Nicholas A Hein
Danstan S Bagenda
Jiangtao Luo

PEPFAR and adult mortality

A replication study of HIV development assistance effects in Sub-Saharan African countries

February 2018

Replication Paper 15

Health

3ie 10 YEARS
International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programs. We are the global leader in funding, producing and synthesizing high-quality evidence of what works, for whom, how, why and at what cost. We believe that using better and policy-relevant evidence helps to make development more effective and improve people's lives.

3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyze the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce or challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The 3ie Replication Program also includes grant-making windows to fund replication studies of papers identified by donors and implementing organizations. Requests for proposals are issued based on support from donors or crowdsourced recommendations. The studies are chosen by a demand-driven process and include published studies that are considered influential, innovative or counterintuitive. The aim of the 3ie Replication Program is to improve the quality of evidence from development impact evaluations for use in policymaking and program design.

About this report

The publication of this report was supported by the Bill & Melinda Gates Foundation. The contents of this report are the sole responsibility of the authors and do not necessarily represent the official views of the Bill & Melinda Gates Foundation or 3ie, its donors or its Board of Commissioners. Please direct any comments or queries to the corresponding author, Jiangtao Luo at jiangtao.luo@unmc.edu

Suggested citation: Hein, NA, Bagenda, DS and Luo, J, 2018. *PEPFAR and adult mortality: a replication study of HIV development assistance effects in Sub-Saharan African countries. 3ie Replication Paper 15.* Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Marie Gaarder
Managing editor: Benjamin D Wood
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill and Akarsh Gupta

PEPFAR and adult mortality: a replication study of HIV development assistance effects in Sub-Saharan African countries

Nicholas A Hein
University of Nebraska Medical Center

Danstan S Bagenda
University of Nebraska Medical Center

Jiangtao Luo
University of Nebraska Medical Center

3ie Replication Paper 15

February 2018

Acknowledgements

The 3ie Replication Program supported this research. We thank the 3ie team members for their help and support, especially Senior Evaluation Specialist Dr. Benjamin Wood, Evaluation Specialist Dr. Eric W Djimeu and Program Manager Scott Neilitz for their suggestions and criticism at each stage. We also thank several anonymous reviewers for their comments and suggestions.

We thank our colleagues, Dr. Jane Meza, Dr. Lynette Smith and Dr. Fang Yu, from the Department of Biostatistics, University of Nebraska Medical Center. Our special thanks go to Kaeli K Samson for proofreading and correcting our typos and grammatical errors.

We thank Dr. Nicole Perfito from Science Exchange for her suggestion and special wisdom.

We are grateful to Dr. Eran Bendavid, Dr. Jay Bhattacharya and Dr. Grant Miller (from Stanford University) and Dr. Charles B Holmes (from Office of the US Global AIDS Coordinator, US Department of State, Washington, DC) for giving us permission to replicate their research. We appreciate the enormous help from Dr. Eran Bendavid in providing the data and original computing programs.

Summary

The US budget for global health funding increased from US$1.3 billion in 2001 to more than US$10 billion in recent years. Since policy decisions often hinge on whether aid allocation has a significant and intended impact, understanding the relationships between these factors is of critical importance. A lot of global health funding was used for the Global Fund to Fight AIDS through the US President's Emergency Plan for AIDS Relief (PEPFAR) in Africa, but results were uncertain. The selected paper, Bendavid and colleagues' *HIV Development Assistance and Adult Mortality in Africa* (2012a), shows positive effects of PEPFAR in reducing adult mortality in Africa. These results are significant, implying that the accumulated effects of PEPFAR had finally reached a detectable, statistically significant level.

We chose to conduct a replication study, given the importance of an independent evaluation of this study to examine the robustness of the results. Bendavid and colleagues' 2012a study was a cross-country, retrospective analysis using the maternal mortality module from the Demographic and Health Surveys of 27 African countries, administered between 1998 and 2008. Nine of these countries (Ethiopia, Kenya, Mozambique, Namibia, Nigeria, Rwanda, Tanzania, Uganda and Zambia) received aid from PEPFAR. The authors compared the adult mortality per 1,000 adults between the ages of 15 and 59 in PEPFAR and non-PEPFAR countries, using a difference-in-difference analysis that included fixed effects for countries and years, as well as personal and time-varying covariates.

Bendavid and colleagues found that in 2003, the age-adjusted adult mortality in PEPFAR countries was 8.3 per 1,000 adults (95% CI, 8.0–8.6) compared with 8.5 per 1,000 adults (95% CI, 8.3–8.7) in non-PEPFAR countries. In 2008, the age-adjusted adult mortality in PEPFAR countries was 4.1 per 1,000 adults (95% CI, 3.6–4.6), compared with 6.9 per 1,000 adults (95% CI, 6.3–7.5) in non-PEPFAR countries. Using a difference-in-difference adjusted analysis, the odds of adult mortality in a PEPFAR country between 2004 and 2008 was 0.84 (95% CI, 0.72–0.99; p=0.03), compared with a non-PEPFAR country. During the years of full PEPFAR implementation (2004 to 2008), the odds of adult mortality were lower in PEPFAR countries than non-PEPFAR countries.

Our first objective of this replication research was to conduct a pure replication of the study –to establish whether the published findings are reproducible using the study's own data and methods. Other than a few minor discrepancies (e.g. typos, different adjustment in using statistical methods, software used) that do not change the conclusions of the study, we replicated the original study.

Duber and colleagues (2010) also examined the effects of PEPFAR on all-cause adult mortality using a different data set and shorter period, 2000 to 2006. Our second objective was to examine whether using Bendavid and colleagues' methods with the Duber and colleagues time frame would affect the original findings of Bendavid and colleagues. We found that Bendavid and colleagues' findings were robust to the time period used but did not match the findings from Duber and colleagues. It is unclear whether the included years and quality of data set explain the contradictory results. Bendavid and colleagues and Duber and colleagues include different focus countries. In particular, South Africa was not included in the Bendavid and colleagues' data set. Both papers highlight South Africa as having a high prevalence of HIV that could skew the results.

Contents

Acknowledgements ..... **i**
Summary ..... **ii**
List of figures and tables ..... **iv**
Abbreviations and acronyms ..... **v**
1. Introduction **1**
2. The pure replication **2**
2.1 The data 2
2.2 Reproducing the summary statistics 4
2.3 Reproducing the main results 11
2.4 Reproducing mortality effects 14
2.5 Reproducing sensitivity analysis 17
2.6 Pure replication conclusions 21
3. Additional analysis **21**
3.1 Measurement and estimation analysis 22
4. Discussion **28**
5. Limitations **29**
6. Conclusions **30**
Appendix A: Summary of differences between pure replication and original findings **31**
Appendix B: Push-button replication final report – *Journal of American Medical Association*: Eran Bendavid **32**
Appendix C: Push-button replication comparison tables and descriptions **34**
Appendix D: List of files received from the original authors **39**
References **40**

List of figures and tables

Figure 1: Replication results of trends in developmental assistance for HIV to focus countries and non-focus countries: mean per-country assistance in 2008 US$, 1998–2008 8
Figure 2: Replication results of age-adjusted mortality trends in the focus and non-focus countries, 1998–2008 10
Figure 3: Replication results of country-level annual adult mortality trends, 1998–2008 11
Figure 4: Replication results for odds of death when using year relative to program initiation as the main predictor variable 14

Table 1: Replication and original results of study countries, participants and group designation 6
Table 2: Replication and original results of comparison of focus countries and non-focus countries with each other 7
Table 3: Replication and original results for regression models estimating the odds ratio of death in study adults in focus countries versus non-focus countries 12
Table 4: Replication and original results for estimation of the number of deaths averted for the period 2004–2008 16
Table 5: Replication and original results for leave-one-out analysis 17
Table 6: Replication and original results for relative odds of death associated with PEPFAR for subsets of countries and surveys 19
Table 7: Replication and original results for sensitivity analysis using linear time trends 20
Table 8: Comparison of focus and non-focus countries with each other for 2000 and 2006 23
Table 9: Regression models estimating the odds of death in study adults in focus versus non-focus countries for 2000 – 2006 24
Table 10: Leave-one-out country analysis for years 2000 – 2006 25
Table 11: Relative odds of death associated with PEPFAR for countries that contain data for 2000–2006 27
Table 12: Sensitivity analysis using linear time trends for 2000 – 2006 28

Appendix figures and tables

Figure C1: Trends in HIV development assistance for focus countries and non-focus countries: mean per-country assistance in 2008 US$, 1998–2008 (PBR) 38
Figure C2: Age-adjusted adult mortality trends in the focus and non-focus countries, 1998–2008 (PBR) 38

Table A1: Summary of the pure replication 31
Table C1: Study countries, participants and group designation (PBR) 35
Table C2: Comparison of focus countries and non-focus countries with each other and with non-study Sub-Saharan African countries (PBR) 36
Table C3: Regression models estimating the odds ratio of death in study adults in focus countries versus non-focus countries (PBR) 37

Abbreviations and acronyms

CI	Confidence interval
DHS	Demographic and Health Surveys
GDP	Gross domestic product
LOWESS	Locally weighted scatterplot smoothing
OR	Odds ratio
PBR	Push-button replication
PEPFAR	United States President's Emergency Plan for AIDS Relief

1. Introduction

The US budget for global health funding has increased from US$1.3 billion in 2001 to more than US$10 billion in recent years (Valentine et al. 2016). Since policy decisions often hinge on whether aid allocation has had a significant and intended impact, understanding the relationships between these factors is of critical importance (Glassman et al. 2013). A significant amount of global health funding has been used for the Global Fund to Fight AIDS, Tuberculosis and Malaria through the United States President's Emergency Plan for AIDS Relief (PEPFAR) in Africa, but results have been uncertain (Bendavid et al. 2012a; Kaiser Family Foundation 2015). The study selected, Bendavid and colleagues' *HIV Development Assistance and Adult Mortality in Africa* (2012a), first showed positive effects of PEPFAR in reducing adult mortality in Africa. The results implied that the effects of PEPFAR had accumulated and finally reached a detectable, statistically significant level.

It is important for us to give an independent evaluation of this study. Replicating this study fit the mission and strategy of the International Initiative for Impact Evaluation. In addition, Bendavid and colleagues' paper, *HIV Development Assistance and Adult Mortality in Africa*, published in *JAMA* in 2012, was ranked in the top five studies according to our impact score criteria, which is based on the frequency with which a journal article is cited in a specific year. This study investigates the relationship between increased funding to countries receiving aid through PEPFAR and adult mortality.

PEPFAR began its first full year of funding in 2003 and provides funding to 15 focus countries for delivery of antiretroviral therapy and other HIV prevention programs (PEPFAR 2015). Funding allocated to PEPFAR countries increased dramatically between 2004 and 2010 (Kaiser Family Foundation 2015), but the effectiveness of the increased funding to these focus countries on adult mortality is under-studied. Previous studies addressing this question either showed no effect of increased PEPFAR funding on adult mortality during a relatively circumscribed time frame, 2000 to 2006 (Duber et al. 2010), or used estimates with modeled data of mortality rates (Bendavid and Bhattacharya 2009). Since mortality is a direct measure for public health, Bendavid and colleagues (2012a) sought to determine whether trends in mortality following PEPFAR funding reflect benefits beyond HIV-related mortality, using a broader time frame as well as survey data from individuals to more directly measure mortality. The authors performed two primary analyses: (1) a cross-country comparison of adult mortality between 1998 and 2008 in 9 African countries receiving PEPFAR funding (focus countries) and 18 African countries that did not receive funding (non-focus countries), and (2) a within-country comparison of the intensity of PEPFAR implementation and adult mortality in 22 districts of Tanzania and 30 districts of Rwanda.

The main finding of Bendavid and colleagues (2012a) was that adult mortality declined more dramatically (8.3 per 1,000 in 2003 (95% confidence interval [CI], 8.0–8.6) versus 4.1 per 1,000 in 2008 (95% CI, 3.6–4.6) in countries receiving PEPFAR funding; however, they could not distinguish between effects on all-cause adult mortality or effects solely on HIV-related mortality. Similarly, they could not detect a difference in adult mortality that was associated with PEPFAR implementation intensity between districts in Tanzania and Rwanda. The main conclusion is "between 2004 and 2008, all-cause adult mortality declined more in PEPFAR focus countries relative to non-focus

countries" (Bendavid et al. 2012a). Additionally, Bendavid and colleagues (2012a) identified two other factors associated with lower adult mortality – the educational level of the female respondents to the individual household surveys and the effectiveness of the government, which is a measure that captures perceptions of the quality of a country's public services, among other things (Kaufmann et al. 2010).

The findings from Bendavid and colleagues (2012a) generated a substantial amount of debate. In a subsequent *JAMA* article, Shelton (2012) challenged the estimate of the association between mortality and PEPFAR funding, suggesting that the estimates of reduction in mortality should have accounted for population size and the prevalence of HIV in each country. In the same issue, Emanuel (2012) wrote an editorial piece about the implications of the Bendavid study for funding world health programs. Bendavid and Battacharya (2014) looked more broadly at clinical outcomes and funding for health aid, and found that life expectancy increased and the mortality rate of children under 5 years of age decreased with investment.

Our first objective in this replication study was to complete a pure replication. We reproduced the results presented in Bendavid and colleagues' (2012a) paper and electronic appendix using the authors' data set and statistical methods presented in the original paper. In addition to the pure replication, we examined the robustness of Bendavid and colleagues' (2012a) findings using Duber and colleagues' (2010) methodology as a guide. The results from both papers are contradictory on whether PEPFAR was associated with a decrease in all-cause mortality.

2. The pure replication

Our pure replication uses the de-identified, merged data from the Bendavid and colleagues (2012b) study to reassess the intervention, PEPFAR implementation. We reconstructed the original results using the de-identified, merged data and the paper as a guide. This approach has strengths and weaknesses. By using this method, we reduce the amount of overhead in preparing the data. However, we are limited to the countries and variables that the original authors selected.

2.1 The data

The data set was from Demographic and Health Survey (DHS)[1] data. The de-identified, merged, person-level data consisted of 38 DHS that span 27 African countries, with 9 focus countries and 18 non-focus countries, from 1998 to 2008 (Bendavid et al. 2012b).

DHS are nationally representative household (weighted and clustered) surveys that provide data for a broad range of indicators. The sample uses a stratified, two-stage cluster design. The first stage is usually enumeration areas drawn from census files. The second stage is a sample of households in each selected enumeration area, drawn from an updated list of households. In selected households, individual women of reproductive age (15 to 49 years) are interviewed after providing voluntary informed consent. These women must be either permanent residents of the household or visitors who were present in the household on the night before the survey. This sampling scheme is nationally representative; however, it does under-sample households that include only

[1] See https://dhsprogram.com/

men. As recognized by Bendavid and colleagues (2012a), this under-sampling could bias the results if sibling mortality rates in households with only men are substantially different from households that have at least one female member. Unfortunately, we cannot verify this possibility.

The DHS contains household and biomarker questionnaires. For information on topics not included in the model questionnaires, optional questionnaire modules are available at the request of host countries, such as the maternal mortality module used in this study. The maternal mortality module has been in the same format and present in every DHS since the late 1980s. The maternal mortality module contains sibling information of the respondent. Siblings must be born to the respondent's natural mother, and siblings can be alive or dead and live with or away from the respondent. Sibling information includes the age of siblings, whether siblings are still alive and, if not alive, when they died and age at death.

2.1.1 The longitudinal data

Using the raw data supplied by the original authors, we independently created a longitudinal data set with repeated observations for the siblings of the respondent, as described in the original paper; the respondents were not included in the longitudinal data set. Note that during the period, from 1998 to 2008, some countries had multiple surveys administered. However, the probability that a respondent participated in multiple surveys is essentially zero.

For each sibling in the data, we created a sequence of dummy variables to indicate whether that sibling was alive or dead between 1998 and 2008. We excluded observations if the sibling was not between the ages of 15 and 59 years and any subsequent years after the death of the sibling. The last year of a sibling's repeated observations is the year before the survey, since the survey year contains incomplete follow-up for the entire calendar year. We calculated the sibling's age using the sibling's date of birth supplied by the respondent for each repeated observation. Along with the above-defined indicator variable and age, we created a recall variable and focus country variable. The recall variable is the difference between the survey year and the repeated observation year. The focus country variable is a dummy variable that indicates whether the survey response originated from a focus country (PEPFAR implemented) or non-focus country (PEPFAR not implemented).

We merged this newly created longitudinal data set with two other data sets provided by the original authors. The two other data sets contained information that varied by country and year for the following: population, HIV prevalence among adults 15 to 49 years old, gross domestic product (GDP) per capita, HIV aid, HIV aid per adult with HIV, antiretroviral coverage, percent of population in urban residence, and government effectiveness. The merged data set was the data set used for analysis.

2.1.2 Statistical methods

The original paper used a logistic regression and a difference-in-difference analysis to evaluate the effects of PEPFAR implementation. Specifically, the original authors compared the odds of adult (defined as men and women aged 15 to 59 years) all-cause mortality in focus and non-focus countries pre- and post-PEPFAR implementation. The original authors defined PEPFAR implementation as post-2003. Bendavid and

colleagues (2012a) compared all-cause mortality at the individual level using a logistic regression model. By examining all-cause mortality at the person level, the original authors could adjust for individual and country-level covariates. The individual covariates were recall period between survey year and repeated observation year, sibling age in years, respondent's education and respondent's place of residence. The time-varying, country-level covariates included HIV prevalence, per capita development assistance for health from sources other than PEPFAR, GDP per capita and index of government effectiveness from the World Governance Indicators. The original authors computed adjusted and unadjusted odds ratios (ORs) using a logit model with robust standard errors clustered by country. All models included year and country fixed effects. The primary variable of interest was whether a sibling lived in a focus country during PEPFAR's implementation. To implement the primary variable of interest, Bendavid and colleagues created another dummy variable indicating pre- or post-PEPFAR implementation. The primary variable of interest was the interaction between the dummy variable for PEPFAR implementation and the dummy variable indicating whether the survey originated from a focus country. Adjusted models included country-level and individual level covariates. The original authors used Stata version 11.2 for the original analysis.

We conducted the replication analysis using the same methods as the original analysis using SAS/STAT software version 9.4 (SAS Institute Inc., Cary, NC, USA) and Stata version 14.1. We used the SAS SURVEYLOGISTIC procedure for this analysis. The SURVEYLOGISTIC procedure allows for clustering by countries, thereby relaxing the assumption of independent and identically distributed errors within a country. This methodology permits the computation of unadjusted and adjusted ORs, with robust standard errors.

The original authors provided a segment of their code to us as a courtesy, so we know exactly how they performed a portion of the original analysis. Some of the analysis requires that the outcome is age adjusted. It was unclear what standard population the original authors used, and the supplied code did not incorporate any age adjustments.

2.1.3 Formatting the data

We obtained the data provided by Bendavid and colleagues in Stata and converted to SAS using Stata version 14.1. If there was a discrepancy in the replication results using SAS software, we reanalyzed the results using the Stata version 14.1, using the code provided by the authors if applicable. Overall, we identified some discrepancies; however, these discrepancies did not have an impact on the main findings. We highlight any difference in the respective table or figure, and we included a summary of the differences in Appendix A, Table A1, as a courtesy for the reader.

2.2 Reproducing the summary statistics

Our pure replication began by reproducing Table 1 of the original paper. In Table 1 we show a summary of the survey fieldwork dates, number of respondents, number of observations after the creation of the longitudinal data set and number of deaths by country. We also stratified the countries by focus and non-focus countries according to the stratification used in the original paper.

When looking at survey fieldwork dates for Kenya, Namibia, Nigeria, Rwanda, Uganda, Zambia, Burkina Faso, Cameroon, Congo, Democratic Republic of Congo, Gabon, Guinea, Lesotho, Madagascar, Malawi, Mali, Niger, Senegal, Sierra Leone, Swaziland and Zimbabwe, we observe a difference in the dates in our replication study versus the original paper. For a majority of the date discrepancies, the survey window is either shorter or longer by one month. In the replication we used the variable *v008* to determine the survey fieldwork dates. The variable *v008* is contained in the maternal mortality module and *v008* is the date of interview, formatted as century-month codes. It is unclear why there is a difference in the survey fieldwork dates, considering that the dates are the only discrepancy between our Table 1 and Table 1 of the original paper. The number of unique adults, number of observations and number of deaths are in agreement with our replication results.

Our results for study countries, participants and group designation can be found in Table 1, and the original paper results are also included alongside in our replication Table 1. We have also highlighted any differences between the two studies in the table.

Similarly, we replicate the Table 2 summary statistics of the original paper in a Table 2 that includes the original paper Table 2 figures alongside. Table 2 is a comparison of focus countries and non-focus countries with each other and with non-study Sub-Saharan countries and includes the means and CIs of various parameters, grouped by focus countries, non-focus countries and other Sub-Saharan countries. Following the method described by the original paper authors, we compare the means of various parameters of focus and non-focus countries, and study countries and other Sub-Saharan countries using two-tailed t-tests, and similarly report the results as p-values. However, we are unable to include the non-study Sub-Saharan African countries and *antiretroviral coverage, %*, since the data the original authors used to produce those was not provided and we are also unable to retrieve the data using the references provided.

We identified some discrepancies in CIs that appear to be a result of rounding. However, the point estimate for *HIV aid per country, millions of $* for the focus countries for 1998 differed and cannot be explained by rounding. This discrepancy leads to a difference in p-values comparing focus and non-focus countries for that particular year. We are unclear of the cause of the discrepancy. Furthermore, the push-button replication (PBR), which used the original authors' supplied code and data with no modification to either, matched our results for *HIV aid per country, millions of $* for 1998. Here, PBR "is checking whether the original authors' code can be run on the original data to reproduce the published results" (Brown and Wood 2016).

We display the replication and original results in Table 2. We have also highlighted all differences between the replication study and original study in Table 2. The notes shown for the table are those that appeared in the original authors Table 2.

Table 1: Replication and original results of study countries, participants and group designation

	Replication study				Original study			
Country	**Survey fieldwork dates**	**No. of unique adults**	**Observations, no.**	**No. of deaths**	**Survey fieldwork dates**	**No. of unique adults**	**Observations, no.**	**No. of deaths**
Focus countries								
Ethiopia	2–6/2000, 4–8/2005	96,980	391,835	2,596	2–6/2000, 4–8/2005	96,980	391,835	2,596
Kenya	4–9/2003, 11/2008–3/2009	73,580	491,521	2,971	4–9/2003, 11/2008–2/2009	73,580	491,521	2,971
Mozambique	8/2003–1/2004	41,103	189,752	1,367	8/2003–1/2004	41,103	189,752	1,367
Namibia	9–12/2000, 11/2006–3/2007	64,382	340,338	3,303	9–12/2000, 10/2006–3/2007	64,382	340,338	3,303
Nigeria	6–11/2008	122,815	1,020,435	4,590	6–10/2008	122,815	1,020,435	4,590
Rwanda	5–11/2000, 2–8/2005	74,818	316,179	2,943	6–8/2000, 2–7/2005	74,818	316,179	2,943
Tanzania	10/2004–2/2005, 12/2009–5/2010	83,992	615,367	2,993	10/2004–2/2005, 12/2009–5/2010	83,992	615,367	2,993
Uganda	9/2000–3/2001, 5–10/2006	62,132	301,234	2,856	9/2000–3/2001, 4–10/2006	62,132	301,234	2,856
Zambia	11/2001–6/2002, 4–10/2007	60,014	328,837	4,228	11/2001–5/2002, 4/2007–1/2008	60,014	328,837	4,228
Non-focus countries								
Benin	8–11/2006	64,463	449,155	1,703	8–11/2006	64,463	449,155	1,703
Burkina Faso	1–3/1999, 6–12/2003	55,416	206,068	1,123	11/1998–3/1999, 6–12/2003	55,416	206,068	1,123
Cameroon	2–9/2004	41,422	222,637	1,550	2–8/2004	41,422	222,637	1,550
Chad	7–12/2004	20,891	111,943	736	7–12/2004	20,891	111,943	736
Congo	7–11/2005	28,305	175,576	1,323	1–11/2005	28,305	175,576	1,323
Congo Dem Rep	1–9/2007	38,637	295,800	1,887	1–8/2007	38,637	295,800	1,887
Gabon	7/2000–2/2001	22,083	43,671	210	7/2000–1/2001	22,083	43,671	210
Guinea	4–8/1999, 2–6/2005	44,848	177,877	977	5–7/1999, 2–6/2005	44,848	177,877	977
Lesotho	9/2004–2/2005, 10/2009–1/2010	47,185	334,908	4,428	9/2004–1/2005, 10/2009–1/2010	47,185	334,908	4,428
Liberia	12/2006–4/2007	23,052	178,489	842	12/2006–4/2007	23,052	178,489	842
Madagascar	11/2003–6/2004, 11/2008–7/2009	107,869	844,146	3,509	11/2003–3/2004, 11/2008–8/2009	107,869	844,146	3,509
Malawi	7–11/2000, 1/2004, 9/2004–2/2005	84,041	305,436	3,945	7–11/2000, 10/2004–1/2005	84,041	305,436	3,945
Mali	1–6/2001, 3–12/2006	92,775	470,612	2,161	1–5/2001, 5–12/2006	92,775	470,612	2,161
Niger	1–6/2006	34,858	243,442	942	1–5/2006	34,858	243,442	942
Senegal	1–6/2005	55,881	347,114	1,096	2–5/2005	55,881	347,114	1,096
Sierra Leone	4–8/2008	19,675	165,810	891	4–6/2008	19,675	165,810	891
Swaziland	6/2006–3/2007	18,458	128,135	1,739	7/2006–2/2007	18,458	128,135	1,739
Zimbabwe	8–12/1999, 8/2005–2/2006, 4/2006	58,937	247,359	3,394	9–12/1999, 8/2005–2/2006	58,937	247,359	3,394

Table 2: Replication and original results of comparison of focus countries and non-focus countries with each other

		Replication study			Original study				
		Mean (95% CI)			**Mean (95% CI)**				
Parameter		**Focus countries**	**Non-focus countries**	**p-value[a]**	**Focus countries**	**Non-focus countries**	**p-value[a]**	**Other Sub-Saharan countries[b]**	**p-value[c]**
Population, millions	1998	33.6 (5.1 to 62.1)	9.8 (4.5 to 15)	0.0164	33.6 (5.1 to 62.1)	9.8 (4.5 to 15)	0.0164	10.3 (2.9 to 17.7)	0.35
	2008	43.4 (7.8 to 79)	12.8 (5.9 to 19.8)	0.0149	43.4 (7.8 to 79)	12.8 (5.9 to 19.8)	0.0149	13.6 (4.8 to 22.4)	0.3
HIV prevalence among adults 15–49 y old, %	1998	8.1 (5 to 11.3)	6.5 (2 to 11)	0.6177	8.1 (5 to 11.2)	6.5 (2 to 11)	0.6177	5.0 (1.1 to 8.9)	0.41
	2008	7.5 (3.9 to 11.1)	5.8 (1.9 to 9.8)	0.5739	7.5 (3.9 to 11.0)	5.8 (1.9 to 9.8)	0.5739	5.0 (1.0 to 9.1)	0.56
GDP per capita, constant $	1998	471.3 (–1.4 to 944.1)	641.8 (98.7 to 1,184.8)	0.6692	471.3 (98.6 to 844.1)	641.8 (98.7 to 1,184.8)	0.6692	767.2 (152.6 to 1,381.9)	0.58
	2008	629.1 (15.1 to 1243.1)	654.5 (180.3 to 1,128.8)	0.9458	629.1 (115.1 to 1,143.1)	654.5 (180.3 to 1128.8)	0.9458	995.1 (148.4 to 1,841.8)	0.34
HIV aid per country, millions of $	1998	7.3 (1 to 13.6)	2 (–0.2 to 4.3)	0.0407	6.3 (0.0 to 14.6)	2 (–0.1 to 4.2)	0.16	1.8 (0.9 to 2.7)	0.25
	2008	240.5 (168.7 to 312.3)	24.6 (10.2 to 39.1)	<0.0001	240.5 (168.7 to 312.3)	24.6 (10.2 to 39.1)	<0.0001	63.1 (–6.1 to 132.4)	0.37
HIV aid per adult with HIV, $	1998	3.8 (1.8 to 5.7)	6.3 (0.2 to 12.3)	0.5548	3.8 (1.8 to 5.7)	6.3 (0.2 to 12.3)	0.5548	18.4 (0.6 to 35.1)	0.11
	2008	171 (75.8 to 266.3)	76.9 (54.9 to 98.9)	0.0074	171 (75.8 to 266.3)	76.9 (54.9 to 98.9)	0.0074	113.1 (40.7 to 185.6)	0.89
Antiretroviral coverage, %					2.6 (–2.1 to 7.4)	1.9 (–3.1 to 7.2)	0.46	1.9 (–2.6 to 6.5)	0.68
					55.6 (38.3 to 73.6)	28.6 (16.1 to 41.2)	0.04	39.6 (26.6 to 52.5)	0.51
Urban residence, %	1998	24 (15.8 to 32.3)	33.7 (25.4 to 42)	0.1273	24 (15.8 to 32.3)	33.7 (25.4 to 42)	0.1273	38.1 (29.6 to 46.5)	0.45
	2008	28.1 (19 to 37.2)	38 (29.1 to 46.9)	0.1451	28.1 (19 to 37.2)	38 (29.1 to 46.9)	0.1451	42.4 (33.5 to 51.3)	0.43

Notes: [a] p-values provided from 2-tailed *t* tests on the data for the specified year in the focus countries compared with the non-focus countries.

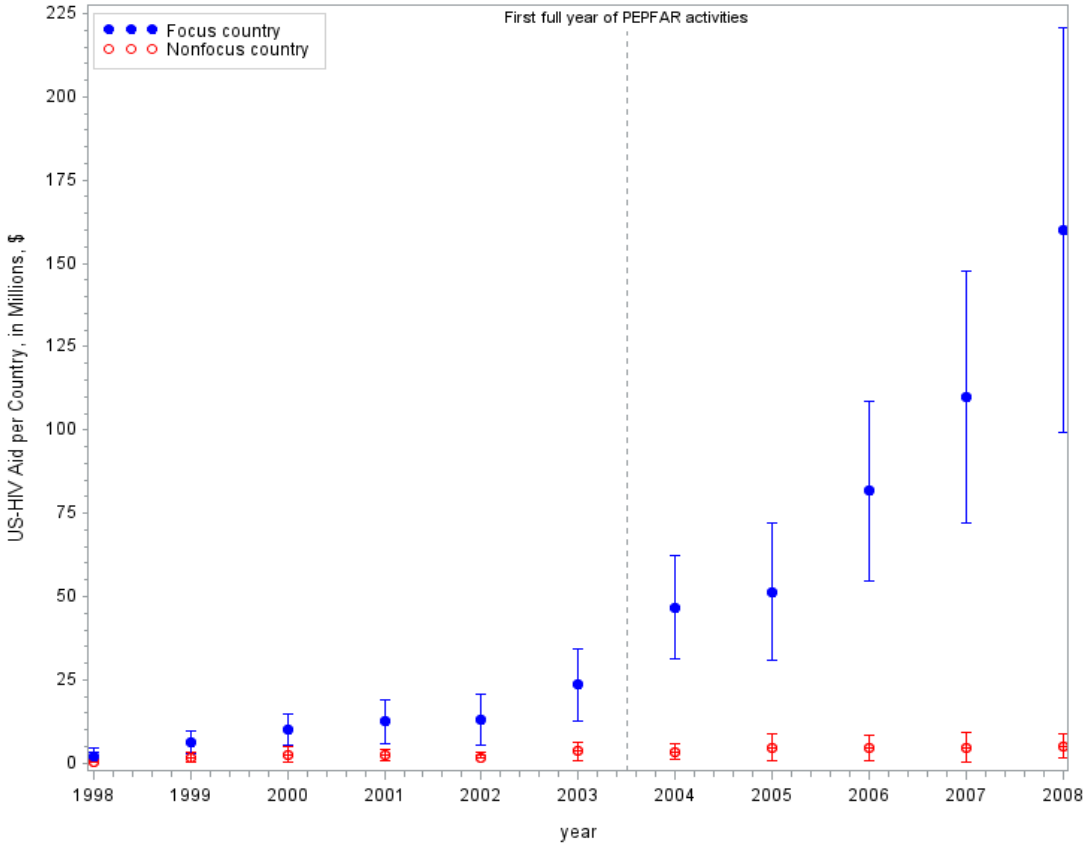
[b] Sub-Saharan African countries not included in this study are Angola, Central African Republic, Burundi, Djibouti, Eritrea, Somalia, Sudan, Botswana, South Africa, Côte d'Ivoire, Ghana, Guinea-Bissau, the Gambia and Togo.

[c] p-values provided for the comparison between the aggregated estimates for all 27 study countries (focus and non-focus countries) and the Sub-Saharan African countries not included in the study. This comparison provides a comparison between the studies excluded from the study and those included.

Next, we reproduced two of the three figures in the original manuscript. We were unable to reproduce Figure 3, “Adult mortality trends in Tanzania separated by PEPFAR activity, 1998–2008,” due to not being able to obtain the appropriate data. Figure 1 is a trend in development assistance for HIV to focus and non-focus countries from 1998 to 2008, while Figure 2 is an age-adjusted adult mortality trend in the focus and non-focus countries for the same years. The original authors considered 2004 to be the first full year of PEPFAR implementation; we indicated the time of full implementation by a vertical dashed line in both figures.

Figure 1 displays the mean US HIV aid in millions of dollars for focus and non-focus countries, along with 95 percent error bars. Our replication results appear to match the original study. Because of copyright concerns, we present only the reproduced figures throughout our replication paper.

Figure 1: Replication results of trends in developmental assistance for HIV to focus countries and non-focus countries: mean per-country assistance in 2008 US$, 1998–2008



Note: Error bars represent 95 percent CIs. A greater increase in assistance to the focus countries is seen between 2003 and 2004. Data is drawn from the Institute for Health Metrics and Evaluation database. We considered calendar year 2004 to be the first full year of PEPFAR’s activities. Error bars indicate 95 percent CIs; PEPFAR, the US President’s Emergency Plan for AIDS Relief.

The last figure we reproduced from the article, excluding the supplementary material, is Figure 2, age-adjusted mortality. The original paper mentions, in the footnote of the figure, that the original authors used United Nations Population Division age-structured

population estimates for each country for age adjustments. We calculated age weights for each 5-year age group between the ages of 15 and 59 years. We applied these weights to the crude mortality estimates, and the point estimates represent the adjusted mortality per 1,000 adults along with 95 percent CI error bars. As per the original study, we used a narrow-bandwidth (0.6) locally weighted scatterplot smoothing (LOWESS) curve to fit the trend. A LOWESS curve is a weighted, least-squares regression fitted locally on subsets of the original data; the parameter determines the percentage of points in the local subsets.

We implemented the age adjustment using the United Nations Population Division (2015) age-structured population estimates from 2005 for the 27 study countries. For each 5-year age group, we summed the population estimates for the 27 study countries, creating a standard population for each age group. The weight for each age group is the standard population of the age group divided by the sum of the standard populations. The age-adjusted rate is the crude mortality rate for a particular age group multiplied by the appropriate weight and normalized per 1,000. Summing the individual age-adjusted rates gives the age-adjusted mortality per 1,000 for adults aged 15 to 59 years. We calculated age-adjusted rates separately for focus and non-focus countries. We calculated the 95 percent CIs separately for the focus and non-focus countries using a method developed in Fay and Feuer (1997).

Figure 2 from the replication study and original paper do not match perfectly; however, they display the same general trend. In the results section of the original paper, the authors provide point estimates and CIs for 2003 and 2008. In 2003, age-adjusted mortality was 8.3 per 1,000 adults in focus countries (95% CI, 8.0–8.6) and 8.5 per 1,000 adults in non-focus countries (95% CI 8.3–8.7). In 2008, the age-adjusted adult mortality declined to 4.1 per 1,000 adults in focus countries (95% CI, 3.6–4.6) and 6.9 per 1,000 in non-focus countries (95% CI, 6.3–7.5).

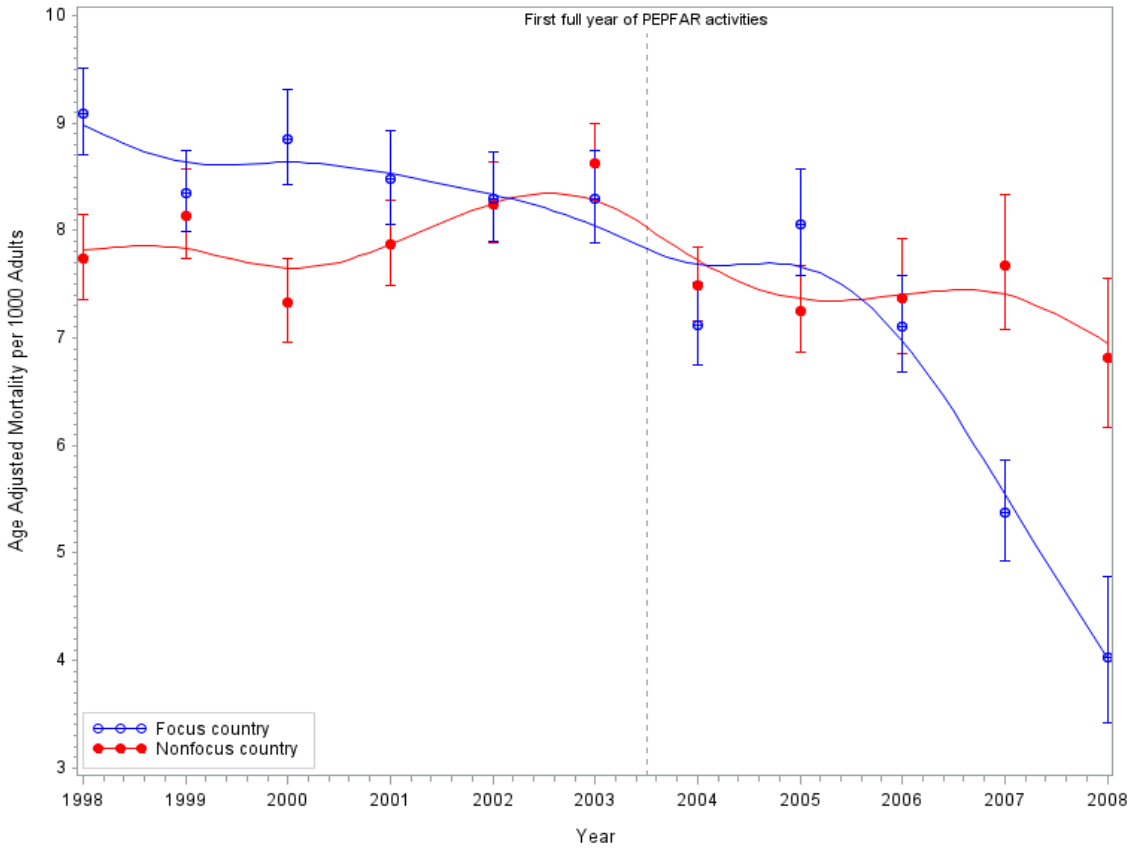
Our replication results indicate that the age-adjusted adult mortality in 2003 was 8.3 per 1,000 adults (95% CI, 7.9–8.7) in focus countries and 8.6 per 1,000 adults (95% CI, 8.3–9.0) in non-focus countries. In 2008, our results indicate that age-adjusted adult mortality per 1,000 adults was 4.0 (95% CI, 3.4–4.8) in focus countries and 6.8 (95% CI, 6.2–7.6) in non-focus countries. The original results and our results both show that the age-adjusted mortality in focus countries has been decreasing more rapidly than in non-focus countries after the implementation of PEPFAR. Prior to PEPFAR implementation, focus countries and non-focus countries had similar age-adjusted mortality rates for most years.

We examined the robustness of our results by using 2000 and 2010 as standard populations and determined that the year used for the standard population did not affect our results. The cause of the discrepancy between our results and the published results is unclear. There is no mention of the year or countries that were used for the standard population in the original paper or in the code provided by the original authors. Additionally, we used a different revision of World Population Prospects. We present our replication results in Figure 2.

In addition to the aggregated, age-adjusted mortality trends by focus and non-focus countries, the original authors provide mortality trends by country, as shown in eFigure1

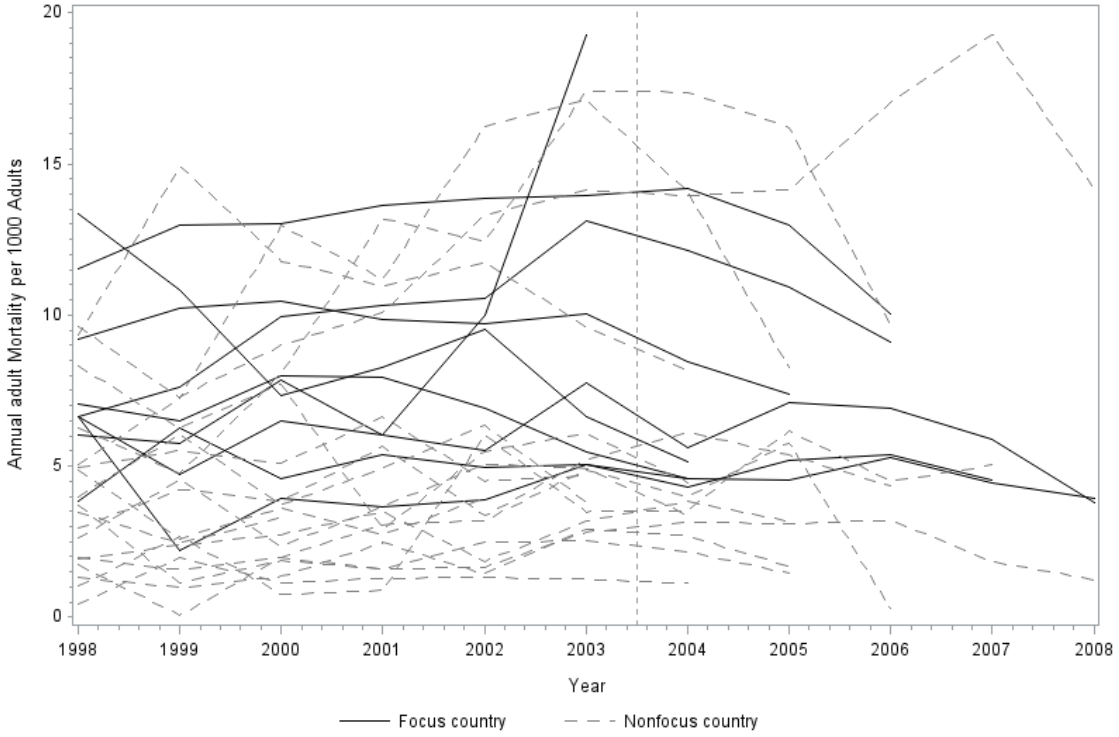
of the appendix to the original paper. Each line in the figure represents the number of adult deaths per 1,000 adults who were alive for any part of a given year. The original authors state that they calculated the mortality trends by country using relevant DHS. We reproduced Figure 1 (Bendavid et al. 2012a) using the longitudinal data set and display our results in Figure 3. Our results appear to be in agreement with the original figure; however, it appears that our results vary slightly for a few countries. For these few countries, the last year of data between our results and the original results do not agree. We are unable to explain the discrepancy.

Figure 2: Replication results of age-adjusted mortality trends in the focus and non-focus countries, 1998–2008



Note: Each point represents the probability that an adult aged 15 to 59 years died during the indicated year per 1,000 in either a focus or non-focus country. The error bars represent 95 percent CIs. Point estimates are age adjusted and age-adjusted CIs are calculated using the method in Fay and Feuer (1997). The trend line is fit by using a smoothing spline.

Figure 3: Replication results of country-level annual adult mortality trends, 1998–2008



Our replication results appear to be in overall agreement with the summary statistics of the original paper. There was no statistical difference in HIV prevalence between focus and non-focus countries (Table 1); the implementation of PEPFAR creates a widening gap of HIV assistance between focus and non-focus countries (Table 1 and Figure 1); and focus countries showed more decline in age-adjusted, all-cause mortality between 2004 and 2008 (Figure 2). The few discrepancies were minor and did not change the interpretation of the results.

2.3 Reproducing the main results

We present the primary results from the original paper in Table 3. In the original Table 3, the original authors presented the ORs, 95 percent CIs and p-values from three regression models: unadjusted, adjusted for country-level covariates and adjusted for country-level and individual-level covariates. The primary outcome was all-cause adult mortality.

In the original paper, unadjusted ORs were calculated using logistic regression, with the primary variable of interest as described in section 2.1.2, and country and year fixed effects. The coefficient of the primary variable of interest represents the difference-in-difference treatment estimator. We computed robust standard errors, clustered by country, to relax the assumption of independent and identically distributed errors within a country. We calculated adjusted ORs in a similar manner, except that country-level and individual-level covariates were included. We present our results for the effects of PEPFAR on adult deaths, as measured by the difference-in-difference indicator, in Table 3, alongside the original results. Notes for Table 3 are from the original authors.

Table 3: Replication and original results for regression models estimating the odds ratio of death in study adults in focus countries versus non-focus countries

	Replication study						Original study					
	Unadjusted OR (95% CI)[a]	**p-value**	**Adjusted OR with country covariates (95% CI)**	**p-value**	**Adjusted OR With country and personal covariates (95% CI)**	**p-value**	**Unadjusted OR (95% CI)[a]**	**p-value**	**Adjusted OR with country covariates (95% CI)**	**p-value**	**Adjusted OR with country and personal covariates (95% CI)**	**p-value**
Adult death[b]	0.80 (0.68–0.95)	0.01	0.82 (0.72–0.95)	0.01	0.84 (0.72–0.97)	0.02	0.80 (0.68–0.95)	0.01	0.83 (0.72–0.95)	0.01	0.84 (0.72–0.99)	0.03
HIV prevalence (per additional 1%)			1.07 (1.00–1.15)	0.04	1.07 (1.01–1.14)	0.03			1.07 (1.01–1.14)	0.03	1.07 (1.01–1.14)	0.03
Non-PEPFAR assistance[c]			1.00 (0.98–1.01)	0.87	1.00 (0.99–1.01)	0.89			0.99 (0.96–1.01)	0.24	0.99 (0.96–1.02)	0.61
GDP per capita (per additional $1)			1.00 (1.00–1.00)	0.82	1.00 (1.00–1.00)	0.63			1.00 (1.00–1.00)	0.65	1.00 (0.99–1.01)	0.58
Government effectiveness (per 1 point increase)[d]			0.62 (0.40–0.95)	0.03	0.58 (0.38–0.89)	0.01			0.62 (0.41–0.95)	0.03	0.58 (0.38–0.89)	0.01
Sibling age (per year)					1.05 (1.04–1.05)	<0.001					1.05 (1.04–1.05)	<0.001
Residence in urban area[e]					0.94 (0.89–1.00)	0.05					0.94 (0.89–0.99)	0.05
Education (per additional year)					0.99 (0.98–1.00)	0.01					0.98 (0.98–0.99)	0.01
Recall (interval between survey and observation, per year)					0.97 (0.95–0.99)	0.01					0.97 (0.95–0.99)	0.006

Notes: [a] All results are exponentiated coefficients on parameters in logistic regression models. Unadjusted model includes the main effect as well as country and year fixed effects. All CIs are calculated using robust standard errors clustered by country.

[b] These ORs represent the relative reduction in mortality among adults living in the focus countries while PEPFAR was implemented compared with adults living in non-focus countries.

[c] All development assistance for health from all donors minus US-funded HIV development assistance per capita.

[d] The index is centered at 0 and each 1 point represents 1 standard deviation, with higher numbers representing greater government effectiveness.

[e] These variables are characteristics of the index women rather than the sibling. The residence status and educational status of the sibling are not known.

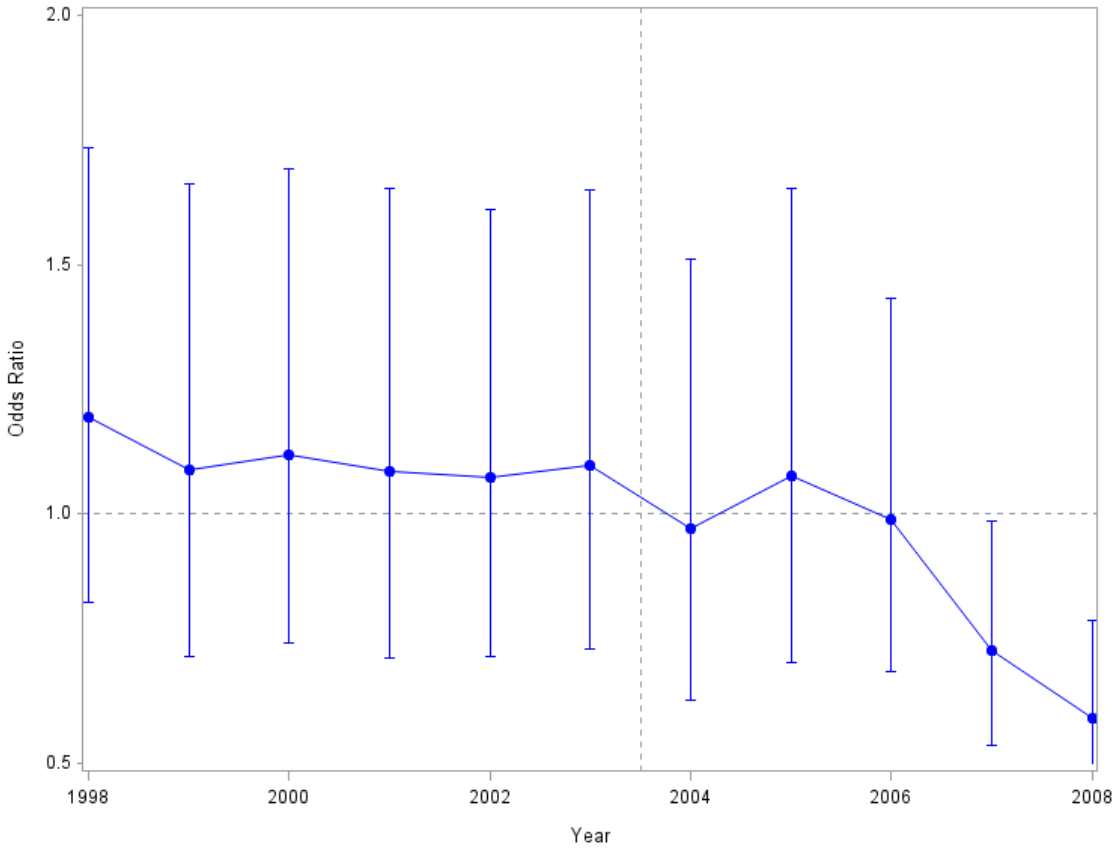
Our replication of the primary results follows the original results. However, there were some differences between the replication and original results. Most differences appeared to be rounding errors in either the point estimate or the CIs. There were two notable differences. The CI for the OR of *adult death* under the model that adjusted for country-level and individual-level covariates is wider in the original paper (0.72–0.99) than in our results (0.72–0.97), which makes our main findings slightly more significant, with a p-value of 0.03 versus 0.02. A similar effect occurs for the ORs of *Non-PEPFAR assistance* in both adjusted models.

To exclude the possibility of software differences, we verified our results by using the supplied code. The results of the provided code matched our replication results. It is unclear why the replication results and the results presented in the paper do not match exactly. In the electronic appendix of the original paper, there is a footnote that mentions that the authors accounted for the DHS sampling weights that account for the sampling scheme. There are no further details on how the authors accounted for the sampling weights. The original authors did not supply sampling weights in the data set, so we were unable to verify if these differences were caused by the inclusion of sampling weights.

The original authors were concerned that the decline in adult mortality in focus countries that implemented PEPFAR could have been a result of preexisting trends in those countries. To examine if this decline was part of a preexisting trend, the original authors used a linear time trend. They transformed the main variable of interest to a set of dummy variables that indicated the focus countries for every year from 1998 to 2008. We implemented the dummy variables by the interaction of the indicator variable for focus country and the year variable. We then performed an unadjusted logistic regression with the newly created main variable of interest as our only covariate. As with previous models, we calculated robust standard errors, clustering by country, and then plotted the ORs and 95 percent CI for each year. We present our results in Figure 4. Our results appear to be in perfect agreement with the original results; there does not appear to have been a preexisting trend in adult mortality prior to the implementation of PEPFAR.

Even with the differences in Table 3, our interpretation of the results matches that of the original authors: between 2004 and 2008, all-cause adult mortality declined more in countries that implemented PEPFAR than countries that did not implement PEPFAR. Our results for the main variable of interest is slightly more significant than the original paper when adjusting for country-level and individual-level covariates. These results do not appear to be the result of a preexisting trend, as shown in Figure 4.

Figure 4: Replication results for odds of death when using year relative to program initiation as the main predictor variable



2.4 Reproducing mortality effects

To examine the number of deaths averted by PEPFAR, we used a three-step process. We first used our results from the logistic regression to predict two quantities for each person-year observation: the predicted mortality of a person if PEPFAR had been in place and the predicted mortality of an individual if PEPFAR had not been in place. We obtained 10 predicted quantities for each observation, two predictions per year for 2004 to 2008. We then calculated the predicted mortality by focus country and year, as follows:

$$\hat{\theta}_j^0(\tau) = \frac{1}{pop_j}\sum P(mortality\ without\ PEPFAR)$$

$$\hat{\theta}_j^1(\tau) = \frac{1}{pop_j}\sum P(mortality\ with\ PEPFAR)$$

where *j* is an index of the focus countries, τ is the index of the year and pop is the number of observations.

Next, we calculated the effects of PEPFAR on the decrease in the mortality rate in each year and each focus country, as follows:

$$d_j(\tau) = \hat{\theta}_j^0(\tau) - \hat{\theta}_j^1(\tau)$$

Finally, we calculated the number of deaths averted by focus country *j*, as follows:

$$DA_j(\tau) = d_j(\tau) \times std_j(\tau)$$

$$Deaths\ averted_j = \sum_{i=1}^{5} DA_j(i)$$

where $std_j(\tau)$ is the number of adults (aged 15 to 59 years) in country *j* and year τ as estimated from the World Population Prospects (United Nations Population Division 2015).

We found that our results were fairly consistent with the original results, except for Mozambique and Rwanda. Point estimates for these two countries differed by a large margin. These different point estimates then affected the calculation of deaths averted.

We were not surprised that our results did not exactly match the original results. When reproducing the main results, it was clear that our logistic regression had slightly different coefficients, which would affect the predicted mortalities. Additionally, the original authors' process was somewhat ambiguous. We are unsure how they predicted their mortalities from the logistic regression. Based on their description, it appears that they predicted only two quantities (mortality with PEPFAR and mortality without PEPFAR) and then limited the predictions to years 2004 to 2008. However, this would not be possible, since Mozambique did not have any surveys administered during this period. It is also unclear what standard population the original authors used. These differences aside, we feel that our replication results support the findings of the original authors for the number of deaths averted. We show our results in Table 4 alongside the original results, along with the original notes.

Table 4: Replication and original results for estimation of the number of deaths averted for the period 2004–2008

	Replication study				Original study			
	Adult mortality with PEPFAR*	Adult mortality without PEPFAR**	Adult population (millions)+	Deaths averted (thousands)++	Adult mortality with PEPFAR*	Adult mortality without PEPFAR**	Adult population (millions)+	Deaths averted (thousands)++
Ethiopia	5.4	6.5	38.8	201.7	5.0 (2–12.8)	5.9 (2.3–15.2)	35.3	161.9 (98–418.6)
Kenya	5.2	6.2	19.3	96.2	4.6 (1.9–11.1)	5.4 (2.2–13.2)	17.6	74.3 (43.8–181.4)
Mozambique	5.9	7.0	10.7	60.3	8.2 (4.4–17.6)	8.4 (4.5–18.1)	9.7	9.4 (7.8–23.1)
Namibia	8.2	9.8	1.1	9.0	9.1 (3.9–22.2)	10.8 (4.7–26.3)	1	8.4 (4.8–20.4)
Nigeria	3.9	4.7	74.2	280.2	3.9 (1.5–9.4)	4.6 (1.8–11.2)	70.3	254.5 (153.7–614.2)
Rwanda	7.6	9.1	5.0	36.2	6.2 (2.5–15.4)	7.4 (3–18.3)	4.2	24.1 (14.4–59.9)
Tanzania	4.2	5.0	20.3	82.2	3.8 (1.6–9.1)	4.6 (1.9–10.8)	18.4	65.7 (38.3–155.5)
Uganda	7.9	9.4	13.6	102.0	7.8 (3.2–18.8)	9.3 (3.8–22.2)	13.4	96.8 (57.2–232.2)
Zambia	10.8	12.9	6.1	62.3	10.5 (4.2–24.1)	12.4 (5–28.5)	5.7	55.1 (33.1–126.3)
Total				930.1				740.8 (443.3–1,808.5)

Notes:

* These are the mean predicted annual adult mortality per 1,000 adults from the fully adjusted logistic regression for the period from 2004 to 2008 under the assumption that PEPFAR had been active in the focus countries ("actual" scenario).

** These are the mean predicted annual adult mortality per 1,000 adults from the fully adjusted logistic regression for the period from 2004 to 2008 under the assumption that PEPFAR had not been active in the focus countries ("counterfactual" scenario).

+ Mean adult population for the period from 2004 to 2008.

++ The number of deaths averted is estimated by multiplying the difference between the predicted and observed adult mortality by the size of the adult population. It is calculated separately for each year and summed over 2004 to 2008. Therefore, it differs somewhat from the product of the mean annual difference, population size and the five years of activity (2004 to 2008, inclusive).

2.5 Reproducing sensitivity analysis

To examine the robustness of the results, the original authors performed several sensitivity analyses. We examined whether any one country unduly affected the main findings by conducting a leave-one-out analysis. We used an unadjusted and an adjusted logistic regression with robust standard errors, where the latter adjusted for country-level and individual-level covariates. All models included the difference-in-difference treatment indicator and country and year fixed effects. Our results are in agreement with the original results; the magnitude and direction of the ORs appear consistent when performing the leave-one-out analysis. There appears to be some rounding errors, which we did not highlight. There are four cases were rounding cannot explain the discrepancy, which we did highlight. These differences did not change the significance of the results. We display our results in Table 5, along with the original results and notes.

Table 5: Replication and original results for leave-one-out analysis

	Replication study			Original study	
Country left out	Unadjusted OR (95% CI)	Adjusted OR (95% CI)*		Unadjusted OR (95% CI)	Adjusted OR (95% CI)*
Benin	0.80 (0.67–0.94)	0.83 (0.71–0.95)		0.80 (0.67–0.94)	0.82 (0.71–0.95)
Burkina Faso	0.80 (0.68–0.95)	0.84 (0.72–0.97)		0.80 (0.68–0.95)	0.83 (0.72–0.97)
Cameroon	0.80 (0.68–0.95)	0.84 (0.72–0.97)		0.80 (0.68–0.95)	0.83 (0.72–0.97)
Chad	0.80(0.68–0.95)	0.84 (0.73–0.97)		0.80(0.67–0.96)	0.84 (0.72–0.97)
Congo	0.79 (0.67–0.94)	0.82 (0.71–0.95)		0.79 (0.67–0.94)	0.82 (0.71–0.95)
Congo Dem Rep	0.80 (0.67–0.95)	0.84 (0.72–0.98)		0.80 (0.67–0.95)	0.84 (0.72–0.97)
Ethiopia	0.82 (0.69–0.97)	0.86 (0.74–1.00)		0.82 (0.69–0.97)	0.86 (0.74–0.99)
Gabon	0.80 (0.68–0.95)	0.85 (0.74–0.99)		0.80 (0.68–0.95)	0.85 (0.73–0.99)
Guinea	0.80 (0.67–0.94)	0.83 (0.72–0.97)		0.80 (0.67–0.95)	0.83 (0.71–0.96)
Kenya	0.80 (0.67–0.96)	0.82 (0.70–0.96)		0.80 (0.67–0.96)	0.82 (0.69–0.96)
Lesotho	0.86 (0.74–0.99)	0.89 (0.78–1.02)		0.86 (0.74–0.99)	0.87 (0.75–1.006)
Liberia	0.80 (0.67–0.95)	0.84 (0.72–0.97)		0.80 (0.67–0.95)	0.84 (0.72–0.97)
Madagascar	0.79 (0.66–0.94)	0.84 (0.71–0.99)		0.79 (0.66–0.94)	0.83 (0.71–0.96)
Malawi	0.79 (0.67–0.94)	0.84 (0.73–0.98)		0.79 (0.67–0.94)	0.84 (0.72–0.98)
Mali	0.80 (0.67–0.95)	0.85 (0.73–0.99)		0.80 (0.67-–0.95)	0.85 (0.73–0.99)
Mozambique	0.80 (0.68–0.95)	0.84 (0.72–0.97)		0.80 (0.68–0.95)	0.83 (0.72–0.97)
Namibia	0.78 (0.65–0.93)	0.82 (0.70–0.96)		0.78 (0.65–0.93)	0.82 (0.70–0.96)
Niger	0.79 (0.67–0.94)	0.83 (0.72–0.96)		0.79 (0.67–0.94)	0.83 (0.71–0.96)
Nigeria	0.76 (0.64–0.90)	0.80 (0.68–0.93)		0.76 (0.64–0.90)	0.78 (0.67–0.91)
Rwanda	0.83 (0.70–0.97)	0.84 (0.72–0.98)		0.83 (0.70–0.97)	0.84 (0.71–0.98)
Senegal	0.80 (0.67–0.94)	0.83 (0.72–0.96)		0.80 (0.67–0.94)	0.83 (0.71–0.96)
Sierra Leone	0.81 (0.68–0.96)	0.85 (0.73–0.99)		0.81 (0.68–0.96)	0.85 (0.72–0.99)
Swaziland	0.81 (0.69–0.96)	0.83 (0.70–0.97)		0.81 (0.69–0.96)	0.83 (0.70–0.97)
Tanzania	0.81 (0.68–0.96)	0.84 (0.71–0.98)		0.81 (0.68–0.96)	0.84 (0.71–0.98)
Uganda	0.82 (0.69–0.97)	0.86 (0.74–0.99)		0.82 (0.69–0.97)	0.85 (0.73–0.99)
Zambia	0.80 (0.67–0.97)	0.84 (0.71–0.99)		0.80 (0.67–0.97)	0.85 (0.73–0.98)
Zimbabwe	0.79 (0.67–0.94)	0.85 (0.73–0.99)		0.79 (0.67–0.94)	0.82 (0.73–0.98)

Note: *Fully adjusted model (personal and country covariates)

To further examine the impact specific countries may have on the results, as per the original sensitivity analysis, we created three subsets of countries. Using these subsets of countries, we performed an unadjusted and two adjusted logistic regressions as described in section 2.3. The first subset of countries comprised those with data before

and after PEPFAR implementation: Benin, Congo, Democratic Republic of Congo, Ethiopia, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Swaziland, Tanzania, Uganda, Zambia and Zimbabwe. The second subset used only the most recent survey for each country. The third subset used only countries with data from 1998 through at least 2007, which consisted of four focus countries (Kenya, Nigeria, Tanzania and Zambia) and three non-focus countries (Lesotho, Madagascar and Sierra Leone).

There are multiple differences between our results and the original results. We only highlight differences if the OR switched from significant to non-significant and vice versa. Our results further support the robustness of the main results, whereas the original results were mixed. We show our results and the original results notes in Table 6.

The last sensitivity analysis performed by the original authors repeated the main analysis, using a linear time trend instead of a binary indicator for the main variable of interest. A linear time trend estimates the effects of PEPFAR for each additional year of implementation as opposed to an overall effect. If PEPFAR had the intended effect, one would expect the linear time trend to be in agreement with the overall effect of PEPFAR in decreasing all-cause adult mortality. The original analysis investigated the mean effect of PEPFAR, whereas the linear time trend examined the effect for each year PEPFAR was implemented. We implemented the linear time trend by creating an interaction between the indicator variable for whether an individual lived in a focus country and a newly created *year_cat* variable. *Year_cat* had a value of 0 for all years before 2004. Starting in 2004, *year_cat* had a value of 1, and then in 2005 *year_cat* had a value of 2, and so on. This interaction represents the difference-in-difference for adult mortality as a linear time trend.

As with the previous sensitivity analysis, our results had multiple discrepancies when compared with the original results. Again, we only highlight results that changed from significant to non-significant and vice versa. Under the fully adjusted regression model, our results indicate that the main variable of interest is now borderline significant; however, there is some loss of power when using a linear time trend. The direction of the ORs are consistent, and we feel that our results still show the robustness of the findings. We display our results in Table 7, along with the original results and notes from the original authors.

It is unclear why our point estimates and CIs differed in the last two sensitivity analyses. We created our longitudinal data set from the cleaned data set provided by the authors. Additionally, the authors provided a Stata do-file that allowed us to verify our results in Stata. When performing the sensitivity analyses using Stata and the provided code, the results matched the results that we report.

Table 6: Replication and original results for relative odds of death associated with PEPFAR for subsets of countries and surveys

	Replication study			Original study		
	Unadjusted (95% CI, p-value)[a]	Adjusted with country covariates (95% CI, p-value)	Adjusted with personal and country covariates (95% CI, p-value)	Unadjusted (95% CI, p-value)[a]	Adjusted with country covariates (95% CI, p-value)	Adjusted with personal and country covariates (95% CI, p-value)
Subset (i)[b]	0.80 (0.68–0.95, 0.01)	0.83 (0.73–0.95, 0.01)	0.85 (0.74–0.99, 0.03)	0.83 (0.70–0.98, 0.03)	0.85 (0.71–1.02, 0.07)	0.87 (0.74–1.03, 0.08)
Subset (ii)[b]	0.90 (0.76–1.08, 0.27)	0.94 (0.79–1.13, 0.52)	0.94 (0.78–1.13, 0.51)	0.84 (0.71–0.99, 0.04)	0.85 (0.71–1.03, 0.08)	0.87 (0.75–1.04, 0.09)
Subset (iii)[b]	0.75 (0.60–0.93, 0.009)	0.73 (0.61–0.88, 0.001)	0.72 (0.59–0.86, 0.0004)	0.75 (0.60–0.93, 0.008)	0.74 (0.60–0.92, 0.007)	0.74 (0.60–0.92, 0.007)

Notes:

[a] Unadjusted model includes the main effect as well as country and year fixed effects. All CIs are calculated using robust standard errors clustered by country.

[b] These ORs represent the relative reduction in mortality among adults living in focus countries while PEPFAR was implemented compared with adults living in non-focus countries.

(i) Countries with data before and during PEPFAR's implementation

(ii) Only the most recent survey from each country

(iii) Countries with data from 1998 through at least 2007

Table 7: Replication and original results for sensitivity analysis using linear time trends

	Replication study			Original study		
	Unadjusted (95% CI, p-value)[a]	Adjusted with country covariates (95% CI, p-value)	Adjusted with country and personal covariates (95% CI, p-value)	Unadjusted (95% CI, p-value)[a]	Adjusted with country covariates (95% CI, p-value)	Adjusted with country and personal covariates (95% CI, p-value)
Adult death[b]	0.93 (0.86–0.99, 0.04)	0.94 (0.89–1.00, 0.05)	0.94 (0.88–1.00, 0.06)	0.96 (0.93–1.00, 0.03)	0.97 (0.94–1.00, 0.05)	0.97 (0.95–1.00, 0.04)
HIV prevalence (per additional 1%)		1.08 (1.00–1.16, 0.04)	1.07 (1.00–1.14, 0.04)		1.07 (1.01–1.13, 0.04)	1.07 (1.00–1.14, 0.06)
Non-PEPFAR assistance[c]		1.00 (0.98–1.01, 0.79)	1.00 (0.99–1.01, 0.95)		1.00 (0.98–1.02, 0.88)	1.00 (0.98–1.02, 0.93)
GDP per capita (per additional $1)		1.00 (1.00–1.00, 0.66)	1.00 (1.00–1.00, 0.54)		1.00 (1.00–1.00, 0.73)	1.00 (1.00–1.00, 0.78)
Government effectiveness (per 1 point increase)[d]		0.61 (0.39–0.94, 0.03)	0.57 (0.37–0.88, 0.01)		0.64 (0.45–0.96, 0.04)	0.60 (0.39–0.91, 0.02)
Sibling age (per year)			1.05 (1.04–1.05, <0.001)			1.05 (1.04–1.05, <0.001)
Residence in urban area			0.94 (0.89–1.00, 0.05)			0.93 (0.88–0.98, 0.01)
Education (per additional year)			0.99 (0.98–1.00, 0.01)			0.98 (0.97–0.99, <0.001)

Notes:

[a] All results are exponentiated coefficients on parameters in logistic regression models. Unadjusted model includes the main effect as well as country and year fixed effects. All CIs are calculated using robust standard errors clustered by country.

[b] These ORs represent the relative reduction in mortality among adults living in the focus countries while PEPFAR was implemented compared with adults living in non-focus countries.

[c] All development assistance for health from all donors minus US-funded HIV development assistance per capita.

[d] The index is centered at 0 and each 1 point represents 1 standard deviation with higher numbers representing greater government effectiveness.

2.6 Pure replication conclusions

The pure replication managed to reproduce the results of the original paper very well with some minor differences. The differences tended to affect the width of the CIs and did not change the significance of the findings. In some cases, our CIs were narrower than the original CIs, which strengthened the significance of the main findings. Our results and the PBR results were an exact match. Since we used SAS and Bendavid and colleagues (2012a) used Stata, we were confident that computational or programming environments were the cause of the discrepancies between our results and the results of the original paper.

The pure replication led to the same findings as the original authors and we found the findings to be robust using the sensitivity analysis, as described in the paper. Between 2004 and 2008, all-cause mortality decreased more in countries that implemented PEPFAR than in countries that did not implement PEPFAR. Additionally, an increase in HIV prevalence increased the odds of adult mortality. The difference in HIV prevalence between focus and non-focus countries was not statistically significant.

3. Additional analysis

Duber and colleagues (2010) examined the effects of PEPFAR in Africa using 14 health indicators from publicly available data. Health indicators for 46 African countries were collected for 2000 and 2006, using the World Health Organization database. As in Bendavid and colleagues (2012a), Duber and colleagues examined whether PEPFAR had a greater effect on decreasing all-cause mortality in focus countries, compared with non-focus countries. Duber and colleagues examined the median fractional change in all-cause mortality from 2000 to 2006. They did not find a statistically significant effect when comparing the median fractional change from 2000 to 2006 in all-cause mortality between focus and non-focus countries. These results contradict the findings from Bendavid and colleagues. However, Duber and colleagues and Bendavid and colleagues used different statistical methods for their analyses.

Bendavid and colleagues used a logistic regression model adjusted for country-level and individual-level covariates. We cannot directly use their methods on the data from Duber and colleagues[2], since the structures of the two data sets are different. The Bendavid and colleagues' method requires individual-level longitudinal data; however, the data used by Duber and colleagues was population-level longitudinal data. Therefore, we will use the Bendavid and colleagues' methods with the Duber and colleagues study period to see if the results are consistent.

The Duber and colleagues (2010) methodology used a population-averaged inference, with data collected at the country level, and considered 2000 as the baseline measurement year. For the 14 health indicators, they calculated the percent change from baseline, aggregated by focus and non-focus country. They used a Wilcoxon rank sum test to compare the fractional change between focus and non-focus countries. Duber and colleagues found that only one health indicator was statistically different between focus and non-focus countries when comparing median fractional change from baseline. They

[2] See http://www.who.int/whosis/data/Search.jsp.

then limited the analysis to the 29 countries in the African region that Bendavid and Bhattacharya (2009) used, which ultimately did not change their results.

Duber and colleagues used a non-parametric approach in analyzing the data. The methodology did not require any distribution assumptions. The lack of distribution assumptions eliminates the possibility of fitting an incorrect model. However, non-parametric tests are less powerful than parametric tests. Thus, a non-parametric test is less likely to identify a significant difference than its parametric counterpart when that difference is small.

Bendavid and colleagues (2012a) used a parametric approach with data gathered at the individual level. Specifically, they used logistic regression with a difference-in-difference study design, which is the parametric equivalent to the Duber and colleagues (2010) methodology. However, data collection occurred at different levels for Bendavid and colleagues and Duber and colleagues. Bendavid and colleagues adjusted their model for country-level and individual-level covariates, and relaxed the assumption of independent and identically distributed errors by calculating robust standard errors clustered by country (i.e. individuals living in the same country are correlated). The Bendavid and colleagues study required more information than the Duber and colleagues study, but the Bendavid and colleagues' methodology provided a more powerful statistical test.

3.1 Measurement and estimation analysis

3.1.1 Methods

In our measurement and estimation analysis (Brown et al. 2014), we used the statistical methods described in Bendavid and colleagues (2012a) and the pure replication from Section 2.1.2. We limited the data set to a subset of the data set used in the pure replication, examining only observations between 2000 and 2006 (inclusive). As in the pure replication, we compared characteristics of the focus and non-focus countries with each other, using a two-tailed t-test. Next, we examined the difference in the odds of all-cause mortality between focus and non-focus countries, using a logistic regression with the difference-in-difference indicator as described in Section 2.1.2. We used the same three regression models as in the pure replication – unadjusted, adjusted for country covariates, and adjusted for country and individual level covariates. Lastly, we performed a sensitivity analysis similar to the sensitivity analysis carried out in the pure replication. We performed a logistic regression on the unadjusted and adjusted model, leaving out any one country at a time, including only countries that had all data for 2000 to 2006, and using a linear time trend, as opposed to a dichotomous indicator, for PEPFAR implementation.

3.1.2 Results

Group comparisons show that the characteristics of the focus and non-focus countries are similar in most respects, except population and HIV aid per country. The mean population of the focus countries was statistically larger than the mean population of the non-focus countries, regardless of year examined. Additionally, in 2000 and 2006, focus countries were receiving significantly more aid (in millions of dollars) than non-focus countries. In 2000, the mean HIV aid in focus countries was US$16.6 million (95% CI [8.3, 24.9]), rising to US$125.8 million (95% CI [97, 154.6]) in 2006. Non-focus countries had an increase in aid from 2000 to 2006, but the increase was quite small compared

with focus countries. In non-focus countries, mean aid changed from US$4.8 million (95% CI [1.2, 8.3]) to US$17.6 million (95% CI [8, 27.1]). However, when examining aid per adult with HIV, the difference between focus and non-focus countries was not significant, regardless of the year. Mean aid per adult living with HIV in 2000 was US$10 million in focus countries (6.2, 13.7) versus US$26.9 million in non-focus countries (–2.3, 56.1). Furthermore, group comparisons show that mean HIV prevalence among adults 15 to 49 years old in 2000 was not statistically different between focus and non-focus countries (p-value 0.67). The mean prevalence of HIV in focus countries was 8.1 percent (95% CI [4.8, 11.5]) as opposed to 6.7 percent (95% CI [2.1, 11.3]) for non-focus countries. See Table 8 for the remaining results.

Table 8: Comparison of focus and non-focus countries with each other for 2000 and 2006

		Mean (95% CI)		
Parameter		**Focus countries**	**Non-focus countries**	**p-value[a]**
Population, millions	2000	34.8 (6.2 to 63.3)	10.4 (4.9 to 16)	0.02
	2006	40.9 (7.6 to 74.2)	12.2 (5.5 to 18.8)	0.01
HIV prevalence among adults	2000	8.1 (4.8 to 11.5)	6.7 (2.1 to 11.3)	0.67
15-49 years old, %	2006	7.7 (3.9 to 11.5)	6.1 (2.1 to 10.1)	0.59
GDP per capita, constant $	2000	480.8 (–2.5 to 964)	609.6 (143.7 to 1,075.5)	0.71
	2006	586.9 (3 to 1170.8)	634.5 (177.7 to 1,091.2)	0.89
HIV aid per country, millions of $	2000	16.6 (8.3 to 24.9)	4.8 (1.2 to 8.3)	0.002
	2006	125.8 (97 to 154.6)	17.6 (8 to 27.1)	<0.0001
HIV aid per adult with HIV, $	2000	10 (6.2 to 13.7)	26.9 (–2.3 to 56.1)	0.40
	2006	104.8 (38.9 to 170.7)	66.5 (42 to 91)	0.15
Urban residence, %	2000	24.8 (16.6 to 33)	34.5 (26 to 43)	0.13
	2006	27.3 (18.5 to 36.1)	37.1 (28.3 to 45.9)	0.15

Note: [a] p-values provided from 2-tailed *t* tests on the data for the specified year in the focus countries compared with the non-focus countries.

Our unadjusted regression analysis indicates that the odds of all-cause mortality after PEPFAR implementation (2003) for individuals living in focus countries is 0.86 (95% CI [0.74–1.00]; p-value 0.04) compared with people living in non-focus countries. This statistically significant reduction in the odds of adult mortality holds in the adjusted model with country covariates but does not hold for the adjusted model with country and individual-level covariates. When examining the adjusted model with country and individual-level covariates, the OR of death is 0.87 (95% CI [0.75–1.01]; p-value 0.07). Additionally, regardless of the model, an increase in HIV prevalence is not associated with a change in the OR of all-cause adult mortality. Of the remaining covariates in the fully adjusted model, only sibling age and education (per additional year) are associated with all-cause adult mortality. For each year increase in sibling age, the OR of death is 1.05 (95% CI [1.04–1.05]; p-value <0.01). Each additional year of education is protective, with an associated OR of death of 0.99 (95% CI [0.98–1.00]; p-value <0.01). Table 9 contains the full results from all regression models.

We further examined the sensitivity of these results by leaving any one country out from the regression models. The results appear robust. For the unadjusted model, the direction and magnitude of the point estimates were consistent and the majority of the point estimates were statistically significant. The non-significant point estimates were non-significant by a small margin. Results for the fully adjusted regression model are

similar, except that most point estimates remained statistically non-significant, as in the original results. Table 10 contains the full results of the leave-one-out analysis.

Table 9: Regression models estimating the odds of death in study adults in focus versus non-focus countries for 2000 – 2006

	Unadjusted OR (95% CI)[a]	**p-value**	**Adjusted OR with country covariates (95% CI)**	**p-value**	**Adjusted OR with country and personal covariates (95% CI)**	**p-value**
Adult death[b]	0.86 (0.74–1.00)	0.04	0.86 (0.76–0.97)	0.02	0.87 (0.75–1.01)	0.07
HIV prevalence (per additional 1%)			1.04 (0.98–1.10)	0.16	1.05 (0.99–1.10)	0.08
Non-PEPFAR assistance[c]			0.99 (0.97–1.00)	0.14	0.99 (0.97–1.00)	0.09
GDP per capita (per additional $1)			1.00 (1.00–1.00)	0.28	1.00 (1.00–1.00)	0.50
Government effectiveness (per 1 point increase)[d]			0.81 (0.60–1.08)	0.16	0.76 (0.53–1.08)	0.12
Sibling age (per year)					1.05 (1.04–1.05)	0.00
Residence in urban area[e]					0.95 (0.88–1.02)	0.18
Education (per additional year)[e]					0.99 (0.98–1.00)	0.00
Recall					0.97 (0.94–1.00)	0.05

Notes: [a] All results are exponentiated coefficients on parameters in logistic regression models. Unadjusted model includes the main effect as well as country and year fixed effects. All CIs are calculated using robust standard errors clustered by country.

[b] These ORs represent the relative reduction in mortality among adults living in the focus countries while PEPFAR was implemented, compared with adults living in non-focus countries.

[c] All development assistance for health from all donors minus US-funded HIV development assistance per capita.

[d] The index is centered at 0, and each 1 point represents 1 standard deviation with higher numbers representing greater government effectiveness.

[e] These variables are characteristics of the index women rather than the sibling. The residence status and educational status of the sibling are not known.

Table 10: Leave-one-out country analysis for years 2000 – 2006

Country left out	Unadjusted (95% CI)	Adjusted with country and individual covariates (95% CI)
Benin	0.86 (0.74–0.99)	0.86 (0.75–1.00)
Burkina Faso	0.86 (0.74–1.00)	0.87 (0.75–1.01)
Cameroon	0.86 (0.74–1.00)	0.87 (0.75–1.01)
Chad	0.86 (0.74–1.00)	0.87 (0.75–1.01)
Congo	0.85 (0.74–0.99)	0.86 (0.74–1.00)
Congo Dem Rep	0.86 (0.74–1.00)	0.87 (0.75–1.01)
Ethiopia	0.88 (0.76–1.02)	0.90 (0.79–1.04)
Gabon	0.86 (0.74–1.00)	0.87 (0.75–1.01)
Guinea	0.86 (0.74–0.99)	0.86 (0.75–1.00)
Kenya	0.85 (0.72–1.00)	0.84 (0.72–0.99)
Lesotho	0.88 (0.76–1.03)	0.90 (0.76–1.06)
Liberia	0.86 (0.74–0.99)	0.86 (0.75–1.00)
Madagascar	0.88 (0.76–1.02)	0.90 (0.78–1.05)
Malawi	0.85 (0.74–0.99)	0.88 (0.76–1.02)
Mali	0.86 (0.74–1.00)	0.87 (0.75–1.02)
Mozambique	0.86 (0.74–1.00)	0.87 (0.75–1.01)
Namibia	0.85 (0.72–1.00)	0.85 (0.73–0.99)
Niger	0.85 (0.74–0.99)	0.86 (0.75–1.00)
Nigeria	0.81 (0.71–0.92)	0.83 (0.70–0.98)
Rwanda	0.88 (0.76–1.01)	0.89 (0.76–1.04)
Senegal	0.86 (0.74–0.99)	0.87 (0.75–1.01)
Sierra Leone	0.87 (0.75–1.01)	0.88 (0.76–1.03)
Swaziland	0.87 (0.75–1.01)	0.87 (0.74–1.02)
Tanzania	0.86 (0.73–1.01)	0.86 (0.74–1.01)
Uganda	0.88 (0.76–1.02)	0.90 (0.77–1.04)
Zambia	0.87 (0.74–1.02)	0.86 (0.72–1.03)
Zimbabwe	0.84 (0.73–0.97)	0.90 (0.78–1.03)

When examining only the countries where data is available for all years from 2000 to 2006, the original findings hold. The unadjusted and adjusted with country covariates regression models indicate that PEPFAR is statistically protective against all-cause adult mortality, whereas there is not a significant association between all-cause adult mortality and PEPFAR when examining the regression model with country and individual-level covariates. When including only countries with data from 2000 to 2006, the CIs of the models for the effects of PEPFAR implementation are narrower than the results when all countries are included for 2000 to 2006.

For the unadjusted model, the results narrow from 0.86 (95% CI [0.74, 1.00]) to 0.87 (95% CI [0.74, 0.95]). With the fully adjusted model, the results narrow from 0.87, 95 percent CI (0.75, 1.01) to 0.88 (95% CI [0.78, 1.01]). HIV prevalence remains a statistically non-significant indicator of all-cause mortality with an OR of 1.06 (95% CI [0.99, 1.14]) when examining countries with complete data for 2000 to 2006. As with the original findings, each additional year of education is protective, with an OR of 0.99 (95% CI [0.98, 1.00]), and sibling's age (per year) increased the odds of all-cause mortality, with an OR of 1.05 (95% CI [1.04, 1.05]). However, in the original results for 2000 to 2006, government effectiveness was not associated with all-cause mortality (OR 0.76;

95% CI [0.53, 1.08]). When limiting the data set to countries with data for all years from 2000 to 2006, government effectiveness is associated with a decreased odds of all-cause mortality (OR 0.59; 95% CI [0.46, 0.76]).

See Table 11 for our full results. The focus countries are Kenya, Namibia, Nigeria, Tanzania and Zambia; non-focus countries are Democratic Republic of Congo, Lesotho, Liberia, Madagascar, Sierra Leone and Swaziland.

Table 11: Relative odds of death associated with PEPFAR for countries that contain data for 2000–2006

	Unadjusted OR (95% CI)[a]	p-value	Adjusted OR with Country Covariates (95% CI)	p-value	Adjusted OR With Country and Personal Covariates (95% CI)	p-value
Adult death[b]	0.84 (0.74 - 0.95)	0.01	0.88 (0.79 - 0.99)	0.04	0.88 (0.78 - 1.01)	0.06
HIV prevalence (per additional 1%)			1.12 (1.03 - 1.21)	0.01	1.06 (0.99 - 1.14)	0.11
Non-PEPFAR assistance[c]			0.98 (0.97 - 1.00)	0.03	0.99 (0.97 - 1.00)	0.02
GDP per capita (per additional $1)			1.00 (1.00 - 1.00)	0.69	1.00 (1.00 - 1.00)	0.78
Government effectiveness (per 1 point increase)[d]			0.68 (0.56 - 0.82)	0	0.59 (0.46 - 0.76)	0.00
Sibling age (per year)					1.05 (1.04 - 1.05)	0.00
Residence in urban area[e]					0.99 (0.87 - 1.12)	0.85
Education (per additional year)[f]					0.99 (0.98 - 1.00)	0.01
Recall					0.96 (0.93 - 0.99)	0.01

Notes: [a] All results are exponentiated coefficients on parameters in logistic regression models. Unadjusted model includes the main effect as well as country and year fixed effects. All CIs are calculated using robust standard errors clustered by country.

[b] These ORs represent the relative reduction in mortality among adults living in the focus countries while PEPFAR was implemented compared with adults living in non-focus countries.

[c] All development assistance for health from all donors minus US-funded HIV development assistance per capita.

[d] The index is centered at 0 and each 1 point represents 1 standard deviation, with higher numbers representing greater government effectiveness.

[e] These variables are characteristics of the index women rather than the sibling. The residence status and educational status of the sibling are not known.

The last sensitivity analysis performed was a linear time trend. This analysis may be underpowered, but it will show if the general trend holds – i.e. if PEPFAR is associated with a decrease in all-cause adult mortality. All three regression models indicate that PEPFAR is associated with a reduction in the odds of all-cause mortality; however, these point estimates are all non-significant, though they all showed consistent direction and magnitude. As with the original 2000 to 2006 results, HIV prevalence was not associated with all-cause mortality (OR 1.05; 95% CI [0.99, 1.11]). Additionally, sibling age (per year) remained associated with all-cause mortality (OR 1.05; 95% CI [1.04, 1.05]) and education (per additional year) offered protection against all-cause mortality (OR 0.99, 95% CI [0.98, 1.00]). Table 12 contains the full results of the linear time trend.

Table 12: Sensitivity analysis using linear time trends for 2000 – 2006

	Unadjusted OR (95% CI, p-value)[a]	**Adjusted OR with country covariates (95% CI, p-value)**	**Adjusted OR with country and personal covariates (95% CI, p-value)**
Adult death[b]	0.95 (0.87-1.03, 0.19)	0.95 (0.88–1.03, 0.21)	0.96 (0.88–1.04, 0.31)
HIV prevalence (per additional 1%)		1.05 (0.98–1.12, 0.16)	1.05 (0.99–1.11, 0.08)
Non-PEPFAR assistance[c]		0.99 (0.97–1.00, 0.17)	0.99 (0.98–1, 0.10)
GDP per capita (per additional $1)		1.00 (1.00–1.00, 0.52)	1.00 (1.00–1.00, 0.77)
Government effectiveness (per 1 point increase)[d]		0.78 (0.57–1.07, 0.13)	0.74 (0.51–1.07, 0.10)
Sibling age (per year)			1.05 (1.04–1.05, 0.00)
Residence in urban area[e]			0.95 (0.88–1.02, 0.18)
Education (per additional year)[e]			0.99 (0.98–1.00, 0.00)
Recall			0.97 (0.94–1.00, 0.04)

Notes: [a] All results are exponentiated coefficients on parameters in logistic regression models. Unadjusted model includes the main effect as well as country and year fixed effects. All CIs are calculated using robust standard errors clustered by country.
[b] These ORs represent the relative reduction in mortality among adults living in the focus countries while PEPFAR was implemented compared with adults living in non-focus countries.
[c] All development assistance for health from all donors minus US-funded HIV development assistance per capita.
[d] The index is centered at 0 and each 1 point represents 1 standard deviation, with higher numbers representing greater government effectiveness.
[e] These variables are characteristics of the index women rather than the sibling. The residence status and educational status of the sibling are not known.

4. Discussion

In this paper, we have performed a replication and comparative analysis on the paper *HIV Development Assistance and Adult Mortality in Africa* by Bendavid and colleagues (2012a). Using the paper and the electronic appendix as a guide, we were able to replicate the results, excluding the subnational analysis of district-level data for Tanzania and Rwanda, as we were unable to obtain the district-level data for these two countries.

In the comparative analysis, we aimed to examine how methodology would affect the results of Duber and colleagues' (2010) paper, in which the authors found no statistical

evidence that PEPFAR influenced the adult mortality rate when comparing focus and non-focus countries (p-value 0.348). These results are not in agreement with an earlier paper by Bendavid and Bhattacharya (2009) and the current paper by Bendavid and colleagues (2012a). Bendavid and Bhattacharya (2009) and Duber and colleagues (2010) used similar methods. We aimed to examine if methodology would affect Duber and colleagues' results. Using the method from Bendavid and colleagues (2012a) and the study period from Duber and colleagues (2010), we found that PEPFAR had a significant impact on all-cause adult mortality when examining the effects of PEPFAR using an unadjusted and adjusted logistic regression with country-level covariates. The unadjusted model, using Bendavid and colleagues' methods, is the parametric equivalent to non-parametric methods of Duber and colleagues.

The fully adjusted model (with country and individual-level covariates) was borderline non-significant when examining the association between PEPFAR and all-cause adult mortality. The results for the fully adjusted model are not surprising, since the sample size was smaller, which would widen the CIs. However, the point estimate for the 2000 to 2006 data does fall within the CI of the original analysis and is close to the original point estimate in the study by Bendavid and colleagues (2012a). Duber and colleagues (2010) assert that a possible explanation for the non-significant results was the fact the PEPFAR activity requires a sufficient amount of time for the effects to accumulate in the focus countries. If this assertion were true, then we would expect more varied point estimates in the fully adjusted models using Bendavid and colleagues' method for the restricted data (2000 to 2006) and original analysis (1998 to 2008).

However, the two studies use different focus countries for analysis. Duber and colleagues (2010) included three additional countries (Botswana, South Africa and Cote d'Ivoire) that Bendavid and colleagues (2012a) did not include. Bendavid and colleagues state that these countries were not included due to unsuitable data sources and that Botswana and South Africa have particularly high HIV-prevalence that could affect the results. Duber and colleagues state that it appears that South Africa showed worsening health indicators during the study period. It is possible that not including South Africa as a focus country biased the results.

It remains unclear whether the methods or included focus countries are causing the discrepancy between the results of Duber and colleagues (2010) and Bendavid and colleagues (2012a). It is clear that in the focus countries used by Bendavid and colleagues, PEPFAR was associated with a decrease in all-cause adult mortality in a short time frame.

5. Limitations

The limitations that Bendavid and colleagues (2012a) reported still apply. They mention two limitations: the fact that the survey used relies on a sibling's recall of deaths in the family, and the potential for these surveys to be biased, because families with high mortality rates tend to be underrepresented. Additionally, because Botswana, South Africa and Cote d'Ivoire were not included in the analysis, the analysis cannot be generalized to all focus countries. The authors excluded these countries because of unsuitable data. Lastly, the original authors recognize that there could be preexisting

trends that could have affected all-cause adult mortality in the focus and non-focus countries, unrelated to PEPFAR implementation.

In addition to the limitations of the original paper, the replication has limitations of its own. Having a data set supplied by the original authors expedites the replication process, but limits the scope of any additional analyses. Furthermore, we cannot replicate any of the data cleaning that the original authors performed. Finally, the choice of variables to include/exclude in the data set is the sole discretion of the original authors.

6. Conclusions

Duber and colleagues (2010) and Bendavid and colleagues (2012a) are not in agreement about the effectiveness of PEPFAR's reducing all-cause adult mortality. Our replication study supports the findings of Bendavid and colleagues. Our additional analysis was unable to answer whether the methods used by Duber and colleagues and Bendavid and colleagues were the cause of the discordant results. Both papers mention South Africa (not included in Bendavid and colleagues' data set) as a possible linchpin in their results. The study period did not have a large impact on the results when using the Bendavid and colleagues' data set.

We cautiously agree with the Bendavid and colleagues' (2012a) findings that PEPFAR is associated with a reduction in all-cause adult mortality in focus countries compared to non-focus countries for 2004 to 2008. We are unable to state whether the reduction in all-cause mortality is because of a reduction in HIV mortality or some other mechanism. Not including Botswana and South Africa in the analysis could affect the significance of the main findings, as both countries have a high prevalence of HIV. If PEPFAR is reducing all-cause mortality by reducing HIV-related mortality, then including Botswana and South Africa may strengthen Bendavid and colleagues' main findings.

Appendix A: Summary of differences between pure replication and original findings

This appendix contains a summary table of the differences found between the original paper and the replication analysis during the pure replication for the convenience of the reader.

Table A1: Summary of the pure replication

	Discrepancy	Replication	Original	Comments
Table 1	Survey fieldwork dates	Several		Survey fieldwork date range is either one month longer or shorter than the original paper
Table 2	HIV aid per country 1998 point estimate for focus countries	7.3	6.3	
	p-value for t test - HIV aid per country 1998	0.0407	0.16	Explained by different point estimates
	Confidence limit differences	Several		Rounding/Typographical?
Figure 2	Point estimates and confidence interval differences	Several		Different standard population could have been used.
Figure 3	Years included in analysis	Several		Different data sets could have been used.
Table 3	Point estimate and confidence interval differences	Several		Most appear to be caused by differences in rounding; does not change significance of main results
Table 4	Point estimates of predictions	Several		Coefficients of logistic regression model slightly different
Table 5	Point estimate and confidence interval differences	Several		Most appear to be caused by differences in rounding; does not change the robustness of main results
Table 6	Point estimate and confidence interval differences	Several		Does not change the robustness of main results
Table 7	Point estimate and confidence interval differences	Several		Does not change the robustness of main results

Appendix B: Push-button replication final report – *Journal of American Medical Association*: Eran Bendavid

Section 1: Basic information

- Original paper citation: Bendavid, E, Holmes, CB, Bhattacharya, J and Miller, G, 2012. *HIV development assistance and adult mortality in Africa*. *JAMA*, 307(19), pp.2060–2067.
- Original authors and contact email addresses:
  - Eran Bendavid, Stanford University Division of General Medical Disciplines and Center for Health Policy and Center for Primary Care and Outcomes Research, ebd@stanford.edu
  - Charles B. Holmes, Office of the US Global AIDS Coordinator, US Department of State, Washington, DC, holmescb@state.gov
  - Jay Bhattacharya, Stanford University Center for Primary Care and Outcomes Research, jay@stanford.edu
  - Grant Miller, Stanford Center for International Development, ngmiller@stanford.edu
- PBR researcher: Jiangtao Luo and Nicholas A. Hein
- List of materials obtained: a Stata file and three data sets used for analyses
- Classification: *Incomplete and Minor differences*

Section 2: Replication process

A replication unified code of do-file and three data files (named *covout_narrow*, *wgi govt effectiveness* and *WideMerge_v1*) were obtained from the first author of the paper, Dr. Bendavid. Before executing the do-file, the directory was appropriately changed and logging was included. The do-file "replication unified code" was used to verify Tables 1 through 3 and Figures 1 and 2. We could not reproduce Figure 3 from the original code.

All other tables and figures were reproduced by the do-file, except Table 1 and Table 2 were incomplete. In Table 1, we had to create the code for the survey fieldwork dates. The century-month code date of interview was included in the dataset. We converted the century-month code date to a month and year and tabulated these dates by country. In Table 2, we could not verify antiretroviral coverage, since this variable was in a different data set that was unavailable. Additionally, in Table 2, we could not reproduce the column *other Sub-Saharan countries*, as there was no data provided for these countries. We could not reproduce these two tables from the original code, so we classified them as *incomplete,* but we could reproduce them by writing our own code.

Section 3: PBR classification justification

Minor differences and incomplete: There were some minor differences in point estimates and confidence intervals in Tables 2 and 3. All p-values led to the same interpretation as the original study except one, the t-test for *HIV aid per country, millions of $ (1998)* in Table 2, which was significant in the PBR but non-significant in the original paper. In Table 3, which displays the main results of the paper, point estimates and CIs never differed by more than three-hundredths of a unit. In Table 1, survey fieldwork dates varied by a month for many of the countries. Figure 2, produced by the PBR, does not

match the original paper. The general trend and CIs seem to match; however, the point estimates do not appear to coincide. Lastly, there were parts of the tables that we could not reproduce (see Section 2), and we could not reproduce Figure 3 in its entirety, since it was not data driven. Figure 3 is a sub-analysis and therefore does not affect the interpretation of the main findings – i.e. all-cause mortality declined more in PEPFAR focus countries relative to non-focus countries.

Appendix C: Push-button replication comparison tables and descriptions

A) Not eligible table or figure: Figure 3, “Adult Mortality Trends in Tanzania Separated by PEPFAR activity, 1998-2008,” was not subject to replication since they were not data-driven.

B) The following is a description of the differences between the PBR and the original results. We stratified differences by table and table classification is italicized after the description.

Table 1: Study countries, participants and group designation *(minor differences and incomplete)*

- We converted the century-month code date to a month and year and tabulated these dates by country.
- Survey fieldwork dates differed by plus/minus one month for most countries.

Table 2: Comparison of focus countries and non-focus countries with each other and with nonstudy Sub-Saharan African countries (*minor differences and incomplete*)

- Different rounding for the upper CI for *HIV prevalence* for focus countries for 1998 and 2008 (not highlighted).
- CIs do not match for *GDP* for focus countries for 1998 and 2008 and cannot be explained by a rounding error.
- *HIV aid per country* point estimates and CIs do not match for focus and non-focus countries for 1998 and led to a significant p-value at the $\alpha = 0.05$ level.
- Rounding error of upper CI for *Urban residence* of focus countries in 1998 (not highlighted).
- *Antiretroviral coverage* data was not provided.
- Data for *Other Sub-Saharan Countries* was not provided.

Table 3: Regression models estimating the OR of death in study adults in focus countries versus non-focus countries (*minor differences*)

- Upper CI for adult death in the fully adjusted model was 0.97 in the PBR versus 0.99 in the original paper.
- Non-PEPFAR assistance point estimates and CIs differed in both adjusted models but did not change statistical significance at the $\alpha = 0.05$ level.
- The remaining differences are of magnitude of one-hundredth of a unit and are not highlighted.

Figure 1: Trends in development assistance for HIV for focus countries and non-focus countries: Mean per-country assistance in 2008 US$, 1998–2008 *(comparable replication*)

- Exact match to original paper.

Figure 2: Age-adjusted adult mortality trends in the focus and non-focus countries, 1998–2008 *(minor differences*)

- Point estimates appear not to match for all years, but overall trend seems the same.

C) PBR Tables

	Comparable
	Minor differences
	Major differences
	No access to data
	Information not reported in table

Table C1: Study countries, participants and group designation (PBR)

Country	Survey fieldwork dates	No. of unique adults	Observations, no.	No. of deaths
Ethiopia	2–6/2000, 4–8/2005	96,980	391,835	2,596
Kenya	4–9/2003, 11/2008–3/2009[a]	73,580	491,521	2,971
Mozambique	8/2003–1/2004	41,103	189,752	1,367
Namibia	9–12/2000, 11/2006–3/2007	64,382	340,338	3,303
Nigeria[b]	6–11/2008	122,815	1,020,435	4,590
Rwanda	5–11/2000, 2–8/2005	74,818	316,179	2,943
Tanzania	10/2004–2/2005, 12/2009–5/2010[a]	83,992	615,367	2,993
Uganda	9/2000–3/2001, 5–10/2006	62,132	301,234	2,856
Zambia	11/2001–6/2002, 4–10/2007	60,014	328,837	4,228
Benin	8–11/2006	64,463	449,155	1,703
Burkina Faso	1–3/1999, 6–12/2003	55,416	206,068	1,123
Cameroon	2–9/2004	41,422	222,637	1,550
Chad	7–12/2004	20,891	111,943	736
Congo	7–11/2005	28,305	175,576	1,323
Congo Dem Rep	1–9/2007	38,637	295,800	1,887
Gabon	7/2000–2/2001	22,083	43,671	210
Guinea	4–8/1999, 2–6/2005	44,848	177,877	977
Lesotho	9/2004–2/2005, 10/2009–1/2010[a]	47,185	334,908	4,428
Liberia	12/2006–4/2007	23,052	178,489	842
Madagascar	11/2003–6/2004, 11/2008–7/2009[a]	107,869	844,146	3,509
Malawi	7–11/2000, 1/2004, 9/2004–2/2005	84,041	305,436	3,945
Mali	1–6/2001, 3–12/2006	92,775	470,612	2,161
Niger	1–6/2006	34,858	243,442	942
Senegal	1–6/2005	55,881	347,114	1,096
Sierra Leone	4–6/2008, 8/2008	19,675	165,810	891
Swaziland	6/2006–3/2007	18,458	128,135	1,739
Zimbabwe	8–12/1999, 8/2005–2/2006, 4/2006	58,937	247,359	3,394

Notes: [a] The fieldwork (data collection) for these surveys was carried out after the end of the study period at the end of 2008 through May 2010. However, the study's last year of analysis was 2008 because surveys from only two countries (Lesotho and Tanzania) contain data for all of 2009, and there was no data from any of the study countries for 2010. Each fieldwork date range represents an individual survey.

[b] Another DHS with data on adult mortality was conducted in Nigeria in 1999. However, the survey's data quality was shown to be poor, and we did not include the survey in analysis.

Table C2: Comparison of focus countries and non-focus countries with each other and with non-study Sub-Saharan African countries (PBR)

		Mean (95% CI)				
Parameter		Focus countries	Non-focus countries	P value[a]	Other Sub-Saharan Countries[b]	P value[c]
Population, millions	1998	33.6 (5.1 to 62.1)	9.8 (4.5 to 15)	0.02	10.3 (2.9 to 17.7)	0.35
	2008	43.4 (7.8 to 79)	12.8 (5.9 to 19.8)	0.01	13.6 (4.8 to 22.4)	0.3
HIV prevalence among adults 15–49 y old, %	1998	8.1 (5.0 to 11.3)	6.5 (2 to 11)	0.62	5.0 (1.1 to 8.9)	0.41
	2008	7.5 (3.9 to 11.1)	5.8 (1.9 to 9.8)	0.57	5.0 (1.0 to 9.1)	0.56
GDP per capita, constant $	1998	471.3 (–1.4 to 944.1)	641.8 (98.7 to 1,184.8)	0.67	767.2 (152.6 to 1,381.9)	0.58
	2008	629.1 (15.1 to 1,243.1)	654.5 (180.3 to 1,128.8)	0.95	995.1 (148.4 to 1,841.8)	0.34
HIV aid per country, millions of $	1998	7.3 (1.0 to 13.6)	2.0 (–0.2 to 4.3)	0.04	1.8 (0.9 to 2.7)	0.25
	2008	240.5 (168.7 to 312.3)	24.6 (10.2 to 39.1)	<0.0001	63.1 (–6.1 to 132.4)	0.37
HIV aid per adult with HIV, $	1998	3.8 (1.8 to 5.7)	6.3 (0.2 to 12.3)	0.55	18.4 (0.6 to 35.1)	0.11
	2008	171 (75.8 to 266.3)	76.9 (54.9 to 98.9)	0.007	113.1 (40.7 to 185.6)	0.89
Antiretroviral coverage, %	2003	2.6 (–2.1 to 7.4)	1.9 (–3.1 to 7.2)	0.46	1.9 (–2.6 to 6.5)	0.68
	2008	55.6 (38.3 to 73.6)	28.6 (16.1 to 41.2)	0.04	39.6 (26.6 to 52.5)	0.51
Urban residence, %	1998	24 (15.8 to 32.3)	33.7 (25.4 to 42)	0.13	38.1 (29.6 to 46.5)	0.45
	2008	28.1 (19 to 37.2)	38 (29.1 to 46.9)	0.15	42.4 (33.5 to 51.3)	0.43

Notes: [a] P values provided from 2 tailed t tests on data for the specified year in the focus countries compared with the non-focus countries.

[b] Sub-Saharan countries not included in this study are Angola, Central African Republic, Burundi, Djibouti, Eritrea, Somalia, Sudan, Botswana, South Africa, Cote d'Ivoire, Ghana, Guinea-Bissau, the Gambia and Togo.

[c] P values provided for the comparison between the aggregated estimates for all 27 study countries (focus and non-focus countries) and the Sub-Saharan African countries not included in the study. This comparison provides a comparison between the countries excluded from the study and those included along the observed metrics.

Table C3: Regression models estimating the odds ratio of death in study adults in focus countries versus non-focus countries (PBR)

	Unadjusted OR (95% CI)[a]	P value	Adjusted OR with country Covariates (95% CI)	P value	Adjusted OR with country and personal covariates (95% CI)	P value
Adult death[b]	0.80 (0.68–0.95)	0.01	0.82 (0.72–0.95)	0.01	0.84 (0.72–0.97)	0.02
HIV prevalence (per additional 1%)			1.07 (1.00–1.15)	0.04	1.07 (1.01–1.14)	0.03
Non-PEPFAR assistance[c]			1.00 (0.98–1.01)	0.87	1.00 (0.99–1.01)	0.89
GDP per capita (per additional $1)			1.00 (1.00–1.00)	0.82	1.00 (1.00–1.00)	0.63
Government effectiveness (per 1 point increase)[d]			0.62 (0.40–0.95)	0.03	0.58 (0.38–0.89)	0.01
Sibling age (per year)					1.05 (1.04–1.05)	<0.001
Residence in urban area[e]					0.94 (0.89–1.00)	0.05
Education (per additional year)					0.99 (0.98–1.00)	0.01
Recall (interval between survey and observation, per year)					0.97 (0.95–0.99)	0.01

Notes: [a] All results are exponentiated coefficients on parameters in logistic regression models. Unadjusted model includes the main effect as well as country and year fixed effects. All CIs are calculated using robust standard errors clustered by country.

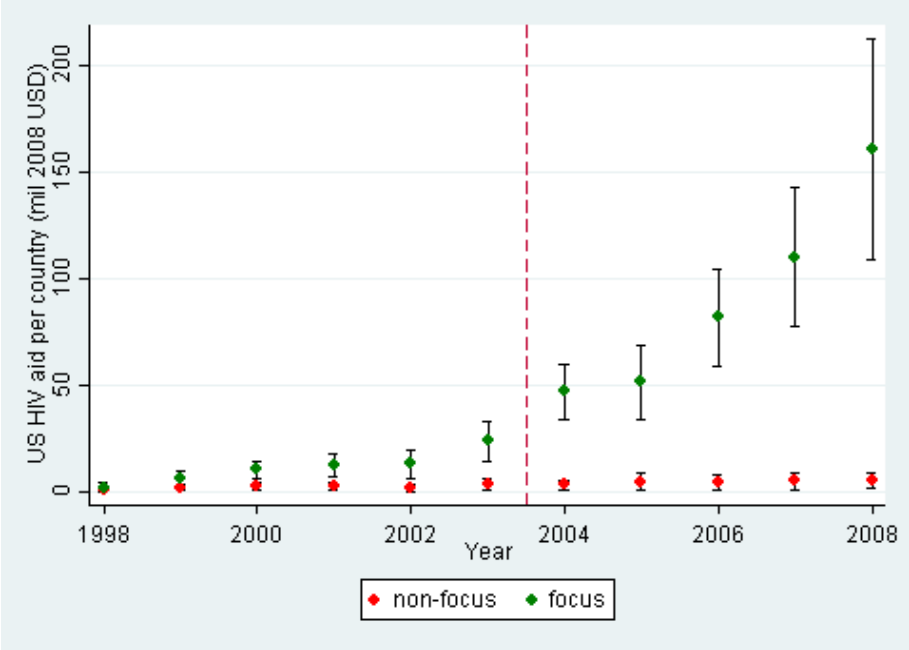
[b] These ORs represent the relative reduction in mortality among adults living in the focus countries while PEPFAR was implemented compared with adults living in non-focus countries.

[c] All development assistance for health from all donors minus US-funded HIV development assistance per capita.

[d] The index is centered at 0 and each 1-point represents 1 standard deviation, with higher numbers representing greater government effectiveness.

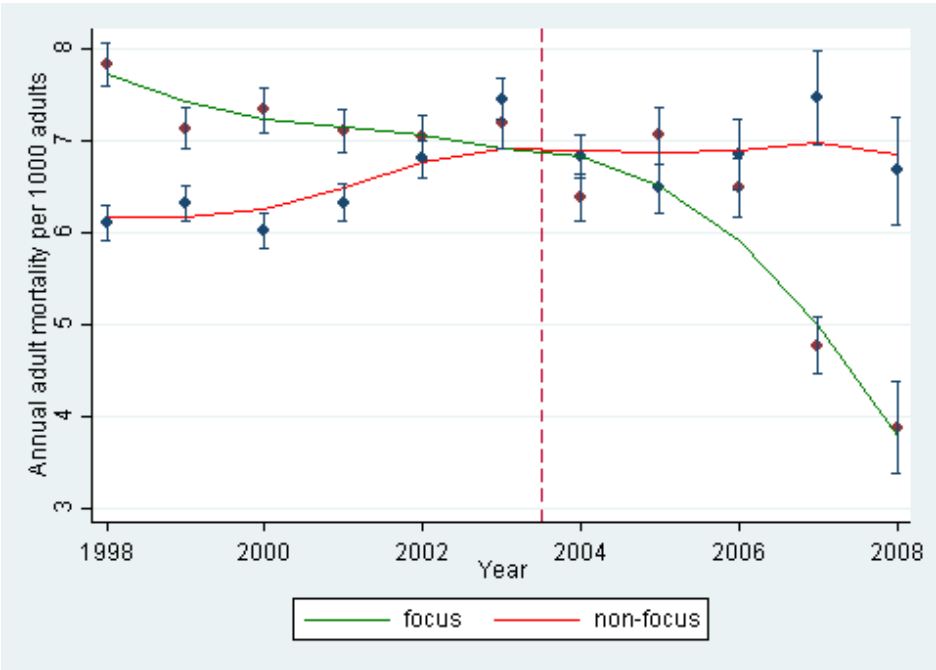
[e] These variables are characteristics of the index woman rather than the sibling. The residence status and educational status of sibling are not known.

Figure C1: Trends in HIV development assistance for focus countries and non-focus countries: mean per-country assistance in 2008 US$, 1998–2008 (PBR)



Note: Data are drawn from the Institute for Health Metrics and Evaluation database. We considered calendar-year 2004 to be the first full year of PEPFAR's activities. Error bars indicate 95 percent CI; PEPFAR, the US President's Emergency Plan for AIDS Relief.

Figure C2: Age-adjusted adult mortality trends in the focus and non-focus countries, 1998–2008 (PBR)



Note: Each point represents the probability that an adult between 15 and 59 years old died during the year per 1,000 adults alive for any part of the year. A narrow-bandwidth (0.6) LOWESS curve is used to fit the trend. LOWESS is a nonparametric method of fitting a curve using local regressions for each point. Error bars indicate 95 percent CI; PEPFAR, the US President's Emergency Plan for AIDS Relief.

Appendix D: List of files received from the original authors

List of do-files received

- replication unified code

List of data files received

- covout_narrow
- wgi govt effectiveness
- WideMerge_v1

References

Bendavid, E and Bhattacharya, J, 2009. The President's Emergency Plan for AIDS Relief in Africa: an evaluation of outcomes. *Annals of Internal Medicine*, 150(10), pp.688–695.

Bendavid, E and Bhattacharya, J, 2014. The relationship of health aid to population health improvements. *JAMA Internal Medicine*, 174(6), pp.881–887.

Bendavid, E, Holmes, CB, Bhattacharya, J and Miller, G, 2012a. HIV development assistance and adult mortality in Africa. *JAMA*, 307(19), pp.2060–2067.

Bendavid, E, Holmes, CB, Bhattacharya, J and Miller, G, 2012b. Replication data for: HIV development assistance and adult mortality in Africa.

Brown, AN and Wood, BDK, 2016. What's first for replication studies is what's next for 3ie's replication programme. Available at: <http://blogs.3ieimpact.org/whats-first-for-replication-studies-is-whats-next-for-3ies-replication-programme/> [Accessed 25 October 2017].

Brown, AN, Cameron, D and Wood, BDK, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness*, 6(3), pp.215–235.

Demographic and Health Surveys (DHS) 2011. ICF International. Available at: <http://www.measuredhs.com> [Accessed 9 January 2012].

Duber, HC, Coates, TJ, Szekeras, G, Kaji, AH and Lewis, RJ, 2010. Is there an association between PEPFAR funding and improvement in national health indicators in Africa? A retrospective study. *Journal of the International AIDS Society*, 13(1), p.1.

Emanuel, EJ, 2012. PEPFAR and Maximizing the Effects of Global Health Assistance. *JAMA*, 307(19), pp.2097–2100.

Fay, MP and Feuer, EJ, 1997. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Statistics in Medicine*, 16(7), pp.791–801.

Glassman, A, Fan, V, Over, M, Silverman, R, McQueston, K and Duran, D, 2013. More health for the money: putting incentives to work for the Global Fund and its partners, [online]. Available at: <http://www.cgdev.org/sites/default/files/More-Health-for-the-Money.pdf> [Accessed 28 April 2015].

Kaiser Family Foundation, 2015. The U.S. President's Emergency Plan for AIDS Relief (PEPFAR). Available at: <http://kff.org/global-health-policy/fact-sheet/the-u-s-presidents-emergency-plan-for/> [Accessed 28 April 2015].

Kaufmann, D, Kraay, A and Mastruzzi, M, 2011. The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2), pp.220-246.

Luo, J, 2016. *HIV development assistance and adult mortality in Africa: a replication study of Bendavid (2012).* Replication plan. Washington, DC: International Initiative for Impact Evaluation (3ie).

Luo, J and Hein, N, 2016. *HIV development assistance and adult mortality in Africa: a replication study of Bendavid (2012)*. Replication plan. Washington, DC: International Initiative for Impact Evaluation (3ie).

SAS Institute Inc. 2015. *Base SAS® 9.4 Procedure Guide, Fifth Edition*. Cary, NC: SAS Institute Inc.

Shelton, JD, 2012. PEPFAR and Adult Mortality. *JAMA*, 308(10), pp.972–973.

United Nations Population Division, 2015. World population prospects, the 2015 revision. Available at: <http://esa.un.org/wpp/> [Accessed 1 June 2016].

United States President's Emergency Plan for AIDS Relief (PEPFAR), 2015. Eleventh Annual Report to Congress on PEPFAR. Available at: <http://www.pepfar.gov/reports/progress/index.htm> [Accessed 28 April 2015].

Valentine, A, Wexler, A and Kates, J, 2016. The U.S. Global Health Budget: Analysis of the Fiscal Year 2017 Budget Request. Available at: https://www.kff.org/global-health-policy/issue-brief/the-u-s-global-health-budget-analysis-of-the-fiscal-year-2017-budget-request/ [Accessed 28 April 2015].

Other publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/

When to start ART? A replication study of timing of antiretroviral therapy for HIV-1-associated Tuberculosis. 3ie Replication Paper 14. Djimeu, EW, 2018.

STRETCHing HIV treatment: a replication study of task shifting in South Africa. 3ie Replication Paper 13. Chen, B and Alam, M, 2017.

Cash transfers and HIV/HSV-2 prevalence: a replication of a cluster randomized trial in Malawi. 3ie Replication Paper 12. Smith, LM, Hein, NA and Bagenda, DS, 2017.

Power to the people?: a replication study of a community-based monitoring programme in Uganda, 3ie Replication Paper 11. Donato, K and Garcia Mosqueira, A (2016)

Fighting corruption does improve schooling: a replication study of a newspaper campaign in Uganda, 3ie Replication Paper 10. Kuecken, M, and Valfort, MA (2016)

The effects of land titling on the urban poor: a replication of property rights, 3ie Replication Paper 9. Cameron, Drew B, Whitney, Edward M and Winters, Paul C (2015)

Male circumcision and HIV acquisition reinvestigating the evidence from young men in Kisumu, Kenya, 3ie Replication Paper 8. Djimeu, EW, Korte, JE and Calvo, FA (2015)

Walking on solid ground: a replication study on Piso Firme's impact, 3ie Replication Paper 7. Basurto, MP, Burga, R, Toro, JLF and Huaroto, C (2015)

The impact of India's JSY conditional cash transfer programme: A replication study, 3ie Replication Paper 6. Carvalho, N and Rokicki, S (2015)

Recalling extra data: A replication study of finding missing markets, 3ie Replication Paper 5. Wood, BDK and Dong, M (2015)

The long and short of returns to public investments in fifteen Ethiopian villages, 3ie Replication Paper 4. Bowser, WH (2015)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses, 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication, 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

TV, female empowerment and demographic change in rural India, 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

Quality evidence for policymaking: I'll believe it when I see the replication, 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

Replication Paper Series

International Initiative for Impact Evaluation
1029 Vermont Avenue, NW
Suite 1000
Washington, DC 20005
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org