www.elsevier.com/locate/worlddev

doi:10.1016/j.worlddev.2011.07.021

Chasing Success: Health Sector Aid and Mortality

SVEN E. WILSON [*]

Brigham Young University, Provo, USA

Summary. — As many cases studies show, successful public health measures are being implemented in many places around the globe, and country-level mortality has fallen significantly in recent decades in all but a few countries. Are the two linked? Does development assistance for health (DAH) improve, on balance, recipient countries' mortality trajectory? Using a new data source containing DAH on 96 high mortality countries, the regression analysis shows no effect of DAH on mortality. Other types of aid, including water development, also have no effect. Economic growth, on the other hand, has a strong negative effect on mortality. These findings confirm and build upon recent work by Williamson (2008) and are shown to be robust to a variety of sensitivity analyses and alternative model specifications and estimation methods.

This analysis also shows that the effectiveness of DAH has not increased over time, even as the level of that funding has increased four-fold, though spending on infectious diseases and family planning may have caused small reductions in mortality. Furthermore, even though it is encouraging that DAH has tended to go where the need is highest, it also goes to states that have experienced the greatest mortality reductions in the recent past. In other words, DAH appears to be *following* success, rather than *causing* it.
© 2011 Elsevier Ltd. All rights reserved.

Key words — aid effectiveness, mortality, health, public health

1. INTRODUCTION

Effective public health measures can save lives. That fact is not in question. The rich, healthy countries of the world were, at one time, as unhealthy as most impoverished nations today and, in many cases, even worse. Clean water, effective sanitation, immunizations, antibiotics, rehydration therapy, malaria prevention and treatment, and better nutrition have, among many other tools, dramatically reduced mortality among infants, children and adults across the globe, both historically and in recent decades. In general, the public health practices necessary to implement wide-scale reductions in mortality are not particularly complex—though they can be multi-faceted—nor particularly expensive.

Health, many believe, is an area where development assistance is likely to see positive results. Bold titles such as *Millions Saved: Proven Success in Global Health* (Levine, 2004) highlight apparent success at virtually eliminating measles in southern Africa, in eradicating smallpox globally, in preventing STDs in Thailand, in reducing child mortality through vitamin A in Nepal, in successfully implementing rehydration therapies and reducing diarrheal deaths in Egypt, and many others. These types of case studies are very encouraging. [1]

But these successes might lead to misplaced faith because they tell us little about whether development assistance for health (DAH), in aggregate, has had a positive impact on health outcomes. To answer that question correctly, we need to look at *all* DAH programs in *all* places, rather than identifying only highly successful cases. When all the DAH projects are added up, do they result in a meaningful increase in public health or health care in the recipient country and is DAH actually improving health beyond where it would be without the assistance? Using a variety of specifications and alternative assumptions, I search diligently for statistical evidence that aggregate DAH reduces mortality. I find—over and over again—no correlation. These investigations point overwhelmingly to noneffectiveness of DAH on mortality, whether using infant mortality (IMR), child (under 5) mortality (CMR), or life expectancy at birth (e_0).

Williamson (2008) was the first to look at DAH and mortality, and she also found no effect of DAH. But the analysis here extends what she did in many ways. The AidData database contains many large donors not found in the standard OECD-CRS data used by Williamson, and with that extended data, I explore extensive sensitivity analyses and alternative specifications she did not use, especially the latent growth model. [2] Furthermore, this analysis is the first to estimate changes in aid effectiveness across time and the effectiveness of the various components of DAH, such as spending on infectious diseases. Encouragingly, there is some evidence that spending on HIV/AIDS, other infectious disease, and family planning have had statistically significant (but very small) effects on mortality.

The most important innovation of this analysis is that I explore the impact that endogenously determined DAH may have on the aid effectiveness story in a manner that goes beyond GMM models. [3] In particular, I estimate a simple aid allocation model and find that reductions in mortality lead to large and significant increases in incoming DAH flows. This finding has two important implications. The first is that DAH effectiveness is not hiding behind the endogeneity problem, since the correlation between DAH and unobserved factors affecting mortality appears to be in the *opposite* direction than would be necessary to undermine the noneffect of DAH. Second, the increasing flows to countries that have *already* experienced reductions in mortality suggest a possible reason for aid ineffectiveness that has received little attention: DAH flows are responsive to country-level variables, but money is flowing to where mortality is improving rapidly for reasons other than effective aid programs.

In sum, even though the mortality trajectories of countries have almost universally and nontrivially improved since 1975 across the globe, countries receiving high levels of DAH have done no better, on average, than countries receiving low levels of DAH. There are certainly many public health programs and projects that reduce mortality *among those treated* (I empha-

---

[*] Final revision accepted: May 6, 2011.

size again that the effectiveness of proven public health programs is not in dispute—rehydration works; vaccinations work; public sanitation works), but at the country level, the effect of DAH on mortality is still missing. And because so many nations have made vast progress in reducing mortality, the continuing presence of preventable deaths of millions of children and adults around the globe each year is particularly damning evidence of political failure, of the inability to get resources and people in place that can address the relatively simple problems that are killing people. Even today, numerous developing countries have infant mortality rates (IMR) well above 100 deaths per 1,000 live births, certainly a travesty when Western Europe has an IMR of less than 5 per 1,000. The successes—and the failures—indicate a need to better understand the effects of development assistance.

2. BACKGROUND

Human health is multi-dimensional and can be measured in many ways. But the most salient and most important indicator of poor health is death. This is particularly true in the developing world where millions die each year from easily preventable causes. Infants are especially vulnerable and reductions in infant deaths have the greatest effect on a population's life expectancy. This analysis uses the three most common measures of mortality—IMR, CMR, and e_0—as measures of population health, but the results presented here are robust to all three measures (since there is such a high correlation between them).

Mortality is also tightly connected with economic growth. However, improving health is driven mostly by technology (in the economist's sense of the word, which has to do with "know-how," rather than equipment) not by wealth. In the late 19th century the richest countries in the world all had IMR rates well over 100 (Woods, Watterson, & Woodward, 1988) because their public health practices were abysmal—mostly because they did not know better, not because they lacked resources. It was not until the 20th century that the germ theory of disease took firm hold on the population. The dramatic 20th century decline in IMR was primarily a result of better medical and public health practices (including immunization, antibiotics, indoor and outdoor sanitation) used to fight the infectious diseases that are so dangerous to children and adults. [4]

(a) Changes in levels and composition of DAH

In the past two decades a lot of attention and funds have been flowing into the health sectors of the world's developing economies. A recent examination of all DAH funds (Ravishankar et al., 2009), shows a fourfold increase between 1990 and 2007 in all public and private funds that can be tracked with available data sources. This increase has been significantly augmented by the entrance of the Bill and Melinda Gates Foundation and the creation of large multilateral organizations such as the Global Fund for Aids, Tuberculosis and Malaria (GFATM) and the Global Alliance for Vaccines and Immunizations (GAVI) as well as increases by traditional sources (developed countries, the World Bank, and regional development banks).

Composition of DAH across health care sub-categories has also changed significantly as well, which is shown in Figure 1 (the data and categories will be described below). [5] This is primarily due to the global AIDS epidemic. Since 1990, expenditures on AIDS and other STDs (almost all of which is targeted at HIV/AIDS) exploded, and spending on HIV/AIDS is now
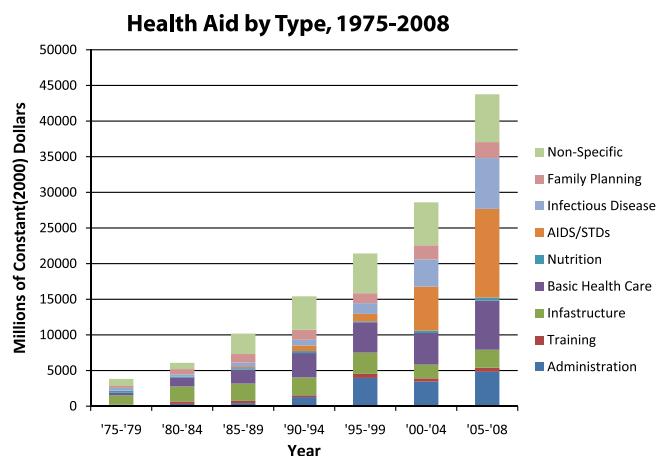

Health Aid by Type, 1975-2008

Figure 1. *Health aid by type, 1975–2008.*

the single biggest category of DAH spending. [6] But AIDS funding accounts for only a little over a third of the increase since 1990. [7] Other sub-sectors have changed significantly as well. Prior to 1990, DAH was dominated by infrastructure, basic health care and family planning. Since then, spending on combating infectious disease has expanded, as has administration, while all other categories have declined in terms of a percentage of total DAH. Battling infectious diseases, including HIV/AIDS, has come to dominate DAH, and new funding sources (GFATM, GAVI, and the Gates Foundation) have had this as their primary focus.

(b) The effectiveness of DAH

This study is not the first to examine the effectiveness of health sector aid. Recent work by Williamson (2008) has a similar approach as the one used here. [8] She finds a negligible impact of health sector aid on a variety of health outcomes, including IMR. Boone (1996) and Kosack and Tobin (2006) also find no impact of development assistance on infant mortality or life expectancy. Negative findings are not universal, however. Gomanee, Morrissey, Mosley, and Verschoor (2005) find that total aid flows (as a percentage of GDP) do lead to higher levels of aggregate welfare, as measured by the Human Development Index (HDI), though the effects are weaker for infant mortality. And Kosack (2003) argues that development aid has a positive effect on HDI when the country is a democracy but a negative effect in an autocracy.

Very little is known about health aid effectiveness at the sub-aggregate level—such as the effect of projects targeted at particular health problems or at particular communities. White (2003) looks at specific health interventions in Bangladesh and finds that health outcomes are not related to health aid but are related to aid in other sectors. In a carefully controlled 15 year study in rural Gambia, Hill, MacLeaod, Joof, Gomez, and Walraven (2000) reports that both villages with assistance in providing primary care and those without assistance experience declines in child health. Some work does show the effectiveness of some times of targeted aid, however. Du Lou, Pison, and Aaby (1995), for instance, evaluate a vaccination program in Senegal and find, unsurprisingly, a relationship between vaccination rates and child mortality.

As discussed earlier, the *Millions Saved* (Levine, 2004) review conducted at the Center for Global Development discusses a large number of cases where public health programs (with a variety of immunization efforts being the most

convincing) were associated with positive effects. Furthermore, a vast body of epidemiological literature shows the effectiveness of various targeted interventions on mortality. These are to be expected because we know very well what will happen if people are given immunizations, antibiotics, clean drinking water, more calories and better nutrition. What is evident, however, is our limited understanding of what happens to the dollars that are allocated by donors for these purposes. As Radelet (2006) concludes, "Beyond specific cases studies, there is little systematic evidence on the relationship between aid and health, education income distribution or other outcomes." Studies that select on the dependent variable to identify successful efforts can contribute little to the larger question at hand.

3. DATA AND METHODS

(a) *Data sources and definitions*

For this analysis I use data from AidData,[9] which has been developed by researchers at the College of William and Mary and at Brigham Young University (Tierney *et al.*, 2011). AidData combines the widely used data from bilateral donors put out by the OECD's Creditor Reporting Service (CRS) with a large number of non-OECD bilateral donors and a variety of multilateral financial institutions including regional development banks (most of which are not in the CRS) as well as the World Bank. In also include health-relate funding from the Bill and Melinda Gates Foundation (BMGF) and from the Global Alliance for Aids and Vaccinations (GAVI). Much research on development in the past uses data consisting solely of Official Development Assistance (ODA), but AidData includes projects, including loans and grants that include both ODA and non-ODA. Furthermore, in this analysis, no distinction is made between loans and grants. Because most loans are long-term with low interest rates, I assume that loans will have, at least in the medium-term, roughly equivalent effects to grants.

All data taken from AidData are in the form of commitments, rather than disbursements. While disbursements are theoretically appealing, the data on disbursed amounts is in many cases spotty or missing entirely in AidData (because it is missing from the sources AidData draws on). Investigations of the literature on this topic (including the working paper by Bulir and Hamann (2001)) show no compelling evidence that disbursed amounts differ markedly from commitments, but I realize this is a complex question. Furthermore, there is variation across projects and donors in how long money takes to "hit the street." For this analysis we assume a 1 year offset between the time a commitment is made and the money is spent, though the findings are robust to different specifications of the time lag between commitment and disbursement.

As noted above, data used here include country level aggregates from two important donors in the health sector: GAVI and the BMGF, which have been very active over the past decade. Data from GAVI and BMGF was obtained from the Institute for Health Metrics and Evaluation[10] (IHME), not from AidData. Currently these funds are not given purpose codes so they cannot be placed into sub-sector categories. Also, the IHME data are in the form of estimated disbursements in the given year, requiring the amounts from those sources to be back-dated by 1 year to match with the methodology discussed above.

Data from multilateral institutions and non-DAC bilaterals have been assigned purpose codes with a new coding scheme that builds on the system of purpose codes developed by the CRS, but one which allows more granularity in the categorization and eliminates the frequent problem in the CRS data whereby projects are lumped into catch-all codes if they have multiple activities within the same sector. By using project descriptions, AidData researchers assigned purpose codes to each project. They have also recoded over 118,000 projects from the health and population sectors in the CRS and assigned them AidData codes. Thus all the projects used here were coded by the same team using the same set of consistent criteria (as opposed to the CRS, where codes are assigned by individual donors scattered across the world using changing criteria).[11]

DAH in this analysis includes the health population sectors. The population sector is included because it captures HIV/AIDS projects and family planning, including reproductive health. DAH is broken up into nine categories that are used to a lesser extent in the analysis:

1. Administration (policy and management).
2. Training (medical training and health education).
3. Infrastructure (medical services, hospitals, specialty clinics).[12]
4. Basic health care (nursing care, drugs, child health).[13]
5. Nutrition.
6. HIV/AIDS and STDs.
7. Other infectious diseases.
8. Family Planning (including reproductive health).
9. Nonspecific (cannot be assigned one of the above categories).

In this analysis we construe health aid to be of two types. In addition to DAH, total aid to the water sector is broken out as a separate category because clean water is such a crucial determinant of health. Of course not all water projects are focused on providing drinking water, so the force of this variable will be somewhat diluted. All other development finance given to the recipient country is included as an additional control.

Other data in this analysis come from the World Bank Development Indicators (WDI) database, including IMR, CMR, e_0, GDP *per capita* (in USD 2000 constant dollars) and population. Data on IMR and CMR are obtained from the World Bank's World Development Indicators (WDI). The democracy variable is the widely used Polity2 score.

The variables in the analysis are calculated at 5 year intervals. The population and mortality measures are for the years indicated, 1975, 1980, etc. Real GDP and DAH are calculated as the sum of total dollars across the 5 year period (1975–79, 1980–84, etc.) in *per capita* terms. Except for the Polity2 democracy score, all dependent and independent variables are estimated in logs, which means that the regression coefficients will represent elasticities. An elasticity is the percentage change in the dependent variable given a percentage change in the independent variable, which will be discussed more later.

The data used in this analysis is relatively complete. In a small percentage of countries, data is missing for a few early years. In these cases, the missing values for independent variables (but not the dependent variable) are constructed based on a linear projection of the other later data points for the country. When data is missing for all years (14 of 98 countries), the country is dropped from the analysis, though sensitivity analysis using multiple imputation methods for these missing countries is employed to check for robustness. The DAH variable is present for all cases.

(b) *Models and estimation methods*

Studies employing panel data can employ a variety of alternative models and estimation methods. Unfortunately, these

alternative approaches often yield drastically different results in panel data studies done at the country level (Wilson & Butler, 2007). Ideally, the choice of a model should be driven by theory and by the features of the data. In this case, a dominant feature of the data is that countries follow widely different mortality trajectories over time. Mortality levels in 1975 are very different from one another, as are the slopes of the time trends (some improve much faster than others) from 1975 to 2005.

The population mortality rate M_{it} follows a trajectory over time, where countries are indexed by i and time periods (in 5 year increments) by t. Other variables shift this trajectory up and down; these include the primary variable of interest, DAH_{it} and control variables, which include other forms of aid, GDP, population, and Polity2 (these are all components of the vector $\mathbf{X}_{it}$). The models employed here are special cases of the following linear trajectory:

$$M_{it} = \beta_0 + \beta_{1i}t + \beta_2 \mathrm{DAH}_{it} + X_{it}\beta_3 + \beta_4 M_{it-1} + u_i + e_{it}$$

Many factors can shift the mortality trajectory, as shown above. In addition to the observables in $\mathbf{X}_{it}$, shifters include the lagged value of the dependent variable, M_{it-1}, and an unobserved country effect, u_i, which varies across countries but is constant over time. This term captures the host of unobserved economic, political and cultural determinants of mortality and significantly reduces problems with omitted variable bias. The residual, e_{it}, is an *i.i.d.* error process. Finally, in most models the slope of the trajectory, β_{1i}, is assumed to be common across countries ($\beta_{1i} = \beta_1$). As specified above, however, it is possible to allow the slope to vary across countries (similar to the way that the u_i term allows for variation in intercepts).

A variety of simple and more complex models are estimated as part of this analysis. More detail on these methods is included in the on-line Appendix A. Two models deserve special attention here:

The Dynamic panel model ($\beta_{1i} = \beta_1$). The model (hereafter, DPM) captures the full dynamic specification above except that the slope is assumed constant. Thus, this model is a generalized version of fixed and random effects models, as well as the frequently employed lagged dependent variable model (LDV) suggested by Beck and Katz (1995) and used widely in the literature. I estimate this model in two ways. The first approach is to use the simple fixed effects model (sometimes called the least squares dummy variable model), with the addition of cluster-robust standard errors, to adjust for the panel structure of the error terms. A problem with this approach is that least squares estimates of the DPM are biased and inconsistent (Nickell, 1981), though Monte Carlo work by Judson and Owen (1999) suggest the main source of bias in cross-country studies concerns the parameter β_4, which is of little interest here.

The second approach is to use the generalized method of moments (GMM) technique for panel data developed by Arellano and Bond (1991). The Arellano-Bond approach takes first differences of the data and then uses lagged values of the variables as instrumental variables. This is a highly effective way to account for the obvious endogeneity of many of the variables in the model, including DAH. However, the Arellano-Bond approach was designed to estimate models with a few time periods and thousands of panels, and thus it is not ideal for country-level panel regressions. This is seen in the very large standard errors that accompany the GMM estimates in this study. Thus the reduction in potential bias of the estimates comes at a cost of considerable loss in precision. Furthermore, the small-sample properties of these estimators are not well known. The Arellano-Bond estimator also treats slope coefficients as fixed, as do the previous models.

The Latent Growth Model ($\beta_4 = 0$) The LGM is part of a class of models that go by many names including random coefficient models, mixed models, hierarchical linear models and multi-level models. Its strength is the highly flexible functional form that allows both random intercepts and random slopes. Furthermore, in the estimation results to follow, arbitrary correlation between the random effects in the model is allowed (and proves to be important). Latent growth models, however, also rely on large sample properties of maximum likelihood estimation for their justification and are not necessary unbiased in small samples. No lagged dependent variable is used in the LGM approach. The LGM is subject to endogeneity bias, but this bias is expected to be much smaller than least squares estimation techniques. Endogenous variables are those that are correlated with the error process. Because the LGM has a very flexible error structure, some of the correlation between observable variables and the explanatory variables is parceled out among random coefficients in the LGM. GMM estimators exploit the correlation between the instruments and the explanatory variables, but this entails losing a lot of information,[14] resulting in high standard errors. The LGM models have no such loss of information and, therefore, have much lower standard errors.

Most of our analytical attention will be focused on the dynamic panel model and the latent growth model, since they are far more robust than simpler models. But since it is useful to see the consequences of moving from simple models to more general ones, the simpler models will be estimated for comparison purposes (some of these results are shown, others are available in the Appendix A). Our main interest is the coefficient β_2, which will be negative if DAH is effective (or positive when using life expectancy as the dependent variable). As noted above, all variables (except Polity2 and time) will be estimated in logs, allowing us to interpret those estimated coefficients as elasticities.

The models above were estimated using Stata 11 software. Each model was estimated for each of the three mortality indicators discussed above. In each case, cluster-robust standard errors[15] were reported along with the coefficient estimates. 84 countries were used in the analysis reported below. Estimates were run only for "high mortality" countries. A high-mortality country was defined as a country that, at any time since 1975, the IMR exceeded 50 deaths per 1,000 live births. 14 countries were not used because of missing data, though imputation techniques for the missing data were employed in the robustness checks discussed below.[16]

A note comparing these methods with those of the recent work by Williamson (2008) is in order, since the general strategy and data share some commonality to the approach here. For reasons not discussed in her paper, Williamson does not use the common Arrelano-Bond estimation strategy (nor other GMM approaches typically used in panel studies). Her GMM approach is a somewhat *ad hoc* (though reasonable) approach of using the lagged aid as an instrument.[17] This approach is similar in spirit but may actually lead to very different estimates. Her "baseline" estimates (to which she compares her GMM estimates) are simple fixed effects models, meaning she lets the intercepts for each country but imposes a common slope across countries. The LGM model discussed above is much more robust because it allows a unique slope for each country in the analysis. I also estimate, for comparison purposes, a variety of different models discussed above and in the Appendix A. Perhaps most important, though, is that

Williamson includes *all* countries in her analysis, including developed countries, thereby assuming not only that the mortality trajectory is constant across all countries in the analysis but that the trajectory is the same for both rich and poor countries alike. I include only poor, high-mortality countries. Her approach yields a higher N but at the cost of imposing a common framework on very different types of countries. Yet, even though there are sharp differences in data and methods across the two studies, both her analysis and mine end up with the same noneffect for DAH.

(c) *Control variables*

The models estimated here are parsimonious but do include control variables for other potential determinants of aid effectiveness. The first is other development assistance. DAH makes up just one part of the aid equation, and other aid, such as education, may have an impact on population health. Furthermore, since clean water is such a crucial determinant of health in the developing world, I separate water aid out as its own category. It is not included as part of health aid because it is often difficult to determine whether aid for water projects is actually focused on making drinking water cleaner (it may just be for agriculture or industry).[18] Thus there are three aid variables: DAH, water aid, and all other development assistance.

Two additional controls play key roles in the aid effectiveness literature. The first is growth in national income. Most aid effectiveness studies look at GDP growth as the dependent variable, but when other outcomes are being studied, it is important to include income as a control, especially when examining changes over time.[19] The strong correlation between income and mortality has been recognized since the pioneering work of Preston (1975).

In this analysis, I use the log of real GDP *per capita* to measure national income.

A second important factor is the political environment. Very little empirical work has been done on the relationship between democracy and health.[20] Acemoglu and Robinson (2005) see democracy as, essentially, a dictatorship of the poor and middle class, who will make public health spending a priority. Besley and Kudamatsu (2006) argue that democratic governments are accountable to a broad group of people who will remove elected officials from office should they fail to address public health issues and that democracies have stronger mechanisms for selecting competent and honest leaders to implement public health policies.

4. RESULTS

(a) *Model estimates*

Three important points can be made about estimated effects for DAH in this analysis. First, and most important, they are *very* small. Table 1 contains the DAH coefficients for each model for each of the 3 mortality measures. All of the coefficients are near zero, especially those coefficients with a negative sign—which is required for DAH to be effective. An elasticity of −0.007 (Model 7b) for instance, means that an increase in 1% in DAH spending is associated with a decrease of approximately 0.007% in IMR. For a concrete example, consider a country that has an IMR = 75 and receives $15 *per capita* over a 5 year period. In order to reduce IMR by 1 unit to 74, the *per capita* DAH would have to increase to $102—almost a 7-fold increase! In short, even dramatic increases in aid would cause barely perceptible decreases in mortality.

Second, the estimated effects are reasonably precise. 95% confidence intervals on the DAH coefficient often contain zero, but the bounds of the interval are not large. The interval for Model 7b, for instance, is −0.0037 to 0.0114. Other inter-

Table 1. *The effect of targeted health aid on mortality*

Model specification	Mortality measure		
	IMR *Elasticity*	CMR *Elasticity*	Life expect. (e_0) *Elasticity*
(1) Bi-variate OLS	−0.0023	−0.0066	−0.0006
	(0.0245)	(0.0309)	(0.0068)
(2) Multi-variate OLS	0.0392**	0.0417**	−0.0052**
	(0.0166)	(0.0201)	(0.0055)
(3) First difference model	0.0066**	0.0072**	−0.0004
	(0.0031)	(0.0035)	(0.0010)
(4) Fixed effects model	0.0290***	0.0319***	−0.0047*
	(0.0087)	(0.0096)	(0.0025)
(5) Random effects model	0.0339***	0.0376***	−0.0048**
	(0.0085)	(0.0095)	(0.0024)
(6) Lagged dependent variable model	0.0053	0.0023	0.0017
	(0.0099)	(0.1225)	(0.0017)
(7a) Dynamic panel model (OLS)	0.0147*	0.0164	0.0008
	(0.0078)	(0.0086)	(0.0020)
(7b) Dynamic panel model (GMM)	−0.0074	−0.0064	0.0039
	(0.0087)	(0.0087)	(0.0059)
(8) Latent growth curve model	0.0038	0.0037	−0.0003
	(0.0039)	(0.0044)	(0.0021)

Notes: Cluster-robust standard errors are in parentheses.
Complete regressions results are found in the on-line Appendix A.
Given the log–log format of these regressions, all coefficient estimates above represent *elasticities* (the percentage change in the dependent variable over the percentage change in the independent variable).
*** *P*-value <.001.
** *P*-value <.05.
* *P*-value <.1.

vals are similarly constrained. The estimates are largely statistically insignificant, but that has less to do with the standard errors than with the fact that the estimates themselves are near zero. The relatively modest standard errors (which would be even smaller without cluster-robust estimation) are the reason several of the estimates are statistically significant at the 5% level. But all the estimates that are significant are in the "wrong" direction! [21]

Third, as suspected, the overall model results are quite sensitive to specification and estimation method. This is disconcerting but not unusual in time-series cross-sectional work when doing country-level analysis (though researchers often fail to conduct or report sensitivity analysis on model specification). This finding is consistent with the work of Roodman (2007), who found that most of the recent empirical studies on development assistance and economic growth in the past decade are highly sensitive to assumptions about model specification. Wilson and Butler (2007), in a sensitivity analysis of cross-country studies in political science find the conclusions are, in general, highly sensitive to model specification.

Table 2 has all the estimated coefficients for a few of the models. For comparison purposes, I include the static OLS model (2), an OLS model with a lagged dependent variable (6) and the DPM model estimated with simple fixed effects (7a) in addition to the Arrellano-Bond (7b) and LGM (8) models. As anticipated, the standard errors of the Arrellano-Bond model are significantly greater than standard errors in the latent growth model.

What both these models show, in addition to the ineffectiveness of aid dollars (of any type), is an important effect of economic growth. The Arrellano-Bond estimate is twice as high as the LGM estimate, but the confidence intervals for the two variables overlap significantly (though neither contains zero). The LGM model assigns much more weight to unobserved factors correlated with time, rather than GDP growth, but in both specifications the effect of GDP is nontrivial and statistically significant, though we note that there is a lot of variation in mortality (among high mortality countries) that cannot be explained by economic growth.

A lot of attention has been paid in the literature to whether the recipient countries are democratic. I find no affect for democracy, nor do I find an interaction effect between democracy and the level of DAH (not shown in Table 2) in any of the estimated models. This is in contrast to the findings of Besley and Kudamatsu (2006), who find significant effects of democracy. This may be due to the fact that their study includes all countries, including many rich democratic ones. [22] Finally, larger countries show a consistent pattern of lower mortality than smaller ones, an effect which is statistically significant in the LGM specification, but which is quite small in all models.

(b) Sensitivity analysis

The models and specifications above make numerous assumptions which can be challenged and examined. The following assumptions were checked for robustness:

- *Definition of high mortality country*: The baseline assumption was IMR = 50. All analyses were redone using cutoffs of 25 (giving more countries) and 75 (giving fewer).
- *Time delay between commitments and disbursements*: The base assumption was 1 year. A delay of 2 years was also tested.
- *Donor type* Rather than including all DAH funding together, donors are divided into two types: multilateral and bi-lateral.
- *Time it takes DAH to be effective*: The base assumption was to look at the effect of 5 years of aid on IMR; for example, mortality in 1985 was assumed to depend on cumulative DAH from 1980 to 1984. Because it might take longer for aid to have an impact, the models were re-estimated adding in the lag of spending (meaning DAH from 1975 to 1979 was also included in the regression in the example above).

Table 2. *Selected Infant Mortality Regressions*

Dependent Var.: $\ln(\text{IMR}_t)$	Model				
Independent variables:	(2) OLS	(6) LDV(OLS)	(7a) DPM (FE)	(7b) DPM(AB)	(8) LGM
$\ln(\text{IMR}_{t-1})$		0.6943***	0.3155***	0.7865***	
		0.10942	(0.0965)	(0.0868)	
$\ln$(DAH per capita)	0.03919**	0.00525	0.0147*	−0.0074	0.0038
	(0.0166)	(0.0099)	(0.0078)	(0.0087)	(0.0039)
$\ln$(Water aid per capita)	−0.0179	−0.0046	−0.0099*	0.0054	0.0014
	(0.0160)	(0.0078)	(0.0058)	(0.0140)	(0.0030)
$\ln$(All other aid per capita)	−0.01399	−0.02459	−0.0289***	0.0158	−0.0044
	(0.0168)	(0.0154)	(0.0097)	(0.0156)	(0.0038)
$\ln$(GDP per capita)	−0.37041***	−0.15586***	−0.0436	−0.3163**	−0.1431***
	(0.0338)	(0.0382)	(0.0427)	(0.1293)	(0.0229)
$\ln$(Population)	−0.05015**	−0.03197***	0.3836*	−0.0506	−0.0481**
	(0.0232)	(0.0117)	(0.2080)	(0.1173)	(0.0221)
Polity2	−0.00878*	−0.0444*	0.0008	−0.0024	0.0007
	(0.0050)	(0.0023)	(0.0022)	(0.0055)	(0.0012)
t	−0.09653***	−0.01948	−0.1284***	−0.0506	−0.1122***
	(0.0128)	(0.0135)	(0.0329)	(0.1173)	(0.0101)
Constant	8.37149***	3.1652***	−2.3735	4.1115**	6.5580***
	(0.4319)	(0.9272)	(3.1457)	(2.0354)	(0.3946)
N	476	470	470	309	476

Notes: Cluster-robust standard errors are in parentheses
*** *P*-value <.001.
** *P*-value <.05.
* *P*-value <.1.

• *Use of multiple imputation.* In the baseline models, countries without sufficient data were dropped. The models were rerun using multiple imputation methods, for the missing cases.[23] This increased the number of high mortality countries from 84 to 96.

• *Contingent effectiveness*: The effectiveness of DAH might depend on the level of democratization in the recipient country. To check this, DAH levels were interacted with the Polity2 score. This type of interaction is suggested by the important (and controversial) findings of Burnside and Dollar (2000) and by Kosack (2003)

• *Nonlinear effectiveness*: Instead of using the baseline log–log specification, the effect of DAH was estimated without the log transformations, and a quadratic term was added to check for nonlinearities.

• *Definition of DAH*: DAH in this baseline models includes the health and population sectors. DAH was examined separating out family planning and HIV/AIDS funding as separate categories (with other DAH aid as a third category).

• *Changing effectiveness over time*: As shown in Figure 1 and elsewhere, DAH dollars are rising rapidly. Furthermore, this new DAH may be more effective than aid in the past. To test this hypothesis, I estimate a lagged dependent variable model for each period,[24] thus allowing the effectiveness of DAH to vary period by period.

• *Sub-sector differences*: DAH may be effective for some sub-sectors, but not for others. It may be that particular types of aid are more effective than others. Since 1990, there has been a dramatic shift in funding towards infectious diseases, including HIV/AIDS. To test effectiveness by sub-sector, I subdivide the DAH variable into nine categories given earlier.

Complete regression results for all models using these alternative assumptions are available in the on-line Appendix A accompanying this article. However, they all tell the same story about DAH. None of the alternative assumptions results in increasing the magnitude of the DAH effect. I highlight below results from the last two mentioned sensitivity checks.

First, Figure 2 shows the results of estimating period-by-period models to test the possibility that DAH may have become more effective over time. Though almost none of the estimates are statistically significant, they do *not* support a pattern of increasing effectiveness over time. If anything, the opposite story is true. In the 1980s, DAH spending had the hoped-for negative signs (though, again, too small to matter much). Since then, however, the estimated coefficients show that DAH increased mortality, though in the most recent period the effects have fallen to zero.

Second, in Table 3 I report latent growth model estimates for IMR using the DAH spending in the 9 categories since it may be that particular types of aid are more effective than oth-
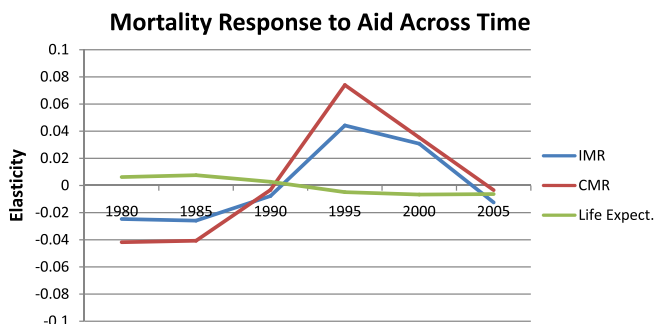
Table 3. *IMR and Specific DAH Categories*

(Latent growth model estimates) Dependent Var.: ln(IMR$_t$) Independent variables:	Coeff.	Std. Error
ln(DAH:Administration)	0.0025	(0.0024)
ln(DAH:Training)	−0.0022	(0.0023)
ln(DAH:Infastructure)	−0.0004	(0.0025)
ln(DAH:Basic health care)	−0.0021	(0.0276)
ln(DAH:Nutrition)	−0.0014	(0.0020)
ln(DAH:HIV/AIDS and STDs)	−0.0048**	(0.0022)
ln(DAH: Other infectious disease)	−0.0069***	(0.0023)
ln(DAH: Family planning)	−0.0095***	(0.0027)
ln(DAH: Non-specific)	−0.0026	(0.0024)
ln(Water aid)	0.0022	(0.0029)
ln(Other aid)	−0.0065	(0.0036)
ln(GDP per capita)	−0.1243***	(0.0224)
ln(Population)	−0.0548**	(0.0223)
Polity2	0.0010	(0.0012)
t	−0.1139***	(0.0104)
Constant	6.5120***	(0.3944)

Notes: Cluster-robust standard errors are in parentheses.
* *P*-value <.1.
** *P*-value <.05.
*** *P*-value <.001.

ers. Since 1990, there has been a dramatic shift in funding towards infectious diseases, including HIV/AIDS. All the DAH categories have small effects that are comparable in magnitude to the DAH spending overall. Again, an estimate this small means the model predicts that a massive increase in aid would have a barely perceptible impact on mortality. However, some of the estimates are statistically significant, even strongly so. In particular, spending on AIDS and other infectious diseases, in addition to family planning, does have a statistically significant effect on mortality, though it is *very* small.

5. ENDOGENEITY AND THE ALLOCATION STORY

Reasonable aid donors want to put their money where it is most needed, where costs are low, and where it is most likely to do the most good. Not following these rules may lead to ineffectiveness. This is another way of saying DAH is likely to be endogenous—correlated with unobserved factors affecting mortality. Countries receive aid or do not receive aid for *reasons* (or at least we hope that is the case). If unobserved factors that influence mortality also affect the amount of aid received, then there is the possibility of endogeneity bias. The Arellano-Bond estimates shown above constitute one approach to addressing endogeneity, but the imprecision of GMM in small samples reduces their usefulness considerably (though they did show the same noneffects as the other models).

Could endogeneity of DAH be masking its effectiveness in the results above? For this to occur, DAH would need to be *positively* correlated with the unobserved factors that *increase* mortality. In other words, endogeneity of DAH might hide its effectiveness if aid flows to countries that are *less* likely to achieve success (after controlling, of course, for level of mortality, GDP and the other observable variables in the model). But this is the opposite of what we would expect. If donors are motivated at all by a desire to actually improve health, aid dollars should flow to where they will be more effective, not less

Figure 2. *Mortality response to aid across time.*

Table 4. *Determinants of Aid Allocation*

Dependent Var.: $ln(DAH_{t+1})$					
Independent variables:	**OLS**	LDV(OLS)	DPM (FE)	DPM(AB)	*LGM*
LDV		0.28069[***]	−0.0673	−0.0639	
		(0.0575)	(0.0417)	(0.1759)	
$ln(IMR_t)$: Need	0.6479[***]	0.4957[***]	0.75276[**]	1.4325[*]	0.655[***]
	(0.1799)	(0.1485)	(0.3310)	(0.8140)	(0.1920)
$ln(IMR_t$-$IMR_{t-1})$: Difficulty	−0.50182[***]	−0.2768[**]	−0.26422	−1.7058	−0.400[**]
	(0.1286)	(0.1201)	(0.2211)	(1.3526)	(0.2012)
$ln(GDP_t)$	0.048035	0.04227[**]	−0.02001	−0.49254	0.024
	(0.0870)	(0.0652)	(0.3156)	(1.1126)	(0.0991)
$ln(Population_t)$	−0.4003[***]	−0.29615[***]	−0.25855	−2.7579[*]	−0.400[***]
	(0.0390)	(0.0333)	(1.0596)	(2.5702)	(0.0454)
$Polity2_t$	0.03382[***]	0.02773[**]	−0.00549	−0.02419	0.026[***]
	(0.0105)	(0.0083)	(0.0100)	(0.0333)	(0.0094)
t	0.28239[***]	0.20112[**]	0.40012[**]	0.80189[**]	0.312[***]
	(0.0385)	(0.0083)	(0.1487)	(0.3453)	(0.0382)
Constant	4.5296[***]	3.3399[**]	2.0357	41.542	4.557[**]
	(1.2096)		(16.9590)	(45.9370)	(1.8524)

Notes: Cluster-robust standard errors are in parentheses.
[***] *P*-value <.001.
[**] *P*-value <.05.
[*] *P*-value <.1.

effective. For example, a reduction in armed conflict should reduce mortality (as people are better able to access health services) and increase aid dollars (as aid workers can more easily deliver services).

We can explore the nature of the allocation process by applying the same models to the allocation process that were used to estimate DAH effectiveness, with DAH in the future period as the dependent variable, and mortality in the current period as the primary independent variable. This will not only tell us more about the possible effects of endogeneity but also provide insight into the factors affecting the allocation of DAH dollars (or at least the commitments to spend).

So where does DAH go? Table 4 shows several estimates of DAH allocation (added over a 5 year period, as before) as a function of infant mortality and other variables. Two variables are of particular interest. First, *need*. Is money being committed to where mortality is the highest. Second, *difficulty*. For difficulty I use the rate of decline in mortality over the preceding 5 year period $ln(IMR_t) - ln(IMR_{t-1})$. The bigger the decline, the less difficult are the challenges and the higher will be the prospects for success, *ceteris paribus*.

Table 4 shows estimates for the same group of models illustrated in Table 2. These models illustrate the sensitivity of results to the method used. They also illustrate the large cost of employing GMM estimates in terms of precision of estimation. But even though there is a large variation in effects across models, the following qualitative results are robust to each alternative shown. First, DAH is going where mortality is highest, indicating that money follows need, at least with respect to mortality. Second, DAH goes where mortality has declined in the recent past. Third, the larger the country, the less it receives in DAH funding (which, to remind the reader, is measured in *per capita* terms). This would be rational if there are significant fixed costs of providing aid or if the donors expected economies of scale to exist with respect to effectiveness (though these scale economies do not seem to exist, given the earlier estimates, but donors perhaps think that they do). [25]

But the effect of two other key variables are less clear and, consequently, puzzling. Most important, I do not find strong evidence that poor countries receive more DAH. The Arellan-

o-Bond estimates suggest that poverty might increase assistance, but the standard error on that estimate is huge and the finding is not robust to other specifications. If need is measured by GDP, then we cannot say with any confidence that aid follows need. Next, the estimated effects of democracy on the allocation process are small and not robust.

I take two important conclusions away from this analysis of the allocation of DAH across countries. First, if recent declines in mortality are used by donors as proxies for the prospects of future success, then these results suggest that donors, as a whole, give more to countries where the prospects for success are greater. This finding undermines the possibility that the endogeneity of DAH spending masks its effectiveness, thereby strengthening the findings above that DAH does not reduce mortality.

Second, in the past four decades we have seen sharp reductions in mortality around the developing world. The results in Tables 2 and 4, taken together, suggest that donors move aid assistance to where declines in mortality have already occurred but, once there, the money has no added benefit in terms of mortality. The strikingly large magnitude of these effects suggests that donor dollars are highly responsive to mortality, but, sadly, these dollars are largely chasing after success, not causing it.

6. DISCUSSION: WHY DO WE NOT SEE EFFECTIVE AID?

The unwelcome conclusion of the preceding analysis is that DAH—when measured in aggregate at the country level—has had no effect on mortality rates in developing countries. The estimates are quite precisely measured, but they are essentially equal to zero. How could this be in light of the well-known cases studies highlighted earlier?

Long and cantankerous debates surround the issue of aid effectiveness in the area of economic development. [26] Critics claim that there are significant leakages in the system, for example, funds allocated by donors that never actually make it to the recipient country; that recipient governments are rid-

dled by corrupt officials that gobble up aid dollars for their personal interests; that aid dollars are fungible and, therefore, will crowd out other government expenditures; that aid undermines the development of sustainable markets and makes countries dependent on aid; and that many aid projects are poorly designed, poorly implemented and poorly evaluated.

Since I do not have data that would speak to these possible explanations, I want to avoid this thicket as much as possible. I note here only that these arguments can apply just as well to DAH as to other types of aid. Whether or not they do is an open question, since there is little sector-specific work on the factors influencing aid effectiveness. An exception to this is the recent important study by Lu *et al.* (2010) showing that DAH *is* highly fungible. According to the authors, every $1 of DAH funding to a recipient government reduces government health expenditures by $0.43–$1.14. [27]

The extensive sensitivity checks conducted previously rule out many possible explanations for why effectiveness does not show up in the estimated models. Perhaps, however, DAH does not show up as effective in empirical models because the level of DAH spending is simply too small to generate any perceptible effects in aggregate measures. This is the explanation aid advocates would most like to believe. According to the most comprehensive set of estimates available (Ravishankar *et al.*, 2009), global spending on development assistance for health has risen from 5.6 billion in 1990 to 21.8 billion in 2007 in real 2007 dollars. Although this is a fourfold increase, currently this amounts to only a few dollars of health aid per person in high mortality countries, including monies that are not actually targeted at the country level. In 1990, funding was about $2 *per capita*, per annum, and in 2007 it was about $8 *per capita*, per annum. [28] Comparing these amounts to the thousands *per capita* spent in the developed world, it is conceivable that they are just too low to matter.

Experimentation with regression specifications that might show DAH effectiveness at higher spending levels was not productive. Figure 3 shows why. It contains a scatter plot of cumulative DAH given to high mortality countries, in constant 2000 dollars, against the change in the infant mortality rate between 1975 and 2005. Even among countries with the highest levels of aid (say among those over $100), there is no obvious relationship with the reduction in infant mortality. About the most optimistic thing feature of this graph is that states that have had resistant or increasing mortality are not those cases that have received a lot of DAH funds. All coun-

tries (with the exception of Zambia) above the 75th percentile (which occurs at about $75 *per capita* in cumulative aid over the period) experienced significant declines in mortality. Of course, the countries experiencing the largest decline in mortality were those that received very little DAH relative to other countries. [29]

Furthermore, many public health interventions are quite inexpensive. For instance, the *per capita* cost of the USAID program to reduce diarrheal deaths in Egypt during the 1980s (Levine, 2004) was only $6. [30] Given the relatively low price of many effective interventions, we should be seeing more of a dent in mortality among countries receiving the most DAH—that is, if the allocated dollars are actually going to effective programs. No one can say how flooding high mortality countries with aid dollars devoted to the health sector at levels much higher than we have seen thus far would affect mortality or other health measures. Perhaps such an effort would generate more of the public health interventions known to save lives. But, unfortunately, nothing in the data suggests this would be the case.

Another reason that estimates of DAH spending might not be accurate is the existence of notable holes in the data, especially as we go further back in time. It seems unlikely, however, that these limitations would account for the estimated ineffectiveness. Imagine what kind of story we would have to tell about patterns of missing data to explain the negative findings above. Essentially, this requires that missing donors from the data have health aid "figured out," while the World Bank, USAID, GFATM, AFDB, GAVI, and BMGF do not.

But there is no obvious reason to think that the missing donors (and missing years of data from some donors) are more effective at health spending than those in the dataset. Indeed, we would expect the opposite: transparent donors are *more* likely to be effective, since their activities are open to public scrutiny, an idea at the heart of the aid transparency movement. Thus missing data from less transparent donors is likely to bias the results towards *more* effectiveness, not less. Additionally, the experience of adding new donors has not changed the basic findings. AidData contains significantly more data (in terms of dollars) than does the OECD-CRS data used in previous studies (such as Williamson (2008)). Adding this new data did not undermine the previous result of ineffectiveness, and, in robustness checks of the results reported here, multilateral DAH is no more effective than bilateral DAH. Moreover, there would have to be extremely large and highly effective donors hiding their data to overturn the ineffectiveness result. This seems highly unlikely, to say the least.

Finally, the lessons from the aid allocation analysis in Section 5 suggest that DAH ineffectiveness may be crucially linked to the allocation process used by aid donors. In addition to political and strategic factors that are often used to explain aid allocation, DAH funding tends to go to countries that have experienced significant improvements, even though those improvements cannot be linked to DAH in the past. In other words, development assistance for health chases after development *that has already occurred*, rather than being put to use where it is likely to be effective.

7. IMPLICATIONS

The central point of this empirical exercise is that although public health measures can be effective (and have been in selected cases), DAH spending from 1975 to 2005 had no discernible effect on country-level mortality rates in high mortality countries. This sobering conclusion is not due to
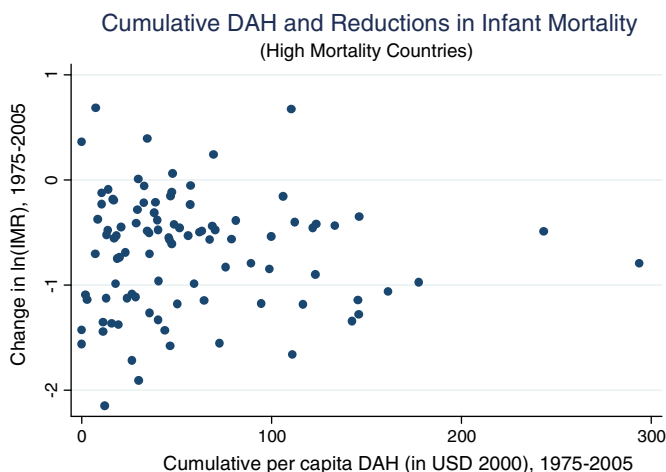

Cumulative DAH and Reductions in Infant Mortality
(High Mortality Countries)

Figure 3. *Cumulative DAH and reductions in infant mortality.*

limited data that allows only imprecise estimates, but comes from relatively precise estimates that are, unfortunately, very close to zero. This finding was subjected to a multitude of alternative specifications and robustness checks, all to the same end: zero. In sum, DAH dollars move strongly towards countries with declining mortality, but they do not generate it.

I have been careful to repeatedly indicate that these results do not mean that public health measures to reduce mortality do not work. Indeed, my assumption at the outset was that they *do* work—which is what makes health such an interesting sector to study. But a rehydration kit administered by a mother to her child in Africa is multiple degrees of separation away from the dollars allocated for that kit by an aid agency in Washington. Not only can significant leakages occur after money is committed, but it is plausible that DAH funds crowd out other public health measures or even work in some cases to undermine progress.

Just because the mortality response to DAH is very small does not, in itself, mean it is not worth the price. Even small marginal returns may be justified when the problems are so deep and the resources of the developed world so vast. This research is certainly not a call to inaction or a reduction in aid.

The findings here should, however, be deeply troubling, especially to those seeking large increases in DAH over a short period of time. What is the source of optimism that a big push will be effective? The most optimistic result from the above analysis is that the thrust to fight HIV/AIDS and other infectious disease may have born some fruit, since the sub-sector analysis shows statistically significant (but very small) effects in the right direction in these categories (and in the area of family planning).

Better information, such as the project-level data in the AidData database and the small but growing volume of careful effectiveness evaluations are of vital importance in increasing effective development assistance, as are more nuanced statistical analyses of aid effectiveness. Some aid advocates at the meetings in Oxford (at which the papers in this volume were presented) were heard to say, "We don't need more regressions." In fact, that seems to be exactly what we *do* need—not only more, but better ones, with better data. The financing of public health measures come out of a larger economic and political system that is responsible for financing development. Thus far, that system has not demonstrated its success in the area of health or, indeed, in many other areas.

NOTES

1. See also Levine (2007).

2. Moreover, Williamson uses both developed and underdeveloped countries in her analysis, though it seems very unlikely that a model that describes mortality in Bangladesh would be similar to a model used for Denmark. The analysis in this paper, on the other hand, is restricted to high mortality countries.

3. Both Williamson and I test GMM models, and these models are the main focus of Williamson's analysis. However, for the case of country-level panel studies, GMM models need vastly bigger datasets than what is available to achieve a convincing level of precision in estimation.

4. The link between national income and mortality is unassailable, though Croghan, Beatty, and Ron (2006) have argued that wealth and governance matter less than targeted health interventions and foreign aid. Their study, however selects on the dependent variable, in that they look at 4 countries that have experienced high reductions in mortality and then tries to identify factors that explain these declines.

5. Total dollars in DAH are much higher in the IHME 2009 report (Ravishankar et al., 2009) on global financing. IHME focuses on tracking as many dollars in possible. Their totals include, for instance, WHO and UN agency spending that is not available at the country-level or by type of funding. They also include totals by NGOs and foundations, which are not part of AidData.

6. This of course depends on how categories are defined, and there are numerous ways to categorize DAH.

7. Shiffman (2008) argues that HIV/AIDS funding has displaced some traditional donor priorities, however, even though other diseases have increased their share. He also argues (2006) that allocation of funds across diseases is not closely tied to disease burden.

8. I estimate a considerable number of additional models beyond those reported by Williamson and use a significantly expanded dataset. I also conduct numerous sensitivity checks, discussed in Section 4 to further support the findings here. Furthermore, estimation of aid allocation models are used here to assess the possible impact of endogeneity as well as provide a new account of why DAH is not effective.

9. This data collection project is ongoing, with new data added when available. I use here the public version available on June 22, 2010. Data are available from Aiddata.org. The actual data files used in the statistical analysis are available from the author.

10. See Ravishankar et al. (2009).

11. The AidData team has developed a detailed coding scheme that includes a dominant purpose code, as well as a set of codes describing the activities associated with the project, to the extent that the donor-provided descriptions allow that coding to be undertaken (Tierney et al., 2011). In this analysis, we do not use the activity codes but focus on the purpose codes. The sectors associate with the AidData purpose code correspond closely with CRS purpose codes.

12. Includes both general health infrastructure and basic health infrastructure. The CRS distinction between "general" and "basic" health is not crystal clear and can be quite arbitrary. General health tends to focus on medical services, especially specialized care, whereas basic health care is typically synonymous with delivery of public health programs, such as infectious disease control.

13. Excluding infrastructure, nutrition, infectious disease, and health education, which are categorized separately.

14. In particular, the portion of X that is correlated with the dependent variable but not correlated with the instrument or the error term is explanatory power lost when using GMM estimation. In very large samples this is not of consequence, but losing information in small samples can be costly in terms of obtaining precise estimates.

15. Standard error adjustment is important when using cross-sectional time-series data because of the dependent nature of the data. Error terms are not independent because they are clustered at the country level. Adjusting for clustering is necessary to not overstate the standard error.

16. Ninety six countries were used in the simple bi-variate models.

17. Williamson also misstates (p. 192) the cause of the endogeneity. She attributes it to the fact that aid might be correlated with health. Because health is observable and controlled in the model, there is no statistical problem with having explanatory variables that are correlated. The real source of the endogeneity is that DAH may be correlated with variables influencing mortality that are observable to the donors but not observed by the econometrician.

18. I do not include variables such as government health expenditures out of a concern for over controlling and because of the obvious endogeneity of government expenditures. The secondary question of whether DAH given through governments or through other channels influences effectiveness is important, but not explored here, where the focus is DAH spending in aggregate, regardless of the mediating variables involved.

19. Even though income is important, the non-effectiveness of aid shows up whether or not income is included in the model.

20. (Besley & Kudamatsu, 2006; Franco, Alvarez-Dardet, & Ruiz, 2004; Lake & Baum, 2001) are studies that have explored democracy and health empirically. Ruger (2005) discusses how the absence of democracy in China had deleterious effects on health in China during the famine of 1958–61 and in the 2003 SARS outbreak.

21. Meaning that DAH increases mortality.

22. If the intent of this study were to investigate the effect of democracy, lopping off all the rich democratic countries from the sample as I do would be a serious weakness. However, democracy in this analysis is a control variable, not a central variable of interest.

23. Imputations were performed in STATA 11 command **mi impute mvn**, which uses an interative MCMC procedure (20 imputations were used following a "burn-in" set of 100), assuming an underlying multivariate normal model. See Gelman, Carlin, Stern, and Rubin (2004).

24. Results are robust to other specifications. Details are available in the Appendix A.

25. It is worth noting that even though the Arellano-Bond estimates of the allocation model are very imprecise, the estimated magnitudes of that model are, in each of the three cases just mentioned, *much* greater (in absolute value) than the magnitudes of the LGM estimates.

26. These debates are reviewed in many places. See Radelet (2006) for a review.

27. DAH given to non-government recipients, however, raises government expenditures on health, according to the study.

28. These are only back-of-the-envelope calculations meant to show the smallness of per capita DAH allocations, not serious estimates.

29. It is worth reminding the reader that the scale on the axis is in terms of changes in natural logs. Thus a decline of 1.0 would represent, for instance, a decline from an IMR of 75 to 27 or from 150 to 55.

30. The cost per death averted was between \$100 and \$200, since kits are distributed to many people who would not have died, even without the program.

REFERENCES

Acemoglu, D., & Robinson, J. A. (2005). *Economic origins of dictatorship and democracy*. Cambridge, UK: Cambridge University Press.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic studies, 58*, 277–297.

Beck, N., & Katz, J. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review, 89*, 634–647.

Besley, T., & Kudamatsu, M. (2006). Health and democracy. *The American Economic Review, 96*, 313–318.

Bulir, A., & Hamann, A. H. (2001). How volatile and unpredictable are aid flows, and what are the Policy Implications. World Institute for Development Economics Research, Discussion Paper No. 2001/143.

Boone, P. (1996). Politics and the effectiveness of foreign aid. *European Economic Review, 40*, 289–329.

Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *The American Economic Review, 90*, 847–868.

Croghan, T. W., Beatty, A., & Ron, A. (2006). Routes to better health for children in four developing countries. *Milbank Quarterly, 84*, 333–358.

Du Lou, A. D., Pison, G., & Aaby, P. (1995). Role of immunizations in the recent decline in childhood mortality and the changes in the female/male mortality ratio in rural Senegal. *American Journal of Epidemiology, 142*(6), 643–652.

Franco, A., Alvarez-Dardet, C., & Ruiz, M. T. (2004). Effect of democracy on health: Ecological study. *British Medical Journal, 329*, 1421–1423.

Gomanee, K., Morrissey, O., Mosley, P., & Verschoor, A. (2005). Aid, government expenditure, and aggregate welfare. *World Development, 33*(3), 355–370.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, UK: Chapman & Hall/CRC.

Hill, A. G., MacLeaod, W. B., Joof, D., Gomez, P., & Walraven, G. (2000). Decline of mortality in children in rural Gambia: The influence of village-level primary health care. *Tropical Medicine & International Health, 5*(2), 107–118.

Judson, R. A., & Owen, A. L. (1999). Estimating Dynamic Panel Data Models: A Guide for Macroeconomists. *Economic Letters, 6*, 9–15.

Kosack, S. (2003). Effective aid: How democracy allows development aid to improve the quality of life. *World Development, 31*, 1–22.

Kosack, S., & Tobin, J. (2006). Funding self-sustaining development: The role of aid, FDI and government in economic success. *International organization, 60*(1), 205–243.

Lake, D., & Baum, M. (2001). The invisible hand of democracy: Political control and the provision of public services. *Comparative Political Studies, 34*(6), 587–621.

Levine, R. (2004). Millions saved: Proven successes in global health. Center for Global Development.

Levine, R. (2007). *Case Studies in global health: Millions saved*. Sudbury, MA: Jones and Bartlett.

Lu, C., Schneider, M., Gubbins, P., Leach-Kemon, K., Jamison, D., & Murray, C. J. L. (2010). Public financing of health in developing countries: A cross-national systematic analysis. *Lancet, 375*, 1375–1387.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*, 1417–1426.

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies, 29*, 231–248.

Radelet, S. (2006). A primer on foreign aid. Center for Global Development Working Paper no. 92. Washington, DC: Center for Global Development.

Ravishankar, N., Gubbins, P., Cooley, R. J., Leach-Kemon, K., Michaud, C. M., Jamison, D. T., *et al.* (2009). Financing Global Health 2009: Tracking Development Assistance for Health from 1990 to 2007. *Lancet, 373*, 2113–2124.

Roodman, D. (2007). The anarchy of numbers: Aid, development, and cross-country empirics. *World Bank Economic Review, 21*(2), 255–277.

Ruger, J. P. (2005). Democracy and health. *QJM, 98*(4), 299–304.

Shiffman, J. (2008). Has donor prioritization of HIV/AIDS displaced aid for other health issues?. *Health Policy and Planning, 23*(2), 95–100.

Tierney, M.J., Nielson, D., Hawkins, D. G., Roberts, J.T., Findley, M.G., Powers, R.M., Parks, B. Wilson, S., & Hicks, R.L. (2011). More Dollars than Sense? Addressing Knowledge Scarcity in Development Finance. World Development Journal, Special Issue.

White, H. (2003). Maintaining momentum to 2015?: An impact evaluation of interventions to improve maternal and child health and nutrition in Bangladesh. World Bank Operations Evaluation Department.

Williamson, C. R. (2008). Foreign aid and development: The impact of foreign aid to the health sector. *Southern Economic Journal, 75*(1), 188–207.

Wilson, S. E., & Butler, D. M. (2007). A lot more to do: The sensitivity of time-series cross-section results to simple alternative specifications. *Political Analysis, 15*(2), 101–123.

Woods, R. I., Watterson, P. A., & Woodward, J. J. (1988). The causes of rapid infant mortality decline in England and Wales, 1861–1921, Part 1. *Population Studies, 42*, 343–366.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.world-dev.2011.07.021.