Int J Public Health (2017) 62:361–370
DOI 10.1007/s00038-016-0935-4

ORIGINAL ARTICLE

Examining the relationship between human resources and mortality: the effects of methodological choices

Pablo Villalobos Dintrans · Claire Chaumont

Received: 16 June 2016 / Revised: 30 November 2016 / Accepted: 2 December 2016 / Published online: 20 December 2016
© Swiss School of Public Health (SSPH+) 2016

Abstract
Objectives Relationship between human resources for health and mortality remains inconclusive despite numerous studies published on the topic in the last decades. This paper investigates how and why methodological trade-offs implicitly made by researchers when using macro-data can in part explains this puzzling lack of agreement.
Methods Using data from the Global Health Observatory, we build a model of the relationship between human resources and mortality, which we progressively alter by changing its scope, variables and analysis period. Then, we compare results among themselves to isolate the impact of methodological choices from other changes in the data.
Results Results demonstrate how methodological choices linked to (1) the analysis period, (2) the definition of health inputs, health outcomes and control variables and (3) the choice of specific variables as proxy for human resources and health outcomes affects the relationship between human resources and health outputs.
Conclusions Results presented highlights the need for complementing existing macro-analysis with other analytical strategies, for better documenting methodological choices in research studies, as well as for further supporting countries' efforts to produce reliable and consistent data.

Keywords Human resources · Health · Mortality · Methodology

Introduction

The relationship between health inputs and health outcomes has been extensively studied to understand—and ultimately justify—how investing in health systems can improve population health. From a health provider's perspective, the link between health inputs and health outcomes can be understood considering a health production function. The latter takes the form of a typical production function:[1]

$$H = h(K, L, E),$$

where health outcomes, H, are "produced" using a combination of inputs, such as capital, K (i.e., health infrastructure), labor, L (i.e., health human resources), and other health-related factors, E (e.g., environmental factors), for a given level of technology, h. The aim of most studies on this topic has been to empirically test this link.

A long tradition of research tries to establish relationships between broad measures of health outcomes and inputs, such as life expectancy at birth and Gross Domestic Product (GDP), which may be affected by determinants beyond the health system (Pritchett and Summers 1996; Bloom and Canning 2000; Case et al. 2002; Deaton 2002; van den Berg et al. 2006; Weil 2007). Other studies use narrower definitions of health outcomes, like child and maternal mortality, which seems particularly adequate, as

Electronic supplementary material The online version of this article (doi:10.1007/s00038-016-0935-4) contains supplementary material, which is available to authorized users.

P. Villalobos Dintrans · C. Chaumont (✉)
Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, USA
e-mail: claire.chaumont@mail.harvard.edu

P. Villalobos Dintrans
e-mail: pav783@mail.harvard.edu

[1] Health production functions from the health provider's perspective. A completely different literature is related to health production from the individuals/consumers/patients' perspective, based on Grossman's (1972, 2000) human capital model.

they are known to be directly impacted by health care provision. For example, availability of skilled health staff is identified as the most important determinant of maternal and newborn health (Dickson et al. 2014). In parallel, many studies have focused on human resources in health on the basis that, although the variety of inputs needed to "produce" health outcomes may be extensive, human resources are not only crucial but also a prerequisite for any type of healthcare (Hertz et al. 1994; Robinson and Wharrad 2001; Anand and Bärnighausen 2004; Ananthakrishnan et al. 2010).

However, despite numerous analyses on the relationship between human resources and health outcomes, no clear agreement emerges from the findings. Previous studies found positive effects of the number of primary care physicians on health outcomes (Franks and Fiscella 1998; Vogel and Ackerman 1998). These results have been replicated for different settings, datasets and measures (James and Cossman 2006; Ananthakrishnan et al. 2010; Castillo-Laborde 2011). Other investigations debate these results, both for primary care and specialist care (Gulliford 2002; Ricketts and Holmes 2007; Starfield et al. 2005). For example, Gulliford (2002) finds no association between the availability of primary care doctors and a series of health outcomes in England. Similarly, using data from 19 OECD countries, Watson and McGrail (2009) find no association between medical supply and avoidable mortality.

The provision of health care services—for which human resources are a prerequisite—is generally considered a key determinant of population health. Many countries invest a significant part of their wealth to develop and maintain such services. Getting a better understanding of this relationship is crucial to ensure that human and financial resources are used in an effective and efficient way. In this context, the lack of agreement on the relationship between human resources and health outputs is puzzling.

The studies reviewed exhibit two common features: (1) the use of aggregate data for empirical work and, (2) the utilization of broad and general definitions of health inputs and health outputs (see Annex for a comparison of studies' main features). This paper assesses how methodological choices around data selection and treatment can impact results, to better interpret the diverse range of results found in the literature. Specifically, it investigates how the (1) the period selection (2) the selection and operationalization of health inputs, health outcomes and control variables, and (3) the choice of specific variables as proxy for human resources and mortality affect the relationship between human resources and health outputs. It focuses on two mechanisms: changes in sample size and changes in observations' values.

The number of observations is crucial to explain results. Many papers deal with this challenge using data quality and availability as criteria for sample selection. Choosing a particular year to estimate cross-section regressions is usually based on data availability for critical variables, completeness in terms of number of observations and how updated the dataset is.

On the other hand, when picking the variables to include in a model, researchers are implicitly deciding the sample's composition and size. This reflects a trade-off between the number of variables and the number of observations included in a model. This problem is especially relevant when dealing with low quality data, such is the case when using aggregated health datasets from different countries.

This trade-off between study period and model specifications will ultimately affect results through two channels: unintended changes in sample size or changes in the observations' values due to time shocks or omitted variable bias. While changes linked to observations' values may reflect the causal relationship between health inputs and health outputs, irrelevant changes such as sample size can obscure this true relationship, ultimately leading to spurious associations between variables.

While decisions regarding such trade-offs are implicitly made by researchers using macro-data, their impact on final results is seldom considered, e.g., when performing robustness checks. They are also rarely presented explicitly as a research limitation, which can prevent less experienced readers from considering them when using research for policy making. The aim of this study is to deliberately show the impact of such choices on a model's results.

Methods

Data sources

The analysis uses data from the Global Health Observatory data repository (GHO). The GHO is an openly available dataset from the World Health Organization (WHO), containing key health indicators, including mortality ratios, human resources and infrastructure for health and other country characteristics from 194 countries for the period 1990–2013. The GHO is the WHO's main health statistics repository, and as such includes the most recent or revised data from member countries, as well as WHO's best estimates to monitor global trends, using methodologies allowing for cross-country and cross-period comparison.

We selected the GHO for several reasons. First, it is the single most comprehensive dataset on health statistics, both in terms of indicators and countries. This increases homogeneity and allows work with similar definitions for the same variables. Second, it is the most reliable and consistent database available at global level. Although differences in data collection procedures and definitions at

the country-level may remain, the WHO validation and triangulation process ensures a high level of standardization for each indicator in the database. Third, as the dataset is publicly available, it is often used by researchers, making our results directly relevant to the discussion on using macro-analysis for policy making. Finally, the use of a single dataset allows us to test the hypothesis that differences in results are explained mainly by the use of different datasets (Speybroeck et al. 2006; Anand and Bärnighausen 2007).

Study design

A four-step approach is used. A simple model is built, using infant mortality rate (IMR) as the dependent variable, health personnel density (including physicians, nurses and midwives) as independent variable and gross national income (GNI) per capita as control. The health personnel density variable—which includes data on doctors' density and nurses and midwives density—is generated by adding both variables, taking missing values as zeros. This strategy permits: (1) use of a more comprehensive measure of human resources for health, avoiding the generation of new missing values; and (2) taking a conservative approach, since the values in this new variable are a lower bound of the real value. While this approach can introduce bias towards increasing a negative relationship between human resources for health and mortality, we consider this limitation acceptable, since the aim of the paper is not to estimate coefficients, but to illustrate how estimations vary when different samples and covariates are used.

Using this simple model, we estimate several Ordinal Least Squares (White-Huber) robust regressions. A first set of estimations are obtained by altering the years of the sample, with 2004, 2009 and 2010 used as examples, as they have the highest number of observations for the independent variable (health personnel density). It is worth noting infant mortality has no missing data for all years included in the dataset.

Second, using 2004 and 2005 as years of reference, we progressively include control variables in our model. Years were chosen based on data availability and as illustrations to highlight the effects previously presented. Control variables are selected based on what has previously been used in the literature, including poverty, health infrastructure, health expenditure, and sanitary conditions (Anand and Bärnighausen 2004; Bryce et al. 2003; Bhargava et al. 2005; Farahani et al. 2009). In a third stage, we compared our results with panel data regressions, using data from 1990 to 2013 to show the effect of exploiting a richer dataset.

Finally, the analysis is reproduced for four different "proxies" of health outcomes: maternal mortality, under five mortality, neonatal mortality and infant mortality. Coefficients are also estimated using three different measures of the independent variable: physician's density, personnel defined as physicians and nurses and midwives, and a broader definition containing physicians, nurses, midwives and community health workers. Results are compared at each step of the model to assess the extent to which choosing specific variables as "proxies" for health outcomes and human resources affects the relationship between them. As the results obtained are similar for the different specifications, only results for infant mortality rate and density of physicians and nurses and midwives are presented in this article.

Results

Initial explorations

A first exploration highlights how year and variable selection drastically affect sample size: while our simple model includes 94 observations in 2004, this number increases to 98 in 2009 but drops to 80 in 2010. On the other hand, adding new variables generates a similar effect: while the more parsimonious model has 59 observations in 2005, a more complete model only includes 13 countries. This drastic attrition appears to be non-random: Table 1 shows the average value of several variables for subsamples of countries with and without information on health personnel, health infrastructure, and poverty, three of the most common variables used in previous studies. In general, countries with available information on health inputs (personnel and beds) have better economic and health conditions than those without data. This suggests that selection of variables in a dataset with information gaps is not innocuous, since data availability can be correlated with one or more variables in the model.

Year selection: detangling the role of time shocks from sample selection

Table 2 shows a set of estimations illustrating the effects of choosing a particular year, using our simple model for three different years (2004, 2009 and 2010). We observe that although coefficients for human resource density are similar in terms of significance and sign, they show important differences in magnitude and overall goodness of fit (columns 1–3). These differences can be attributed to differences in the sample or differences in the variables' values.

To disentangle these effects, the regression is estimated using the same sample (countries) for different years (columns 4–6). These estimations are based on a sample of

Table 1 Average values of gross national income per capita, health expenditure, access to water, and mortality measures for countries with and without data on health human resources, bed per population and poverty, Global health repository dataset, 194 countries, 1990–2013 Source: Authors' elaboration

	Health human resources			Beds		
	No	Yes	Diff	No	Yes	Diff
GNI per capita	9926.72	14,459.06	−4532.34 ***	8952.90	13,926.05	−4973.15***
Health expenditure per capita (purchasing power parity)	814.8538	1012.60	−197.74***	703.95	1122.07	−418.12***
Water access (%)	82.47	88.45	−5.98***	80.14	88.29	−8.15***
Maternal mortality	286.04	133.70	152.33***	318.52	159.63	158.89***
Infant mortality	40.68	28.28	12.40***	46.29	25.09	21.19***
Under-five mortality	59.13	38.48	20.65***	68.24	33.50	34.75***
Neonatal mortality	20.85	15.83	5.02***	23.36	14.20	9.16***

While the whole dataset ($N = 194$) was used for this analysis, missing variables will affect the analytical sample size used for each average value

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2 Effect of year selection on the relationship between infant mortality rate and health density, controlling for income

Dependent variable: infant mortality rate	1 2004: full sample	2 2009: full sample	3 2010: full sample	4 2004: common sample	5 2009: common sample	6 2010: common sample
Gross national income per capita (PPP)/1000	−0.795*** [0.000]	−0.456*** [0.000]	−0.423** [0.000]	−0.528** [0.000]	−0.662*** [0.000]	−0.601*** [0.000]
Doctors + nurses and midwives	−3.807*** [0.909]	−1.254*** [0.442]	−1.496*** [0.399]	−2.796*** [0.870]	−1.491** [0.561]	−1.485*** [0.467]
Constant	67.543*** [3.640]	40.849*** [3.652]	38.640*** [4.335]	50.232*** [6.115]	42.413*** [4.862]	41.516*** [4.842]
Observations (number of countries)	94	98	80	40	40	40
Adjusted R^2	0.479	0.309	0.349	0.425	0.469	0.48

Ordinary Least Squares estimations, Global health repository dataset, 2004, 2009 and 2010

Robust standard errors in brackets

Full sample includes all the countries with data for each year. Common sample includes only countries with data in all the three years (2004, 2009, 2010)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

40 countries, which represents the "greatest common sample", i.e., all countries with non-missing values for the three selected years. These new estimations show the temporal effect (change in values) for this particular group of countries: in 2004 the relationship between health personnel and infant mortality rate seems particularly strong, although part of the effect can be attributable to the countries included in the sample (columns 1 and 4). For year 2009 the opposite effect is observed: when using the same sample, the coefficients for 2009 and 2010 are practically identical (columns 5 and 6), but the sample effect reduces the coefficient showing a weaker relationship between both variables (columns 2 and 5). On the other hand, the estimations for 2010 are not affected by differences in sample, even though the full sample is twice the reduced one.

In conclusion, while our initial comparison suggests time shocks may explain differences in coefficients between years, an analysis using the "greatest common sample" for 2004, 2009 and 2010 shows how changes in sample composition also partially drive results.

Variables selection: detangling the role of omitted variable bias from sample selection

As discussed before, the addition of variables to the model has two effects in the context of this dataset: first, it allows controlling for confounders in the relationship between human resources and mortality; second, it changes the sample used to calculate these effects (coefficients).

The effects of changes in the sample are summarized in Table 3. Again starting from a simple model including one

Table 3 Effect of variable selection on the relationship between infant mortality rate and health density, controlling for income

Dependent variable: infant mortality rate	1 GNI per capita	2 GNI per capita + poverty	3 GNI per capita, sample restricted to countries with poverty value
Gross national income per capita (PPP)/1000	−0.795*** [0.000]	−4.756*** [0.001]	−5.833*** [0.001]
Doctors + nurses and midwives	−3.807*** [0.909]	−0.342 [1.780]	−0.193 [1.963]
Poverty headcount ratio		0.271* [0.142]	
Constant	67.543*** [3.640]	67.982*** [8.973]	80.108*** [4.908]
Observations (number of countries)	94	18	18
Adjusted R^2	0.479	0.85	0.827

Ordinary Least Squares estimations, Global health repository dataset, 2004

Robust standard errors in brackets

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

health input and one control (column 1), the addition of an extra control (poverty) (column 2) changes the coefficients and their interpretation. When moving from column 1 to column 2, it is clear that the new added variable is relevant; on the other hand, when adding poverty to the model the effect of health human resources on infant mortality rate becomes non-significant.

The usual interpretation for this result is that health human resources act as a confounder in the relationship between poverty and mortality (omitted variable bias): the effect of human resources on mortality (column 1) would be considered spurious, because it simply reflects the influence of poverty. However, including poverty to the model not only adds a new source of variation, but also changes the sample. If the observations dropped when adding this new variable are not random, then the results can be explained not because of the covariation between independent and dependent variables, but simply by the change in the number and characteristic of the sample. Columns 2 and 3 show this. Column 3 exhibits the coefficient of health human resources in a regression without poverty as explanatory variable, but using only observations with values for poverty in that year, showing that adding poverty (column 2), does not lead to substantially different results when using the same sample (column 3). Although this subsequent result might not mean much in itself due to the small sample size used (18 observations), it still demonstrates how the change observed between column 1 and column 2 mostly results from a change in sample.

A second effect of changing the model's specification is that the addition of new variables can help in understanding the effect of health personnel on infant mortality rate. Table 4 shows the effect of adding new variables to the initial simple model (column 1). Column 2 shows the effect of adding an extra measure of health input: per capita health expenditure. This variable has data for every observation, which implies that changes in coefficients are not due to changes in the sample (as described before). The variable, however, proves to be not significant and the changes observed between columns 1 and 2 are negligible.

Column 3 adds another health input: health infrastructure, measured as number of beds available in the country, plus an interaction term (personnel and infrastructure). Unfortunately, when including this variable some observations are lost, which raise concerns about the effect of the change in the sample on the coefficients. Column 4 exhibits the same regression than column 2, but using the same sample used in column 3 (effect of sample selection). Despite some differences in the magnitude of the coefficient for health human resources, both columns (2 and 4) look very similar. After adding beds per 1000 people as explanatory variable, the effect of health personnel on mortality is more pronounced, with an important increase in the model's adjusted R^2. These results imply that infrastructure has an independent effect on mortality, and this change is not simply due to changes in the sample.

Similarly, columns 5 and 6 show the effect of adding new controls: water and sanitation conditions. Again, the results of just changing the sample composition (column 6) are small, while adding the variables contributes to increasing the explanatory power of the model.

Finally, columns 7 and 8 show the effect of including poverty as control. Unfortunately, adding this variable reduces the sample size considerably. As before, it shows to be not significant, and the effect of including poverty on the coefficient of health personnel can be interpreted as due to changes in the sample composition.

For all estimations, including new variables change the results because of the addition of a relevant (previously omitted) variable but also because of a potential change in

Table 4 Variable selection effect (omitted variables) on the relationship between infant mortality rate and health density

Dependent variable: infant mortality rate	1	2	3	4	5	6	7	8
Doctors + nurses and midwives	−2.89*** [0.78]	−2.444** [1.064]	−4.407* [2.369]	−2.741* [1.369]	−0.593 [1.672]	−4.412* [2.375]	−7.283 [12.387]	−7.51 [11.405]
Beds per 1000 people			−6.663 [3.961]		−1.382 [2.070]	−6.663 [3.969]	−6.937 [13.093]	−6.849 [12.458]
Doctors + nurses and midwives × beds per 1000 people			0.767 [0.616]		0.094 [0.355]	0.766 [0.617]	0.944 [2.024]	0.997 [1.902]
Gross national income per capita (PPP)/1000	−0.40*** [0.00]	−0.352** [0.000]	−0.233** [0.000]	−0.232* [0.000]	−0.048 [0.000]	−0.239** [0.000]	5.352 [0.009]	4.646 [0.008]
Per capita health expenditure (total, PPP)		−0.002 [0.004]	−0.001 [0.002]	−0.002 [0.004]	0 [0.001]	−0.001 [0.003]	−0.01 [0.049]	−0.006 [0.050]
Use of improved water source (total)					−0.912*** [0.302]		−0.161 [2.418]	−0.418 [2.078]
Use of improved sanitation facilities (total)					−0.32 [0.190]		−0.903 [1.877]	−0.895 [1.642]
Poverty headcount ratio							0.239 [0.786]	
Constant	47.62*** [4.86]	47.37*** [4.922]	56.39*** [10.806]	42.58*** [6.355]	138.7*** [16.732]	56.41*** [10.837]	91.126 [162.411]	120.855 [133.812]
Observations (number of countries)	59	59	41	41	40	40	13	13
Adjusted R^2	0.37	0.358	0.428	0.352	0.669	0.417	−0.361	−0.106

Ordinary Least Squares estimations, Global health repository dataset, 2005

Robust standard errors in brackets

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the sample composition. Given the importance of both effects, they should be addressed separately.

Panel estimations

In a third step, to benefit from the rich dataset, a set of regressions is estimated using data from countries in several years. Panel data combines time series and cross-section, following the same units of observation throughout time. Given this feature, it has several advantages over cross-section studies: more accurate inference of parameters, greater capacity for capturing complexity, and simplifying computation and statistical inference (Hsiao 2007). Table 5 shows the results of the fixed effects estimations using the panel data. Fixed effects estimations are estimated based on the results of the Hausman's test for both set of regressions.

Using this different method, coefficients shows no effect of health inputs on infant mortality. These results hold when using different measures of human resources for health and mortality.

Variable definition: choice of variables as proxy for human resources and mortality

Finally, each step of both models is reproduced using different measures of human resources (combinations of physicians, nurses and midwives and community health workers density) and mortality (maternal, neonatal or under five mortality). Results do not change when using different measures of health, suggesting they are not driven by which variables are chosen as "proxy" for human resources and mortality. However, it also suggests that the challenges identified above hold true across all variables. For simplicity purposes, details of these different models are not presented here but are available upon demand.

Discussion

Results presented show the relevance of understanding several methodological choices when dealing with incomplete and low quality data. Several conclusions can be drawn from this experimentation:

Table 5 Relationship between infant mortality rate and health density

Dependent variable: infant mortality rate	1	2	3	4	5
Doctors + nurses and midwives	−0.174 [0.132]	0.0003 [0.068]	0.034 [0.083]	−0.211 [0.201]	−0.039 [0.128]
Beds per 1000 people		0.302 [0.716]	0.245 [0.781]	−0.114 [0.793]	−1.063 [0.711]
Doctors + nurses and midwives × beds per 1000 people				0.051 [0.041]	0.021 [0.031]
Gross national income per capita (PPP)	−0.000*** [0.000]	−0.000** [0.000]	0.000 [0.000]	0.000 [0.000]	0.000** [0.000]
Per capita health expenditure (total, PPP)		0.001 [0.001]	−0.001 [0.001]	−0.001 [0.001]	0.001 [0.001]
Use of improved water source (total)			−1.581** [0.647]	−1.598** [0.656]	−1.364*** [0.465]
Use of improved sanitation facilities (total)			0.178 [0.381]	0.187 [0.387]	0.334 [0.291]
Year dummy	No	No	No	No	Yes
Constant	35.651*** [1.834]	25.991*** [4.541]	151.528*** [29.625]	153.489*** [30.164]	124.774*** [20.831]
Observations	1100	504	134	134	134
Number of countries	178	135	89	89	89
Adjusted R^2	0.099	0.119	0.749	0.749	0.833

Fixed effect estimations, Global health repository dataset, 1990–2013

Robust standard errors in brackets

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

- Differences in coefficients observed for both simple and more complex models are in part driven by changes in sample composition, affected by selection of years and/or variables. The weights of these changes can and should be disentangled from changes related to time shocks or the inclusion of new variables dealing with omitted variable bias.
- Using panel data can help deal with these challenges, by limiting variations in sample composition—ultimately improving the results' robustness.
- Results do not change when using different measures of health human resources (different combinations of physicians, nurses and midwives and community health workers density) or mortality (maternal, neonatal or under five mortality), suggesting challenges with year and sample selection are consistent regardless of which variables are used to define what "health outcomes" and "human resources" mean.

These findings demonstrate how methodological choices affect the relationship between human resources and health. Most of these results comply with traditional "robustness checks" such as using alternatives measures for the independent variable (results are the same when using maternal mortality, neonatal mortality, and under five mortality) or the dependent variables (health personnel defined as physicians only or adding community health workers). However, our estimations showed to be sensitive to changes in specification and years used, highlighting how choices made during the analytical stage greatly affect the estimated relationship between human resources and mortality. This helps explaining the puzzling lack of agreement in earlier analyses regarding the link between human resources and health outcomes. Previous studies highlighted how combining different aggregated databases present methodological challenges (Speybroeck et al. 2006). This study shows how important challenges remain even when using a combined dataset.

These problems are inherent to the use of macro-data in health. But although their impact on results may be well understood by researchers, they are seldom reported as an explicit limitation, and their impact on results is rarely explored in depth—maybe because researchers consider them as the "necessary evil" of using macro-data.

Macro-data studies are often used by policy-makers to guide policy (see Annex). A simple citation count on Google scholar shows the combined eight studies reviewed for this exploration were quoted more than 1000 times since their publication, including in the World Health Report 2006, for advocacy purposes or as references in

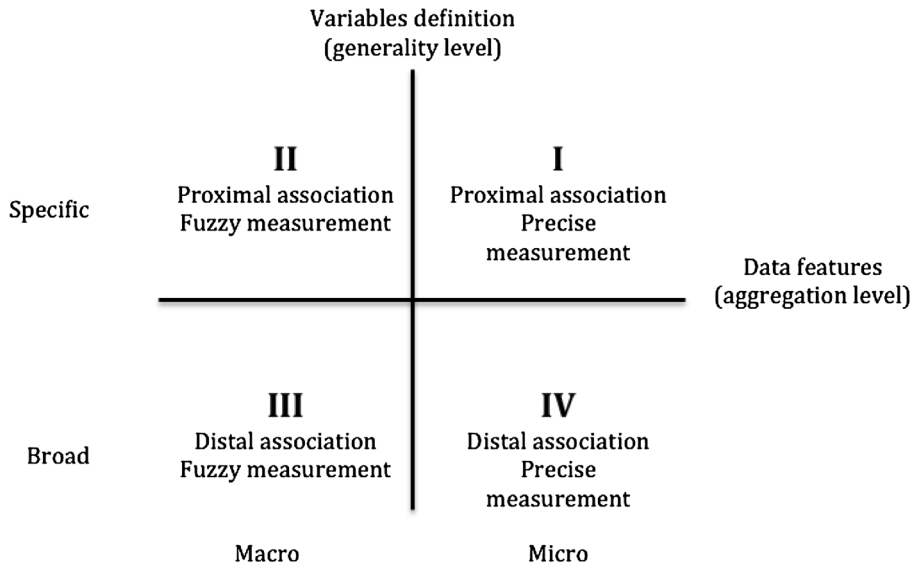

Fig. 1 Definition, measurement and association between variables

textbooks (Ricketts and Holmes 2007; Starfield et al. 2005; Anand and Bärnighausen 2004). Increased access to macro-level databases provides a great opportunity for research which would be absurd to dismiss. But the weight that macro-data studies can have in policy making means we must better acknowledge their limitations and their implications in terms of drawing evidence. Besides, it is worth noting that while we chose human resources and health for our analysis, these challenges likely hold true for other disagreements in empirical health studies, such as the link between health infrastructure and health outcomes, or the relationship between economic growth and health.

In the short-run, these shortcomings can be addressed by two alternative analytical strategies, illustrated in Fig. 1: alternative (1) attempts to improve results by changing the inputs (move from Quadrant III to I or IV) or establishing closer causal relationships between variables; alternative (2) seeks to improve estimations with a given dataset (Quadrant III).

(i) use microdata and other econometric techniques, such as instrumental variables to deal with endogeneity and measurement problems (Strauss 1986; Deolalikar 1988; Thomas and Strauss 1997; Schultz and Tansel 1997; Croppenstedt and Muller 2000; Adams et al. 2003; Rivera and Currais 2005; Schultz 2002, 2005; Cutler et al. 2006; Michaud and van Soest 2008). This will address the fact that causality using aggregate data is frequently diffuse and potentially spurious (Easterlin 2013). The use of randomized experiments can also help with endogeneity problems (Thomas et al. 2006; Dillon et al. 2012)

(ii) take advantage of richer datasets to estimate panel data regressions, as shown in this study. The dynamic nature of panel data can be useful in improving future estimations, to get more conclusive results (Hsiao 2007; Farahani et al. 2009).

In the long run these results also emphasize the need for systematic and standardized data collection, to improve completeness and comparability. International organizations such as WHO have a key role in developing standardized data gathering and control at international level, but also promoting the development of strong Monitoring and Evaluation (M&E) systems at country-level.

Finally, researchers also play a role, by more systematically reporting and documenting the impact of methodological choices on studies' results, including sample and variable selection, but also broader methodological issues not directly addressed in this exploration, such as the impact of variable categorization or the number of variables used for the analysis, etc.

Conclusion

Improved data collection and M&E processes in recent decades mean macro-analysis can now be performed using data across most countries in the world. This has led researchers to use aggregated databases to investigate the crucial relationships between human resources and health outcomes. As this study has shown, the lack of agreement found across published work can in part be explained by methodological choices made regarding study period and model specifications. While these challenges should not prevent researchers from using aggregated databases, this paper highlights the need to have an open discussion about their potential limitations. In addition, while overcoming these limitations might be a distant goal, it is essential to complement these macro-analysis with other strategies

such as microanalysis, panel data or the use of advanced econometric techniques (such as instrumental variables or quasi-experimental methods), as well as to support countries' efforts to produce reliable and consistent data.

Acknowledgements We thank the 2015 class of "Concepts and Methods in Global health" at Harvard T.H. Chan School of Public Health, led by Till Bärnighausen and Barry Bloom for their constructive feedback throughout the class, as well as Chris Wheelahan, and two anonymous reviewers for their careful reading of the manuscript.

Compliance with ethical standards

No funding was received for this study.
This article does not contain any studies with human participants or animals performed by any of the authors. No ethical approval was obtained for this study, as it uses publically available data.

References

Adams P, Hurd M, McFadden D, Merril A, Ribeiro T (2003) Healthy, Wealthy, and Wise? Test for Direct Causal Paths between Health and Socioeconomic Status. J Econom 112:3–56

Anand S, Bärnighausen T (2004) Human resources and health outcomes: cross-country econometric study. Lancet 364:1603–1609

Anand S, Bärnighausen T (2007) Health workers and vaccination coverage in developing countries: an econometric analysis. Lancet 369:1277–1285

Ananthakrishnan AN, Hoffmann RG, Saeian K (2010) Higher physician density is associated with lower incidence of late-stage colorectal cancer. J Gen Intern Med 25(11):1164–1171

Bhargava A, Chowdhury S, Singh KK (2005) Healthcare infrastructure, contraceptive use and infant mortality in Uttar Pradesh, India. Econ Hum Biol 3:388–404

Bloom DE, Canning D (2000) The health and wealth of nations. Science 287:1207–1209

Bryce J, el Arifeen S, Pariyo G, Lanata CF, Gwatkin D, Habicht JP (2003) Reducing child mortality: can public health deliver? Lancet 362(9378):159–164

Case A, Lubotsky D, Paxson C (2002) Economic status and health in childhood: the origins of the gradient. Am Econ Rev 92(5):1308–1334

Castillo-Laborde C (2011) Human resources for health and burden of disease: an econometric approach. Hum Resour Health 9:4

Croppenstedt A, Muller C (2000) The Impact of Farmer's Health and Nutritional Status on Their Productivity and Efficiency: Evidence from Ethiopia. Ec Devel Cult Change 48(3):475–502

Cutler D, Deaton A, Lleras-Muney A (2006) The determinants of mortality. The J Econ Perspect 20(3):97–120

Deaton A (2002) Policy implications of the gradient of health and wealth. Health Aff 21(2):13–30

Deolalikar AB (1988) Nutrition and Labor Producvivity in Agriculture: Estimates for Rural South India. Rev Econ Stats 70(3): 406–414

Dickson KE, Simen-Kapeu A, Kinney MV, Huicho L, Vesel L, Lackritz E, de Johnson Graft J, von Xylander S, Rafique N, Sylla M, Mwansambo C, Daelmans B, Lawn JE (2014) Every Newborn: health-systems bottlenecks and strategies to accelerate scale-up in countries. Lancet 384:438–454

Dillon A, Friedman J, Serneels P (2012) Experimental estimates of the impact of malaria treatment on agricultural worker productivity, labor supply and earnings. International Bureau of Labor Affairs

Easterlin RA (2013) Cross-Sections Are History. Popul Dev Rev 38:302–308

Farahani M, Subramanian SV, Canning D (2009) The effect of changes in health sector resources on infant mortality in the short-run and the long-run: a longitudinal econometric analysis. Soc Sci Med 68(11):1918–1925

Franks P, Fiscella K (1998) Primary care physicians and specialists as personal physicians. Health care expenditures and mortality experience. J Fam Pract 47(2):105–109

Grossman M (1972) On the concept of health capital and the demand for health. J Political Econ 80(2):223–255

Grossman M (2000) The human capital model. In: Culyer AJ, Newhouse JP (eds) Handbook of health economics, vol 1. Elsevier, Amsterdam

Gulliford MC (2002) Availability of primary care doctors and population health in England: is there an association? J Public Health Med 24(4):252–254

Hertz E, Hebert JR, Landon J (1994) Social and environmental factors and life expectancy, infant mortality, and maternal mortality rates: results of a cross-national comparison. Soc Sci Med 39:105–114

Hsiao C (2007) Panel data analysis: advantages and challenges. TEST 16(1):1–22

James WL, Cossman JS (2006) Does Regional Variation Affect Ecological Mortality Research? An Examination of Mortality, Income Inequality and Health Infrastructure in the Mississippi Delta. Popul Res Policy Rev 25:175–195

Michaud PC, van Soest A (2008) Health and wealth of elderly couples: Causality tests using dynamic panel data models. J Health Econ 27(5):1312–1325

Pritchett L, Summers LH (1996) Wealthier is healthier. J Hum Resour 31(4):841–868

Ricketts TC, Holmes GM (2007) Mortality and physician supply: does region hold the key to the paradox? Health Serv Res 42(6 Pt 1):2233–2251 **(discussion 2294–323)**

Rivera B, Currais L (2005) Individual Returns to Health in Brazil. In: López-Casasnovas G, Rivera B, Currais L (eds) Health and Economic Growth: Findings and Policy Implications. MIT Press, Cambridge

Robinson J, Wharrad H (2001) The relationship between attendance at birth and maternal mortality rates: an exploration of United Nations' data sets including the ratios of physicians and nurses to population, GNP per capita and female literacy. J Adv Nurs 34:445–455

Schultz TP (2002) Wage Gains Associated with Height as a Form of Health Human Capital. Am Econ Rev 92(2):349–353

Schultz TP (2005) Productive Benefits of Health: Evidence from Low-Income Countries. Economic Growth Center Discussion Paper N° 903

Schultz TP, Tansel A (1997) Wage and Labor Supply Effects of Illness in Côte d'Ivore and Ghana: Instrumental Variable Estimates for Days Disabled. J Dev Econ 53(2):251–286

Speybroeck N, Kinfu Y, Dal Poz MR, Evans DB (2006) Reassessing the relationship between human resources for health, intervention coverage and health outcomes. Background paper prepared for The World Health Report 2006

Starfield B, Shi L, Grover A, Macinko J (2005) The effects of specialist supply on populations' health: assessing the evidence. *Health Aff 24*: W5–97

Strauss J (1986) Does Better Nutrition Raise Farm Productivity? J Polit Econ 94(2):297–320

Thomas D, Strauss J (1997). Health, Wealth, and Wages: Evidence on Men and Women in Brazil. J Econom 77(1):159–185

Thomas D, Frankenberg E, Friedman J, Habicht JP, Hakimi M, Ingwersen N, Jasgawi, Jones N, McKelvey C, Pelto G, Sikoki B, Seeman T, Smith JP, Sumantri C, Suriastini W, Wilope S (2006). Causal effect of health on labor market outcomes: Experimental evidence. California Center for Population Research 070–06

van den Berg GJ, Lindeboom M, Portrait F (2006) Economic conditions early in life and individual mortality. Am Econ Rev 96(1):290–302

Vogel RL, Ackermann RJ (1998) Is primary care physician supply correlated with health outcomes? Int J Health Serv 28(1):183–196

Watson DE, McGrail KM (2009) More doctors or better care? Healthc Q 12(4):101–104

Weil DN (2007) Accounting for the effect of health on economic growth. Q J Econ 122(3):1265–1306