WDI | 1990-2010 | Infant mortality; U-5 mortality; maternal mortality | Govt expenditure & govt expenditure as % of THE; Private expenditure & OOP expenditure as % of THE (separately) | Govt and private expenditure improve outcome variables | Housman-Taylor panel model | Random effect variable | govt expenditure on health as percantage of total exp on health; physicians density; hospital beds; population with access to safe drinking water; birth attended by skilled health personnel; adult literacy rate |
| Anyanwu & Erhijakpor | 2009 | Africa, 47 countries | WHO, WDI, African Development Bank's database | 1999-2004 | U-5 mortality rate; infant mortality rate | Total health expenditure; public health expenditure | Health expenditure reduces infant mortality and U-5 mortality | 2SLS | Consumption-investment rate (for income); Military expenditure of neighbouring countries and membership of the Franc zone (for expenditure) | Ethnolinguistic fractionalization; female literacy; physicians per 100,000 population; urban population; per capita GDP (HIV prevalence, not included due to data limitation) |
| Bokhari et al | 2007 | All, 127 countries | WB's in-house data, WDI, Millennium Indicators database, World Health Report, Human Development Report, CIA fact book, UN websites, Penn World Tables | 2000 | U-5 mortality; maternal mortality (MDG outcomes) | Public expenditure; donor funding | Health expenditure reduces U5MR and MMR | GMM-H2SL | Military expenditure of neighouring countries and its iteractions (for govt health expenditure); WB annual assessments on the quality of policies and institutions | GDP per capita (endogenous); % population aged 15 or above not literate; paved roads per unit area; % population with access to improved sanitation; donor funding per capita for basic health |

**Table D 1: Review of related studies using cross-country data (continued)**

| Author | Publication year | Region | Data source | Data years | Outcome | Expenditure | Main result (N.B. impacts are not always statistically or economically significant) | Model | IV | Covariates |
|---|---|---|---|---|---|---|---|---|---|---|
| Farag et al | 2013 | LMIC, 133 countries | NHA of WHO; UNICEF; World Rsources Institute; World Databank; WB | 1995, 2000, 2005, 2006 | U-5 mortality; infant mortality | Gvt health spending | Health expenditure reduces U5MR and Infant mortality | Fixed effect (no IV) | NA | per capita GDP; total health spending; govt health spending; donor health spending; fertility rate; urban population; female labor; school enrollment; improved water source; improved sanitation facilities; share of women employed in the nonagriculture sector; TB detection rate; female literacy; physicians per 100,000 |
| Filmer & Pritchett | 1999 | All, 119 countries | UNICEF; ministries of Health data; WB | 1990 | U-5 mortality; infant mortality | Public expenditure on health | Public health expenditure has small and insignificant impacts | 2SLS | Average public sector health spending as a share of GDP; average defense spending as a share of GDP of neighbouring countries | GDP per capita; public exp on health (share of GDP); female education; income inequality; predominantly muslim; ethno-linguisticc fractionalisation; percentage urban; tropical country; access to safe water |
| Gani | 2009 | Pacific Island, 7 countries | WHO, WB, Asian Development Bank | 1990, 2002 | U-5 mortality; infant mortality; crude death rate | Total health expenditure per capita | Health expenditure reduces U5MR and Infant mortality | Fixed effect controlling for AR(1), no IV | NA | per capita income; immunization; urbanisation; calorie intake |
| Hu & Mendoza | 2013 | All, 136 countries | Globalization and Health Nexus Database; WDI; International Country Risk Guide | 1960-2005 | U-5 mortality; infant mortality | Public spending on health | Public spending reduces child mortality | 2SLS | Legal origin of country (UK law, French law, socialist law, German law, Scandinavian law); democracy (democratic accountability index by International Country Risk Guide) | Volatility; Gini coefficient; female literacy rate; public spending on health (% GDP); quality of bureaucracy; control of corruption |

**Table D 1: Review of related studies using cross-country data (continued)**

| Author | Publication year | Region | Data source | Data years | Outcome | Expenditure | Main result (N.B. impacts are not always statistically or economically significant) | Model | IV | Covariates |
|---|---|---|---|---|---|---|---|---|---|---|
| Moreno-Serra & Smith | 2015 | All, 153 countries | WHO, WB, IHME | 1995-2008 | U-5 mortality; adult mortality (female and male) | Public expenditure; OOP expenditure; private insurance | Health expenditure reduces mortality | 2SLS | CO2 emission and battle related deaths (*reverse causality estimation) | GDP per capita; primary education enrolment rate; population aged under 14; pupulation aged over 65; CO2 emmission & conflict deaths (IVs) |
| Rajkumar & Swaroop | 2008 | All, 91 countries | WDI, and other sources (see Appandix A) | 1990, 1997, 2003 | U-5 mortality | Public health expenditure (ln of share of GDP) | Public health spending reduces mortality only for countries with good governance | 2SLS | Legal origin of country | GDP per capita; index of corruption; quality of bureaucracy; female education; income inequality; predominantly muslim; ethnolinguistic fractionalisation; access to safe water; degree of urbanization; percentage of population under 5; distance from the Equator |
| Wagstaff & Claeson | 2004 | All | WB, and other sources (see Table A.1 pp 171) | 1990s | U-5 mortality | Public health expenditure | Govt health spending modestly improves health for well governed countries | 2SLS | Average defense and health spending on neighbouring countries; average value of an index of "voice" among neighbours | Not known |