Empirical Evidence on Development Effectiveness: From Macroeconomic Structures to Micro-Level Implementation

Dissertation

zur Erlangung des wirtschafts- und sozialwissenschaftlichen Doktorgrades

„Doctor rerum politicarum“

der Ruprecht-Karls-Universität Heidelberg

vorgelegt von

Katharina Richert

Heidelberg

August 2018

EMPIRICAL EVIDENCE ON DEVELOPMENT EFFECTIVENESS: FROM MACROECONOMIC STRUCTURES TO MICRO-LEVEL IMPLEMENTATION

Katharina Richert

BETREUER UND ERSTGUTACHTER DER DISSERTATION

Prof. Dr. Axel Dreher

Professor of Economics
Chair of International and Development Politics
Alfred-Weber-Institute for Economics
Ruprecht-Karls-University Heidelberg

ZWEITGUTACHTER DER DISSERTATION

Prof. Dr. Dr. Till Bärnighausen

Medical Specialist in Family Medicine
Director of the Heidelberg Institute of Public Health
Ruprecht-Karls-University Heidelberg

DRITTGUTACHTERIN DER DISSERTATION

Prof. Dr. Jale Tosun

Professor of Political Science
Chair of Political Science
Institute of Political Science
Ruprecht-Karls-University Heidelberg

Für Oma

Danksagung

Ich danke meinem Doktorvater Axel sehr herzlich dafür, mir die Chance zu dieser Promotion gegeben zu haben und währenddessen wirklich jederzeit verfügbar gewesen, mit Rat und Tat zur Seite gestanden und damit eine enorme Betreuungsleistung erbracht zu haben. Ich habe sehr viel von ihm lernen und profitieren können und bin dafür sehr dankbar. Auch möchte ich mich bei Till Bärnighausen und Jale Tosun dafür bedanken, als Gutachter und Gutachterin meiner Promotion zur Verfügung zu stehen. Außerdem danke ich besonders Andreas Fuchs, der meine Dissertation nicht nur als Koautor, sondern auch als eine Art Mentor jederzeit motivierend und mit hilfreichen Einschätzungen und Kommentaren von Beginn an begleitet hat. Ebenso danke ich vor allem meiner durchgehenden Wegbegleiterin Sarah, ohne die die Promotion sicher nur halb so viel Spaß gemacht hätte. Genauso wie allen bereits Genannten danke ich weiterhin den besten Bürokollegen Diego und Lennart und den besten LehrstuhlkollegInnen Alex, Valentin, Angelika, Gerda, Sven, John, Zain, Marta, Kai und Vera, die man sich wünschen könnte, für inspirierende Gespräche, herausragende Hilfsbereitschaft und all die vielen durchdachten Kommentare zu meinen Forschungsarbeiten. Darüber hinaus bin ich sehr froh, Lennart bei meinen Forschungsaufenthalten an meiner Seite gehabt zu haben und mit ihm durch die Höhen und Tiefen der Feldforschung in Aceh gegangen zu sein. Auch danke ich meinen Koautoren Andi, Axel, Peter, Anna, Lennart, Jana und Sebastian für die tolle und konstruktive Zusammenarbeit und für alles, was ich von ihnen lernen konnte. Für ausgleichende Pausen bin ich meinen LehrstuhlkollegInnen ebenfalls dankbar sowie Alex Glas, Sebastian und Freya. Einen herzlichen Dank auch an Heike für die alltägliche, enthusiastische Unterstützung und Schaffung einer sehr angenehmen Arbeitsatmosphäre. Für die ähnlich wichtige moralische Unterstützung danke ich meiner Familie und meinen Freunden sowie insbesondere meinem Bruder Ole und Leif von Herzen für viel entgegengebrachtes Verständnis, Unterstützung und schönen Freizeit-Ausgleich neben der Dissertation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Considering the large aid effectiveness literature, contradicting findings of scholarly work (e.g., Burnside and Dollar 2000; Easterly et al. 2004) and new rigorous evidence rather shows that the aggregated effect of development assistance is not significantly measurable (Dreher and Langlotz 2017). The picture looks more nuanced, however and unsurprisingly, if we disentangle development assistance into its different delivery modes (World Bank 2006), types (Dreher et al. 2008; Clemens et al. 2012) or specific deliverers such as individual bilateral donor countries (e.g., Berthélemy 2006; Minoiu and Reddy 2010 or the first Chapter). However, what exactly is it that makes development cooperation effective in contributing to improved living conditions in developing countries? The present dissertation contributes to the literature which is motivated by this question and seeks to produce scientific responses. In this vein, I address the role of personal characteristics and commitment of individuals in charge of development cooperation, namely development ministers and power structures in multilateral organizations dedicated to fostering development, namely the World Bank's International Finance Corporation (IFC). Subsequently, I look closer into the domains where success or failure of development cooperation is determined, i.e., at the level of project implementation. Jointly with my co-authors, we examine our own development intervention, which we conducted as an impact evaluation in Indonesia, to identify determinants of successful intervention design, which may be useful to the research community and implementers of development interventions. Such impact evaluations of development projects generally have the sole purpose of increasing the effectiveness of development assistance through identifying what works, and what does not, and why. Their immanent success in actually fulfilling this goal and measurably contributing to increased development effectiveness was not quantitatively examined so far, however, and constitutes the final Chapter of this dissertation.

Undoubtedly, development cooperation in practice is aware of its challenges and tackles them through initiatives like the Paris Declaration from 2005 or the Accra Agenda for Action from 2008 (OECD, 2008). Especially the Paris Declaration was a milestone in acknowledging the importance of ownership of recipients, harmonization among donor countries, alignment of recipient and donor systems, mutual accountability for development access and measurement of results as preconditions and pathways for effective development assistance. In light of descriptive statistics that indeed suggest an increasing aid quality after the Paris Declaration in 2005, me and my co-authors were interested to see whether this aid of higher quality was then capable of leading to improved living conditions measured through growth. This is indeed what our difference-in-differences analysis presented in Chapter 1 (Minasyan, Nunnenkamp and Richert 2017) suggests.

Naturally then appears the subsequent question. If we find that good quality assistance can affect growth, as opposed to aid of average quality – what determines the seemingly crucial provision of such good quality aid? Andreas Fuchs and I approached the determinants of bilateral aid volumes and their quality by examining the characteristics of the individuals, who decide over them – the development ministers. Evidence is presented in Chapter 2 (Fuchs and Richert 2018). Within a sample of more than 300 development ministers from 23 OECD member countries of the Development Assistance Committee (DAC), we find that ministers with more experience in office are better suited to fight for higher budgets, and that female ministers are associated with Official Development Assistance (ODA) of higher quality as well as the ones with specific prior professional experience in development cooperation. While it does not come at a surprise that adequately qualified individuals are better at doing their jobs (which is in line with the literature on other political positions), it is disappointing that only a mere 16 percent of the ministers in our sample possessed such a specific qualification when taking office. Potentially, our findings can contribute to overthink common tendencies to allocate the aid ministry more randomly than other issue areas due to the low domestic salience of aid policy. Why female ministers seem to give ODA of better quality is less obvious and we give some potential explanations in the chapter. Nevertheless, the analysis indicates that specific commitment of individuals can make a difference, which is also in line with anecdotal evidence (Day, 2016). Our identification strategy is rigorous, but cannot establish causality, however.

While we have one person in charge for the respective bilateral development assistance of individual countries, multilateral aid agencies are governed by their member states through boards. One of the major multilateral aid agencies is the World Bank. Within the World Bank Axel Dreher and me focus on its private sector lending arm IFC, which is the largest global development institution focused exclusively on the private sector in developing countries (International Finance Corporation 2016). The IFC's day-to-day business is largely governed by its Board of Executive Directors, which consists of 25 mainly rotating member country representatives and the World Bank President. The Executive Directors have to approve every IFC project and therefore possess substantial leverage over the Corporation's business. This institution attracts our attention, as the bulk of IFC's lending benefits companies from industrialized countries, which receive loans from the IFC and use them mainly to operate projects in developing countries with middle income (Ellmers et al. 2010). Building an identification strategy which largely eliminates endogeneity concerns, we find that power structures actually significantly determine project destinations and volumes. If recipient and donor countries are jointly represented at the board of directors or at the politically influential position of the United Nations Security Council (UNSC), they get more, which benefits donor countries' operating companies as well as the countries receiving the projects. This practice comes with the concern that accordingly not the countries in most need and best project operators are supported by the IFC, but those of the most powerful members. This is particularly striking as Dreher et al. (2018) show that politically motivated aid tends to be less effective. What is more, the regulatory structure of the IFC makes the observed practice unlikely to change for good, as the ones who benefit from the system in operation are unlikely to alter it.

Independent from which fortunate or unfortunate circumstances may have led to development projects, specifically and ultimately, the implementers of such projects have their individual capacities and opportunities to run interventions in an effective way. Many development projects aim at behavioral change. My co-authors Lennart Kaplan, Sebastian Vollmer and I conducted such an intervention and, at the same time, evaluated its effectiveness. In the framework of a randomized controlled trial (RCT), we evaluated in Indonesia, whether a checklist with the essential practices to be done during deliveries lead to safer deliveries.[1] The precondition of this intervention to work is thus, health personnel sustainably changing their behavior towards using the checklist. However, there is not much general guidance available in what actually drives these so-called take-up rates. We therefore

[1] The idea is based on promising experience that had been made with checklists for surgeries implemented all over the world, in countries of any income group (Gawande, 2009).

examined the determinants of take-up rates with the example of our specific intervention to inform the general project implementation literature about incentivizing factors that can be included in the implementation design. Alongside group pressure, pressure from superiors and believe in the own capabilities to perform the task, especially the general attitude towards the behavior appeared to be crucial. This *attitude* determinant describes whether participants judged the checklist to be useful for their work. The *attitude* accordingly depends on motivating the respective behavior and is hence particularly well in control of project implementers. We elaborated further on this determinant within a field experiment and found that framing can influence the intention to use the checklist. When the role of international researchers for the project design was highlighted, health personnel were more likely to support the checklist, than when the local effort was accentuated.[2] We randomized treatment groups, which reduces concerns about the internal validity of our experiment, however, we refrain from claiming external validity. In fact, we conclude that framing can be a cost-effective way of improving take-up rates in interventions, particularly about implementers' characteristics. Which highlights will be effective, however, depends on the context.

There is currently much debate on whether such impact evaluations – as we conducted one on Safe Childbirth – are actually a meaningful solution to improve overall development effectiveness or whether they rather divert too many resources from other important areas in development economics. However, evidence on the overall impact of impact evaluations is generally scarce, despite their long history of 40 years in practice, and mainly limited to qualitative research. My final dissertation chapter contributes first quantitative findings to the debate. In a cross-country setting, I find positive correlations between the number of impact evaluations conducted in health and health outcomes. However, endogeneity concerns are particularly difficult to disentangle in the cross-country setting. The subnational level offers improved conditions for directive linkages of impact evaluations in health leading to better health conditions via corresponding development projects. Also sub-nationally, my results robustly indicate that impact evaluations in maternal and newborn health are associated with reduced infant mortality through improved corresponding projects. Impact evaluations,

[2] Local and international researchers were involved in the implementation design, we simply highlighted the one more than the other and vice versa.

accordingly, seem like a meaningful, statistically measurable option to improve development effectiveness.

In a nutshell, my research shows that there is still room for improvement – to be filled with very tangible suggestions, and this dissertation hints at some factors which might lead to positive change. I hope the presented empirical evidence is informative to policy and can contribute to the ongoing debate on development effectiveness.

Moreover, the dissertation contributes several databases to the research community. Specifically, I created three new databases jointly with my co-authors and research assistants and enhanced two further datasets. In the first Chapter, we disentangle the aggregated measure for quality-aid developed by Roodman into its components and make it available to the research community. The second Chapter builds a new and comprehensive dataset on all development ministers and government heads from 1967 to 2012 for 23 member countries of the OECD's DAC and collects information on their personal characteristics with regard to gender, political ideology, prior professional experience in development cooperation, and time in office. For the third Chapter, we collected data on the base countries of companies, which conduct IFC projects in recipient countries. While the recipients of all projects are available via the IFC website, the home countries of operating companies were less obvious. We collect information for 4557 projects over 23 years from 1994-2016. The dataset on Safe Childbirth Checklist use is also completely new as it was created through our intervention. Besides data on health personnel's motivation or attitudes towards work, which is quite specific to our research question, the survey-based dataset entails information on hospital staffing, equipment, number of patients, mortality or morbidity rates. This information can be particularly useful to future research as administrative data for Aceh is generally scarce. The final Chapter made excessive use of a new database on all impact evaluations from the International Initiative for Impact Evaluation's repository and significantly enhanced it for the country Uganda. For all 100 impact evaluation studies in health conducted in Uganda between 1995-2014 and included in the repository, I collected further information based on the original studies. I categorized the interventions into eight specific sub-categories and coded their outcomes, treatment duration and location. This dataset is available for future research.

The dissertation employs different empirical methods depending on the research question. My work is mainly based on panel-data regression analysis across countries or within a quasi-panel with georeferenced data for the country Uganda, using ordinary least squares (OLS) with different kinds of fixed effects. The application of an adequate set of fixed effects is neat as it can reduce the endogeneity concern of omitted-variable-bias substantially and leads to meaningful correlations. In combination with a source of exogenous variation as used in the third Chapter, this method can largely eliminate endogeneity concerns. For the fourth Chapter, we were (partly) able to apply the so-called "gold standard" of causal inference and used randomized data, which we collected within the RCT and evaluated simple differences between treatment and control group. A differences-in-differences approach was used for the first Chapter on the effects of quality-adjusted ODA.

A compact overview of each chapter's specifics is given in the following:

CHAPTER 1

It has been intensively and controversially discussed whether 'good' economic policies and governance in the recipient countries render foreign aid more effective in alleviating poverty and stimulating economic growth. By contrast, the question of whether aid recipient countries would benefit from stronger income effects if foreign donors provided higher quality aid has received scant attention so far. In Chapter 1, we make use of the index of donor performance from the Center for Global Development to compare the effects of quality-adjusted aid and unadjusted aid on changes in GDP per capita. Our difference-in-differences analysis reveals significant and quantitatively important treatment effects for quality-adjusted aid after the introduction of the Paris Declaration on Aid Effectiveness in 2005, while we do not find significant treatment effects for unadjusted aid. This implies that only recipient countries with increased aid inflows of high quality benefit in terms of increasing GDP per capita. The quality of aid matters most when accounting for delayed effects. However, our results depend on the sample of recipient countries.

CHAPTER 2

Over 300 government members have had the main responsibility for international development cooperation in 23 member countries of the OECD's Development Assistance Committee since the organization started reporting detailed ODA data in 1967. Understanding their role in foreign aid giving is crucial since their decisions can influence aid effectiveness and thus economic development on the ground. Chapter 2 examines whether development ministers' personal characteristics are associated with aid budgets and aid quality. To this end, we create a novel database on development ministers' gender, political ideology, prior professional experience in development cooperation, education, and time in office over the 1967-2012 period. Results from fixed-effects panel regressions show that some of the personal characteristics of development ministers matter. Most notably, we find that more experienced ministers with respect to their time in the development office obtain larger aid budgets. Moreover, our results suggest that female ministers as well as officeholders with prior professional experience in development cooperation and a longer time in office provide higher-quality ODA.

CHAPTER 3

The bulk of IFC lending benefits companies from rich countries, and projects in countries with middle income. Large conglomerates such as Lidl or Mövenpick have been among its direct beneficiaries. This contrasts to some extent with the IFC's official mandate, which is to finance poverty-reducing projects for which private capital is not available on reasonable terms. In Chapter 3, we investigate the drivers of this mismatch. According to our theory, the governments of industrialized countries where borrowing companies are based form coalitions with governments of middle-income countries where the projects are implemented. We therefore expect preferential treatment to be most pronounced when the representatives of both the recipient's and the company's countries are best able to collude in exerting their influence. We argue that this will be the case when both countries' governments are represented among the IFC's Board of Executive Directors, and when they have extraordinary clout with major IFC shareholders. Using data for more than 3,000 IFC projects over the 1995-2015 period we show that the (joint) influence of these countries helps them to receive a disproportional share of IFC funding.

CHAPTER 4

The success of development interventions crucially depends on their uptake in the targeted population. We investigate incentives for uptake of those interventions, making use of a framework grounded in psychological theory, "The Theory of Planned Behavior". The framework suggests three determinants for intervention uptake: personal attitudes, the social influence of important others and the perceived ease of intervention use. We use the setup of two randomized controlled trials in Indonesia and Pakistan to test the theory for development intervention purposes. Our findings show that the proposed determinants are indeed associated with increased uptake. We investigate further on the determinant personal attitudes, which is most relevant to our intervention, by conducting a framed field experiment in Indonesia. The experiment shows that the study population in the Indonesian context exhibits higher levels of support for the project if the participation of international actors is highlighted. Consequently, our results encourage international research and development cooperation, first, to consider the determinants suggested by the "Theory of Planned Behavior" in the design of projects in order to increase intervention uptake. Second, depending on the country context, explicitly framing international participation in the conducted project might be a cost-effective way to increase supportive behavior towards the intervention.

CHAPTER 5

This chapter assesses the impact of impact evaluations based on impact evaluation studies in developmental health interventions, collected through the International Initiative for Impact Evaluation's repository. Considering the steady increase in impact evaluations and financial resources already spent, policy and research are demanding evidence on whether the growing effort indeed leads to learning, improved projects and finally, improved outcomes. I analyze this empirical question for the period 1995-2014 in a cross-country panel for general health outcomes and at a sub-national level for the country Uganda with regard to infant mortality rates per districts. Results are very robust and show a significant correlation between impact evaluations in health and improved outcomes on a cross-country level. Based on more precise data measurement and linkages, also sub-national findings indicate that impact evaluations in maternal and neonatal health conditional on increased corresponding ODA disbursements lead to reduced infant mortality.

ACKNOWLEDGEMENTS

CHAPTER 1 – DOES AID EFFECTIVENESS DEPEND ON THE QUALITY OF DONORS?

(Co-authored with Anna Minasyan and Peter Nunnenkamp)

We thank Axel Dreher and Lennart Kaplan for valuable comments and suggestions, as well as participants of the seminar at Heidelberg University and two anonymous reviewers.

CHAPTER 2 – DEVELOPMENT MINISTER CHARACTERISTICS AND AID GIVING

(Co-authored with Andreas Fuchs)

We are grateful for valuable comments by Axel Dreher and our colleagues at his chair, Emma Mawdsley, Pierre-Guillaume Méon, Sarmistha Pal, Francisco J. Pino, Lionel Roger, seminar participants at Heidelberg University, conference participants of the Annual Meeting of the European Public Choice Society (University of Groningen, Netherlands, April 2015), the International Conference on Globalization and Development (University of Goettingen, Germany, May 2015), the Spring Meeting of Young Economists (Ghent University, Belgium, May 2015), the Annual Conference of the Verein für Socialpolitik Research Group on Development Economics (Kiel Institute for the World Economy, Germany, June 2015), the Beyond Basic Questions Workshop (University of Hanover, Germany, July 2015), the Annual Congress of the European Economic Association (University of Mannheim, Germany, August 2015), the CSAE Conference "Economic Development in Africa" (University of Oxford, UK, March 2016), and the Workshop "The Domestic Dimensions of Development Cooperation" (University of Antwerp, Belgium, October 2016), as well as two anonymous reviewers. Excellent research assistance was provided by Laura Barros, Patrick Betz, Jan-Christian Hansen, Merle Kreibaum, Alexander Leinemann, Lisa Oberländer, Samuel Siewers, Stephan Schneider, Franziska Volk, and Helge Zille. We further thank Jamie Parsons for proof-reading of earlier versions of this chapter.

Chapter 3 – The Political Economy of International Finance Corporation Lending

(Co-authored with Axel Dreher)

We thank Lennart Kaplan, Erasmus Kersting, Christopher Kilby, Valentin Lang, Nicola Limodio, Randall Stone, Konstantin Wacker, participants of seminars at Heidelberg University (2016 and 2017), the International Political Economy Society conference (Durham 2016), the ZEW Public Finance Conference (Mannheim 2017), the Annual International Conference of the Research Group on Development Economics (Göttingen 2017), the 10th Beyond Basic Questions Workshop (Gargnano 2017), the Transformations in Global Economic Governance conference (Cambridge 2017), the 3rd FHM Development Workshop (Heidelberg 2017), the workshop Regionalism and Regional Organizations in a Fragmented World (Oxford 2017), the Political Economy of International Organizations (Wisconsin-Madison 2018), and Deakin University (2018) for valuable comments on an earlier version of this chapter, Christian Thiele for programming the download of IFC-project data, Franziska Volk for extensive research assistance, and Jamie Parsons for proof-reading.

Chapter 4 – What Makes a Successful Development Intervention? The Theory of Planned Behavior – An Application to Implementation Research

(Co-authored with Lennart Kaplan, Jana Kuhnt and Sebastian Vollmer)

We thank Holger Rau, York Hagmayer, Kerstin Grosch, Ghida Karbala, Hendriek Yopin, Friederike Lenel, Robert Schmidt, Jan Priebe, Anna Merkel, Martin Bruder, Simone Dietrich, Dirk Landmann, David McKenzie, Miriam Romero, Marcela Ibañez, Gert Pönitzsch, Rossa O'Keeffe-O'Donovan, Santiago Saavedra, Axel Dreher, Sebastian Schneider, Reinhard Weisser, Atika Pasha, Britta Augsburg, Wändi Bruine de Bruin, Erwin Bulte and colleagues from the chair in Heidelberg as well as the participants of seminars at Heidelberg University and the University of Göttingen and the members of the Research Training Group 1723 "Globalization and Development" for valuable feedback on the design and implementation of the field experiment conducted in Indonesia. Moreover, we are grateful for excellent research assistance in Indonesia and Pakistan.

Chapter 5 – The Impact of Impact Evaluations: Cross-Country and Sub-National Evidence

(single-authored)

I thank Axel Dreher and my colleagues Gerda Asmus, Lennart Kaplan, Valentin Lang, Sarah Langlotz, Angelika Müller as well as participants of the Chair Seminar at Heidelberg University for their excellent ideas, suggestions and comments. Moreover, I am grateful to John Cruzatti and Gerda Asmus for generously sharing georeferenced data with me and Theresa Geyer for excellent research assistance.

CHAPTER 1

Does Aid Effectiveness Depend on the Quality of Donors?[3]

Co-authored with Anna Minasyan and Peter Nunnenkamp

1.1. INTRODUCTION

It has been intensively and controversially discussed whether 'good' economic policies and governance in the *recipient* countries render foreign aid more effective in alleviating poverty and stimulating economic growth. The role of the recipient countries' policies and institutions has been stressed by the World Bank (1998) and Burnside and Dollar (2000). Skeptical assessments include Easterly et al. (2004) and Rajan and Subramanian (2008). Considerably less attention has been paid to the question of how *donor* policies could enhance the effectiveness of aid, even though the recent literature offers several indications that the source of funding matters. Specifically, it appears that the effects of aid depend on political characteristics of the donors, the motives underlying their aid, and complementary (non-aid) policies.

Concerning political characteristics, Bermeo (2011) finds that aid from democratic donors promotes democratization in the recipient countries, whereas aid from authoritarian donors is negatively associated with democratization. According to Dreher et al. (2015), political

[3] Chapter available as published article: Minasyan, Anna, Peter Nunnenkamp and Katharina Richert (2017). Does Aid Effectiveness Depend on the Quality of Donors? *World Development* 100: 16-30.

misalignment and greater ideological distance between donor and recipient governments reduces the growth effects of aid by adding to transaction costs and eroding trust.

Focusing on trade-related motives of aid, Berthélemy (2006) differentiates between selfish and altruistic donors. Based on such classifications from the aid allocation literature, Minoiu and Reddy (2010) distinguish the growth effects of so-called developmental and non-developmental aid. These authors find that only developmental aid promotes economic growth in the recipient countries. In a similar vein, Kilby and Dreher (2010) find that the growth impact of aid that addresses recipient needs differs from the impact of aid that is motivated by donor interests. However, the donor classifications used by Minoiu and Reddy (2010) tend to be ad hoc and are typically time invariant. Strategic and geopolitical motives of donors are also likely to erode the effectiveness of aid. Headey (2008), Bearce and Tirone (2010), and Bermeo (2016) consider aid to be geopolitically motivated under Cold War conditions. Dreher et al. (2018) show that aid is less effective in promoting growth when recipient countries are strategically important as temporary members of the UN Security Council (UNSC). Dreher et al. (2013) come to similar conclusions when investigating the effect of UNSC membership on the evaluation of World Bank projects.

The effectiveness of aid may also depend on complementary donor policies. For instance, empirical evidence presented by Minasyan and Nunnenkamp (2016) suggests that higher remittances paid by donor countries, proxying for worker mobility and migration, strengthen the growth effects of aid. Instead of assessing conditional aid effects, Gary and Maurel (2015) construct a measure of donors' policy consistency which includes aid as one of seven elements.[4] They find that more consistent donor policies are associated with higher growth in the recipient countries.

In contrast to Gary and Maurel (2015), we focus on the quality of the donors' *aid policies* in the following. Our analysis thus resembles the distinction between developmental and non-developmental aid by Minoiu and Reddy (2010). However, we employ a time varying measure of the quality of donors to address the question of whether the income effects of aid depend on the

[4] All elements (aid, trade, migration, investment, technology, security, and environment) are part of the Center for Global Development's Commitment to Development Index (for details see: http://www.cgdev.org/cdi-2015).

source of funding. Specifically, we make use of the periodical ranking of donors by Roodman (2012) which is available from the Center for Global Development on an annual basis since 1995.[5] Roodman's ranking covers various aspects of aid-related policies, e.g., by discounting tied aid, adjusting for selective aid allocation, penalizing project proliferation, and rewarding tax policies to support private giving.

Accounting for aid-related policies and adjusting nominal aid disbursements accordingly results in a measure of 'effective' donor support that may deviate substantially from the aid figures typically applied in previous studies on the aid-growth nexus. At the same time, the deviation between quality-adjusted aid and nominal aid varies considerably across donors and over time (for details see Section 1.2.). Consequently, the income effects in the recipient countries are likely to depend on the composition of donors contributing to overall aid inflows. Against this backdrop, we hypothesize that it is quality-adjusted aid, rather than nominal aid, from which recipient countries might benefit in terms of higher GDP per capita. Based on Roodman (2012) we calculate the effective amount of quality-adjusted aid that recipient countries receive from donors of the Development Assistance Committee (DAC).

The Paris Declaration (PD) on Aid Effectiveness from 2005 marked a political turning point in the aid effectiveness debate focusing on improved donor policies. Our descriptive data indeed show an increase in quality-adjusted aid thereafter. Accordingly, we run a difference-in-differences analysis using the year 2005 as dividing line between two 6-year periods before and after the PD. The approach reveals significant treatment effects for quality-adjusted aid, while we do not find significant treatment effects for unadjusted aid. Hence, we offer new empirical evidence for the notion that even though nominal amounts of aid might not affect growth, development cooperation can still have positive income effects if donors improve the quality of their aid.

1.2. THE QUALITY OF AID – MEASUREMENT AND RELEVANCE

Developmental aid can be defined as the part of aid that actually has the potential to foster economic growth in the recipient countries, in contrast to non-developmental aid. Previous

[5] We are most grateful to David Roodman for sharing his data.

studies typically distinguish between the two types of aid by separating donors into distinct categories of 'superior' versus 'inferior', or 'altruistic' versus 'selfish' and comparing the effects of total aid from these groups (e.g., Berthélemy 2006 and Minoiu and Reddy 2010). With this strategy, studies might miss the fact that aid by a specific donor is not homogenous (Kilby and Dreher 2010). However, surprisingly few research tools are available to calculate the amount of effective development aid within a donor country's total aid disbursements and even less so within a recipient country's total aid receipts. A reliable measure of effective aid would empower the ongoing research on the link between aid and economic growth and help identify ways to increase the effectiveness of aid.

The Center for Global Development (CGD) provides such a tool from the donor perspective for total aid disbursements: By assessing the quality of donors' aid policies, the CGD derives the total amounts donors give in effective development aid. Specifically, the CGD applies donor-specific and time-varying discount factors to each donor's nominal aid. For calculations of the discount factor, Roodman (2012) includes only the net amounts of aid after subtracting interest and principal payments, debt relief and rescheduled debt from gross disbursements. Roodman penalizes aid allocated to richer and to more corrupt countries. He further penalizes tied aid and project proliferation. More precisely, tied aid is expected to increase the administrative burden for recipient countries and to reduce the value for money, as donor-produced products have to be imported which are not necessarily the most competitive products. These direct and opportunity costs are estimated to reduce the value of disbursements by 13-23 percent. Similarly, a large number of small projects per donor, known as project proliferation, add to administrative costs and are therefore penalized. Finally, Roodman (2012) rewards fiscal incentives for charitable giving. Tax policies that encourage private donations reduce the discount factor as more funds for civil society organizations increase the amount of money that can be used for development purposes in recipient countries.

Roodman's account of aid-related policies affects effective donor support to varying degree. Comparing the 28 DAC donor countries in our sample, Sweden provides aid of highest quality, whereas the Slovak Republic represents the taillight. Net aid disbursements[6] are on

[6] In calculating the discount rate we follow Roodman (2012) and use his net aid measure.

average discounted by 44 percent for Sweden, compared to 77 percent for the Slovak Republic (Figure 1.1.). Moreover, as can be seen from Figure 1.2., the discount factor varies considerably over time for individual donors.

Figure 1.1. Mean discount factors per donor country (1995-2011)

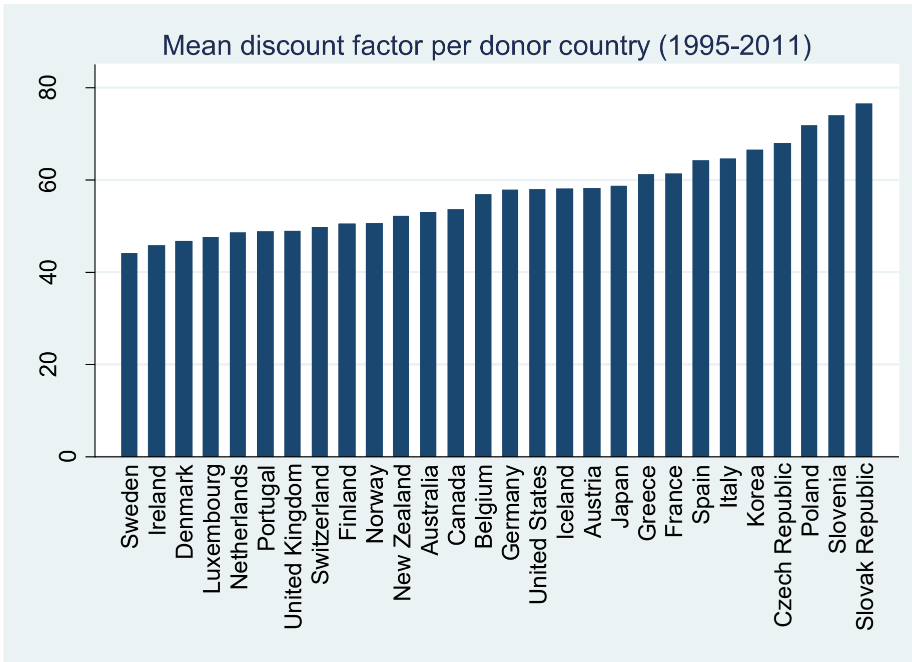
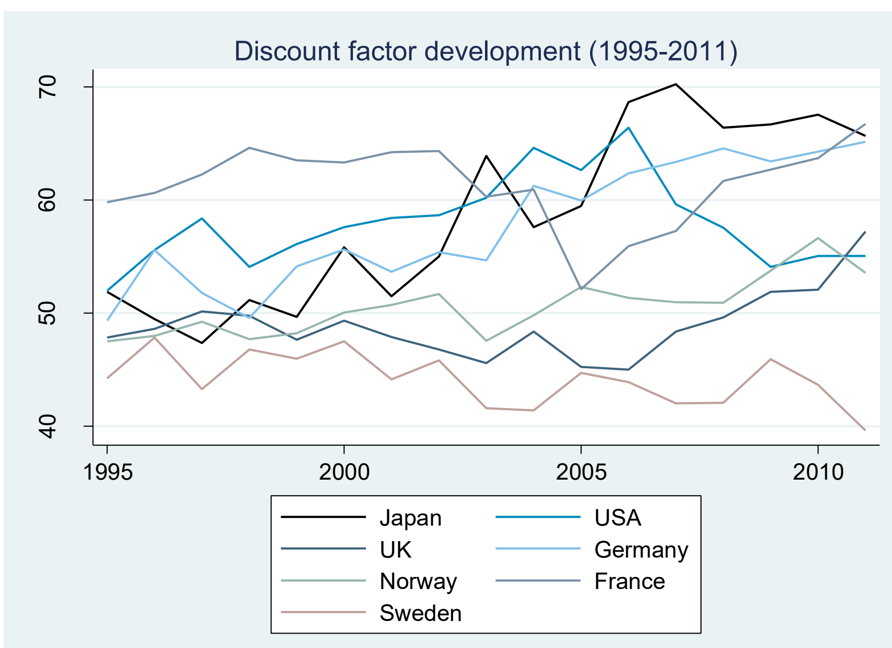


Figure 1.2. Discount factor development for selected donors (1995-2011)



Based on Roodman's (2012) approach we calculate the amount of effective development aid that recipient countries finally receive. While the discount factor varies across donors and over time, it applies to all recipient countries.[7] Our measure of quality-adjusted aid *(QualAid)* uses the donor-specific discount factors *(d)* at time t to adjust each donor *j*'s bilateral aid disbursements *(Aid)* to particular recipient countries i at time t. By aggregating all quality-adjusted aid disbursements from the 28 DAC-donors per recipient, we calculate the total amount of quality-adjusted aid that each recipient receives per year *(QualAid$_{it}$):*

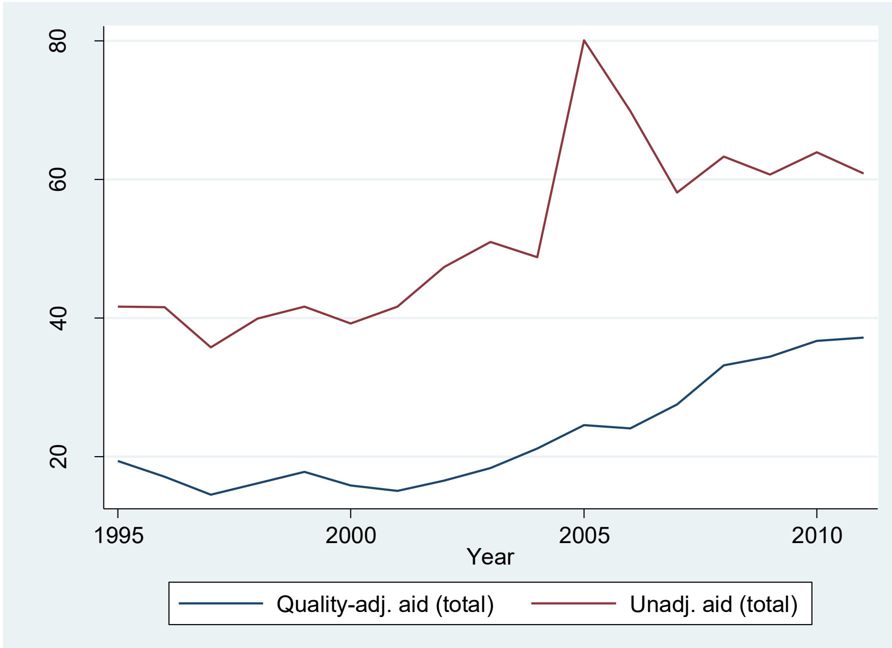
$$QualAid_{it} = \sum_{j=1}^{28} QualAid_{ijt} = \sum_{j=1}^{28} (Aid_{ijt} * d_{jt}) \quad with\ 0 \leq d_{jt} \leq 1 \tag{1}$$

As a result, the variation of quality-adjusted aid across recipient countries stems from varying contributions of specific donors to the recipient's overall aid inflows at different points in time.[8] Figure 1.3. compares the sum of quality-adjusted aid and unadjusted aid for all recipient countries over our period of observation from 1995-2011.

[7] Clearly, it would be preferable to calculate the discount factors bilaterally for each donor-recipient pair. To the best of our knowledge, the data situation does not allow for a comprehensive calculation of the quality of aid at the bilateral level. Moreover, for some quality aspects of Roodman's measure, the principally superior bilateral approach appears to be infeasible for conceptual reasons. This applies especially to Roodman's (2012) account of the tax policies of donor countries that reward charitable private giving in general, i.e., to all recipient countries alike. Note however, that we present complementary estimations at the bilateral level based on so-called country-programmable aid (CPA), which excludes debt relief and other items that cannot be used productively in recipient countries – similar to the first steps of Roodman's calculation of donor discounts. See Table 1.3. in Section 1.4. for details.

[8] The limitation of this approach would be less serious if donors gave aid of similar quality to all recipient countries. We address this issue in detail in Appendix A1. Indeed, we find very few significant differences in the quality of aid from specific DAC donors between recipient countries in the treatment and control groups.

Figure 1.3. Quality-adjusted aid versus unadjusted aid, total US$ for all recipient countries per year (1995-2011)



Interestingly, total quality-adjusted aid disbursements rise notably from 2007 onwards. A change in policies takes time until it can result in actual change on the ground. Considering the time usually required to disburse commitments due to administrative delay and existing contracts, the increase of quality-adjusted aid matches the onset of the Paris Declaration in 2005 quite well.[9] With regard to aid effectiveness, a clear change in donor policies was attempted to be induced by the PD. The signatories of the PD committed themselves to improve the quality and, thus, the effectiveness of aid (OECD 2008).[10] What is more, additional initiatives in the same year such as the UN Millennium Project (UNDP 2005) and the Commission for Africa (2005) strengthened the case for intensified donor efforts and a concentration of aid on the particularly needy. Also, at the G8 Summit in Gleneagles in 2005, political leaders agreed to substantially increase aid by about US$ 50 billion per annum (by 2010) and to double aid to Africa (http://www.unmillenniumproject.org/press/g8overview.htm).

---

[9] In Appendix B1., we present a placebo test by hypothetically pre-dating the Paris Declaration by five years to assess its relevance for explaining the nexus between quality-adjusted aid and changes in GDP per capita.
[10] For details of the Paris Declaration see: http://www.oecd.org/dac/effectiveness/34428351.pdf (accessed: May 2016).

The commitments made by donors in 2005 have been met to varying degree, and the contribution of donors to the increase in quality-adjusted aid differs between the aid policy aspects considered by Roodman (2012) to calculate the above noted discount factors. Various donors contributed to the increase in quality-adjusted aid by raising the share of funds that actually reach the recipient countries and are available there for a productive use. In particular, most donors narrowed the gap between gross and net aid by lowering the repayment obligations of recipients.[11] Furthermore, nominal aid items not involving actual financial transfers to recipient countries played a minor role after the PD. While debt relief and rescheduling accounted for more than one fifth of all donors' aid commitments in 2005, this share dwindled to less than three percent in 2009-2011.[12]

Moreover, most donors improved the quality of aid by increasingly untying aid. For the overall sample of donor countries, Roodman's (2012) 'penalty' for tying aid declined from 7.1 percent of total net aid in 1999-2004 to 3.8 percent in 2006-2011. Untying was most pronounced by this measure for Canada, Italy, Spain and the United States, whereas two major donors – Germany and Japan – counteracted this trend by higher penalties for tying aid after the PD.

Fewer donors helped improve the quality of aid by less proliferation and greater selectivity. Proliferation, as reflected in Roodman's (2012) project size weights, hardly changed when comparing the periods 1999-2004 and 2006-2011 for all donors taken together. On the one hand, ten donor countries improved the quality of aid by reducing proliferation – most notably France, Australia, Denmark and Sweden. On the other hand, proliferation increased for some major donors such as Germany and Japan. Finally, the evidence from Roodman's selectivity weights suggests that the quality of aid hardly improved through a more selective aid allocation. Some donors, notably the United States and Japan, became more selective compared to 2005. However, selectivity was exceptionally weak at the time of the PD.

Considering the year 2005 and the conclusion of the PD as a quasi-experiment, we use the Roodman measure as proxy for improved quality of aid after 2005 and test whether the increase

---

[11] Japan represents a major exception in this respect. Excluding Japan, the share of repayments in gross aid declined to less than eight percent in 2006-2011.

[12] For details see: http://stats.oecd.org/index.aspx?DataSetCode=CRS1

in quality-adjusted aid translated into more effectiveness on the ground, in terms of increased recipient countries' GDP per capita (see next section for details). In the remainder of this section, we argue that policy changes induced by the PD should be captured by Roodman's measure of quality-adjusted aid. Furthermore, we argue that this measure of quality-adjusted aid should be more capable of affecting economic growth than unadjusted aid.

Roodman's (2012) index was developed two years before the PD was signed and can apparently not have been intended to measure potential impacts thereof. All the same, the elements of Roodman's index are closely related to relevant aspects of the donors' aid policies that are also highlighted in the PD, which raised overall awareness of quality differences in aid by clarifying norms and focusing on the quality of aid partnerships (Wood et al. 2011). Roodman's points on the need to reduce tied aid and project proliferation are reflected in two out of the twelve PD indicators to measure progress (OECD 2008).[13] Furthermore, Roodman's selectivity weights for poorer and well-governed recipients as well as the reward for increased charitable giving are in line with the recommendation by Wood et al. (2011) to enforce the implementation of the PD by focusing on the poorest population, highlighting the impact of corruption, and encouraging stronger engagement with civil society organizations. Arguably, Roodman's academic index construction influenced the political progress of concluding important parts of the PD. Consequently, we can reasonably expect that our measure of quality-adjusted aid is superior to unadjusted nominal aid in capturing economic growth effects if the PD had an impact on the effectiveness of aid.

Concerning the aid-growth link, we advance two arguments why quality-adjusted aid should matter, rather than unadjusted aid. First, the quality-adjusted measure takes into account that only funds that actually reach the recipient countries can be used there productively. Most importantly, the quality-adjusted measure nets out principal and interest payments and excludes items such as debt relief and rescheduling that do not involve actual financial transfers to the

[13] In the PD, the OECD echoes Roodman's call for untying aid, in order to get better value for money (OECD 2008: 5). Indicator 8 tracks the degree to which aid is untied, under the second PD principle "Alignment." The need to reduce project proliferation for more aid quality is addressed by the OECD under the principle "Harmonisation." Indicator 9 encourages program based approaches, instead of many small and distinct donor projects (OECD 2008: 6). Moreover, the call for less project proliferation and reduced donor fragmentation is taken up in several parts of the PD (see, e.g., paragraphs 6, 21, and 33).

recipient country. Moreover, it discounts tied aid appropriately and penalizes project proliferation, but also rewards tax policies that encourage charitable giving to arrive at financial transfers actually available for productive use. Second, the quality-adjusted aid measure accounts for development-oriented donor motives and allocation mechanisms. Roodman's selectivity weights for poorer and better governed countries are particularly important in this respect. As argued by Kilby and Dreher (2010), more development-oriented donor motives can lead to more effective use of aid on the ground. Development-oriented donors are more likely to allocate aid to countries that credibly show efforts for effective use and to withhold tranches if recipients do not stick to the committed use of funds. Development-oriented donors are also more likely to adhere to mechanisms that increase aid effectiveness, for instance by applying the commitments made in the PD such as using the partners' accounting mechanisms, embedding projects sustainably, and enforcing monitoring and control mechanisms (OECD 2008).

1.3. Approach and data

We perform a difference-in-differences (DD) analysis to assess whether the income effects of aid depend on the quality of donors providing aid to recipient country *i* at time *t*.[14] Our dependent variable is the difference in GDP per capita (*GDPpc*) between two points in time, as reported in the World Bank's World Development Indicators database.[15] *GDPpc* is logged before differencing. The DD approach combines before-after and with-without comparisons of the difference in *GDPpc*. This combination helps mitigate the limitations of both types of comparison when used in isolation. The simple before-after approach would compare the difference in *GDPpc* in aid recipient countries prior and subsequent to a distinct change in donor behavior. Clearly, the implicit assumption that no other omitted variable might have affected our dependent variable over time is unlikely to hold. The simple with-without alternative of comparing the difference in *GDPpc* between countries benefitting from the change in donor behavior and those not benefitting

[14] The description of the DD approach draws on Nunnenkamp and Öhler (2011).

[15] Alternatively, we consider the growth rate of GDP per capita as well as the level of GDP per capita as dependent variables. See Section 1.4. for details.

would ignore that our dependent variable might have developed differently in the treatment and control groups due to time-varying factors unrelated to aid.

The DD estimator removes any fixed country effects and any fixed time trends. According to Imbens and Wooldridge (2009: 67), the DD approach "is often associated with so-called 'natural experiments', where policy changes can be used to effectively define control and treatment groups."[16] However, in this study, we do not have the ideal DD setting, as there is no group of developing countries that is exogenously excluded from receiving quality-adjusted aid. We will describe in detail below how this identification challenge is solved based on Abadie (2005). Moreover, concerns about causal inference are not completely resolved in the present context. While first-differencing removes omitted-variable bias from time-invariant factors, it does not address omitted-variable bias from time-variant characteristics of recipient countries (Clemens et al. 2012; Roodman 2015). For instance, donors may grant more quality-adjusted aid to new governments embarking on growth-promoting reform programs.[17] This would violate our implicit assumption that quality-adjusted aid is distributed randomly across recipients, based on previously established bilateral aid relations with donors taking autonomous decisions on the quality of their aid.[18] Consequently, higher growth could then be spuriously attributed to quality-adjusted aid. In this study, we control for some of the apparent time-varying factors, such as trade and governance measures. However, factors that are not consistently observed and measured remain unaddressed.

Our period of observation is 1999-2011. The limiting factor is the assessment of the donors' quality which is not available for more recent years. We divide this period into two equally long sub-periods, i.e., 1999-2005 and 2005-2011 ('before' and 'after'). As explained in more detail below,

[16] The great appeal of DD analysis "comes from its simplicity as well as its potential to circumvent many of the endogeneity problems that typically arise when making comparisons between heterogeneous individuals" (Bertrand *et al*. 2004: 249).
[17] Likewise, reverse causality may be an issue when donors grant more quality-adjusted aid to where they expect higher economic growth. We are grateful to an anonymous reviewer for alerting us to unresolved endogeneity concerns. See also Appendix B1. where we present findings from a simple aid allocation model in order to address endogeneity concerns at least tentatively.
[18] In other words, we implicitly assume that a donor country's decision to improve the quality of its aid does not affect this donor's aid shares in given bilateral aid relationships. However, the donor's decision to improve the quality of aid may coincide with a modified distribution of this donor's aid, e.g., when donors with improved aid quality prefer better governed recipient countries.

we also account for delayed effects of the aid treatment by using more recent years for calculating the change in *GDPpc*. We calculate the difference in *GDPpc* between the first and the last year of each period and compare these differences. The outcome for the first period is thus $GDPpc_{2005} - GDPpc_{1999}$ and for the second period $GDPpc_{2011} - GDPpc_{2005}$.

We then consider the difference in time of our outcome variable, i.e., before the treatment and after the treatment. Taking 2005 as the dividing line between 'before' and 'after' appears to be most plausible (as outlined in Section 1.2.), considering the PD and other initiatives being issued this year, which committed the donors to improve the quality and, thus, the effectiveness of aid.

Finally, we consider two different country groups, of which one is receiving the treatment and the other one is not. We define our treatment and control groups based on our quality-adjusted aid measure per recipient and follow the approach of Öhler and Nunnenkamp (2011). Our with-without dimension of the DD approach distinguishes between recipient countries with high increases in quality-adjusted aid after 2005 and recipient countries with low increases or declines in quality-adjusted aid (treatment group, *T*, versus control group, *C*). To identify both groups, we first calculate mean inflows of quality-adjusted aid per recipient country for the first period from 1999-2004 and for the second period from 2006-2011. We then compute the increase in mean quality-adjusted aid between the two periods by taking the difference *($Difference_i$)*:

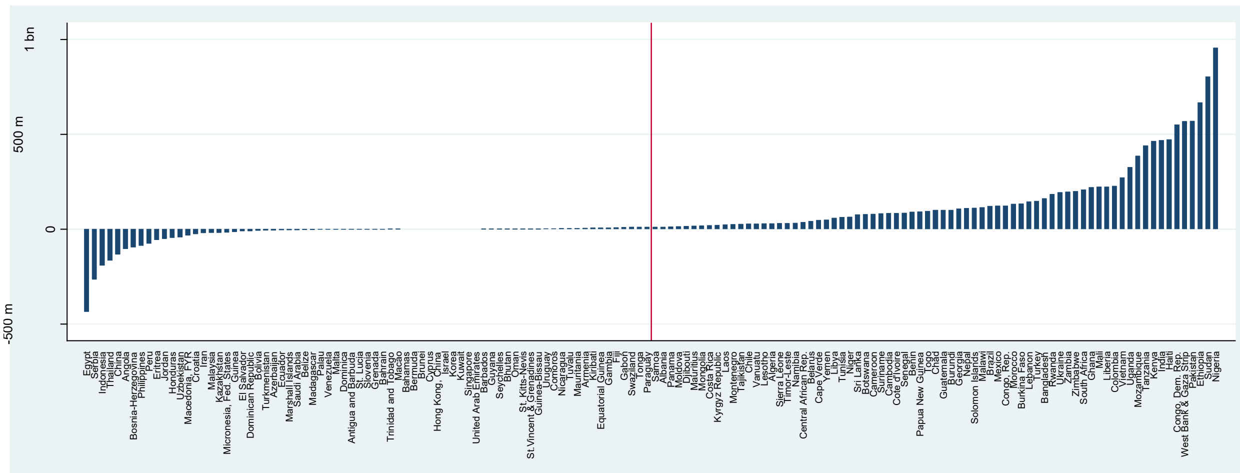
$$Difference_i = \frac{\sum_{t=2006}^{2011} QualAid_i}{6} - \frac{\sum_{t=1999}^{2004} QualAid_i}{6} \qquad (2)$$

We use the median of this measure and select all countries with above-median increases in quality-adjusted aid into the treatment group, and countries with increases at the median or below into the control group. Countries in the treatment group, on average, experienced a considerable increase in quality-adjusted aid, whereas countries in the control group suffered a decline (see Table 1.10.). In per-capita terms, the increase in quality-adjusted aid amounted to 28 US$ for the treatment group when comparing the second period with the first period, compared to a decline by about 9 US$ for the control group.[19] Figure 1.4. shows the difference-values for all

[19] Note that we exclude 2005 for the calculation of average annual aid flows since this year is the dividing line between the sub-periods before and after the change in donor behavior.

recipient countries and the selection into treatment and control group accordingly.[20] Alternatively, we exclude countries around the median and compare the top and bottom tercile of our sample. In both variants of our approach we do not simply differentiate between countries with high and low quality-adjusted aid inflows, but between countries that benefited more (treatment) or less (control) from the policy change induced by the PD.

Figure 1.4. Differences in mean quality-adjusted aid (total US$) between the 1st and the 2nd period per recipient country



Note: The vertical red line illustrates the dividing line between treatment and control group. Countries on the left side are selected into the control group, countries to the right form the treatment group.

Clearly, our treatment and control groups are not selected exogenously. This may violate the identifying assumption of the DD approach: the common trends assumption states that, in the absence of the (aid-related) treatment, the difference in the dependent variable between the two sub-periods would have been the same for the treatment and the control groups. The plausibility of this assumption in the present context is assessed tentatively in Appendix C.1. The trends shown there for the differences in GDP per capita are not perfectly parallel for the two groups of recipient countries, even though it is reassuring that the treatment and control groups seem to have parallel trends after 1995.[21] Abadie (2005) argues that the plausibility of the common trends assumption of DD estimators is open to question if the treatment and control groups differ with

[20] Paraguay is the country which sets the median at 11.6 million US$. All countries with higher differences are selected into the treatment group. The dividing line between treatment and control group is indicated by the red vertical.

[21] Subsequent to the PD in 2005, i.e., the year that divides our two sub-periods of observation, the treatment group shows larger differences in GDP per capita than the control group.

regard to factors that may be associated with the dynamics of the dependent variable. This risk often prevails when evaluating real-world policy interventions. However, Abadie (2005) offers a procedure to overcome this challenge. In addition to control variables that might affect the outcome differently for both groups, Abadie (2005) suggests including interaction terms between these controls and the dummy variable for the treatment period. This procedure accounts for time-varying effects of factors influencing the outcome and, thereby, mitigates identification problems due to trend deviations between the treatment and control group.

For further strengthening the validity of our results with respect to quality-adjusted aid, we perform the same procedure of identifying treatment and control groups for unadjusted aid flows and run the same DD regressions. This gives us the opportunity to compare whether potential effects can be attributed to differences in quality-adjusted aid but not to increases of unadjusted aid (similar to a placebo test). Unadjusted aid flows to country i at time t are (net) disbursements in constant prices of 2012,[22] as available from the OECD-DAC aid database.[23] The aid variables are defined in total US$ or, alternatively, per capita of the recipient countries' population (see below for details). Formally, the DD estimator for our baseline specification is as follows:[24]

$$\begin{aligned} DD = &((GDPpc^T_{2011} - GDPpc^T_{2005}) - (GDPpc^C_{2011} - GDPpc^C_{2005})) \\ &- ((GDPpc^T_{2005} - GDPpc^T_{1999}) - (GDPpc^C_{2005} - GDPpc^C_{1999})) \end{aligned} \tag{3}$$

The estimator corresponds to the coefficient of the interaction term between the dummy variable for the treatment group and the dummy variable for the second period in the basic regression specification without additional control variables. It should be noted that our DD approach does not take into account that the outcome variable for one group of recipient countries may be linked to the outcome for the other group. The differences in GDP per capita for the treatment and control groups are not necessarily independent from each other.[25] In the case of

---

[22] All monetary variables in our dataset are defined as constant 2012 US$.

[23] http://stats.oecd.org/index.aspx?DataSetCode=CRS1 (accessed: May 2016).

[24] While our aid measure is based on average annual flows during 6-year sub-periods, our dependent variable is the difference in the level of GDP per capita during the 6-year sub-periods 1999-2005 and 2005-2011.

[25] We are grateful to an anonymous reviewer for having alerted us to this limitation.

quality-adjusted aid, we believe that the possible bias resulting from this limitation would work against us finding significantly different effects for the treatment group as compared to the control group. In other words, we tend to err on the conservative side: reported differences, as reflected in the coefficient on the interaction between the dummy variables for the treatment group and the second period, tend to be underestimated.[26]

In extended specifications, we add the (logged) level of *GDPpc* at the beginning of the first and second periods. In this way we take into account that changes in *GDPpc* may depend on initial levels. Furthermore, we control for the (logged) rate of inflation as an indicator of economic stability, the ratio of imports plus exports over GDP as an indicator of openness to trade, and control of corruption as an indicator of good governance – all at the beginning of the first and second periods.[27] Another advantage of our DD approach may be noted in this context: we do not face the usual problem of including variables as controls which may actually represent channels through which aid could affect *GDPpc*. As we only include control variables for the first year of both periods, we do not run into the risk of closing channels during the period in question. Moreover, we account for disbursements of multilateral aid as a share of total aid, since our measure of quality-adjusted aid covers only bilateral DAC donors. Finally, we include interaction terms between the control variables and the dummy variable for the second period as outlined above. The fully specified regression equation reads as follows:

$$\Delta \mathrm{GDPpc}_{it} = \alpha_0 + \beta \mathrm{Treatment}_i + \gamma \mathrm{2nd\ period}_t + \delta(\mathrm{Treatment*2nd\ period})_{it} + \lambda \mathbf{X}_{it} + \varrho(\mathbf{X}\mathrm{*2nd\ period})_{it} + \varepsilon_{it} \quad (4)$$

ΔGDPpc is the difference in GDP per capita for recipient i between the first and the last year of each period t. *Treatment* is a dummy variable equal to 1 if a recipient is in the treatment group, and 0 otherwise. *2nd period* is a dummy variable equal to 1 for the treatment period (2005-

[26] In addition, we calculated the correlation of changes in GDP per capita between recipient countries in the treatment and control groups to test at least tentatively for the strength of interdependencies in the outcome variable. We considered differences in GDP per capita over six years, on a rolling annual basis, for this test. This exercise revealed that the correlation is relatively weak (0.3) during our period of observation (1999-2011), compared to earlier and extended periods (0.61 for 1970-2011; 0.68 for 1980-2011; and 0.6 for 1990-2011). Taken together, we consider it unlikely that our results suffer from serious bias of the DD estimator.

[27] We used the first non-missing value at the beginning of the first and second period, to be precise. The data on inflation and trade are from the World Bank's World Development Indicators. Control of corruption is from the World Bank's Worldwide Governance Indicators.

2011), and 0 otherwise. *X* denotes a set of control variables for the recipient country, including (first non-missing values of) the inflation rate, the share of trade in GDP, control of corruption and initial GDP per capita for each period and recipient country.

The coefficient δ of the interaction between the dummy variables of the treatment group and the second period is of principal interest to assess our hypothesis that it is quality-adjusted aid, rather than unadjusted nominal aid, from which recipient countries benefit in terms of increases in *GDPpc*. The hypothesis implies that δ is significantly positive when the treatment is based on the difference of quality-adjusted aid, while δ would be statistically insignificant when the treatment is based on the difference of unadjusted aid as reported by the OECD's Development Assistance Committee. In contrast, δ would be statistically insignificant independent of whether the treatment is based on the difference of quality-adjusted aid or unadjusted aid if it made no difference for recipients that aid is granted according to the criteria used in Roodman's (2012) measure. Standard errors are clustered by recipient country.

To summarize our approach, the PD serves to motivate our work of evaluating potential effects of donor quality on recipient countries' *GDPpc*, and specifically the choice of 2005 as the dividing line between 'before' and 'after'. The treatment group is defined as those countries receiving considerably higher quality-adjusted aid in 2006-2011, compared to 1999-2004 (i.e., we compare the average level of quality-adjusted aid in 2006-2011 with the average level in 1999-2004 to define the increase in quality-adjusted aid). In other words, the treatment is the change in donor behavior after 2005, resulting in a marked increase of quality-adjusted aid (see Figure 1.3.); recipients in the treatment group benefited from this change, countries in the control group did not.

1.4. RESULTS

Baseline results

Table 1.1. presents our baseline results on the effects of aid on the change in GDP per capita for the full sample of recipient countries.[28] We use the median of the difference in aid between the first and second periods to separate the treatment and control groups as outlined in Section 1.3. The difference in aid is based on inflows in total US$ in the upper panel of Table 1.1. and, alternatively, on inflows per capita in the lower panel of Table 1.1. We prefer the first measure since donors are expected to decide on absolute aid amounts when distributing their overall aid budget across recipient countries.[29] Since small recipient countries typically receive higher aid inflows per capita (e.g., Neumayer 2003), they are more likely to fall into the treatment group when donors scale up aid and the treatment is defined in per-capita terms.

We proceed in three steps to compare the treatment effects of quality-adjusted aid (columns 1-3) with the treatment effects of unadjusted aid (columns 4-6). In the first step, we perform the basic DD estimations without any additional control variables (columns 1 and 4). We then include the control variables introduced in Section 1.2. (columns 2 and 5). Finally, we also interact the control variables with the dummy variable for the second period (columns 3 and 6); this way we attempt to control for the effects of omitted time-varying and country-specific variables that may be correlated with increased quality aid (the treatment). For the sake of brevity, we do not show the coefficients on the control variables and their interactions with the dummy variable for the second period. While the signs of the control variables are generally as expected, they typically do not reach statistical significance at conventional levels.[30] In other words, the

[28] It should be noted that we excluded two outliers throughout our empirical analysis: Libya and Equatorial Guinea. While Libya's GDP per capita suffered an exceptionally steep decline in 2011 (Libyan Civil War), the increase in Equatorial Guinea's GDP per capita was exceptionally large in 2008, i.e., shortly after the discovery of large oil reserves.
[29] Our results stay robust if we control for the population of recipient countries. Results are available upon request.
[30] However, the share of multilateral aid proves to be statistically significant in the fully specified model in columns (3) and (6). Specifically, the change in GDP per capita was positively associated with multilateral aid, though only in the first period.

control variables do not have a strong impact on the dynamics of our dependent variable during the period of observation.

Table 1.1. Effects of aid on the change in GDP per capita: full sample, median of difference in aid to separate treatment and control groups

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Treatment based on change in aid inflows, total US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period	-0.025	-0.013	0.024	-0.006	0.013	0.137
	(0.216)	(0.576)	(0.872)	(0.760)	(0.581)	(0.347)
Treatment	-0.004	-0.009	-0.013	-0.008	-0.026	-0.028
	(0.881)	(0.800)	(0.724)	(0.764)	(0.437)	(0.416)
2nd period * treatment	0.067**	0.055*	0.063**	0.029	0.014	0.013
	(0.017)	(0.066)	(0.047)	(0.297)	(0.653)	(0.667)
Constant	0.150***	0.037	0.030	0.152***	0.102	0.055
	(0.000)	(0.790)	(0.865)	(0.000)	(0.440)	(0.734)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	288	238	238	288	238	238
R^2	0.018	0.037	0.055	0.003	0.030	0.049
	Treatment based on change in aid inflows, per capita US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period	-0.019	-0.006	0.074	-0.012	0.006	0.126
	(0.321)	(0.795)	(0.605)	(0.554)	(0.796)	(0.350)
Treatment	-0.029	-0.049	-0.050	-0.020	-0.029	-0.038
	(0.299)	(0.143)	(0.138)	(0.485)	(0.350)	(0.247)
2nd period * treatment	0.055**	0.048	0.044	0.040	0.026	0.035
	(0.048)	(0.107)	(0.134)	(0.154)	(0.386)	(0.237)
Constant	0.162***	0.129	0.105	0.158***	0.087	0.043
	(0.000)	(0.342)	(0.513)	(0.000)	(0.479)	(0.770)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	288	238	238	288	238	238
R^2	0.008	0.038	0.056	0.004	0.031	0.051

Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level.

The insignificant coefficients on the dummy variable for the second period indicate persistent dynamics of our dependent variable throughout the period of observation when considering the whole sample of recipient countries. Likewise, there is no evidence that the dynamics in GDP per capita differed between the treatment and control groups prior to the scaling up of aid, as reflected in the insignificant coefficients on the dummy variable for the treatment group. These results hold independently of whether we consider quality-adjusted aid in columns (1)-(3) or unadjusted aid in columns (4)-(6); they are also independent of whether we use our preferred measure of aid inflows in total US$ in the upper panel of Table 1.1. or the alternative measure of aid inflows per capita in the lower panel.

However, the upper panel of Table 1.1. provides strong indications that the treatment effect of increased aid depends on whether aid is quality-adjusted or not. The interaction between the dummy variables for the second period and the treatment group proves to be significant at least at the ten percent level for quality-adjusted aid in columns (1)-(3). What is more, the average treatment effect is quantitatively important. According to the results in column (3), the average effect of 0.063 amounts to 37 percent of the mean change in GDP per capita for the treatment group (see Table 1.10.).[31] In contrast, the interaction falls considerably short of reaching significance for unadjusted aid in columns (4)-(6). Furthermore, the *p*-values for the tests shown in Table D1.1. (see Appendix D1.) indicate statistically significant differences between the impact of quality-adjusted and unadjusted aid on the change in GDP per capita when the treatment is based on the change in total aid inflows.[32] Taken together, the quality of donors appears to matter for aid effectiveness. Only recipient countries with increased aid inflows of high quality, as revealed by Roodman's (2012) measure, benefit in terms of increasing GDP per capita.

The evidence on the treatment effect of quality-adjusted aid is weaker when considering aid inflows per capita in the lower panel of Table 1.1. The interaction between the dummy

[31] However, as we show in the next sub-section, our results on the effects of quality-adjusted aid are not out of sync with previous studies finding positive growth effects of aid, including Hansen and Tarp (2001) and Clemens et al. (2012).

[32] As explained in more detail in Appendix D1., we test whether the coefficient on the interaction term *2nd period*treatment* differs significantly between quality-adjusted aid and the corresponding estimation with unadjusted aid.

variables for the second period and the treatment group continues to be positive and significant (at the five percent level) in column (1). However, the interaction becomes insignificant when entering the control variables in columns (2) and (3). Moreover, the tests shown in Table D1.1. point to insignificant differences between the impact of quality-adjusted and unadjusted aid in per-capita terms.[33] As noted above, small countries are more likely to fall into the treatment group when the treatment is defined in per-capita terms. This could affect our results if the dynamics in GDP per capita differ between small and large recipient countries. Indeed, closer inspection reveals that particularly small recipient countries (with a population of less than one million) are over-represented in the treatment group when defining the treatment in per-capita terms. At the same time, the difference in GDP per capita was clearly below average for the sub-group of particularly small recipient countries. This provides a first indication that our baseline findings are largely driven by specific groups of recipient countries.[34]

Alternative dependent variables

In Table 1.2., we modify the definition of our dependent GDP variable. We use compound growth rates of GDP per capita (in percent) during the first and second periods in the upper two panels of Table 1.2. In the lower two panels, we use period averages of the logged level of GDP per capita as the dependent variable. In all other respects, the specification of the estimation equation is the same as in Table 1.1.[35]

---

[33] This does not necessarily imply, however, that the effectiveness of quality-adjusted aid strongly depends on whether it is measured in absolute or per-capita terms. In Table D1.2., we report another set of tests of differences in coefficients focusing on the statistical significance of the difference between estimates with total and per-capita aid. The *p*-values suggest that the coefficients on the interaction term *2nd period*treatment* do not differ significantly between the two measures of quality-adjusted aid. The same applies to unadjusted aid.

[34] We return to this issue below when interpreting the results for restricted samples in Table 1.7.

[35] The exception is that we have to drop GDP per capita in the initial year of sub-periods as a control variable when considering the level of GDP per capita as the dependent variable.

Table 1.2. Effects of aid on alternative GDP variables: full sample, median of difference in aid to separate treatment and control groups

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: compound growth rate of GDP per capita (%)					
	Treatment based on change in aid inflows, total US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	1.104**	0.916*	1.082**	0.452	0.196	0.227
	(0.020)	(0.077)	(0.046)	(0.345)	(0.709)	(0.664)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	288	236	236	288	236	236
R^2	0.017	0.037	0.057	0.002	0.031	0.051
	Dependent variable: compound growth rate of GDP per capita (%)					
	Treatment based on change in aid inflows, per capita US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.901*	0.795	0.733	0.633	0.414	0.614
	(0.058)	(0.122)	(0.142)	(0.185)	(0.429)	(0.233)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	288	236	236	288	236	236
R^2	0.007	0.038	0.057	0.004	0.032	0.053
	Dependent variable: log level of GDP per capita (period average)					
	Treatment based on change in aid inflows, total US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	-0.014	0.206**	0.231*	-0.017	0.106	0.115
	(0.628)	(0.034)	(0.051)	(0.550)	(0.287)	(0.332)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	284	236	236	284	236	236
R^2	0.268	0.518	0.524	0.191	0.474	0.479
	Dependent variable: log level of GDP per capita (period average)					
	Treatment based on change in aid inflows, per capita US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	-0.034	0.110	0.098	-0.027	-0.023	-0.008
	(0.231)	(0.253)	(0.308)	(0.355)	(0.819)	(0.941)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	284	236	236	284	236	236
R^2	0.188	0.493	0.497	0.098	0.437	0.443

Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level.

The effects of aid on the growth rate of GDP per capita closely resemble the effects on the change in GDP per capita shown in Table 1.1. Again, the contrast between quality-adjusted aid and unadjusted aid is most pronounced when the treatment is based on aid inflows in total US$. Our previous findings are also corroborated by the tests for differences in coefficients in Table D1.1.[36] The treatment effect of quality-adjusted aid amounts to about one percentage point of the annual growth rate of GDP per capita. This effect is comparable to our baseline results, considering that the increase in GDP per capita by 5.7-6.9 percent in Table 1.1. spans over six years. The growth effect may appear to be surprisingly large compared to previous aid-growth studies, including Hansen and Tarp (2001) and Clemens et al. (2012). All the same, our result is not out of sync once it is taken into account that the aid variable is defined as a percentage of the recipients' GDP in these studies. The above noted studies tend to find that an increase in the aid-to-GDP ratio by one percentage point is associated with an increase in growth by around 0.2 percentage points.[37] OECD-DAC statistics on net aid receipts suggest that the growth effect of an increase in the aid-to-GDP ratio by one percentage point should be compared to the effect of an increase in per-capita aid by 7.8 US$.[38] Recalling the difference of 37 US$ in per-capita aid between our treatment and control group, it appears plausible that the treatment effect of our DD analysis is about one percentage point of annual growth.

The evidence on the treatment effect of quality-adjusted aid is somewhat weaker when considering the logged level of GDP per capita as the outcome variable. Specifically, quality-adjusted aid loses its significance in the lower panels of Table 1.2. in the specification without control variables (column 1).[39] On the whole, however, the evidence in Table 1.2. suggests that our core finding, according to which recipient countries benefit only from increased aid inflows of high quality, is fairly robust to the exact definition of the GDP-related outcome variable.

[36] For details, see the *p*-values shown in Table D1.1. for the regressions with the compound growth rate as dependent variable in Table 1.2.

[37] However, Dalgaard et al. (2004: F208) find that "outside the tropics a one percentage point increase in the aid/GDP ratio leads to an increase in the growth rate in the neighbourhood of one percentage point."

[38] In the case of Sub-Saharan Africa, net aid receipts amounted to 5.5 percent of GDP or about 43 US$ per capita at the beginning of the second (treatment) period in 2005.

[39] Note also that the corresponding tests in Table D1.1. point to statistically insignificant differences between the coefficients on the interaction terms in columns 1 and 4 of Table 1.2.

Bilateral country programmable aid

As noted in Section 1.2., it is hardly possible to calculate the overall quality of aid at the bilateral level, i.e., for donor-recipient pairs. However, for the estimations reported in Table 1.3. we make use of the (limited) data on some specific aspects of the quality of aid at the bilateral level.[40] Specifically, we use bilateral aid data from the OECD-DAC on so-called country programmable aid (CPA). CPA resembles the first steps of Roodman's (2012) calculation of donor discounts insofar as CPA provides an estimate of the volume of aid resources actually transferred to recipient countries, e.g., by excluding debt relief and other items that cannot be used productively in aid recipient countries.[41]

Table 1.3. Effects of country programmable aid (CPA) on the change in GDP per capita: full sample, median of difference in CPA to separate treatment and control groups

Variables	Treatment based on change in CPA inflows, total US$			Treatment based on change in CPA inflows, per capita US$		
	(1)	(2)	(3)	(4)	(5)	(6)
2nd period	0.003	0.014	-0.048	0.001	0.011	-0.025
	(0.849)	(0.504)	(0.759)	(0.971)	(0.575)	(0.855)
Treatment	0.017	0.013	0.009	0.027	0.028	0.026
	(0.494)	(0.665)	(0.775)	(0.296)	(0.325)	(0.389)
2nd period *	0.016	0.011	0.021	0.022	0.016	0.022
treatment	(0.500)	(0.675)	(0.531)	(0.369)	(0.539)	(0.449)
Constant	0.119***	0.078	0.123	0.114***	0.082	0.116
	(0.000)	(0.498)	(0.422)	(0.000)	(0.445)	(0.396)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	272	235	235	272	235	235
R^2	0.011	0.042	0.061	0.021	0.055	0.074

Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level.

[40] The period of observation is shorter in Table 1.3., compared to Table 1.1., due to restricted data availability on CPA. Consequently, the first period had to be reduced to 2000-2005 and the second period to 2005-2010. In other respects, we followed the same procedures to identify treatment and control groups.

[41] More precisely, CPA subtracts from gross aid disbursements unpredictable flows of development assistance like humanitarian aid or debt relief, aid that remains in the donor country like administrative costs, student costs or refugee spending and aid that is not negotiated between donor and recipient governments like food aid. However, unlike the Roodman measure, CPA does not net out loan repayments (for further details see http://www.oecd.org/dac/aid-architecture/cpa.htm).

We employ bilateral CPA data to replicate the baseline estimations for quality-adjusted aid calculated according to Roodman's concept. The results on the effect of CPA are in contrast to the baseline findings reported for quality-adjusted aid in columns (1)-(3) of Table 1.1. In particular, the interaction between the dummy variables for the treatment group and the second period proves to be statistically insignificant in Table 1.3. However, this does not necessarily invalidate our finding that quality-adjusted aid tends to be more effective. Rather, it appears that the CPA-related adjustments to arrive at 'true' aid transfers cannot explain why we find quality-adjusted aid to be more effective.

Decomposing Roodman's measure of aid quality

Recalling that Roodman's concept of quality-adjusted aid is based on net aid plus several discounts, we tentatively assess in the following whether our baseline results on the effectiveness of quality-adjusted aid are driven by one particular element. To this end, we replicate the baseline estimations on quality-adjusted aid in Table 1.1. in several ways: (i) by considering Roodman's measure of net aid (this measure excludes repayments as well as debt relief which is part of aid according to OECD-DAC guidelines), without any further discounts; (ii) by considering quality-adjusted aid, but excluding the discount for tying aid; (iii) by considering quality-adjusted aid, but excluding the selectivity-related discount; and (iv) by considering quality-adjusted aid, but excluding the proliferation-related discount. The results of these estimations are summarized in Table 1.4.

We find that it is already Roodman's measure of net aid that renders aid effective. The interaction terms, *2^{nd} period*treatment*, prove to be significantly positive in columns 1-4 in the upper panel of Table 1.4. This is perhaps surprising, recalling the statistically insignificant results for CPA in Table 1.3. However, this difference can be attributed to Roodman's measure of net aid, while CPA still includes repayments. It is also interesting to note that the exclusion of specific discounts related to aid tying, selectivity and proliferation hardly affects our baseline results on the effectiveness of quality-adjusted aid. In other words, there is no evidence in the lower panels of Table 1.4. that a particular discount factor is driving our baseline results.

Table 1.4. Effects of aid on the change in GDP per capita: decomposition of Roodman's quality measure

Variables	Treatment based on change in aid inflows, total US$			Treatment based on change in aid inflows, per capita US$		
	(1)	(2)	(3)	(4)	(5)	(6)
2nd period * treatment						
	Net aid, no further discounts					
	0.077***	0.063**	0.070**	0.060**	0.045	0.041
	(0.006)	(0.032)	(0.020)	(0.033)	(0.124)	(0.146)
	Quality-adjusted aid, excluding discount for tied aid					
	0.077***	0.068**	0.079**	0.069**	0.064**	0.065**
	(0.006)	(0.023)	(0.012)	(0.013)	(0.036)	(0.036)
	Quality-adjusted aid, excluding selectivity-related discount					
	0.076***	0.063**	0.073**	0.055**	0.040	0.039
	(0.006)	(0.036)	(0.015)	(0.049)	(0.162)	(0.169)
	Quality-adjusted aid, excluding proliferation-related discount					
	0.080***	0.066**	0.075**	0.055**	0.041	0.039
	(0.004)	(0.024)	(0.011)	(0.048)	(0.153)	(0.170)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes

Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level.

Excluding donor countries

In the next step, we check whether our baseline finding of effective quality-adjusted aid, compared to ineffective unadjusted aid, is driven by the bilateral aid relations of one particular donor. This is done by excluding one donor at a time from the estimations on the effects of quality-adjusted and unadjusted aid on the change in GDP per capita. That is, we replicate the calculation of quality-adjusted aid 28 times after excluding one particular donor from our sample of 28 donors. In all other respects the estimations summarized in Table 1.5. resemble those reported in column (3) of the upper part of Table 1.1. For the sake of brevity we only show the interaction of the dummy variables for the treatment group and the second period when the treatment is based on absolute inflows of aid. The results are similar when the treatment is based on per-capita inflows of aid.[42]

[42] The full set of estimation results is available from the authors on request.

Table 1.5. Effects of aid on the change in GDP per capita: excluding one donor at a time

Excluded donor:	Quality-adjusted aid		Unadjusted aid	
Sweden	0.080**	(0.013)	0.026	(0.409)
Ireland	0.064**	(0.044)	0.015	(0.623)
Denmark	0.068**	(0.034)	0.027	(0.384)
Luxembourg	0.064**	(0.044)	0.017	(0.586)
Netherlands	0.073**	(0.020)	0.023	(0.458)
Portugal	0.063**	(0.048)	0.014	(0.642)
United Kingdom	0.071**	(0.022)	0.042	(0.163)
Switzerland	0.079**	(0.013)	0.019	(0.544)
Finland	0.064**	(0.045)	0.015	(0.628)
Norway	0.065**	(0.041)	0.030	(0.354)
New Zealand	0.087***	(0.006)	0.015	(0.615)
Australia	0.068**	(0.032)	0.026	(0.406)
Canada	0.067**	(0.035)	0.024	(0.424)
Belgium	0.068**	(0.033)	0.022	(0.485)
Germany	0.070**	(0.030)	0.043	(0.174)
United States	0.039	(0.211)	0.021	(0.488)
Iceland	0.064**	(0.044)	0.015	(0.625)
Austria	0.063**	(0.046)	0.021	(0.508)
Japan	0.077**	(0.018)	0.065**	(0.042)
Greece	0.074**	(0.019)	0.015	(0.627)
France	0.060*	(0.079)	0.014	(0.664)
Spain	0.063*	(0.050)	0.026	(0.403)
Italy	0.079**	(0.013)	0.014	(0.658)
Korea	0.060*	(0.060)	0.013	(0.673)
Czech Rep.	0.064**	(0.044)	0.016	(0.616)
Poland	0.064**	(0.044)	0.023	(0.468)
Slovenia	0.064**	(0.044)	0.015	(0.625)
Slovak Rep.	0.064**	(0.044)	0.015	(0.625)

The specification of estimations shown in this table corresponds to the full specification in column (3) of Table 1.1. Excluded donors are listed in the order of Figure 1.1. on average discount factors. For the sake of brevity we only show the interaction of the dummy variables for the treatment group and the 2nd period when the treatment is based on absolute inflows of aid. The results are similar when the treatment is based on per-capita inflows of aid. Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level.

As can be seen, the exclusion of individual donors typically affects our baseline results on the effectiveness of quality-adjusted aid only marginally. However, it appears that the United States is largely driving our results.[43] The interaction between the dummy variables for the

[43] It should also be noted that the exclusion of Japan results in a significantly positive interaction between the dummy variable for the second period and the treatment dummy based on unadjusted aid. The effect of quality-adjusted aid on the change in GDP per capita is just slightly higher in this case.

treatment group and the second period proves to be statistically insignificant at conventional levels when excluding the United States – not only for unadjusted aid, but also for quality-adjusted aid.[44] The important role of the United States is not particularly surprising: the United States is not only the largest donor by far in terms of absolute aid flows; the US also stands out in terms of a considerable decline in the discount factor after the Paris Declaration in 2005, as shown in Figure 1.2.

Modified definition of treatment and control groups

In Table 1.6., we modify the definition of treatment and control groups and then re-estimate the same set of model specifications as in Table 1.1. Instead of using the median as the dividing line between treatment and control groups, we widen the gap between treatment and control groups by excluding countries with a change in aid inflows relatively close to the median and comparing the remaining top tercile of sample countries with the bottom tercile. The dummy variable for the treatment group proves to be significantly negative in several estimations in Table 1.6., suggesting that the dynamics of GDP per capita were weaker for countries in the treatment than in the control group during the first period, i.e., before donors scaled up (quality-adjusted) aid. Most importantly, however, previous results on the interaction between the two dummy variables are hardly affected by this modification. As before, the treatment effect of unadjusted aid proves to be consistently insignificant in columns (4)-(6) of Table 1.6. The contrast to quality-adjusted aid is again particularly pronounced in the upper panel of Table 1.6. when using aid inflows in total US$ to separate treatment and control groups.[45] The results on the interaction terms for quality-adjusted aid in per-capita terms are slightly stronger in the lower panel of Table 1.6., compared to the corresponding estimations in the lower panel of Table 1.1.

---

[44] However, the interaction is just marginally insignificant for quality-adjusted aid in the basic specification without control variables (corresponding to column (1) in Table 1.1.; not shown in Table 1.5.).

[45] However, the *p*-value of the test reported in Table D1.1. does not point to a statistically significant difference between quality-adjusted and unadjusted aid for the fully specified model when comparing columns 3 and 6 in the upper panel of Table 1.6.

Table 1.6. Effects of aid on the change in GDP per capita: full sample, top and bottom terciles as treatment and control groups

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Treatment based on change in aid inflows, total US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period	-0.044*	-0.025	0.062	0.006	0.014	0.240
	(0.067)	(0.407)	(0.777)	(0.782)	(0.575)	(0.158)
Treatment	-0.060*	-0.084*	-0.095*	-0.033	-0.040	-0.055
	(0.089)	(0.073)	(0.080)	(0.360)	(0.341)	(0.240)
2nd period * treatment	0.095***	0.076*	0.096**	0.038	0.039	0.064
	(0.006)	(0.057)	(0.040)	(0.242)	(0.278)	(0.120)
Constant	0.179***	0.046	0.010	0.170***	0.042	-0.056
	(0.000)	(0.788)	(0.966)	(0.000)	(0.795)	(0.793)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	198	154	154	192	166	166
R^2	0.020	0.096	0.121	0.011	0.070	0.120
	Treatment based on change in aid inflows, per capita US$					
	Quality-adjusted aid			Unadjusted aid		
2nd period	-0.032	-0.009	0.295	0.010	0.010	0.139
	(0.151)	(0.744)	(0.120)	(0.642)	(0.694)	(0.453)
Treatment	-0.069**	-0.086**	-0.077*	-0.012	-0.032	-0.053
	(0.039)	(0.049)	(0.093)	(0.746)	(0.446)	(0.241)
2nd period * treatment	0.095***	0.075*	0.055	0.013	0.027	0.052
	(0.007)	(0.057)	(0.176)	(0.707)	(0.492)	(0.212)
Constant	0.170***	0.185	0.073	0.156***	0.107	0.069
	(0.000)	(0.271)	(0.738)	(0.000)	(0.530)	(0.736)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	204	156	156	190	161	161
R^2	0.024	0.064	0.098	0.003	0.045	0.084

Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level.

Restricted samples

The estimations reported in Table 1.7. are based on restricted samples to assess more systematically whether specific groups of recipient countries are driving our baseline results. For the sake of brevity, we use the preferred measure of aid inflows in total US$ in these estimations to separate the treatment and control groups.[46] In the first panel at the top of Table 1.7., we exclude 33 countries with a population of less than one million. The results of this exercise underscore the important role of small recipient countries for our baseline findings. The size of the coefficients on the interaction between the dummy variables for the second period and the treatment group still appear to be higher in columns (1)-(3) for quality-adjusted aid than in columns (4)-(6) for unadjusted aid. However, the treatment effects fail to reach statistical significance for quality-adjusted aid, too. This suggests that mainly small recipient countries benefit from being treated with increases in quality-adjusted aid. In sharp contrast, the exclusion of the 30 largest recipient countries in the second panel of Table 1.7. hardly affects the baseline findings shown in Table 1.1.[47]

Table 1.7. Effects of aid on the change in GDP per capita: restricted samples

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Smallest countries (population < 1 million) excluded					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.039	0.039	0.053	0.016	0.023	0.035
	(0.221)	(0.213)	(0.118)	(0.624)	(0.459)	(0.279)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	221	192	192	221	192	192
R^2	0.017	0.092	0.125	0.003	0.078	0.112
	Largest countries (population > 23 million) excluded					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.094***	0.073*	0.081**	0.047	0.019	0.023
	(0.004)	(0.052)	(0.035)	(0.160)	(0.620)	(0.549)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes

46 The median to separate treatment and control groups is calculated based on the restricted sample, after the exclusion of countries.

47 While Table 1.7. does not show the results when using the alternative measure of aid inflows in per-capita terms, we tested for significant differences between total and per-capita aid within the sub-samples of smaller and larger countries. Similar to the tests for differences in coefficients reported in Table D1.2. for the full sample, the corresponding tests for these sub-samples suggest that the coefficients on the interaction term *2nd period*treatment* do not differ significantly. In particular, it appears that the effectiveness of quality-adjusted aid in the sub-sample of smaller countries does not depend strongly on the measurement of aid (see Table D1.3. for details).

Observations	228	183	183	228	183	183
R^2	0.022	0.060	0.082	0.005	0.049	0.072
	Countries with particularly good governance excluded					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.060*	0.067**	0.075**	0.011	0.022	0.026
	(0.053)	(0.042)	(0.040)	(0.717)	(0.520)	(0.447)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	228	200	200	228	200	200
R^2	0.012	0.050	0.089	0.007	0.048	0.088
	Countries with particularly bad governance excluded					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.068**	0.046	0.066*	0.030	-0.001	0.014
	(0.027)	(0.194)	(0.085)	(0.328)	(0.976)	(0.708)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	228	186	186	228	186	186
R^2	0.051	0.035	0.058	0.026	0.018	0.038
	Part II countries excluded; median to separate treatment and control groups					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.037	0.046	0.068**	0.015	0.030	0.045
	(0.195)	(0.124)	(0.043)	(0.592)	(0.315)	(0.162)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	250	223	223	250	223	223
R^2	0.010	0.039	0.056	0.007	0.036	0.052
	Part II countries excluded; top and bottom terciles as treatment and control group					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.048	0.051	0.077*	0.029	0.056	0.069
	(0.180)	(0.180)	(0.070)	(0.415)	(0.160)	(0.114)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	170	149	149	168	146	146
R^2	0.016	0.081	0.096	0.013	0.096	0.136

Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level. We do not show results for the dummy variables for the 2nd period and the treatment group and the constant term to avoid clutter.

On the contrary, we do not find compelling evidence that our baseline results are driven by higher effectiveness of quality-adjusted aid in well governed recipient countries.[48] In the third and fourth panel of Table 1.7., we exclude 30 recipient countries which performed either

[48] We are grateful to an anonymous reviewer for having alerted us to this possibility.

particularly poorly or particularly well in terms of controlling for corruption.[49] On the one hand, the interaction between the dummy variables for the second period and the treatment group is slightly stronger in column (1), without control variables included, for the reduced sample when excluding the 30 countries with particularly bad governance. On the other hand, the interaction is slightly weaker in columns (2) and (3) when excluding the 30 countries with particularly good governance. Importantly, the higher effectiveness of quality-adjusted aid, compared to unadjusted aid in columns (4)-(6), holds independently of whether recipient countries with particularly good or bad governance are excluded from the sample.

In the fifth and sixth panel of Table 1.7., we exclude 19 so-called Part II countries from the sample.[50] This group was introduced by the OECD's DAC after the end of the Cold War in 1993; it consisted of relatively advanced developing countries and transition economies in Eastern Europe. When the DAC reverted to a single list of aid recipients in 2005, most Part II countries did no longer receive official development assistance and some of them had become new donors. Hence, former Part II countries typically fall into the control group in our estimations. At the same time, the average change in GDP per capita was relatively small for this group.

When excluding Part II countries, the interaction terms between the dummy variables for the second period and the treatment with quality-adjusted aid lose their significance in columns (1) and (2). Importantly, however, the treatment effect for quality-adjusted aid proves to be significantly positive in column (3) where the specification includes both the control variables and their interaction with the dummy variable for the second period.

To summarize, Table 1.7. reveals that our baseline results are sensitive to sample selection. This is not exceptional in the aid effectiveness literature. For instance, Easterly et al. (2004) challenged the major result of the seminal paper of Burnside and Dollar (2000) that aid has a positive impact on economic growth in recipient countries pursuing sound economic policies. Easterly et al. (2004) show that this result no longer holds when adding additional countries and

[49] Results are similar when we use democracy instead of control for corruption.

[50] In this specification, we only keep countries in our sample that were on the official DAC list of ODA recipients in 2011. This list is revised every three years (for details see http://www.oecd.org/dac/stats/daclist.htm).

observations to the dataset of Burnside and Dollar (2000). Clearly, sample selection also deserves closer attention in further research on the role of donor quality for the effectiveness of aid.

Delayed effects

In the final step of our analysis, we consider delayed effects of quality-adjusted aid and unadjusted aid on the change in GDP per capita in the recipient countries. Again, we only report the results based on the preferred measure of aid inflows in total US$. Table 1.8. shows the results with lags of one year, two years, and three years. Put differently, the change in GDP per capita during the second period no longer relates to 2005-2011 as in our baseline estimations; instead it relates to 2006-2012, 2007-2013, and 2008-2014, respectively. Estimations in the upper three panels of the table are based on the full sample, while estimations in the lower three panels are based on the reduced sample after excluding Part II countries.

Table 1.8. Effects of aid on the change in GDP per capita: delayed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Total sample, lagged by one year					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.094***	0.091***	0.075**	0.051*	0.046	0.026
	(0.002)	(0.004)	(0.019)	(0.092)	(0.158)	(0.410)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	286	238	238	286	238	238
R^2	0.039	0.047	0.066	0.009	0.027	0.053
	Total sample, lagged by two years					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.096***	0.098***	0.074**	0.056*	0.054	0.029
	(0.003)	(0.005)	(0.031)	(0.084)	(0.122)	(0.383)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	286	238	238	286	238	238
R^2	0.049	0.056	0.078	0.020	0.037	0.066
	Total sample, lagged by three years					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.095***	0.106***	0.078**	0.057*	0.067*	0.039
	(0.005)	(0.004)	(0.030)	(0.090)	(0.072)	(0.272)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	283	237	237	283	237	237
R^2	0.049	0.069	0.097	0.022	0.048	0.083
	Part II countries excluded; lagged by one year					

	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.057*	0.070**	0.071**	0.026	0.048	0.047
	(0.069)	(0.025)	(0.031)	(0.399)	(0.136)	(0.148)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	248	223	223	248	223	223
R^2	0.011	0.038	0.053	0.004	0.030	0.048
	Part II countries excluded; lagged by two years					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.058*	0.071**	0.067**	0.023	0.043	0.038
	(0.087)	(0.037)	(0.049)	(0.500)	(0.224)	(0.266)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	248	223	223	248	223	223
R^2	0.010	0.040	0.059	0.002	0.031	0.054
	Part II countries excluded; lagged by three years					
	Quality-adjusted aid			Unadjusted aid		
2nd period * treatment	0.066*	0.079**	0.076**	0.034	0.055	0.050
	(0.063)	(0.028)	(0.036)	(0.339)	(0.141)	(0.167)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	247	222	222	247	222	222
R^2	0.014	0.047	0.074	0.005	0.038	0.068

Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level. We do not show results for the dummy variables for the 2nd period and the treatment group and the constant term to avoid clutter.

The treatment effects of aid on the change in GDP per capita generally become stronger, compared to the baseline results, when accounting for lags. Some interaction terms even turn out to be significant at the ten percent level when the treatment relates to unadjusted aid. However, this applies only to estimations for the full sample in columns (4) and (5). The treatment with unadjusted aid is no longer significant for the fully specified model in column (6), and after excluding Part II countries in the lower panels of Table 1.8.

What is more, the results in Table 1.8. indicate consistently that the quality of donors matters for the effectiveness of aid. The size of the coefficients on the interaction terms as well as their significance typically increases compared to the corresponding baseline results. The average treatment effect reaches its maximum of 0.078 with three-years lags for the total country sample in column (3), compared to the corresponding treatment effect of 0.063 in Table 1.1. This underscores the quantitative importance of the treatment with quality-adjusted aid.

Table 1.9. Definition of variables and data sources

Variable	Description	Source
2nd period	Dummy for treatment period (2005-2011)	Own definition
Treatment	Dummy for treatment group; see text for alternative definitions	Own definition
Unadjusted aid	Bilateral net ODA disbursements in constant 2012 US$	OECD DAC Statistics 2014
Quality-adjusted aid	Quality-adjusted aid in constant 2012 US$, as defined by Roodman (2012)	Own calculation based on Roodman (2012)
Change in (log) GDP per capita	Six year difference in logged GDP per capita in period 1: 2005-1999 and period 2: 2011-2005	Own calculation based on World Development Indicators (2015)
(log) GDP per capita	Natural log of GDP per capita (constant 2012 US$), first non-missing value in each period	World Development Indicators (2015)
Corruption	Control of corruption, first non-missing value in each period	World Development Indicators (2015)
(log) Inflation	Natural log of inflation, first non-missing value in each period	World Development Indicators (2015)
Openness	(exports + imports)/GDP, first non-missing value in each period	World Development Indicators (2015)
Multilateral aid	Share of multilateral aid in total aid	OECD DAC

Table 1.10. Descriptive statistics by treatment (T) and control (C) group (based on Table 1, columns (1)-(3), upper panel)

Variable	Observations	Mean	Std. dev.	Min	Max
2nd period T	144	1	1	0	1
2nd period C	144	1	1	0	1
Treatment T	144	1	0	1	1
Treatment C	144	0	0	0	0
Quality-adjusted aid (mean) T	144	218000000	246000000	0	1070000000
Quality-adjusted aid (mean) C	144	72700000	130000000	0	730000000
Quality-adjusted aid (difference) T	72	170000000	198000000	11700000	955000000
Quality-adjusted aid (difference) C	72	-26000000	69000000	-434000000	11600000
Change in (log) GDP per capita T	144	0.17	0.15	-0.44	0.55
Change in (log) GDP per capita C	144	0.14	0.18	-0.60	0.69
(log) GDP per capita T	142	7.27	1.04	5.43	9.43
(log) GDP per capita C	142	8.63	1.23	6.11	11.43
Corruption T	142	-0.54	0.58	-1.57	1.54
Corruption C	143	-0.06	0.80	-1.52	2.29
(log) Inflation T	140	2.04	1.07	-1.13	5.71
(log) Inflation C	121	1.52	0.99	-1.70	5.52
Openness T	140	74.06	37.44	20.98	270.36
Openness C	137	101.35	58.36	30.12	422.33
Multilateral aid T	143	33.83	18.48	0	100
Multilateral aid C	123	32.39	24.33	0	100

1.5. SUMMARY AND CONCLUSION

Compared to the intensive debate on the role of policies in the recipient countries for the effectiveness of foreign aid, it has received only scant attention whether aid could induce stronger income effects if donors provided their assistance in higher quality. As argued by Roodman (2012), donors could improve the quality of aid in various ways, inter alia, by untying aid, allocating aid selectively, avoiding project proliferation, and supporting private giving. We use Roodman's ranking of donors along these dimensions to derive quality-adjusted aid inflows for a large sample of recipient countries during the period 1999-2011. We compare the effects of quality-adjusted aid and unadjusted aid as reported by the donors on changes in GDP per capita in the recipient countries by performing a difference-in-differences (DD) analysis.

We find fairly strong indications that quality-adjusted aid is more effective than unadjusted aid in inducing income gains in the recipient countries. The DD analysis for the overall country sample reveals statistically significant and quantitatively important treatment effects for quality-adjusted aid, while we do not find significant treatment effects for unadjusted aid. Only recipient countries with increased aid inflows of higher quality benefit in terms of increasing GDP per capita. Allowing for a delayed impact on changes in GDP per capita, the treatment effects become stronger, compared to the baseline results, corroborating that the quality of aid inflows matters for aid effectiveness.

However, our baseline results are sensitive to sample selection. In particular, it appears that mainly small recipient countries benefit from being treated with increases in quality-adjusted aid. Arguably, relatively altruistic donors providing higher quality aid focus on smaller and better governed recipient countries to achieve more impact. In contrast, relatively selfish donors providing lower quality aid may prefer larger and richer recipient countries to foster their own trade-related interest. Tentative findings based on simple aid allocation models (see Appendix B1. for details) do not provide compelling evidence to this effect. In particular, the small country bias of aid allocation by donors providing lower quality aid appears to be just slightly weaker than that of donors providing higher quality aid. Yet the role of donor characteristics clearly deserves more attention. Another possibility is that large recipient countries see no choice but to accept aid from all available sources, independent of the donors' quality ranking, in order to meet their perceived financing needs. Such an approach appears to be shortsighted, however, as long as various sources supply low quality aid. Then again, small countries might not have much bargaining power resulting in only limited choice. Theoretical arguments hence do not seem to offer fully conclusive explanations and future research may help clarify why it seems to be easier for small recipient countries to benefit from higher quality aid.

Nevertheless, the necessity for the donors obviously is to improve the quality of their aid, rather than 'only' scaling up nominal aid budgets. The considerable discounts shown for various donors in Figure 1.1. and the limited progress, if any, made so far in reducing these discounts reveal the untapped potential for improving the effectiveness of aid in this way. At the same time, further research could pursue two avenues to provide a fuller account of aid-related donor policies. On the one hand, a refined analysis of specific elements of Roodman's donor ranking may offer deeper insights even though we did not find that particular discounts (related to aid tying, donor selectivity and aid proliferation) are driving our results on the effectiveness of quality-adjusted aid. On the other hand, a broader view on donor policies, e.g., with regard to trade liberalization and market access, seems to be required to identify complementarities that may render aid more effective.

APPENDIX FOR CHAPTER 1

Appendix A1. Testing for differences in the quality of aid between the treatment and control group of recipient countries

As noted in Section 1.2., the data situation as well as conceptual issues do not allow for a comprehensive calculation of the quality of aid at the bilateral level, i.e., for donor-recipient pairs. This is why we draw on Roodman's (2012) aggregate measure of a donor's aid quality. The limitation of this approach would be less serious if donors gave aid of the same quality to all recipient countries. In the present context, it appears to be of particular importance that the quality of aid from specific donors does not differ significantly between recipients in the treatment group and recipients in the control group. In this appendix, we evaluate the justification of using an aggregate, donor-based measure of aid quality in several steps.

First of all, we perform group comparison t-tests for significant differences between the means of the donors' quality-related aid shares for the treatment group versus the control group (in the first and second sub-periods). Table A1.1. presents the results of this exercise for the bilateral shares of tied aid as well as the bilateral shares of general budget support. The share of tied aid in net aid proxies the tying status in bilateral aid relations.[51] We consider general budget support as a proxy of (the inverse of) aid proliferation, assuming that aid proliferation is less likely when donors grant mainly general budget support.[52] As can be seen, significant differences in these dimensions of aid quality between the treatment and control group are rare exceptions. This provides a first indication that our approach of applying an aggregate measure of aid quality may not be unduly restrictive.

Second, as for proliferation we also checked whether donor countries that tackled this issue by reducing the number of so-called non-significant aid relations did so in their bilateral relations with recipient countries in both the treatment and the control group. The OECD's Development Assistance Committee (DAC) distinguishes significant and non-significant aid relations by two criteria: (i) whether a donor provides more than its global aid share to a particular recipient country and (ii) whether a donor is among the top donors that cumulatively provide 90 percent

[51] The calculation of this share is based on bilateral aid data drawn from AidData, 2.0 release. Note that this share is not directly comparable with the tying penalty as provided by Roodman (2012).

[52] The calculation of this share is based on DAC data on bilateral aid commitments in the Creditor Reporting System.

of total aid to a particular recipient country. Bilateral aid relations in which a donor fails to meet both criteria at the same time are classified to be non-significant. The DAC argues that the fragmentation of aid is closely related to the number of non-significant aid relations.

The change in the number of non-significant aid relations is available from the DAC in a consistent way for country programmable aid during the period 2007-2014.[53] We draw on this source to assess whether donors that improved the quality of their aid by reducing the number of non-significant relations did so in both sub-groups of recipient countries. This indeed appears to be the case when considering the change in non-significant aid relations between 2007 and 2014 for all 28 donors in our sample taken together. The number of non-significant relations with recipients in the treatment group declined only modestly from 450 to 416 during this period (i.e., by about 8 percent). Importantly, the decline in the number of non-significant relations (from 249 to 192) was no less pronounced for recipients in the control group.

Figure A1.1. portrays the change in the number of non-significant relations for individual donor countries with recipients in the treatment group and those in the control group. The first thing to note is that few donors in our sample helped improve the quality of aid by reducing the number of non-significant aid relations considerably. Canada clearly stands out in this respect. Importantly, the strong decline in Canada's non-significant relations was almost equally split between recipients in the treatment group and those in the control group. Likewise, Spain and Denmark which came closest to Canada in terms of reducing fragmentation did so in both groups of recipient countries. In striking contrast, fragmentation further increased for some donors such as Australia and the United Kingdom. Again, however, recipients were treated similarly by these donors; the number of non-significant aid relations increased in both groups. Considerable changes of fragmentation going in opposite directions for the two recipient groups are rare.[54]

Third, we estimate donor-specific aid allocation models to assess the selectivity of their bilateral aid relations and, thus, the justification to apply the same discounts for granting aid to richer and

[53] The data are available from: https://www.oecd.org/dac/aid-architecture/fragmentation-orphans.htm (accessed: December 15, 2016).

[54] Where they occurred, as for the Netherlands and Sweden, it is interesting to note that fragmentation was reduced in the control group and not – as one could have expected – in the treatment group of recipient countries.

relatively corrupt countries in the treatment and control group. More precisely, we estimate a simple aid allocation model with (logged) GDP per capita, governance, population, and dummy variables for the second period as well as the treatment group of recipients on the right-hand side for (almost) all the donor countries in our sample.[55] Importantly, we interact the standard determinants of aid allocation with the dummy variables to assess whether their impact differed between the treatment and control group of recipient countries. This does not appear to be the case, again with very few exceptions (Table A1.2.).[56] Hence, it seems to be justified to apply the same penalty for granting aid to richer and relatively corrupt countries in both groups.

Taken together, we find surprisingly strong evidence that donors give aid of similar quality to recipients in the treatment and control group. We conclude that the limitation of employing the aggregate measure of aid quality is less serious than one could have expected.

Table A1.1. – *t*-tests (*p*-values) for significant differences between the means of donors' aid shares: treatment versus control group of recipient countries in the first and second sub-periods

	Share of tied aid		Share of general budget support	
	1st sub-period	2nd sub-period	1st sub-period	2nd sub-period
Australia	0.814	0.719	0.195	0.904
Austria	0.364	0.005**	0.219	0.437
Belgium	0.722	0.273	0.076	0.356
Canada	0.639	0.072	0.195	0.081
Czech Rep.	0.827	0.302	0.967	0.196
Denmark	0.127	0.044*	0.144	0.306
Finland	0.620	0.577	0.361	0.247
France	0.830	0.172	0.156	0.059
Germany	0.293	0.005**	0.564	0.309
Greece	0.127	0.964	0.447	0.149
Ireland	-	-	0.368	0.262
Italy	0.118	0.993	0.535	0.297
Japan	0.210	-	0.045*	0.663
Korea, Rep.	n.a.	0.479	n.a.	-
Luxembourg	-	0.518	-	0.487
Netherlands	0.672	0.853	0.185	0.303
New Zealand	0.298	0.448	0.847	0.783
Norway	0.188	0.428	0.812	0.110
Portugal	0.520	0.099	0.290	0.740
Spain	0.838	0.340	0.209	0.664
Sweden	0.017*	0.446	0.142	0.070

[55] The estimations could not be run for a few donors with insufficient data.

[56] We performed the same regression with control of corruption, instead of democracy, as an indicator of the quality of institutions (not shown). Results are affected only modestly by this modification.

United Kingdom	-	-	0.103	0.032*
United States	0.597	0.362	0.489	0.630

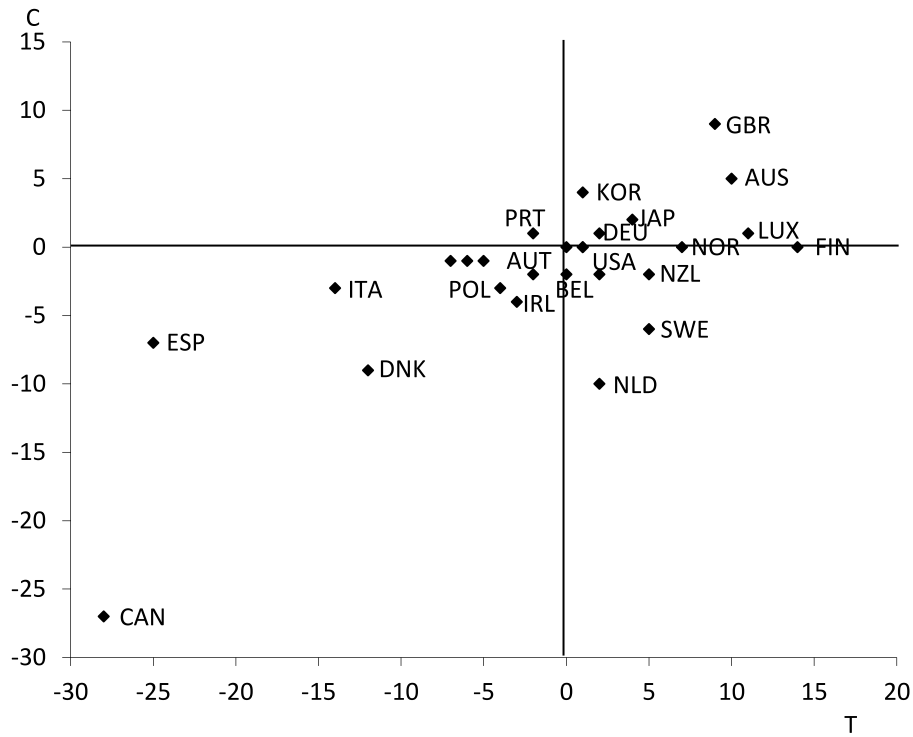
Notes: Number of observations differs between listed donors and types of aid in sub-periods, depending on availability of bilateral aid data. Some (non-traditional) donor countries not listed due to missing bilateral aid data. *, ** significantly different means test in aid shares at the 5-percent and 1-percent level, respectively. "-" when aid shares are 0.000 for both the treatment and control group of recipient countries and *p*-values are not calculated by Stata.

Table A1.2. Aid allocation: regression results for DAC donor countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	Australia	Austria	Belgium	Canada	Denmark	Spain	Finland	Germany	France	United Kingdom	Greece	Switzer-land	Ireland	Italy	Japan	Korea	Luxem-bourg	Nether-lands	Norway	New Zealand	Portugal	Sweden	USA	Czech Rep.
GDPpc (log)	0.027	-0.436*	-0.823***	-0.522*	-0.613	-0.827**	-0.508*	-0.361**	-0.185	-0.476**	-0.029	-1.024***	-0.621**	-0.623**	-0.683**	-0.311	-0.264	-0.940**	-0.406	-0.371	-0.408	-0.612	-0.995***	0.134
	(0.287)	(0.183)	(0.209)	(0.227)	(0.441)	(0.249)	(0.231)	(0.137)	(0.142)	(0.172)	(0.262)	(0.234)	(0.230)	(0.216)	(0.213)	(0.171)	(0.293)	(0.299)	(0.309)	(0.224)	(0.384)	(0.314)	(0.276)	(0.275)
Population (log)	0.676*	0.368	0.284*	0.663***	0.386**	0.625**	0.676***	1.027***	0.567***	0.510**	0.075	0.339*	0.316***	0.372*	0.955***	0.670***	0.213	0.462**	0.482***	0.327	-0.144	0.371**	0.695***	0.202
	(0.287)	(0.213)	(0.128)	(0.118)	(0.122)	(0.197)	(0.087)	(0.095)	(0.105)	(0.169)	(0.142)	(0.156)	(0.087)	(0.162)	(0.142)	(0.125)	(0.150)	(0.169)	(0.119)	(0.188)	(0.198)	(0.119)	(0.189)	(0.163)
Democracy (polity)	0.025	0.138	0.145	0.159**	-0.033	0.168	0.078	0.140**	0.021	0.093	0.006	0.085	0.070	0.056	0.087	0.035	0.047	0.140	0.106	0.068	-0.139	0.219**	0.210**	0.058
	(0.107)	(0.077)	(0.075)	(0.061)	(0.090)	(0.098)	(0.062)	(0.048)	(0.051)	(0.059)	(0.073)	(0.071)	(0.065)	(0.076)	(0.054)	(0.055)	(0.077)	(0.089)	(0.068)	(0.073)	(0.106)	(0.079)	(0.080)	(0.049)
Treated * GDPpc (log)	-0.815	0.225	0.537	0.024	-0.096	1.174**	-0.153	0.139	0.373	-0.460	0.138	0.350	-0.212	0.529	0.358	0.279	-0.067	0.085	-0.389	0.095	-0.105	0.192	0.298	0.088
	(0.450)	(0.277)	(0.318)	(0.273)	(0.507)	(0.352)	(0.294)	(0.190)	(0.259)	(0.301)	(0.329)	(0.305)	(0.331)	(0.314)	(0.310)	(0.235)	(0.355)	(0.380)	(0.366)	(0.311)	(0.498)	(0.380)	(0.357)	(0.368)
Treated * Population (log)	-0.448	0.163	0.173	0.188	-0.045	-0.098	-0.246	-0.239	0.009	0.447*	-0.039	0.185	0.155	0.094	-0.189	-0.180	-0.240	0.000	0.127	-0.217	-0.284	0.156	0.010	-0.060
	(0.385)	(0.263)	(0.185)	(0.158)	(0.213)	(0.262)	(0.162)	(0.153)	(0.165)	(0.212)	(0.205)	(0.199)	(0.170)	(0.210)	(0.183)	(0.172)	(0.218)	(0.241)	(0.180)	(0.251)	(0.329)	(0.180)	(0.229)	(0.248)
Treated * Democracy (polity)	0.060	-0.121	-0.171	-0.111	0.019	-0.077	0.016	-0.087	-0.148	0.067	-0.006	0.013	0.003	-0.151	0.115	-0.063	-0.021	0.018	-0.027	-0.038	0.319*	-0.257*	-0.071	-0.094
	(0.184)	(0.099)	(0.118)	(0.086)	(0.118)	(0.134)	(0.087)	(0.064)	(0.093)	(0.098)	(0.091)	(0.089)	(0.108)	(0.104)	(0.097)	(0.085)	(0.108)	(0.108)	(0.096)	(0.115)	(0.145)	(0.105)	(0.111)	(0.088)
2nd Period * GDPpc (log)	-0.362	0.615*	0.336	0.005	0.463	0.442	0.056	0.289	-0.033	0.426	0.314	0.253	0.024	0.070	0.123	0.126	-0.117	0.374	-0.121	-0.196	-0.070	0.017	0.490	0.388
	(0.367)	(0.311)	(0.306)	(0.311)	(0.631)	(0.469)	(0.295)	(0.255)	(0.270)	(0.241)	(0.316)	(0.337)	(0.358)	(0.346)	(0.301)	(0.329)	(0.344)	(0.427)	(0.418)	(0.574)	(0.578)	(0.449)	(0.372)	(0.581)
2nd Period * Population (log)	-0.049	0.244	0.240	-0.165	0.073	0.039	-0.140	0.113	0.076	0.098	0.372*	0.198	-0.191	-0.093	-0.143	0.133	0.225	-0.023	0.001	0.028	0.268	-0.012	0.156	0.174
	(0.380)	(0.314)	(0.178)	(0.152)	(0.278)	(0.266)	(0.117)	(0.144)	(0.159)	(0.217)	(0.178)	(0.205)	(0.156)	(0.235)	(0.189)	(0.181)	(0.187)	(0.247)	(0.206)	(0.264)	(0.275)	(0.168)	(0.248)	(0.240)
2nd Period * Democracy (polity)	-0.044	0.021	0.073	0.034	0.131	0.021	0.073	-0.022	0.020	-0.098	0.011	0.090	0.051	0.144	-0.025	0.038	0.101	-0.018	0.083	-0.024	0.235	-0.031	-0.094	-0.038
	(0.150)	(0.123)	(0.107)	(0.088)	(0.133)	(0.145)	(0.081)	(0.079)	(0.077)	(0.083)	(0.102)	(0.109)	(0.091)	(0.105)	(0.078)	(0.099)	(0.107)	(0.122)	(0.107)	(0.159)	(0.148)	(0.103)	(0.109)	(0.106)
Treated * 2nd Period * GDPpc (log)	0.371	-0.552	-0.702	-0.221	-0.549	-0.916	0.006	-0.166	0.139	-0.274	0.008	-0.180	-0.154	-0.100	-0.465	-0.295	0.180	-0.508	0.320	0.324	0.536	-0.251	-0.272	-0.920
	(0.560)	(0.417)	(0.455)	(0.380)	(0.714)	(0.569)	(0.380)	(0.316)	(0.401)	(0.386)	(0.408)	(0.426)	(0.472)	(0.471)	(0.404)	(0.408)	(0.454)	(0.544)	(0.518)	(0.677)	(0.736)	(0.553)	(0.468)	(0.722)
Treated * 2nd Period * Population (log)	0.094	-0.289	0.022	0.125	0.274	0.053	0.307	-0.033	-0.043	-0.071	-0.025	-0.234	0.277	-0.016	0.034	0.071	-0.132	-0.178	0.022	-0.117	-0.076	-0.112	-0.192	-0.430
	(0.497)	(0.376)	(0.269)	(0.220)	(0.391)	(0.344)	(0.207)	(0.226)	(0.243)	(0.267)	(0.259)	(0.269)	(0.251)	(0.305)	(0.239)	(0.260)	(0.307)	(0.350)	(0.291)	(0.379)	(0.487)	(0.259)	(0.315)	(0.384)
Treated * 2nd Period * Democracy (polity)	-0.052	-0.023	-0.106	-0.075	-0.047	-0.089	-0.162	-0.041	-0.034	-0.031	-0.023	-0.204	-0.087	-0.120	-0.077	-0.104	-0.145	-0.032	-0.102	-0.097	-0.352	0.048	0.007	0.058
	(0.235)	(0.149)	(0.162)	(0.120)	(0.169)	(0.186)	(0.117)	(0.098)	(0.133)	(0.128)	(0.129)	(0.132)	(0.143)	(0.140)	(0.127)	(0.141)	(0.152)	(0.158)	(0.148)	(0.220)	(0.206)	(0.150)	(0.146)	(0.174)
Number of countries	110	109	113	116	95	117	112	120	120	114	113	106	105	116	120	117	100	113	112	87	81	108	120	63
R2	0.22	0.24	0.40	0.58	0.31	0.30	0.46	0.68	0.35	0.54	0.13	0.39	0.39	0.23	0.55	0.43	0.14	0.32	0.32	0.17	0.12	0.31	0.50	0.10
Observations	184	207	208	219	159	214	206	228	229	215	195	200	188	212	229	222	180	205	214	148	132	203	229	94

Notes: Four donors of our overall sample are missing due to insufficient number of observations (Iceland, Poland, Slovenia, and Slovak Republic). Constant term as well as dummy variables for recipient countries in the treatment group, the second period, and the interaction between treatment group and second period are included in all regressions but not shown for the sake of brevity. ***, **, * if significant at the one, five and ten percent level, respectively.

Figure A1.1. Change in the number of insignificant aid relations, 2007-2014



Source: OECD-DAC, https://www.oecd.org/dac/aid-architecture/fragmentation-orphans.htm

Appendix B1. Donor-specific regressions on aid allocation and placebo test

As discussed in Section 1.2., our DD estimator does not necessarily resolve endogeneity concerns. Therefore, this appendix presents donor-specific regressions on aid allocation and a placebo test by 'pre-dating' the Paris Declaration.

The findings from a simple aid allocation model help evaluate the seriousness of some endogeneity concerns at least tentatively. The results reported in Table B1.1. are based on donor-specific regressions, considering the recipient countries' GDP per capita, governance in terms of control of corruption, and population as the widely used standard determinants of aid allocation across recipient countries. We also include dummy variables for the second period and the treatment group of recipients (not shown). DAC donors with sufficient data to assess the allocation of their aid are listed in ascending order of the discounts for quality-adjusted aid (as shown in Figure 1.1.). This ranking of donors helps us compare the coefficients on the standard determinants of aid allocation between donors that provide higher quality aid (upper half of the list) and those providing lower quality aid (bottom half).

From this simple exercise we do not find evidence suggesting that richer recipient countries get more aid from donor countries granting higher quality aid. Rather, the poverty orientation of such donors, as reflected in the negative coefficients on GDP per capita in the corresponding aid allocation equations, appears to be somewhat stronger than the poverty orientation of donors granting lower quality aid. The coefficients on GDP per capita are -0.54 and -0.23, on average, for the upper and lower half of donors in the table, respectively. Neither do we find evidence that weaker governance in the recipient countries leads to inflows of lower quality aid. It rather appears that donors typically pay little attention to governance, in terms of controlling for corruption, when deciding on the allocation of their aid. Most of the coefficients on governance prove to be statistically insignificant, independently of whether the specific donor country grants higher or lower quality aid. Hence, it seems unlikely that our findings reported in the main body of the chapter are seriously distorted by these dimensions of potential endogeneity.

Table B1.2. reports the results of a placebo test to assess whether the Paris Declaration in 2005 was really relevant for higher effectiveness of quality-adjusted aid. For this placebo test, we assume that the Paris Declaration would have been agreed five years earlier in 2000. Since the data for quality-adjusted aid is available only from 1995, we split the placebo period 1995-2005 into two 5-year sub-periods: 1995-1999 and 2001-2005. In other respects, the placebo test follows the procedures underlying the baseline estimations reported in the upper part of Table 1.1.

According to the results in Table B1.2., both types of aid have been effective in increasing GDP per capita for the treatment group in 2001-2005. This surprising finding may be attributed to the increase in aid after the 9/11 terrorist attacks. It appears that this event affected the aid-growth nexus for both types of aid. In the present context, however, it is more important that the placebo test does not point to higher effectiveness of quality-adjusted aid, compared to unadjusted aid, after the placebo treatment in 2000. This contrast to our baseline findings suggests that the Paris Declaration in 2005 was a relevant event indeed to explain the significantly higher effects of quality-adjusted aid on the change of GDP per capita for the treatment group in the aftermath of the Paris Declaration.

Table B1.1. – Aid allocation by specific DAC donors: significant coefficients on GDP per capita, control of corruption and population

	GDP per capita	Control of corruption	Population
Sweden	-0.529***		0.433***
Ireland	-0.825***	0.488*	0.363***
Denmark	-0.701***		0.502***
Luxembourg	-0.442***	0.549*	0.218**
Netherlands	-0.868***		0.404***
Portugal	-0.393*		
United Kingdom	-0.574***		0.692***
Switzerland	-0.722***		0.485***
Finland	-0.540***		0.548***
Norway	-0.563***		0.587***
New Zealand			
Australia	-0.372*		
Canada	-0.565***	0.309*	0.611***
Belgium	-0.595***		0.475***
Germany	-0.224**		1.044***
United States	-0.455***	-0.533*	0.609***
Austria	-0.240*		0.498***
Japan	-0.663***	0.568**	0.606***
Greece	0.312**		0.272***
France			0.749***
Spain			0.630***
Italy	-0.319*		0.331***
Korea		-0.367*	0.517***
Czech Rep.			

Notes: DAC donors listed in ascending order by discounts for quality-adjusted aid (see Figure 1.1.). DAC donors with insufficient number of observations not listed (Iceland, Poland, Slovak Rep. and Slovenia). Statistically insignificant coefficients are not shown; ***, **, * significant at the one, five and ten percent, respectively. All regressions include dummy variables for the treatment group of recipients and the second period (not shown).

Table B1.2. Effects of aid on the change in GDP per capita: placebo test for the full sample, median of difference in aid inflows (total US$) to separate treatment and control groups

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Quality-adjusted aid			Unadjusted aid		
2nd period	0.025	0.020	-0.270**	0.010	0.011	-0.301***
	(0.187)	(0.352)	(0.025)	(0.637)	(0.614)	(0.008)
Treatment	0.012	-0.005	-0.011	0.014	0.023	0.019
	(0.675)	(0.840)	(0.962)	(0.640)	(0.340)	(0.423)
2nd period * treatment	0.053	0.091***	0.087***	0.083**	0.095***	0.101***
	(0.135)	(0.002)	(0.004)	(0.020)	(0.001)	(0.001)
Constant	0.098***	0.103	0.218**	0.097***	0.074	0.203*
	(0.000)	(0.292)	(0.038)	(0.000)	(0.457)	(0.057)
Controls	No	Yes	Yes	No	Yes	Yes
2nd period * controls	No	No	Yes	No	No	Yes
Observations	290	233	233	290	233	233
R^2	0.043	0.115	0.147	0.065	0.152	0.184

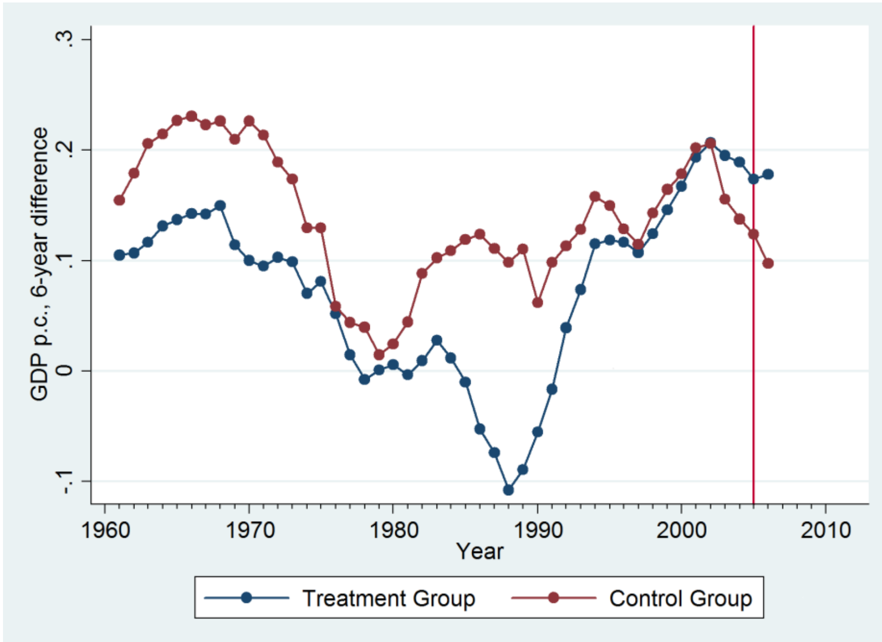
Standard errors are clustered at the recipient country level. Robust p-values in brackets. * (**, ***) indicates statistical significance at the ten (five, one) percent level.

Appendix C1. Trends of the change in GDP per capita for the treatment and control group

The DD approach assumes that, in the absence of the (aid-related) treatment, the difference in the dependent variable between the two sub-periods would have been the same for the treatment and the control groups. In this appendix, we tentatively assess the plausibility of this assumption. Specifically, we portray in Figure C1.1. how the differences in GDP per capita for the treatment and control groups of recipient countries evolved since 1960. Differences in GDP per capita are calculated over six years, on a moving annual basis. Each dot in Figure C1.1. shows the average change in GDP per capita for countries in the treatment group or countries in the control group, respectively, over the period from *t* to *t*+5; e.g., the two dots for 2005 relate to the difference of GDP per capita in 2011 minus GDP per capita in 2005.

As can be seen, the trends are not perfectly parallel for the two groups of recipient countries. However, both groups seem to have parallel trends after 1995. See Section 1.3. for a discussion of methodological implications.

Figure C1.1. Trends of the differences in GDP per capita since 1960



Appendix D1. Tests for statistically significant differences between the coefficients of the interaction terms, *2nd period*treatment*

We report two types of tests in Table D1.1. and Tables D1.2. and D1.3. The *p*-values of these tests are briefly interpreted in the main body of the chapter.

Table D1.1. shows tests for statistically significant differences between the impact of quality-adjusted aid and unadjusted aid on the change in GDP per capita in Tables 1.1. and 1.6., or the alternative definitions of the dependent variable in Table 1.2. More precisely, we test whether the interaction term *2nd period*treatment* differs significantly between quality-adjusted aid and the corresponding estimation with unadjusted aid. In the front column of Table D1.1., we specify the interaction terms compared by noting the relevant table, the panel (in Tables 1.1. and 1.6.) or the definition of the dependent variable (in Table 1.2.), and the columns where the two interaction terms are based on the same specification of the estimation equation. For instance, the *p*-value of 0.061 in the first line of Table D1.1. compares the coefficients on the interactions terms shown in columns 1 and 4 of the upper panel of Table 1.1. (0.067 and 0.029, respectively).

Tables D1.2. and D1.3. show tests for significant differences between the impact of aid measured in total US$ with that of aid measured in per-capita terms. The tests reported in Table D1.2. are based on the interaction terms, *2nd period*treatment*, for the two aid measures shown in Table 1.1. for the full sample of recipient countries; i.e., we compare the interaction terms for either quality-adjusted aid or unadjusted aid in the upper panel of Table 1.1. (where the treatment is based on the change in aid inflows in total US$) with the corresponding interaction terms in the lower panel of Table 1.1. (where the treatment is based on the change in aid inflows in per-capita terms). For instance, the *p*-value of 0.573 compares the coefficients on the interaction terms shown in column 1 of the upper and the lower panel of Table 1.1. (0.067 and 0.055, respectively). The tests reported in Table D1.3. compare the interaction terms (for either quality-adjusted aid or unadjusted aid) between the two measures of aid after excluding the smallest recipient countries or, alternatively, the largest recipient countries. The interaction terms are shown in Table 1.7. when the treatment is based on the change in aid inflows in total US$, while the corresponding interaction terms with the treatment based on the change in aid inflows in per-capita terms are available on request. For instance, the *p*-value of 0.902 reported in the first line of Table D1.3. compares the coefficients on the interaction term shown in column 1 of the top panel of Table 1.7. (0.039) and the corresponding interaction term for quality-adjusted aid in per-capita terms (not shown in Table 1.7.).

Table D1.1. Tests for statistically significant differences between the coefficients of the interaction terms, 2nd *period*treatment*, for quality-adjusted aid and unadjusted aid

Coefficients	p-value
Table 1.1., upper panel, column 1 & 4	0.061
Table 1.1., upper panel, column 2 & 5	0.079
Table 1.1., upper panel, column 3 & 6	0.060
Table 1.1., lower panel, column 1 & 4	0.459
Table 1.1., lower panel, column 2 & 5	0.381
Table 1.1., lower panel, column 3 & 6	0.785
Table 1.6., upper panel, column 1 & 4	0.010
Table 1.6., upper panel, column 2 & 5	0.067
Table 1.6., upper panel, column 3 & 6	0.210
Table 1.6., lower panel, column 1 & 4	0.005
Table 1.6., lower panel, column 2 & 5	0.097
Table 1.6., lower panel, column 3 & 6	0.963
Table 1.2., compound growth, total aid, column 1 & 4	0.059
Table 1.2., compound growth, total aid, column 2 & 5	0.063
Table 1.2., compound growth, total aid, column 3 & 6	0.039
Table 1.2., compound growth, per capita aid, column 1 & 4	0.450
Table 1.2., compound growth, per capita aid, column 2 & 5	0.366
Table 1.2., compound growth, per capita aid, column 3 & 6	0.739
Table 1.2., log GDP, total aid, column 1 & 4	0.871
Table 1.2., log GDP, total aid, column 2 & 5	0.102
Table 1.2., log GDP, total aid, column 3 & 6	0.070
Table 1.2., log GDP, per capita aid, column 1 & 4	0.666
Table 1.2., log GDP, per capita aid, column 2 & 5	0.059
Table 1.2., log GDP, per capita aid, column 3 & 6	0.169

Table D1.2. Tests for statistically significant differences between the coefficients of the interaction terms, 2nd *period*treatment*, for the aid measure in total US$ and in per-capita terms, full sample in Table 1.1.

	p-value
Table 1.1., Column 1, U&L (quality aid total & quality aid per capita)	0.573
Table 1.1., Column 2, U&L (quality aid total & quality aid per capita)	0.769
Table 1.1., Column 3, U&L (quality aid total & quality aid per capita)	0.486
Table 1.1., Column 4, U&L (unadj. aid total and unadjusted aid per capita)	0.509
Table 1.1., Column 5, U&L (unadj. aid total and unadjusted aid per capita)	0.432
Table 1.1., Column 6, U&L (unadj. aid total and unadjusted aid per capita)	0.204

Note: U&L denotes upper and lower panel.

Table D1.3. Tests for statistically significant differences between the coefficients of the interaction terms, *2nd period*treatment*, for the aid measure in total US$ and in per-capita terms, excluding smallest or largest countries in Table 1.7.

	p-value
Quality aid (smallest countries excluded)	
Total vs per capita, column 1	0.902
Total vs per capita, column 2	0.513
Total vs per capita, column 3	0.759
Unadjusted aid (smallest countries excluded)	
Total vs per capita, column 4	0.312
Total vs per capita, column 5	0.334
Total vs per capita, column 6	0.216
Quality aid (largest countries excluded)	
Total vs per capita, column 1	0.580
Total vs per capita, column 2	0.590
Total vs per capita, column 3	0.274
Unadjusted aid (largest countries excluded)	
Total vs per capita, column 4	0.509
Total vs per capita, column 5	0.344
Total vs per capita, column 6	0.266

CHAPTER 2

Development Minister Characteristics and Aid Giving[57]

Co-authored with Andreas Fuchs

2.1. INTRODUCTION

During the run up to Germany's 2009 General Elections, the Free Democratic Party campaigned for the abolishment of the Federal Ministry for Economic Cooperation and Development. When Chancellor Angela Merkel formed a coalition with the Free Democrats later that year, it was a frontrunner of the Free Democratic Party who took office in the ministry: Dirk Niebel, then the party's General Secretary. Rather than sticking to the announcement to dissolve the ministry, the new minister asked for an increase in the ministry's budget.[58] Niebel did not have any professional experience in development cooperation when he took office, leading the German news magazine *Stern* to conclude: "Nobody can really say what actually qualifies Dirk Niebel as development aid minister."[59]

---

[57] Chapter 2 is available as published article: Fuchs, Andreas and Katharina Richert (2018). Development Minister Characteristics and Aid Giving. *European Journal of Political Economy* 53: 186-204.

[58] DIE WELT, "Neuer Minister: Niebel verlangt mehr Geld für Entwicklungshilfe," *WELT.de*, 23 November 2009, available at http://www.welt.de/politik/deutschland/article5297548/Niebel-verlangt-mehr-Geld-fuer-Entwicklungshilfe.html (accessed 26 November 2014).

[59] Christ, Sebastian and Hans-Peter Schütz, "Entwicklungshilfeministerium: Dirk Niebel, Minister auf Bewährung," *stern.de*, 29 October 2009 (own translation), available at http://www.stern.de/politik/deutschland/entwicklungshilfeministerium-dirk-niebel-minister-auf-bewaehrung-1517745.html (accessed 26 November 2014).

Does it matter who is in charge of development cooperation? Over 300 ministers responsible for development aid have entered (and left) office in 23 member countries of the OECD's Development Assistance Committee (DAC) since this institution started reporting detailed aid flows in 1967.[60] 30 percent of the ministers are explicitly "Ministers for Development Cooperation" (or have similar titles), while in most cases development aid has been the responsibility of the foreign minister. 18 percent of the ministers have been women and a mere 16 percent possessed any professional experience in development cooperation when they took office. While it is highly disputed whether (and how) aggregate aid affects the economic growth of developing countries (e.g., Burnside and Dollar 2000; Easterly et al. 2004), there is evidence that certain types of aid have positive effects on development outcomes (Dreher et al. 2008; Clemens et al. 2012; Bjørnskov 2013; but see also Roodman 2015).[61] Moreover, scholarship has shed light on some unwelcome side effects that aid might have on conflict, governance, and sustainable development in general (e.g., Bjørnskov 2010; Nunn and Qian 2014). Understanding the role played by development ministers in foreign aid is crucial since their decisions might influence both the *quantity* and the *quality* of aid and thus impact aid effectiveness and aggravate or mitigate the potential side effects of aid.

In order to study whether the personal characteristics of development ministers matter for donors' aid giving, we build a novel database covering all ministers of OECD-DAC countries responsible for development cooperation since 1967. The study covers all country-years for which detailed aid flows have been reported to the OECD-DAC (as of July 28, 2014). Using panel econometric models, we then estimate the link between development ministers' personal characteristics and (1) aid quantity, i.e., the size of aid budgets in terms of Official Development Assistance (ODA),[62] as well as (2) aid quality, i.e., the share of aid budgets that

[60] As of the end of 2012, 23 countries (and the European Commission) were members of the OECD-DAC: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, South Korea, Spain, Sweden, Switzerland, the United Kingdom, and the United States. We only cover aid provided by DAC donors in order to analyze a comparatively homogenous set of donor countries, which have agreed to follow the same guidelines on the management of development aid. This is not to say that DAC donors would not show considerable variation with respect to their aid motives (see, for example, Doucouliagos and Manning 2009).

[61] See Doucouliagos and Paldam (2008, 2010, 2011) for meta studies of the aid effectiveness literature.

[62] ODA is defined by the OECD (2008) as "those flows to countries and territories on the DAC List of ODA Recipients (available at www.oecd.org/dac/stats/daclist) and to multilateral development institutions which are: i. provided by official agencies, including state and local governments, or by their executive agencies; and ii. each transaction of which: a) is administered with the promotion of the

is expected to be particularly conducive to achieving developmental goals as operationalized by the foreign-assistance component of the Commitment to Development Index (CDI) (Roodman 2012).[63] Specifically, we test whether these two variables are associated with ministers' gender, ideology, prior professional experience in development cooperation, university education, and years in office.

Our chapter combines two strands of the literature. First, it contributes to the empirical aid literature (e.g., Dudley 1979; Alesina and Dollar 2000; Kuziemko and Werker 2006) and to the scholarly work on aid budgets in particular (e.g., Bertoli et al. 2008; Tingley 2010; Dreher and Fuchs 2011; Brech and Potrafke 2014; Fuchs et al. 2014). Second, the chapter adds to the burgeoning literature on the effects of political leaders' personal characteristics on economic outcomes. Previous research has focused on the role of gender (e.g., Chattopadhyay and Duflo 2004; Koch and Fulton 2011), political ideology (e.g., Neuenkirch and Neumeier 2015), educational and professional background (e.g., Göhlmann and Vaubel 2007; Spilimbergo 2009), regional and ethnic origin (e.g., Hodler and Raschky 2014; De Luca et al. 2015), socioeconomic status (Hayo and Neumeier 2012, 2014, 2016), and time in office (e.g., Jochimsen and Thomasius 2014; Moessinger 2014). Contributions cover the role played by country leaders (e.g., Dreher et al. 2009; Besley et al. 2011), foreign ministers and defense ministers (Koch and Fulton 2011), finance ministers (e.g., Moser 2007; Chatagny 2015), central bankers (e.g., Göhlmann and Vaubel 2007; Neuenkirch and Neumeier 2015), heads of subnational regions (Hayo and Neumeier 2012, 2014), and mayors (e.g., Ferreira and Gyourko 2014; Freier and Thomasius 2016), among others.[64]

The development minister offers a particularly interesting case to reinvestigate the role of leadership since this position receives relatively little (domestic) attention compared to other cabinet members, such as the head of government or the minister of finance, despite its global importance. Development ministers have a low profile at home, usually being either

---

economic development and welfare of developing countries as its main objective; and b) is concessional in character and conveys a grant element of at least 25 per cent (calculated at a rate of discount of 10 per cent)."

[63] As there is a lack of consensus as to what constitutes high-quality aid, we will also show results where we remove controversial subcomponents of this aggregate indicator.

[64] A related literature analyzes how leadership *changes* (rather than personal characteristics) affect economic outcomes (McGillivray and Smith 2004; Jones and Olken 2005; Moser and Dreher 2010; Dreher and Jensen 2013).

annexed to their respective foreign ministry or having a low rank in cabinet.[65] Given the low issue salience of aid policies (e.g., Lundsgaarde 2013; Szent-Iványi and Lightfoot 2015), it appears likely that foreign ministers assuming the function of a development minister are selected based on their stance on foreign policy issues rather than on their development profile. All of this is beneficial for our research design as the selection of the officeholder should thus be more independent from policy positions in the issue area he or she oversees than in the case for ministers in other issue areas. Nevertheless, decisions made by development ministers are of high importance. In sharp contrast to the disinterest in development issues in many governments, parliaments and the public, aid decisions taken at donor ministries can have huge impacts on the ground as DAC countries alone provide more than US$ 100 billion annually to the developing world.

The previous literature barely touches on the role of the decision-makers responsible for the provision of development assistance. The existing papers that cover the role of development ministers only analyze the impact of the ministers' gender. Dreher et al. (2015a) find that female development ministers are more responsive to gender issues when allocating aid than their male counterparts. Kleemann et al. (2016) discover only minor gender differences in the allocation of aid for education. A systematic analysis of development minister characteristics is still lacking. Additionally, since the above two papers do not control for female heads of government, their empirical strategies come with the drawback of not capturing the pure effect of the *minister*'s gender. Appointing a female development minister could just be a proxy for women having control over government in general.[66] In order to approach the identification of a genuine effect stemming from the personal characteristics of the development minister, we control for the personal characteristics of the respective head of government and donor-country-fixed effects, or, alternatively, government-head-fixed effects. The inclusion of government-head-fixed effects also helps mitigate concerns that the observed effects might be driven by the strategic appointment of ministers by heads of governments as these regressions only exploit variation within governments over time. However, we cannot

[65] For example, the development ministers are not full-ranked ministers in France (*ministre délégué*). The German development minister has full cabinet rank but is the lowest ranked line minister according to German protocol.

[66] Lu and Breuning (2014) control for the gender composition of governments in their analysis of the role of gender for aid generosity. However, they do not include donor-country-fixed effects, meaning that the observed effects may be driven by unobserved country characteristics.

control for changes in the focus of a given head of government while in office. Since we lack adequate external instrumental variables for the appointment of development ministers, we do not claim that the observed coefficients represent causal effects.

Our chapter provides the first test for the role of the personal characteristics, beyond gender, of development ministers for a large sample of donor countries.[67] Additionally, we also offer the first quantitative analysis of the ministers' role for the "development-friendliness" of donors' aid giving ("aid quality") and introduce a rigorous empirical strategy that can be applied in the context of research on the role of government members more generally. The results show that some personal characteristics of development ministers matter. Most notably, more experienced ministers with respect to their time as development minister obtain larger aid budgets: One additional year in the development office increases total ODA commitments by 0.7 percent. Additionally, we find that the share of quality ODA increases by 1.1 percentage points if development ministers possess prior professional experience in the field of development cooperation and by 0.2 percentage points for each additional year in office. Moreover, aid quality is on average one percentage point higher if development ministers are female.

We proceed as follows. Section 2.2. discusses potential linkages between the personal characteristics of development ministers and the quantity and quality of ODA. Section 2.3. presents our novel dataset covering the characteristics of the 320 ministers that have been responsible for the OECD's development aid since 1967. Section 2.4. introduces the empirical approach and presents our results. The final section concludes and discusses avenues for future research.

2.2. THEORETICAL CONSIDERATIONS

Development ministers negotiate at the cabinet table about the size of their budget and then assume the leadership role to distribute the aid money across recipient countries, sectors, and types of aid. Theory suggests that minister characteristics are particularly salient in the case of decision-making by development ministers compared to other government members. In contrast to government expenditures in most other political spheres (e.g., spending by the

[67] Dreher et al. (2015b) analyze the role of the political color of both the Ministry of Economic Cooperation and Development and the Foreign Federal Office on German aid allocation.

ministry of transport), taxpayers cannot directly observe how their government allocates the country's aid budget in recipient countries (e.g., Martens et al. 2002).[68] This absence of the usual control mechanism between taxpayers and politicians creates a principal-agent problem and development ministers should thus have sufficient discretion to influence aid giving to further their own interests. These can be either altruist motives to promote development in recipient countries, or self-interests such as their intention to get promoted within the ministerial hierarchy or to obtain side-payments from lobby groups. Early work indeed suggests that the development minister has "considerable latitude within the bounds of general policy directives" (Breuning 1999: 732).

Nevertheless, there are also good reasons to believe that development ministers do not have the power to significantly influence aid giving. Most fundamentally, the minister's power is limited due to shared competencies within the government and parliamentary oversight (Dreher et al. 2015b). Other forces that reduce the minister's room to maneuver include a powerful aid bureaucracy – whether in the form of the ministry itself or an aid agency – that follows its own agenda (Easterly 2002a; Copelovitch 2010), special interest groups that may dominate aid decisions (Anwar and Michaelowa 2006; Hicks et al. 2010), and traditions and other patterns of persistence in aid relations, including the emergence of "lead donors" (Steinwand 2015). If development ministers possess sufficient power to overcome the constraining influences of other actors, their personal characteristics will affect the quantity and quality of aid. In what follows, we discuss how the gender, political ideology, and experience of a development minister could affect aid giving.

Gender

Women and men show significant differences in their preferences.[69] For example, Togeby (1994) identifies a gender gap in foreign policy attitudes: women are on average more supportive of development aid (and less supportive of military interventions). However, it is

---

[68] Although some large (e.g., infrastructure) projects receive some media coverage in donor countries, taxpayers typically do not directly observe the benefits of aid. Indirect effects on the domestic sphere, such as reduced refugee flows or a possible reduction in transnational terrorist attacks, are difficult to quantify.

[69] Croson and Gneezy (2009) review experimental evidence on gender-specific differences in risk preferences, social preferences, and competitive preferences. Econometric studies show, for example, that, as women obtain the right to vote, government size and social spending increase (Lott and Kenny 1999; Aidt and Dallal 2008).

unclear whether such gender differences in individuals' preferences also translate into politicians' actual decision-making. Strategic considerations, including party pressure or log-rolling, may prevent these differences from affecting policies (Funk and Gathmann 2015). Scholarly evidence is mixed on whether the described overall larger support of women for development aid is reflected in legislators' decisions on aid policies. Empirical studies do not show a consistent positive relationship between the strength of female representation in parliament and the size of aid budgets (Breuning 2001; Lundsgaarde et al. 2007; Olsen-Telles 2013; Fuchs et al. 2014; Lu and Breuning 2014; Hicks et al. 2016).

Predictions are also not straightforward with respect to total ODA budgets when it comes to the role of female development ministers rather than parliamentarians. Independently of gender-specific differences in the ministers' stance towards foreign aid, ministers have a vested interest in maximizing their respective budget as a greater budget increases their chances of success as a minister (and should thus affect their chance of being promoted and re-elected). Based on this, one would expect to observe no significant difference in the size of aid budgets between female-led and male-led aid ministries. However, empirical evidence points at systematic gender differences in negotiation outcomes, such as in salary negotiations, in which women are worse off than men (e.g., Gerhart and Rynes 1991). Explanations include, among others, women's lower willingness to self-promote, and negotiation partners, both male and female, who make lower offers to women since they assume that female negotiators will give in more easily than their male counterparts (Solnick 2001). Applying these findings to political negotiations, it could be argued that male ministers more successfully negotiate for larger aid budgets than female ministers.

Turning to the specific use of the aid money, which should affect the quality of aid, Hicks et al. (2016) find strong evidence of gender differences arising from larger female representation in parliaments: the level of the flows going to education, health and social capital projects as well as to least-developed countries appear to increase with stronger representation of women in national parliaments. One could thus expect that donor countries with female development ministers also provide more aid to social sectors and poor countries, which is – as some argue (e.g., Mosley 1985) – of higher ODA quality. This would be consistent with evidence of gender differences in foreign policy attitudes discussed above (e.g., Togeby 1994).

However, there are reasons to believe that such "female behavior" does not hold at the level of political leaders. Analyzing the role of gender in foreign policy, Koch and Fulton (2011) show that female representation in parliaments causes a decrease in defense spending and conflict behavior but they find the opposite effect for female defense ministers and government heads. In the words of Koch and Fulton (2011), "[w]hen it comes to masculinized leadership positions, like executive office, this challenge to gain credibility may lead women to present themselves as more masculine, in an attempt to combat the stereotype." Given that women face more barriers to access leadership positions, Jochimsen and Thomasius (2014: 394), referring to Eagly et al. (1995), note that "[i]f a woman must be 'twice as good as a man' in order to be appointed to a leadership position [...] then women may be more effective leaders and superior performers compared to their male colleagues." Similarly, empirical evidence on monetary policy suggests that women take more "hawkish" decisions than their male counterparts (Farvaque et al. 2009). Translating this to the case of development ministers, one might expect "tougher" behavior from female development ministers in negotiations over budgets and in their usage of budgets in the sense that they pursue a more self-interested development policy at the detriment of aid quality.

Political ideology

Scholarship has scrutinized the effects of government ideology on aid giving. First, it is argued that, in analogy to domestic social welfare transfers, left-wing governments provide more aid than right-wing governments since the former are supposedly more "altruistic" and "pro-poor" (Thérien and Noël 2000; Round and Odedokun 2004). Moreover, the left's stronger tendency to interfere in market mechanisms is put forward as an explanation (Tingley 2010). Second, right-wing governments are said to provide larger aid budgets as they see aid as a tool to promote commercial and geostrategic interests (Round and Odedokun 2004; Bertoli et al. 2008). Finally, it might again just be that ministers aim to maximize their budgets independent of their respective political ideology as argued above. While Thérien and Noël (2000) and Tingley (2000) provide evidence for the first argument, Lundsgaarde et al. (2007) find political ideology to be insignificant, and, according to Bertoli et al. (2008) and Dreher et al. (2015b), right-wing governments provide more aid.

Based on the same lines of argumentation, there are also reasons to believe that development ministers' political orientation affects the quality of aid. Brech and Potrafke

(2014) find that left-wing governments experience stronger increases in bilateral grant aid and grant aid to least developed and lower middle-income countries. This might suggest that left-wing governments provide in general higher aid quality compared to right-wing governments as such an allocation pattern leads to lower future financial obligations for recipient countries and is more need-oriented. This is in line with the argument that right-wing decision-makers are guided to a greater extent by domestic political and commercial interests than their left-wing counterparts who are expected to believe more in the merits of redistribution from the rich to the poor. However, empirical evidence in this regard is mixed. While Fleck and Kilby (2006) find that development concerns in the United States matter more under a Democratic president and Congress compared to when the president and/or Congress are Republican, Dreher et al. (2015b) reject claims that Germany's aid allocation under conservative governments is guided to a larger extent by commercial and politico-strategic interests than under left-wing administrations.

Experience

Prior research shows that the professional background of political decision-makers affects their decisions while in office. For example, former central bank staff prefer lower inflation rates than former politicians after being appointed to central bank councils (Göhlmann and Vaubel 2007; Farvaque et al. 2009) and former entrepreneurs are more likely to implement market-liberalizing reforms when they lead a government (Dreher et al. 2009). Based on these results, one could also argue that relevant experience in development cooperation, including work experience in aid agencies, development NGOs and developing-country embassies, influences the work of the development minister. There are at least two reasons why such technical experience could be positively related to aid quality. First, ministers who worked in development cooperation prior to assuming office should have had the chance to observe the differential effects of aid interventions and determinants of success on the ground. Such field experience could have stimulated a learning process on how to provide effective development aid.[70] Second, ministers who have been engaged in the field of development cooperation before assuming office are more likely to be intrinsically motivated to foster development than those

[70] Jacqmin and Lefebvre (2016) make a similar argument in favor of learning by doing for education ministers.

who have not. Intrinsically motivated ministers are expected to shift the balance away from donor self-interests towards need orientation. While this explains differences in the quality of aid, it is unlikely that prior experience in development cooperation influences the quantity of aid, as greater development affinity does not necessarily translate into better negotiation skills at the cabinet table.

Turning to education, the empirical evidence that educational backgrounds matter for political decision-making is much weaker than for specific prior occupations (Göhlmann and Vaubel 2007; Dreher et al. 2009; Jochimsen and Thomasius 2014; Moessinger 2014). Still, there are reasons to believe that a training in economics can make a difference when there is strong need for economic expertise such as in the context of economic development. In this regard, Hallerberg and Wehner (2012) find that countries with a high frequency of financial crises, such as Greece and Portugal, are more likely to appoint economists as economic policymakers than other OECD countries. Dreher et al. (2009: 170) identify a potential advantage of trained economists "in implementing reforms as they are more likely to distinguish good from bad advice and might be more able to resist the pressure of lobbying groups preferring the status quo." Similarly, economics-trained development ministers might be in a better position to implement effective development policies as they better understand the market mechanisms and market failures at play in developing countries and are thus better able to identify successful development measures. If this is true, we should observe higher aid quality when the development minister has obtained a degree of higher education in economics. At the same time, however, economists are found to be more selfish (e.g., Frey and Meier 2003) and might thus exhibit a stronger focus on personal or donor-country self-interests to the detriment of aid quality. To the extent to which economists are better bargainers, aid budgets of economics-trained development ministers should be larger. In line with this, Jochimsen and Thomasius (2014: 394) evoke the possibility that "trained economists […] are more successful in convincing their cabinet colleagues of sound budgets with low deficits."

Beyond ministers' technical experience acquired prior to taking office, their experience on the job could also affect aid giving. Usually, the political power of ministers increases with their time in office as they accumulate experience. Referring to finance ministers, Feld and Schaltegger (2010: 509) argue that a "minister who succeeds in remaining a long time in office usually enjoys a politically powerful position towards the parliament, the administration and the interest groups." Moessinger (2014: 185) suggests that "an experienced finance minister

[…] know[s] more about the schemes of his cabinet colleagues in attracting additional funds for their respective ministries." Along similar lines, more experienced development ministers should better know how to successfully secure funds for their own ministry as they can more forcefully oppose the finance minister and more successfully compete against other cabinet colleagues. Moreover, more experienced development ministers should have acquired more knowledge over time on the types of aid that work. Assuming that development ministers want to increase the impact of development aid (out of humanitarian motives or career concerns), we expect that development ministers learn over time how to provide more effective aid and shift resources accordingly. On the contrary, the longer a minister is in office the higher the chances that he or she gets "captured" by the aid industry, which could reduce the quality of aid. For example, ministers captured by a vivid NGO community are more likely to hand out many small projects to satisfy their various needs, which amplifies the problem of project proliferation.

2.3. DEVELOPMENT MINISTER DATABASE

We define "development minister" as the donor country's government member that holds the main responsibility for development cooperation.[71] We first identify the names and governing periods of all development ministers for the years in which the respective OECD-DAC donor reports detailed aid flows to the OECD's Creditor Reporting System (i.e., since 1967 at the earliest).[72] We collect the required data through internet research from publicly available sources, including government websites, the personal websites of the ministers, Political Data Yearbook interactive,[73] and Wikipedia, among others. Where necessary, we contacted the ministries or other government institutions via e-mail to gather additional information. Following the described procedure, we obtain a dataset with 957 observations containing 320 ministers for 23 OECD-DAC countries between 1967 and 2012, i.e., for a maximum of 46 years per country.

[71] The respective cabinet member is either a minister, minister of state, or secretary of state. In what follows, we use the term "development minister" for the sake of brevity.
[72] We attribute years during which two or more development ministers are in office to the minister who is longest in power during that year. In six cases, however, two ministers were equally long in office (6 months). We then keep the minister being in office for the last six months in our dataset.
[73] See http://www.politicaldatayearbook.com/ (last accessed 11 December 2014).

To be able to study the relationship between the personal characteristics of development ministers and donors' aid giving, we collect information on five personal characteristics of development ministers – mirroring the hypotheses introduced in the previous section. First, we collect information on the ministers' genders. The binary variable for a minister's gender is coded as one for women. Across all OECD-DAC donors under analysis, a female minister is in charge of development cooperation in one fifth of all country-years. For comparison, only 5 percent of all heads of governments are women in our dataset. Sweden shows the largest proportion of female-led development cooperation with women being in power over 27 of 46 years, closely followed by Canada with 26 years. In Australia, Italy, and South Korea, the position of the development minister has never been assigned to a woman (as of 2012).[74] Analyzing the gender distribution over time, Figure 2.1.a shows a sharp increase in the number of female development ministers starting with the turn of the century. While only 14 percent of ministers are female by 1990, the share of women increases to 43 percent in 2000. In the peak years of 2001, 2005 and 2006, the gender distribution is almost balanced with a total of 11 female ministers in 23 countries.

Second, we gather data on ministers' political ideologies measured on a five-tier left-right scale. Bjørnskov and Potrafke (2011) use the social democratic party as an "anchor party," following the idea that its national branches are broadly comparable on the international level.[75] They assign a value of 0 to social democratic parties and classify the remaining political parties accordingly. Following their approach, we code the political ideology of development ministers with regards to the economic policy position of the political party they are affiliated with. Specifically, a value of -1 is assigned to "unreformed socialist and communist," -0.5 to "modern socialist," 0 to "social democratic," 0.5 to "conservative," and 1 to "liberalist economic policy."[76] We find that the position of the development minister is – with 48 percent

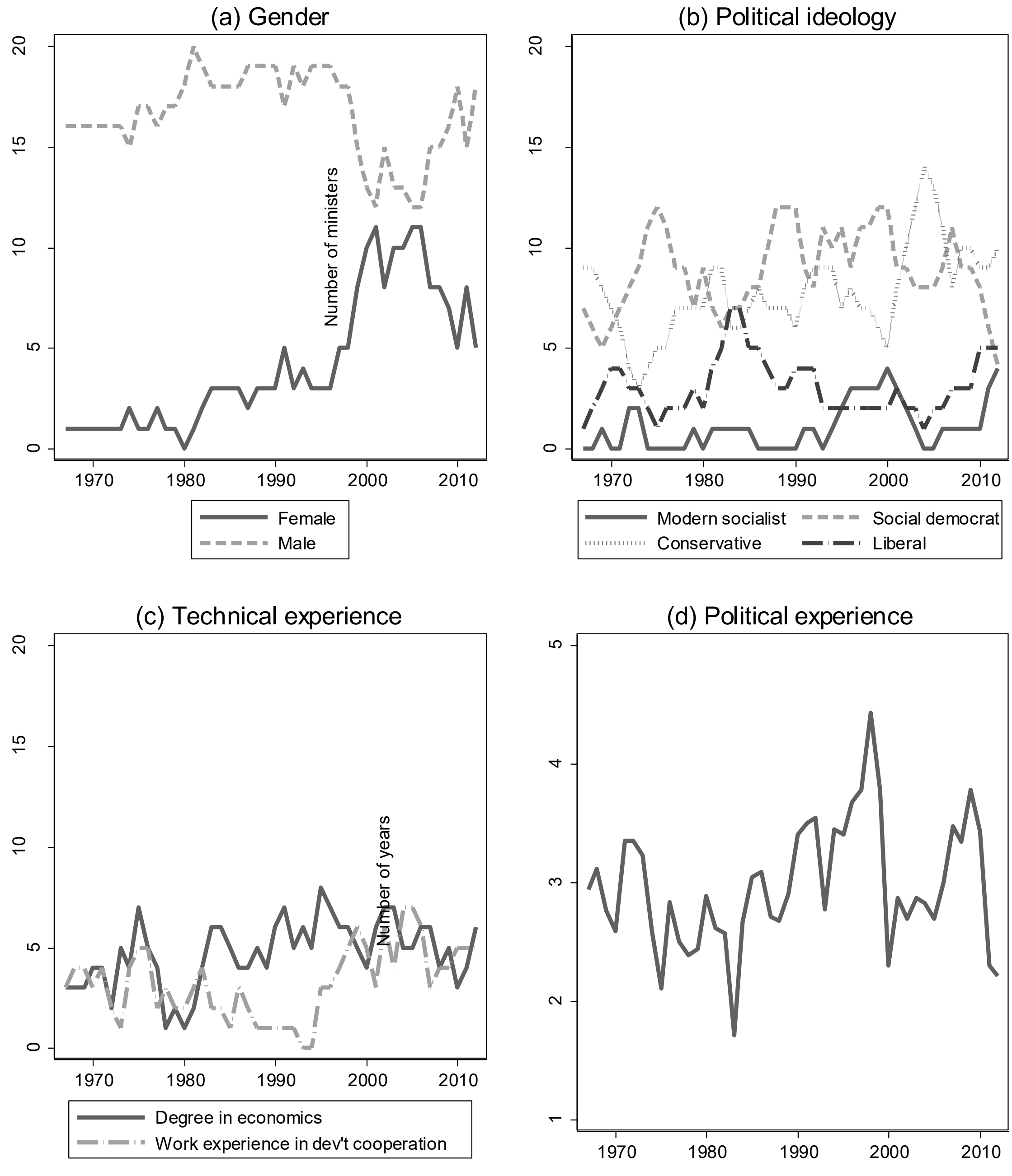
[74] Julie Bishop assumed office in Australia's Department of Foreign Affairs and Trade in 2013 and thus became Australia's first female development minister.

[75] As a test of robustness, we also compare our results below with a right-left coding scheme based on parties' election manifestos.

[76] We are grateful to Christian Bjørnskov for having generously provided us with their raw database that enabled us to translate their categorization of the ideologies of political parties to the case of development ministers. In countries without a social democratic party, such as France for instance, we follow Bjørnskov and Potrafke (2011) and code ministers' parties relative to a fictional central party to keep the classification pattern consistent. In cases where ministers are not party members, we code their economic-policy orientation based on other relevant information provided in their CVs (such as

of all country-years covered – almost as equally often assigned to left-wing politicians (including social democrats) than to right-wing politicians (52 percent; see Figure 2.1.b for details). No unreformed socialist or communist has been appointed over the time period under study.

Figure 2.1. Personal characteristics of development ministers (1967-2012)



---

memberships in relevant associations). In the absence of such information, we code them in line with the respective head of government that selected the respective minister.

Third, we collect information on whether the development ministers possess development-specific work experience when they take office. Specifically, we code a binary variable that takes a value of one if the minister has gained professional experience in the field of development cooperation.[77] As can be seen in Figure 2.1.c, a large majority of development ministers lack relevant work experience in the development context upon assuming office. Only 16 percent possess any prior work experience in development cooperation before coming into power.

Fourth, the database includes information on whether the ministers have obtained a degree of higher education in economics or business.[78] 23 percent of all ministers have received such training (see again Figure 2.1.c).

Fifth, in addition to ministers' experience prior to taking office, we also examine their political experience gained on the job. Specifically, we calculate the number of years a development minister holds office in a given year, irrespective of whether the period in office was interrupted by another minister's term or not. Almost 15 percent of the ministers hold office for only one year, which demonstrates a relatively large fluctuation in the position. The average tenure of a development minister is 3 years. There is much more fluctuation in the office of the development ministers (320 ministers overall) compared to heads of governments (207). Luxembourg's Jacques Poos is the minister that gained the most experience in office (15 years). Figure 2.1.d plots the average tenure of ministers in power over time.

To sum up, the typical minister is male, stays in power for three years and cannot be clearly attributed to one of the two political camps of left or right. He has neither received economics training nor gained prior professional experience in development cooperation. Appendix A2.1. lists all development ministers covered by our database.

---

[77] We code professional experience in development cooperation as one if ministers have worked for international development organizations (e.g., the United Nations Development Programme), national development agencies (e.g., the *Agence Française de Développement*), or non-governmental organizations addressing development concerns (e.g., *Médecins sans Frontières*). Additionally, we code this variable as one if the development minister led the development ministry in a prior term.

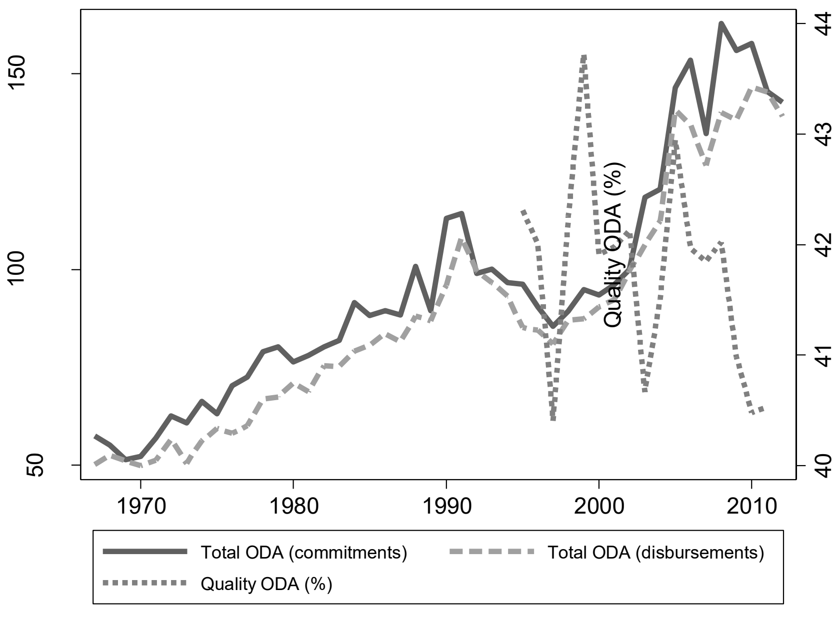
[78] We analyze business degrees jointly with economics as these study programs typically convey a comprehensive understanding of economic principles. Our definition covers degrees in the related fields of commerce, finance, political economy, and alumni of "Philosophy, Politics and Economics," an interdisciplinary study program with a strong economics component. Later we will also analyze five other educational backgrounds.

2.4. ECONOMETRIC ANALYSIS

Empirical Approach

In order to analyze the role of development ministers in shaping the size and "development-friendliness" of a donor's development cooperation, we estimate the association of the development ministers' personal characteristics with (1) the size of aid budgets (*TotalODA*), and (2) the quality of aid (*QualityODA*). To account for *TotalODA*, we use a donor country's total amount of ODA in logarithms and measured in constant 2012 US$ (OECD 2014).[79] We analyze both ODA commitments and disbursements as both measures come with their respective advantages and disadvantages. Commitments allow us to capture the impact that the development minister exerts directly on development policy. Disbursements in a particular year on the contrary may already have been committed under a minister's predecessor and thus falsely be attributed to the successor. The use of disbursements, however, comes with the advantage that one accounts for the development minister's influence on the current spending process. Moreover, Roodman (2012) points to the potential risk of overestimating aid when using aid commitments if ministers knowingly or unknowingly over-promise aid. Thus, only disbursements mirror the actual effort of donor countries. As can be seen from Figure 2.2., annual ODA disbursements are systematically smaller than commitments.

Figure 2.2. ODA quantity (1967-2012) and ODA quality (1995-2011)



[79] Note that we add a value of US$ 1 to all ODA values before taking logarithms so as not to lose zero values in our sample.

The measurement of *QualityODA* is not straightforward. Although several comprehensive indices propose various ways to measure the quality of ODA (Easterly 2002b; Easterly and Pfutze 2008; Birdsall and Kharas 2010; Knack et al. 2011), the quality-adjusted aid measure developed by Roodman (2012) is the only one that goes back as far as 1995.[80] Roodman discounts gross ODA disbursements for several factors that are judged as reducing the effectiveness of aid. He first subtracts debt forgiveness grants and rescheduled debt from OECD-defined ODA to obtain his measure of "gross aid." He then adjusts the amount of "gross aid" by the extent to which a donor's aid is tied, by principal and interest payments, and by administrative costs. Finally, Roodman also rewards policies that are expected to increase a donor's development impact. Specifically, he implements a selectivity weight for ODA given to poorer and to well-governed countries. The resulting measure of "quality-adjusted aid" is the amount of ODA that is estimated to be effective. As our measure of aid quality, we divide Roodman's quality-adjusted bilateral aid disbursements by his total bilateral gross aid disbursements. The resulting average values of *QualityODA* range between 18 percent for Japan and 56 percent for Sweden (see also Figure 2.2.).[81] Figure 2.3. shows for four important donors how both *TotalODA* and *QualityODA* evolve over the terms of development ministers.[82]

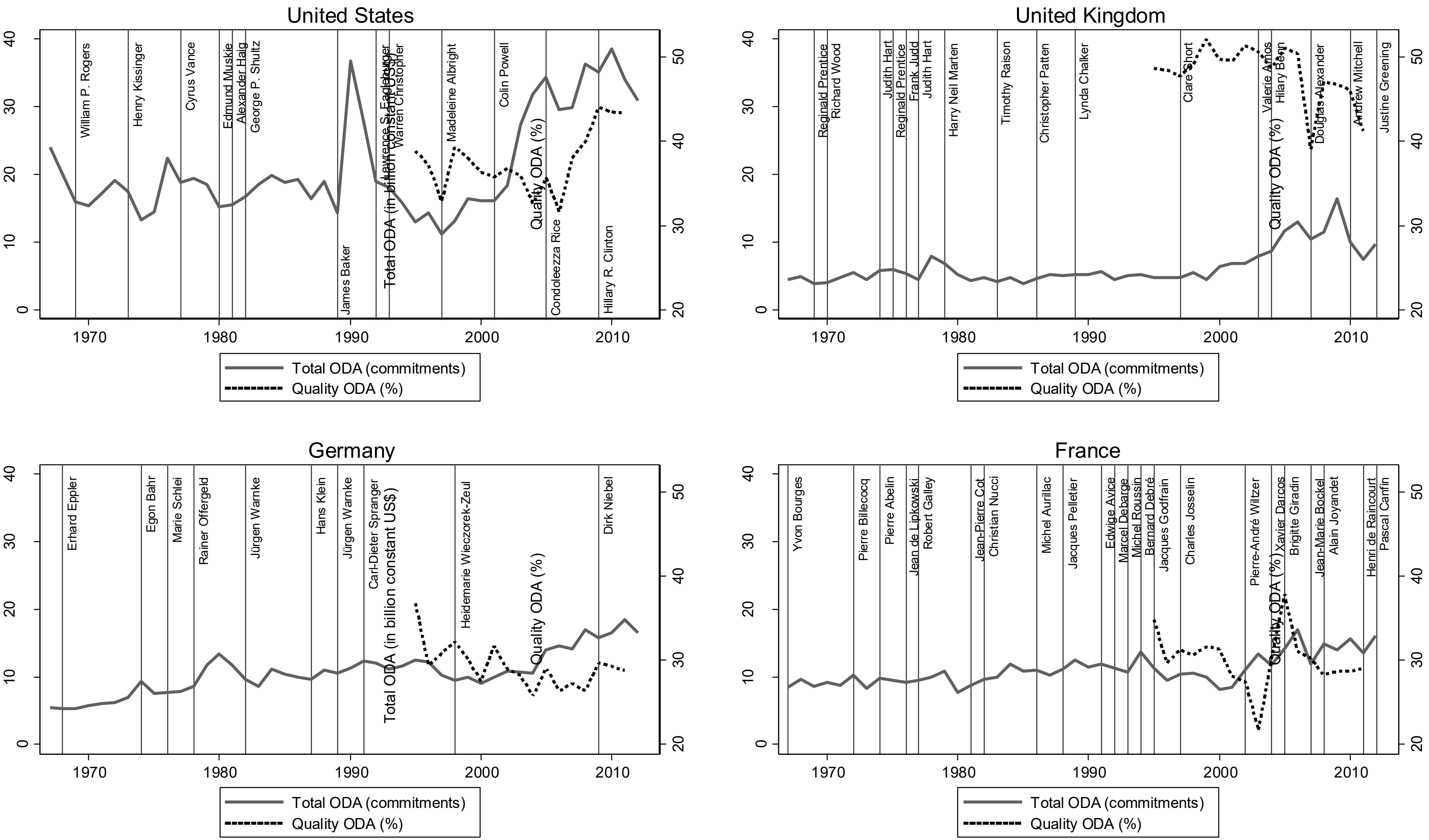
---

[80] We are grateful to David Roodman for having generously provided us with access to an extended version of his dataset.

[81] In our dataset, Portugal constitutes an outlier as its aid quality is highly volatile and it is the only donor that shows a negative quality-adjusted ODA value in a year (1997). This occurs as the country disproportionally supports richer recipients in that year. Hence, we set Portugal's value of *QualityODA* to zero in 1997. Note that our results below are robust to the exclusion of Portugal from our regressions.

[82] In Appendix A1.3., we show for each individual donor how aid quantity and quality developed since 1967.

Figure 2.3. ODA quantity (1967-2012) and ODA quality (1995-2011) of four important donor countries by development minister



Of course, this indicator is only one of many ways to measure ODA quality. However, it is important to note that there is a large overlap of Roodman's measure with alternative indicators of aid quality (Easterly 2002b; Easterly and Pfutze 2008; Birdsall and Kharas 2010; Knack et al. 2011). They are similar to Roodman (2012) as they all contain measures for selectivity towards poorer and better-governed countries and account for the degree to which a country's aid is untied. While most components are not subject to a larger debate, the benefits of aid selectivity are controversial. Most importantly, although several scholars argue that aid is more effective in better-governed countries (e.g., Svensson 1999; Burnside and Dollar 2000), there is no robust evidence that aid promotes growth if given to countries with good policies or institutions (Easterly et al. 2004; Doucouliagos and Paldam 2010). This is why we also show results below where we remove selectivity weights for good governance (and income). Moreover, we show results when removing the penalties for tied aid and project proliferation.

We regress our two dependent variables on the same set of independent variables, testing for a potential role of the ministers' gender, their political ideology, and their experience as captured by their prior professional experience in development cooperation, education in economics or business, and time in office.[83] Our regression equations read as follows:

(1) $$\log(TotalODA_{it}) = \beta_1 \log(TotalODA_{it-1}) + \beta_2\, Gender_{it-1} + \beta_3\, Ideology_{it-1} + \Sigma_l\, \beta_{4l}\, Experience_{ilt-1} + \Sigma_m\, \beta_{5m}\, Controls_{imt-1} + \eta_i + \mu_t + u_{it}$$

(2) $$QualityODA_{it} = \gamma_1\, QualityODA_{it-1} + \gamma_2\, Gender_{it} + \gamma_3\, Ideology_{it} + \Sigma_l\, \gamma_{4l}\, Experience_{ilt} + \Sigma_m\, \gamma_{5m}\, Controls_{imt} + \eta_i + \mu_t + v_{it}$$

The index *i* refers to the respective donor country, *t* stands for the respective year, *l* allows for the three different measures of experience, and *m* identifies the 15 variables that form our set of control variables described below.

While we lag all independent variables in the *TotalODA* regression by one year, the variables enter simultaneously in the *QualityODA* regression. This is because budget negotiations in national parliaments typically take place in the previous year. Once the budget is determined, ministers' decisions can influence the allocation of aid across recipients, sectors

[83] Although there are no clear expectations on how minister age could affect aid quantity and quality, we also explored a potential role of age. Since age never turned out to have a significant impact on either dependent variable, we excluded age from our specifications.

and types of aid throughout the year, all of which can affect ODA quality. To test the robustness of our results, we also show regression results below based on alternative timing decisions.

In order to approach the identification of a genuine effect stemming from the personal characteristics of the development minister, we additionally control for the corresponding personal characteristics of the respective head of government.[84] The inclusion of these control variables prevents us from falsely attributing the influence of heads of government to development ministers. Such a correlation is most obvious for political ideology as the selection of the development minister by the head of government will be a function of their respective political orientations.[85] Female heads of government might also be more likely to appoint female ministers and a similar argument can easily be made for heads of government with professional experience in development cooperation or training in economics due to networks or affinity towards candidates with similar characteristics. Moreover, government-head and minister characteristics might also be interlinked in more complex ways. For example, to the extent to which left-wing heads of government are more likely to appoint female ministers (Escobar-Lemmon and Taylor-Robinson 2005), a significant effect of the minister's gender might be driven by the political orientation of the head of government instead. Applying the same logic to the legislative, we additionally include the share of women in parliament (data from Brady et al. 2014 and World Bank 2014) and the mean ideological orientation of parliament members (data from Bjørnskov and Potrafke 2011) as control variables.

As we argue above, the low salience of aid policy provides reasons to believe that the selection of development ministers is more random than in other issue areas. Still, a strategic selection of development ministers is of concern.[86] For example, a head of government who believes that aid is important may on the one hand appoint a more experienced development

[84] We build on the dataset in Bjørnskov and Potrafke (2011), which in turn is based on Woldendorp et al. (2000), to create a dataset on the characteristics of donor countries' heads of government.

[85] This also applies to coalition governments since parties with a similar political orientation are more likely to build a coalition. The correlation between the ideology of ministers and heads of government in our sample is 0.656.

[86] To give an example, Day (2016: 21) notes that UK Prime Minister Tony Blair's "personal championing of developing issues afforded Clare Short, the Development Secretary, political cover to reform British aid policy. Short also enjoyed the active support of Chancellor Gordon Brown, who also desired to see the UK aid budget increase."

minister and on the other hand push for higher ODA budgets. If we then observe a positive correlation between the minister's experience and the size of ODA budgets, we may falsely attribute this to the minister's performance. To mitigate this concern, we also estimate regressions that include government-head-fixed effects instead of donor-country-fixed effects. Although this does not fully eliminate our concerns if heads of government change their preferences during their time in office, results based on this more conservative specification raise our confidence that the observed effects are indeed driven by differences in development minister characteristics rather than the strategic appointment of ministers by heads of governments as we only exploit variation within governments over time.

In the selection of the remaining control variables, we follow Fuchs et al. (2014) and Brech and Potrafke (2014). First, we include the lagged dependent variable to account for aid inertia. Bertoli et al. (2008: 24) argue that this is an important variable in aid budget regressions "since the persistence in budgetary allocations determines a significant path-dependence in the evolution of aid effort." Second, we control for (logged) per-capita GDP, trade openness, government expenditure as a share of GDP (data from World Bank 2014) and the debt-over-GDP ratio (Abbas et al. 2010) to capture the donor country's (international) economic and fiscal situation. Third, we include a donor country's level of political globalization (Dreher 2006; Dreher et al. 2008) to account for the transmission of ideas through networks of intergovernmental and non-governmental organizations (Lundsgaarde et al. 2007). Fourth, we add a binary variable for the existence of an aid agency in the donor country (Fuchs et al. 2014). As Bertoli et al. (2008: 15) argues, "an independent aid agency may be able to preserve an appropriate aid level and allocation regardless of the political orientation and aid preferences of the newly elected government." Finally, we control for the logged size of the population living in the donor's former colonies (Mayer and Zignago 2006; World Bank 2014) to account for the donor country's historical past. Table 2.1. provides descriptive statistics and Appendix A2.2. gives an overview of all variables used, their definitions and sources.

Table 2.1. Descriptive statistics

Variable name	**Obser-vations**	**Mean**	**Sd**	**Min**	**Max**
(log) ODA commitments	919	21.50	1.40	17.12	24.37
(log) ODA disbursements	957	21.30	1.51	16.46	24.21
Quality ODA	389	41.74	11.31	0.00	69.24
Quality ODA (no selectivity reward)	389	66.51	13.92	0.00	86.23

Quality ODA (no proliferation penalty)	389	53.68	14.17	10.21	78.16
Quality ODA (no tied aid penalty)	389	43.95	10.64	0.00	75.14
(log) GDP per capita	931	10.41	0.39	9.11	11.53
Openness	940	69.14	45.22	9.68	352.90
Gov. expenditure	940	18.71	3.69	8.09	28.06
Debt	957	52.15	32.90	0.00	238.03
Political globalization	883	85.85	11.64	45.34	98.43
Aid agency	957	0.42	0.49	0	1
(log) Colonial history	957	10.33	8.70	0.00	21.56
Female minister	957	0.20	0.40	0	1
Female gov. head	957	0.06	0.24	0	1
Female parliament	949	16.66	11.91	0.00	47.30
Right-wing minister	957	0.31	0.40	-0.50	1.00
Right-wing gov. head	957	0.31	0.35	-0.50	1.00
Right-wing parliament	934	0.27	0.17	-0.43	0.78
Ideological difference	957	0.27	0.44	0	1
Ideological difference (more right-wing)	957	0.14	0.35	0	1
Ideological difference (more left-wing)	957	0.13	0.34	0	1
Right-wing minister (election manifesto)	953	-3.56	19.61	-58	48.46
Prof. dev. coop. minister	957	0.16	0.36	0	1
Prof. dev. coop. gov. head	957	0.04	0.20	0	1
Economics & business minister	957	0.22	0.42	0	1
Economics & business gov. head	934	0.31	0.46	0	1
Economics minister	957	0.19	0.40	0	1
Economics gov. head	934	0.23	0.42	0	1
Business administration, commerce, or finance minister	957	0.05	0.22	0	1
Business administration, commerce, or finance gov. head	934	0.09	0.28	0	1
Law minister	957	0.38	0.48	0	1
Law gov. head	957	0.46	0.50	0	1
Medicine, public health, or pharmacy minister	957	0.02	0.13	0	1
Medicine, public health, or pharmacy gov. head	957	0.01	0.11	0	1
Nat. sciences, mathematics, engineering, or agr. science minister	957	0.05	0.22	0	1
Nat. sciences, mathematics, engineering, or agr. science gov. head	957	0.05	0.21	0	1
Political science or other social sciences minister	957	0.27	0.44	0	1

Political science or other social sciences gov. head	957	0.22	0.41	0	1
Teaching, social work, or pedagogics minister	957	0.09	0.29	0	1
Teaching, social work, or pedagogics gov. head	957	0.02	0.14	0	1
Tenure minister	957	2.99	2.35	1	15
Tenure gov. head	957	4.34	3.12	1	18
General ministerial experience minister	952	19.27	32.88	0	219
Tenure minister + General ministerial experience minister	952	55.06	43.97	12	252
Head of government in the future	957	0.06	0.23	0	1

Moreover, we construct a binary variable that takes a value of one if the political orientation of the development minister and the head of government is different. This is to account for government fractionalization: a larger number of parties involved in decision-making results in more compromises and more concessions being granted to each party (Dreher and Langlotz 2015). One could thus expect that the quantity of ODA increases if the development minister and the head of government have different political ideological orientations.[87]

We estimate our baseline models with ordinary least squares (OLS) using donor-country- and year-fixed effects and standard errors that are robust to heteroskedasticity and clustered at the government-head level. Such a fixed-effects estimation that includes a lagged dependent variable may lead to inconsistent estimators and induce the so-called Nickell bias through the correlation of the lagged dependent variable with the error term (Nickell 1981). However, with an average number of time periods per donor of over 35 years, the problem should be negligible in our *TotalODA* regressions and OLS appropriate. Since the time period covered in our *QualityODA* regressions is much shorter with a maximum of 17 years, we also show results from two further specifications: first, we exclude the lagged dependent variable; second, we exclude country-fixed effects (and keep the lagged dependent variable). As noted by Angrist and Pischke (2008), the first approach may overestimate the true effect, while the latter may lead to an underestimation. The true effect will thus lie in between this upper and lower bound.

[87] Beyond this set of personal characteristics, we further constructed a binary variable that takes a value of one if the development minister is head of a separate ministry for development cooperation. Since this variable never showed up significantly in our regressions, we decided to drop it from the analysis below. Results are available upon request.

The Quantity of ODA

Table 2.2. presents the results for aid budgets. We start by analyzing the results of a specification that excludes government-head characteristics (columns 1 and 2) and then add the personal characteristics of the heads of government as well as variables capturing the gender and ideological composition of national parliaments to approach the identification of a genuine effect stemming from the characteristics of development ministers (columns 3 and 4). We examine both total ODA commitments (columns 1 and 3) and total ODA disbursements (columns 2 and 4). Our model has large explanatory power as evidenced by the R-squared of more than 80 percent. It is also reassuring that the results for the control variables are largely in line with expectations: we find evidence for persistent ODA budgets and for donors providing more ODA as they grow richer, as shown by the positive and highly significant coefficients on *lagged DV* and *(log) GDP per capita*. *Openness* and *government expenditure* enter with the expected positive sign but do not reach statistical significance at conventional levels in each specification. A higher debt-over-GDP ratio is associated with lower total ODA levels, at the one-percent level of significance. *Political globalization* shows the expected significant positive relationship with the quantity of aid, at least at the ten-percent level. The existence of an *aid agency* in the donor country is positively linked with total ODA disbursements (but not with commitments), at the one-percent level of significance. In line with Bertoli et al. (2008), ODA budgets appear to work as substitutes for a colonial legacy, at least at the ten-percent level of significance.

Table 2.2. Development minister characteristics and total ODA budgets (1971-2012)

	(1) (log) Total ODA com.	(2) (log) Total ODA disb.	(3) (log) Total ODA com.	(4) (log) Total ODA disb.	(5) (log) Total ODA com.	(6) (log) Total ODA disb.
Lagged DV	0.4750***	0.6984***	0.4774***	0.6566***	0.0979	0.2927***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.100]	[0.000]
(log) GDP per capita	1.3020***	0.8093***	1.3170***	0.9722***	2.2188***	1.9817***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Openness	0.0012	0.0008	0.0016	0.0016***	-0.0008	-0.0016*
	[0.237]	[0.112]	[0.108]	[0.003]	[0.525]	[0.070]
Gov. expenditure	0.0176**	0.0100	0.0159*	0.0102	0.0362**	0.0391***
	[0.036]	[0.145]	[0.055]	[0.117]	[0.023]	[0.004]
Debt	-0.0014***	-0.0009***	-0.0012***	-0.0013***	-0.0010	-0.0004
	[0.001]	[0.001]	[0.005]	[0.000]	[0.272]	[0.456]
Political globalization	0.0050**	0.0024*	0.0060**	0.0043**	0.0053	0.0059*
	[0.021]	[0.086]	[0.024]	[0.029]	[0.236]	[0.063]

Aid agency	0.0300	0.0765***	0.0162	0.0827***	0.0210	0.2088**
	[0.371]	[0.004]	[0.614]	[0.003]	[0.827]	[0.012]
(log) Colonial history	-0.2252***	-0.1266**	-0.1414*	-0.1114*	-0.4366	-0.5402***
	[0.004]	[0.029]	[0.054]	[0.061]	[0.234]	[0.003]
Female minister	-0.0355	-0.0215	-0.0289	-0.0274	-0.0381	-0.0434
	[0.107]	[0.266]	[0.186]	[0.189]	[0.227]	[0.150]
Female gov. head			0.0088	0.0226		
			[0.791]	[0.370]		
Female parliament			0.0064***	0.0039**		
			[0.002]	[0.022]		
Right-wing minister	0.0075	-0.0063	0.0008	-0.0356*	0.0175	0.0113
	[0.713]	[0.661]	[0.979]	[0.088]	[0.739]	[0.722]
Right-wing gov. head			-0.0062	0.0112		
			[0.883]	[0.678]		
Right-wing parliament			-0.0134	0.2299**		
			[0.880]	[0.014]		
Ideological difference			0.0693***	0.0371**		
			[0.001]	[0.015]		
Prof. dev. coop. minister	0.0160	0.0265	0.0017	0.0243	-0.0496	-0.0034
	[0.505]	[0.143]	[0.944]	[0.164]	[0.172]	[0.910]
Prof. dev. coop. gov. head			-0.0012	0.0043		
			[0.968]	[0.841]		
Economics & business minister	0.0085	0.0157	-0.0050	0.0081	-0.0156	-0.0019
	[0.625]	[0.205]	[0.753]	[0.508]	[0.621]	[0.931]
Economics & business gov. head			0.0377*	0.0390***		
			[0.066]	[0.005]		
Tenure minister	0.0074**	0.0047*	0.0068*	0.0054**	0.0088	0.0070**
	[0.044]	[0.069]	[0.059]	[0.047]	[0.134]	[0.046]
Tenure gov. head			-0.0023	-0.0034*		
			[0.342]	[0.078]		
Country FE	Yes	Yes	Yes	Yes	No	No
Gov. head FE	No	No	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	836	866	808	838	836	866
R-squared (within)	0.813	0.917	0.818	0.916	0.459	0.647
Number of countries	23	23	23	23	23	23
Number of gov. heads	182	187	180	185	182	187
Average number of years	36.4	37.7	35.1	36.4	4.6	4.6

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Turning to the interpretation of the results for our variables of interest, we do not find a significant relationship between the gender of development ministers (and heads of government) and *TotalODA*. However, the coefficient on *female parliament*, i.e., the share of women in national parliaments, is positive and statistically significant at least at the five-percent level – in line with the findings in Hicks et al. (2016). A one-percent increase in the share of female deputies in parliament raises ODA commitments by 0.6 percent and ODA disbursements by 0.4 percent on average (columns 3 and 4). This finding shows that female representation matters, but at the *legislative* not *executive* level (see also Lu and Breuning 2014).

We also do not find a significant link between ODA volumes and development ministers' political orientation on a left-right scale of economic policy orientation.[88] This is in line with our expectations: ministers fight for an increase in their budgets independently of where they stand ideologically. This also holds when we replace our ideology measure with a right-left scale based on parties' statements in election programs.[89] Also, the political ideology of the heads of government does not appear to be associated with aid budgets – an unsurprising result given prior empirical research summarized in Fuchs et al. (2014). Again, we find differences between the executive and parliament. More economically liberal parliaments are associated with larger total ODA disbursements at the five-percent level of significance.[90] Moving one ideological unit to the right – such as from social democratic to economically liberalist – increases disbursements by 23 percent. When we split the sample into the Cold War and post-Cold War period (see Appendix B2.1., columns 3-4, for details), we find that the effect stems from the Cold War era. This finding casts doubts on widespread expectations that more right-wing parliaments provide less aid and thus corroborates similar results in Dreher et al. (2015b). This could also hint at right-wing politicians using aid during the Cold War more intensively as a geostrategic instrument than left-wing politicians.

The binary variable *ideological difference,* indicating that the head of government and development minister have different political orientations, turns out to be positive and statistically significant at least at the five-percent level. Divided governments show an increase of total ODA commitments by 7 percent and of total ODA disbursements by 4 percent. We check whether this is indeed driven by diverging interests rather than by development ministers being more left-wing and thus potentially more aid-supporting than their heads of government. In order to do so, we construct two new binary variables for ideological

[88] Our findings are similar when we replace the five-tier-scaled variable with a simple binary variable for right-wing ministers (conservative or economically liberal parties). Results are available upon request.

[89] Specifically, we use the Manifesto Project Dataset (Volkens et al. 2016) and match the right-left position of each party to our development minister database. The correlation with our baseline measure of minister ideology over our sample period is 52.2 percent. See Appendix B2.2. for detailed regression results.

[90] This finding is not robust for commitments but holds when we exclude the lagged dependent variable (Appendix B2.1., columns 1 and 2).

differences. The first takes a value of one if the development minister is more left-wing than the head of government. The second takes a value of one if the development minister is more right-wing than the head of government. When we replace the baseline measure of ideological differences with these two new variables, both variables are jointly significant at the one-percent level (ODA commitments) and at the ten-percent level (ODA disbursements), respectively. However, their coefficients are not significantly different from one another at conventional levels of significance (p-values of 0.660 and 0.646; see Appendix B2.3. for details). This supports the view that diverging interests within the government causes each partner to grant the other partner more concessions (e.g., Dreher and Langlotz 2015).

With respect to our experience variables, we find that ODA budgets increase with the tenure of the development minister, at conventional levels of significance (columns 1-4). More precisely, each additional year of experience as development minister increases ODA commitments by 0.7 percent and disbursements by 0.5 percent on average (columns 3 and 4).[91] Taking this at face value, Luxembourg's Jacques Poos in his 15th year would have been able to secure an aid budget that is roughly 10 percent larger than that of a newcomer. As hypothesized, political experience as development minister appears to provide ministers with an advantage in their fight for higher budgets. There is also some evidence that a longer tenure of heads of government has the opposite, negative effect on aid budgets, as evidenced by the negative significant coefficient on *tenure gov. head* in column 4 (at the ten-percent level). More experience in office seems to enable heads of government to better defend against demands for budget increases. An additional year of experience as head of government is associated with a reduction of ODA disbursements by 0.3 percent.[92] In contrast to on-the-job experience, development ministers' prior professional experience in development cooperation *(prof. dev. coop. minister*) and training in economics or business (*economics & business minister*) do not seem

---

[91] When we omit the lagged dependent variable, the effect increases to 1.0 and 1.2 percent, respectively (Appendix B2.1., columns 1-2). The finding appears to be driven by the post-Cold War period (Appendix B2.1., columns 3-6). Given the inclusion of the lagged dependent variable, the values above correspond only to the short-run effect of tenure. Taking account of the coefficient of the lagged dependent variable, we obtain long-run effects of 1.3 and 1.7 percent, respectively.

[92] Non-linearities in the accumulation of political experience do not appear to play a significant role. Our results also show that it is the experience on the development office rather than general ministerial experience that matters for ODA budgets. Finally, as a proxy of unobserved ability, we also code a binary variable that takes a value of one if the development minister will become head of government in the future. However, ministers who make it to the head of government do not appear to attract more aid money. See columns 1 and 2 of Appendix B2.4. for detailed results.

to be associated with ODA budgets as none of the corresponding coefficients reaches statistical significance at conventional levels.[93] On the contrary, we find evidence for larger aid budgets when the government is led by a trained economist.

Columns 5 and 6 of Table 2.2. present regressions with government-head-fixed effects rather than government-head characteristics and donor-country-fixed effects. In this very strict specification, we only identify the possible effects of development minister characteristics on aid budgets through variation of development ministers *within* the tenure of each particular head of government. Focusing on the variables of interest for the sake of brevity, the results confirm our main findings from above. Development ministers' experience in office appears to be the only relevant minister characteristic that plays a role for aid budgets. *Tenure minister* remains positive and statistically significant at the five-percent level for ODA disbursements, while all other development minister characteristics do not reach statistical significance in this conservative specification. Even when we add minister-fixed effects to our baseline model, i.e., identify the effect of tenure on ODA budgets only through variation within ministers over their time in office, we still find support for this hypothesis (see panel E of Appendix B2.4.).[94]

To sum up, the development ministers' personal characteristics do not seem to matter much with regards to the quantity of ODA. Only one finding appears to be largely robust: a longer time in the development office strengthens the ministers' ability to negotiate higher ODA budgets. This also holds when we use contemporaneous rather than lagged explanatory variables (Appendix B2.6.) or replace our dependent variable by ODA disbursements as a share of government expenditures (Appendix B2.7.).[95] The appointment of more experienced

---

[93] We also analyzed economics and business separately as well as five other educational backgrounds of development ministers, i.e., university degrees in (1) law, (2) medicine, pharmacy, or public health, or, (3) sciences, mathematics, engineering, or agricultural science, (4) political science or other social sciences, and (5) teaching, social work, or pedagogics. None of them appears to play a robust role for ODA budgets. Appendix A2.4. provides descriptive statistics on the frequency of these degrees and columns 1 and 2 of Appendix B2.5. present detailed regression results.

[94] Given that this is a very demanding test, it is not surprising that the corresponding coefficients have p-values of 0.062 and 0.102 and are only weakly significant or even marginally insignificant.

[95] Since the inclusion of a lagged dependent variable can do harm to the estimates of other explanatory varables (e.g., Achen 2010), it is important to highlight that our qualitative results are largely unaffected when we remove the lagged dependent variable (Appendix B2.1.). One may also argue that political globalization and the existence of an aid agency constitute "bad controls" in the terminology of Angrist and Pischke (2008). Development ministers' characteristics may affect the decisions to enter international organizations and sign treaties, i.e., they could affect a country's degree of political

ministers can thus help countries to achieve the UN target to provide 0.7 percent of GNI as development aid. Political experience seems to pay off.

The Quality of ODA

Table 2.3. shows our results for ODA quality. We begin our analysis with specifications that exclude the lagged dependent variable (columns 1 and 2). While the regression in column 1 additionally excludes government-head and legislative controls, these variables are included in column 2. Starting with the interpretation of the results for the control variables, the coefficient on *(log) GDP per capita* is positive and reaches statistical significance at the five-percent level. High-quality ODA thus seems to be a luxury good: as countries become richer, incentives to provide self-interested aid shrink. In line with this explanation, a larger *debt*-to-GDP ratio is associated with significantly lower ODA quality (at the one-percent level). *Government expenditure* on the other hand is positively associated with higher ODA quality, suggesting that governments with higher expenditures have more experience with redistribution. *Openness* does not have a significant relationship with ODA quality. ODA quality decreases with the degree of *political globalization* and the existence of an *aid agency*, both at the ten-percent level of significance. While the effect of political globalization is surprising as one would expect more globalized countries to embrace a greater role in international development, aid agencies have higher administration costs which might harm the share of quality ODA. Finally, the positive and highly significant coefficient on *(log) colonial history* suggests that stronger ties with developing countries could give former colonial powers insights into where aid will be most effective and how to channel their aid more efficiently.

Table 2.3. Development minister characteristics and ODA quality (1995-2011)

	(1) Quality ODA	(2) Quality ODA	(3) Quality ODA	(4) Quality ODA	(5) Quality ODA
Lagged DV			0.2449**	0.7029***	0.0293
			[0.038]	[0.000]	[0.735]
(log) GDP per capita	15.2759**	16.9968**	11.1843**	-0.2558	26.9395**
	[0.026]	[0.014]	[0.026]	[0.826]	[0.030]
Openness	0.0348	0.0354	0.0164	0.0203*	0.0264
	[0.242]	[0.213]	[0.524]	[0.076]	[0.357]
Gov. expenditure	1.1572***	0.9757**	0.5387	0.2746	1.4362

globalization. In addition, the institutional design of aid agencies may depend on the respective development minister in charge. When we remove these potential "bad controls" in addition, our results remain similar (Appendix B2.8.).

	[0.003]	[0.024]	[0.256]	[0.161]	[0.156]
Debt	-0.1178***	-0.1009***	-0.0634***	-0.0560**	-0.0480
	[0.000]	[0.000]	[0.008]	[0.032]	[0.133]
Political globalization	-0.2675*	-0.2317*	-0.1476	0.1131	-0.0967
	[0.055]	[0.077]	[0.199]	[0.149]	[0.417]
Aid agency	-1.7096*	-1.8045*	-1.6019**	-1.3811*	-2.4474
	[0.070]	[0.073]	[0.040]	[0.069]	[0.233]
(log) Colonial history	16.9263***	16.7040***	17.3956***	-0.0761	20.4548**
	[0.001]	[0.002]	[0.001]	[0.134]	[0.039]
Female minister	0.8782	1.0448*	1.2345**	0.9802*	0.7638
	[0.118]	[0.099]	[0.010]	[0.064]	[0.185]
Female gov. head		-2.4266**	-3.1056**	-0.8441	
		[0.046]	[0.024]	[0.354]	
Female parliament		0.1193	-0.0432	-0.0161	0.0726
		[0.231]	[0.605]	[0.728]	[0.609]
Right-wing minister	0.4247	-0.6945	-0.6600	-0.7011	-0.8947
	[0.624]	[0.529]	[0.421]	[0.504]	[0.623]
Right-wing gov. head		2.1887	2.0506*	0.7247	
		[0.122]	[0.054]	[0.551]	
Right-wing parliament		-5.1830	-4.3708	4.6252	5.8445
		[0.134]	[0.161]	[0.192]	[0.358]
Ideological difference		0.1781	-0.1828	-0.0312	-1.2962
		[0.801]	[0.752]	[0.963]	[0.263]
Prof. dev. coop. minister	1.2172**	1.1406*	1.1491**	-0.5053	1.3063*
	[0.047]	[0.088]	[0.044]	[0.493]	[0.097]
Prof. dev. coop. gov. head		-2.5437	-2.0237	-0.4762	
		[0.120]	[0.113]	[0.560]	
Economics & business minister	-0.7844	-0.9859	-0.5221	0.1794	0.6708
	[0.161]	[0.137]	[0.304]	[0.774]	[0.213]
Economics & business gov. head		-0.1709	-0.4718	0.0914	
		[0.828]	[0.472]	[0.874]	
Tenure minister	0.1207	0.1693*	0.1703**	0.1135	0.1658
	[0.241]	[0.089]	[0.040]	[0.136]	[0.268]
Tenure gov. head		0.0945	0.0758	-0.0039	
		[0.345]	[0.331]	[0.963]	
Country FE	Yes	Yes	Yes	No	No
Gov. head FE	No	No	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	389	387	365	365	365
R-squared (within)	0.231	0.263	0.288	0.807	0.159
Number of countries	23	23	23	23	23
Number of gov. heads	88	88	82	82	82
Average number of years	16.9	16.8	15.9	15.9	4.5

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Turning to our variables of interest, we find a positive association between *female minister* and the quality of ODA, which is statistically significant at the ten-percent level in column 2. Interpreting the size of the effect, we find that ODA quality increases by one percentage point when a woman directs development policy. Even though this effect is not of a large magnitude and only corresponds to US$ 28 million of additional quality ODA annually when holding gross aid constant, the result supports the idea that women and men differ in their preferences. Strikingly, however, the coefficient on *female gov. head* has the opposite sign and is statistically significant at the five-percent level. Specifically, we find that ODA quality decreases by 2.4 percentage points when a woman leads government.[96] Women in the highest political position might feel the urge to overcome the gender stereotype as discussed by Koch and Fulton (2011). Consequently, female heads of government might support more self-interested aid giving than their male counterparts.[97] The insignificant coefficient on *female parliament* suggests that no comparable gender difference exists at the legislative level. We conclude that gender differences in aid quality cannot be generalized and depend on the specific position under analysis.

With respect to political ideology, we find no robust significant relationship between our variables and the quality of ODA. This non-finding applies to development ministers, heads of government, the ideological differences of the former and latter, and also to the ideology of parliamentarians. We also obtain a non-finding when we replace our baseline measure of minister ideology with one based on parties' statements in election programs (column 3 of Appendix B2.2.). We conclude that the quality of ODA is independent of the political ideology of the relevant decision-makers.

Continuing with development ministers' experience, we find that ministers' professional experience in development cooperation as well as their experience in office matter for the quality of ODA. According to column 2, ministers with specific development experience succeed in increasing aid quality by 1.1 percentage points. Assuming gross aid to be constant, the average donor provides US$ 30 million of additional quality ODA annually if a development minister with such a background comes to power. Accordingly, we find some support for our hypothesis that ministers with development experience deliver higher quality

---

[96] Only 11 of 210 heads of government in our sample are female, accounting for 6.3 percent of the minister-years. Hence, we cannot rule out that this finding is driven by individual personalities and that we thus only capture something like a Thatcher or Merkel effect. Regression results from specifications that individually drop one of the ten countries that ever had a female government head from the estimation sample suggest that the effect is largely driven by Australia and New Zealand.

[97] The comparison of regressions that exclude (column 1) and include (column 2) government-head and legislative controls supports our decision to include these variables. In column 1, *female minister* supposedly absorbs part of the negative effect of *female gov. head*, leaving *female minister* with an insignificant and less positive coefficient than in column 2.

ODA – either because they have acquired knowledge of how aid can be more effective or because they have a higher intrinsic motivation to give quality ODA.[98] Also, development ministers shift more resources to "development-friendlier" activities as they gain experience on the job. One additional year in office raises the share of quality ODA given on average by 0.2 percentage points, amounting to US$ 4.7 million for the average donor (column 2). Among the three variables capturing the ministers' experience, *economics minister* is the only one not to reach statistical significance at conventional levels. Economists thus neither appear to be more selfish in the sense that they provide lower quality aid, nor more able to raise aid quality.[99] Also, the corresponding government-head experience variables do not reach statistical significance in our regressions.[100]

The *lagged DV* reaches statistical significance at the five-percent level when we include it in column 3. The quality of aid thus appears to be path dependent although the relationship is weaker than for aid quantity. The results for most control variables are similar compared to the specifications excluding the *lagged DV* in columns 1 and 2. Only *gov. expenditure* and *political globalization* lose statistical significance at conventional levels. Most notably, our findings for the minister characteristics are robust and the statistical significance of *female minister*, *prof. dev. coop. minister* and *tenure minister* actually increase from the ten-percent to the five-percent level.[101]

Since the average number of years per panel is 16 when using donor-country-fixed effects with lagged DV (column 3), our estimates may suffer from the Nickell bias. As discussed above, Angrist and Pischke (2008) recommend regressions that only include either

[98] We cannot control for experience in domestic regional development, which could affect ODA quality in a similar way. However, this should – if at all – downward bias our findings for experience in international development and thus work in our favor.

[99] Analyzing the six other educational backgrounds of development ministers described in footnote X above, we do not find most of them to affect ODA quality. Only ministers with degrees in medicine, public health, or pharmacy appear to provide lower-quality aid than ministers with other degrees or no degree. See again Appendix B2.5.

[100] When we remove *political globalization* and *aid agency*, two potentially "bad controls" as discussed above, results are similar (columns 7-3 of Appendix B2.8.).

[101] This is the only specification in which *right-wing gov. head* becomes statistically significant. Governments led by right-wing politicians seem to provide an aid quality that is 2 percentage points higher according to column 3. This finding is the opposite of our expectation that right-wing politicians might link more domestic economic benefits to aid, leading to lower ODA quality. However, we do not put much emphasis on this finding as the coefficient is only significant at the ten-percent level and does not hold in any other specification.

the lagged dependent variable (column 4) or country-fixed effects (column 2) to bracket the true effect between an upper and lower bound. We find a significantly positive coefficient on *female minister* in all specifications, but our findings for *prof. dev. coop. minister* and *tenure minister* are not robust to the exclusion of country-fixed effects.

As we discussed above, Roodman's decision to reward recipient countries with good institutions is particularly debatable. Removing the aid selectivity adjustment, i.e., excluding the punishment for aid flows to richer and poorly governed countries, our finding that a development background is associated with higher ODA quality becomes stronger (see column 2 of Table 2.4.). However, we do not find anymore that ODA quality is higher if a female minister is in office and that ODA quality improves with a longer time in office. Both earlier findings thus appear to be driven by these ministers' focus on aid selectivity. Ministers with development experience seem to be most suited to reduce problems associated with project proliferation as the respective coefficient becomes smaller and insignificant once we remove the penalty for proliferation (column 3 of Table 2.4.).

Table 2.4. Development minister characteristics and ODA quality (1995-2011, alternative adjustments)

	(1) Quality ODA	(2) Quality ODA No selectivity reward	(3) Quality ODA No proliferation penalty	(4) Quality ODA No tied aid penalty
Lagged DV	0.2449**	0.2933**	0.3152***	0.1857*
	[0.038]	[0.029]	[0.003]	[0.072]
(log) GDP per capita	11.1843**	7.8445	11.5880*	7.0939
	[0.026]	[0.323]	[0.060]	[0.161]
Openness	0.0164	-0.0288	0.0364	0.0136
	[0.524]	[0.420]	[0.215]	[0.628]
Gov. expenditure	0.5387	0.2850	0.7181	0.4291
	[0.256]	[0.619]	[0.135]	[0.374]
Debt	-0.0634***	-0.0829**	-0.0504**	-0.0607***
	[0.008]	[0.011]	[0.026]	[0.005]
Political globalization	-0.1476	-0.0863	-0.0830	-0.1198
	[0.199]	[0.544]	[0.490]	[0.328]
Aid agency	-1.6019**	-1.6809	-2.4307***	-1.1258
	[0.040]	[0.210]	[0.003]	[0.184]
(log) Colonial history	17.3956***	14.5201**	21.8077***	15.9239***
	[0.001]	[0.010]	[0.001]	[0.004]
Female minister	1.2345**	0.7243	1.6455***	0.8830*
	[0.010]	[0.293]	[0.001]	[0.056]
Female gov. head	-3.1056**	-1.9396**	-2.1522	-2.7389**
	[0.024]	[0.042]	[0.167]	[0.037]
Female parliament	-0.0432	0.0322	0.0043	-0.0636

	[0.605]	[0.792]	[0.965]	[0.430]
Right-wing minister	-0.6600	1.5051	-1.2285	-0.7472
	[0.421]	[0.157]	[0.159]	[0.343]
Right-wing gov. head	2.0506*	1.1456	2.5154**	2.1377**
	[0.054]	[0.387]	[0.026]	[0.040]
Right-wing parliament	-4.3708	-11.4162*	-3.6732	-6.1191*
	[0.161]	[0.059]	[0.262]	[0.052]
Ideological difference	-0.1828	1.0236	-0.6830	-0.2660
	[0.752]	[0.139]	[0.295]	[0.640]
Prof. dev. coop. minister	1.1491**	2.1488***	0.5334	1.3129**
	[0.044]	[0.009]	[0.383]	[0.031]
Prof. dev. coop. gov. head	-2.0237	-2.1776	-3.6360***	-0.9674
	[0.113]	[0.124]	[0.004]	[0.394]
Economics & business minister	-0.5221	-0.9598	-0.8006	-0.0185
	[0.304]	[0.186]	[0.164]	[0.971]
Economics & business gov. head	-0.4718	0.0266	-0.5163	-0.4827
	[0.472]	[0.978]	[0.370]	[0.521]
Tenure minister	0.1703**	0.1260	0.2321**	0.0977
	[0.040]	[0.216]	[0.030]	[0.205]
Tenure gov. head	0.0758	0.0612	0.0697	0.1110
	[0.331]	[0.603]	[0.414]	[0.179]
Country FE	Yes	Yes	Yes	Yes
Gov. head FE	No	No	No	No
Year FE	Yes	Yes	Yes	Yes
Observations	365	365	365	365
R-squared (within)	0.288	0.344	0.361	0.252
Number of countries	23	23	23	23
Number of gov. heads	82	82	82	82
Average number of years	15.9	15.9	15.9	15.9

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Summing up, we find evidence that the ministers' gender and experience matter for ODA quality. This also holds when we use lagged rather than contemporaneous explanatory variables (Appendix Table B2.9.). As hypothesized, development ministers' experience, in the form of prior professional experience in development cooperation and within their office, increases the quality of ODA. Although female ministers appear to provide higher-quality ODA, the opposite is true for aid given during the terms of female heads of government. This finding thus cautions against generalizations about gender differences in aid giving and highlights that such differences depend on the specific position in the political hierarchy being considered. We find no evidence that ministers' ideology or an education in economics or business are linked with ODA quality. However, our findings are contingent on the chosen definition of ODA quality and some results lack robustness to specifications that exclude country-fixed effects.

2.5. CONCLUSIONS

Development ministers play an important role in determining aid outcomes but the importance of their role in this office often goes unnoticed by their home country. According to a poll conducted in September 2014 by Forsa, only 2 percent of Germans know that Gerd Müller is their development minister.[102] Almost two years after the French development minister Pascal Canfin took office in May 2012 (and shortly before he was removed again), a French news magazine still listed him among the "unknown ministers."[103] In sharp contrast to the disinterest in development issues in many governments, parliaments and the public, aid decisions taken at donor ministries can have huge impacts on the ground as DAC countries alone provide more than US$ 100 billion annually to the developing world. Building on the burgeoning literature on political leadership, this chapter is an attempt to assess the link between the personal characteristics of the government member responsible for development cooperation ("development minister" in short) and the quantity and quality of ODA.

To examine this research question, we introduce a novel database covering all development ministers of OECD-DAC countries since the OECD started reporting detailed ODA flows in 1967. The outcome is a dataset covering 320 ministers, active in 23 countries over 46 years. Using panel econometric models, we then estimate the link between development ministers' personal characteristics and (1) the size of aid budgets, and (2) aid quality as operationalized by the foreign-aid component of the Commitment to Development Index (CDI). Specifically, we test the role of the minister's gender, political ideology, prior professional experience in development cooperation, university education, and time in office. In order to approach the identification of a genuine effect stemming from the personal characteristics of the development minister, we control for the corresponding personal characteristics of the respective head of government as well as donor-country- and time-fixed effects. In order to mitigate selection effects, we also run regressions with government-head-fixed effects and thus identify effects only through variation within governments over time.

---

102 See poll "stern-RTL Wahltrend" available at http://www.stern.de/politik/deutschland/stern-rtl-wahltrend-das-milieu-der-afd-waehler-ist-rechtspopulistisch-bis-rechtsradikal-2137035.html (accessed 27 November 2014).

103 "Ces ministres dont on ne connait toujours pas le nom," *Valeursactuelles.com*, 3 March 2014, available at http://www.valeursactuelles.com/politique/ces-ministres-dont-on-ne-connait-toujours-pas-le-nom-43891 (last accessed 8 December 2014).

Our results show that experience matters. In line with our expectations, we find that the minister's political experience, as measured by their time in the development office, is positively associated with the quantity of aid: one additional year in office increases total ODA volume by 0.7 percent (controlling for the total ODA volume of the previous year). If one wants to raise the likelihood that a country's aid level achieves the UN target to provide 0.7 percent of GNI as development aid, one should hire an experienced development minister. If one believes that aid is detrimental to development outcomes, one should favor a rookie instead. With respect to the quality of aid, one additional year in office raises the share of quality ODA by 0.2 percentage points. Ministers with prior work experience in the field of development cooperation provide on average 1.1 percentage points larger share of quality ODA than ministers that lack such experience. Moreover, female ministers appear to be more successful in providing quality ODA than their male counterparts: the share of quality ODA increases by 1 percentage point when women lead the ministry responsible for development cooperation. However, we cannot fully rule out reverse causation and our findings for aid quality lack robustness to specifications that exclude country-fixed effects.

Taken together, our results show that some of the development ministers' personal characteristics are associated with aid giving but they do not matter much overall. This finding stands in contrast to the significant impact that scholars have found for the characteristics of heads of government, central bank governors, and finance ministers. Why are development ministers different? There are several potential explanations for this. First, an extensive strand of literature has shown that development aid is allocated due to national geostrategic and commercial interests (e.g., Alesina and Dollar 2000; Younas 2008). The defense of national interests ties too closely into the agenda of heads of government and parliamentarians and this might limit the room for development ministers to maneuver. This explanation would be in line with our finding that the gender and ideological composition of parliaments and some of the government-head characteristics are statistically significant in our regressions. Second, our analysis shows that it is not the ideology of ministers or heads of governments that directly matters for the quantity of ODA, but rather diverging ideologies that play a role. This finding hints at the important role of the negotiation process within governments that deserves closer investigation.

We leave several questions for future research. Since the importance of non-DAC donors in international development cooperation is rapidly growing (e.g., Dreher et al. 2013), future research should analyze the role of development minister characteristics in these donors' aid policies. Given that most of the big emerging donors are non-democratic and have weaker institutions than their DAC counterparts, the role of minister characteristics in these countries might be larger as a result of the fewer checks and balances that they have in place. Moreover, it would be necessary to test our explanations as to why ministers' characteristics do not show the expected effects by including variables capturing, for instance, donors' geostrategic or commercial interests. This undertaking would require a dyadic study design, including donor and recipient countries. It also appears fruitful to investigate the role of department heads within ministries. They are less likely to fluctuate and it would be interesting to learn more about the power struggles and negotiations within ministries. Finally, future research could delve deeper into the role of ministers' professional backgrounds and test whether ministerial experience in regional development is associated with ODA quality or whether their prior professions affect the sectoral allocation of ODA.

APPENDIX FOR CHAPTER 2

Appendix A2.1. List of development ministers (23 DAC countries, 1967-2012)

Country	Year	Name	Gender	Ideology	Experience dev. coop.	Economics education	Tenure
Australia	1967	Paul Hasluck	male	0.5	1	0	4
Australia	1969	Gordon Freeth	male	0.5	0	0	1
Australia	1970	William McMahon	male	0.5	0	1	1
Australia	1971	Leslie Bury	male	0.5	1	1	1
Australia	1972	Nigel Bowen	male	0.5	0	0	1
Australia	1973	Gough Whitlam	male	0	0	0	1
Australia	1974	Donald Robert Willsee	male	0	0	0	1
Australia	1976	Andrew Peacock	male	0.5	1	0	1
Australia	1981	Anthony Austin Street	male	0.5	0	0	1
Australia	1983	Bill Hayden	male	0	0	1	1
Australia	1989	Gareth Evans	male	0	0	1	1
Australia	1996	Alexander Downer	male	0.5	0	1	1
Australia	2008	Stephen Smith	male	0	0	0	1
Australia	2011	Kevin Rudd	male	0	1	0	1
Australia	2012	Bob Carr	male	0	0	0	1
Austria	1967	Lujo Toncic-Sorinj	male	0.5	0	0	2
Austria	1968	Kurt Josef Waldheim	male	0.5	1	0	1
Austria	1970	Rudolf Kirchschläger	male	0	0	0	1
Austria	1974	Erich Bielka-Karltreu	male	0	1	0	1
Austria	1977	Willibald Pahr	male	0	0	0	1
Austria	1983	Erwin Lanc	male	0	0	0	1
Austria	1985	Leopold Gratz	male	0	0	0	1
Austria	1986	Peter Jankowitsch	male	0	1	0	1
Austria	1987	Alois Mock	male	0.5	0	0	1
Austria	1995	Wolfgang Schüssel	male	0.5	0	0	1
Austria	2000	Benita-Maria Ferrero-Waldner	female	0.5	1	0	1
Austria	2005	Ursula Plassnik	female	0.5	0	0	1
Austria	2009	Michael Spindelegger	male	0.5	0	0	1
Belgium	1967	Pierre Harmel	male	0.5	0	0	2
Belgium	1969	Raymond Scheyven	male	0.5	1	0	1
Belgium	1972	Lucien Harmegnies	male	-0.5	0	0	1
Belgium	1973	Guy Cudell	male	-0.5	0	0	1
Belgium	1974	Renaat van Elslande	male	0	0	0	1
Belgium	1977	Lucien Outers	male	1	0	0	1
Belgium	1979	Mark Eyskens	male	0	0	1	1
Belgium	1981	Daniel Coens	male	0	0	0	1
Belgium	1982	Jacqueline Mayence-Goossens	female	1	0	0	1
Belgium	1983	Francois-Xavier de Donnea	male	1	0	1	1
Belgium	1986	André Kempinaire	male	1	0	0	1
Belgium	1988	André Geens	male	0	0	1	1
Belgium	1992	Erik Derycke	male	0	0	0	1
Belgium	1995	Reginald Moreels	male	0	1	0	1
Belgium	2000	Eddy Boutmans	male	-0.5	0	0	1
Belgium	2004	Marc Verwilghen	male	0.5	0	0	1
Belgium	2005	Armand de Decker	male	0.5	0	0	1
Belgium	2008	Charles Michel	male	0.5	0	0	1
Belgium	2011	Olivier Chastel	male	0.5	0	0	1
Belgium	2012	Paul Magnette	male	0	0	0	1
Canada	1967	Paul Joseph James Martin	male	0	0	0	5
Canada	1968	Mitchell Sharp	male	0	0	0	1
Canada	1975	Allan MacEachen	male	0	0	1	1
Canada	1977	Donald Jamieson	male	0	0	0	1
Canada	1979	Flora MacDonald	female	0.5	0	0	1
Canada	1980	Mark MacGuigan	male	0	0	0	1
Canada	1983	Allan MacEachen	male	0	0	1	3
Canada	1985	Monique Vézina	female	0.5	0	0	1
Canada	1986	Monique Landry	female	0.5	0	0	1

Country	Year	Name	Gender	Ideology	Experience dev. coop.	Economics education	Tenure
Canada	1993	Monique Vézina	female	0.5	0	0	2
Canada	1994	André Ouellet	male	0	0	0	1
Canada	1996	Pierre Pettigrew	male	0	0	0	1
Canada	1997	Diane Marleau	female	0	0	1	1
Canada	2000	Maria Minna	female	0	0	0	1
Canada	2002	Susan Whelan	female	0	0	0	1
Canada	2004	Aileen Carroll	female	0	0	0	1
Canada	2006	Josée Verner	female	0.5	0	0	1
Canada	2008	Beverley Oda	female	0.5	0	0	1
Switzerland	1967	Willy Spühler	male	0	0	1	2
Switzerland	1970	Pierre Graber	male	0	0	1	1
Switzerland	1978	Pierre Aubert	male	0	0	0	1
Switzerland	1988	René Felber	male	0	0	0	1
Switzerland	1993	Flavio Cotti	male	0.5	0	0	1
Switzerland	1999	Joseph Deiss	male	0.5	0	1	1
Switzerland	2003	Micheline Calmy-Rey	female	0	0	1	1
Switzerland	2012	Didier Burkhalter	male	1	0	1	1
Germany	1967	Hans-Jürgen Wischnewski	male	0	0	0	1
Germany	1969	Erhard Eppler	male	0	0	0	1
Germany	1975	Egon Bahr	male	0	0	0	1
Germany	1977	Marie Schlei	female	0	0	0	1
Germany	1978	Rainer Offergeld	male	0	0	0	1
Germany	1983	Jürgen Warnke	male	0.5	0	0	1
Germany	1987	Hans Klein	male	0.5	0	0	1
Germany	1989	Jürgen Warnke	male	0.5	0	0	5
Germany	1991	Carl-Dieter Spranger	male	0.5	0	0	1
Germany	1999	Heidemarie Wieczorek-Zeul	female	0	0	0	1
Germany	2010	Dirk Niebel	male	1	0	0	1
Denmark	1967	Jens Otto Krag	male	0	0	1	6
Denmark	1968	Poul Hartling	male	1	0	0	1
Denmark	1972	Knud Børge Andersen	male	0	0	0	1
Denmark	1974	Ove Guldberg	male	1	0	0	1
Denmark	1975	Knud Børge Andersen	male	0	0	0	3
Denmark	1979	Henning Christophersen	male	1	0	1	1
Denmark	1980	Kjeld Olesen	male	0	0	0	1
Denmark	1983	Uffe Ellemann-Jensen	male	1	0	1	1
Denmark	1993	Helle Degn	female	0	0	0	1
Denmark	1995	Poul Nielson	male	0	1	0	1
Denmark	2000	Jan Trøjborg	male	0	0	0	1
Denmark	2001	Anita Bay Bundegaard	female	0.5	0	0	1
Denmark	2002	Per Stig Møller	male	0.5	0	0	1
Denmark	2005	Ulla Pedersen Tørnæs	female	1	0	0	1
Denmark	2010	Søren Pind	male	1	0	0	1
Denmark	2012	Christian Friis Bach	male	0.5	1	1	1
Spain	1980	Marcelino Oreja Aguirre	male	0.5	0	0	5
Spain	1981	José Pedro Pérez-Llorca	male	0.5	0	0	1
Spain	1983	Fernando Morán	male	0	0	0	1
Spain	1986	Francisco Fernández Ordóñez	male	0	0	0	1
Spain	1992	Javier Solana	male	0	0	0	1
Spain	1996	Abel Matutes	male	0.5	0	1	1
Spain	2000	Josep Piqué	male	0.5	0	1	1
Spain	2003	Ana Palacio	female	0.5	0	0	1
Spain	2004	Miguel Ángel Moratinos	male	0	1	0	1
Spain	2011	Trinidad Jiménez	female	0	0	0	1
Spain	2012	José Garcia-Margallo y Marfil	male	0.5	0	0	1
Finland	1967	Ahti Kalle Samuli Karjalainen	male	0.5	0	0	4
Finland	1970	Väinö Olavi Leskinen	male	0	0	0	1
Finland	1972	Taisto Kalevi Sorsa	male	0	0	0	1
Finland	1973	Ahti Kalle Samuli Karjalainen	male	0.5	0	0	7
Finland	1975	Olavi Johannes Mattila	male	0.5	1	1	1

Country	Year	Name	Gender	Ideology	Experience dev. coop.	Economics education	Tenure
Finland	1976	Taisto Kalevi Sorsa	male	0	0	0	2
Finland	1977	Paavo Matti Väyrynen	male	0.5	0	0	1
Finland	1982	Pär Olav Mikael Stenbäck	male	0.5	0	0	1
Finland	1983	Paavo Matti Väyrynen	male	0.5	0	0	6
Finland	1987	Taisto Kalevi Sorsa	male	0	0	0	3
Finland	1989	Pertti Kullervo Paasio	male	0	0	0	1
Finland	1991	Toimi Olavi Kankaanniemi	male	0.5	0	0	1
Finland	1994	Pekka Olavi Haavisto	male	-0.5	0	0	1
Finland	1999	Satu Maijastiina Hassi	female	-0.5	0	0	1
Finland	2002	Suvi-Anne Siimes	female	-0.5	0	1	1
Finland	2003	Paula Ilona Lehtomäki	female	0.5	0	1	1
Finland	2007	Paavo Matti Väyrynen	male	0.5	0	0	10
Finland	2011	Heidi Hautala	female	-0.5	1	0	1
France	1967	Yvon Bourges	male	0.5	1	0	1
France	1973	Pierre Billecocq	male	0.5	0	0	1
France	1974	Pierre Abelin	male	0.5	1	0	1
France	1976	Jean de Lipkowski	male	0.5	1	0	1
France	1977	Robert Galley	male	0.5	0	0	1
France	1981	Jean-Pierre Cot	male	-0.5	0	0	1
France	1983	Christian Nucci	male	-0.5	1	0	1
France	1986	Michel Aurillac	male	0.5	1	0	1
France	1988	Jacques Pelletier	male	0	0	0	1
France	1991	Edwige Avice	female	-0.5	0	1	1
France	1992	Marcel Debarge	male	-0.5	0	0	1
France	1993	Michel Roussin	male	0.5	0	0	1
France	1995	Jacques Godfrain	male	0.5	0	1	1
France	1997	Charles Josselin	male	-0.5	0	0	1
France	2002	Pierre-André Wiltzer	male	0.5	0	0	1
France	2004	Xavier Darcos	male	0.5	0	0	1
France	2005	Brigitte Giradin	female	0.5	1	0	1
France	2007	Jean-Marie Bockel	male	0	0	0	1
France	2008	Alain Joyandet	male	0.5	0	0	1
France	2011	Henri de Raincourt	male	0.5	0	0	1
France	2012	Pascal Canfin	male	-0.5	0	0	1
UK	1967	Arthur Bottomley	male	0	0	0	1
UK	1968	Reginald Prentice	male	0	0	1	1
UK	1970	Richard Wood	male	1	0	1	1
UK	1974	Judith Hart	female	0	1	1	1
UK	1975	Reginald Prentice	male	0	1	1	3
UK	1977	Judith Hart	female	0	1	1	2
UK	1979	Harry Neil Marten	male	1	0	0	1
UK	1983	Timothy Raison	male	1	0	0	1
UK	1987	Christopher Patten	male	1	0	0	1
UK	1990	Lynda Chalker	female	1	0	0	1
UK	1997	Clare Short	female	0	1	0	1
UK	2003	Valerie Ann Amos	female	0	1	0	1
UK	2004	Hilary Benn	male	0	1	0	1
UK	2007	Douglas Alexander	male	0	0	0	1
UK	2010	Andrew Mitchell	male	1	1	0	1
Greece	1996	Theodoros Pangalos	male	0	0	1	1
Greece	1999	Georgios A. Papandreou	male	0	0	0	1
Greece	2004	Petros Molyviatis	male	0.5	1	0	1
Greece	2006	Dora Bakoyannis	female	0.5	0	0	1
Greece	2010	Georgios A. Papandreou	male	0	0	0	6
Greece	2011	Stavros Lambrinidis	male	0	0	1	1
Greece	2012	Dimitris Avramopoulos	male	0.5	1	0	1
Ireland	1974	Garret FitzGerald	male	0.5	0	1	2
Ireland	1978	Michael O'Kennedy	male	0	0	0	1
Ireland	1980	Brian Lenihan	male	0	0	0	1
Ireland	1982	Gerry Collins	male	0	0	0	1

Country	Year	Name	Gender	Ideology	Experience dev. coop.	Economics education	Tenure
Ireland	1983	Peter Barry	male	0.5	0	0	1
Ireland	1987	Brian Lenihan	male	0	0	0	3
Ireland	1990	Gerry Collins	male	0	0	0	2
Ireland	1992	David Andrews	male	0	0	0	1
Ireland	1993	Tom Kitt	male	0	0	0	1
Ireland	1995	Joan Burton	female	-0.5	0	1	1
Ireland	1997	Liz O'Donnell	female	1	0	0	1
Ireland	2002	Tom Kitt	male	0	0	0	3
Ireland	2005	Conor Lenihan	male	0	0	1	1
Ireland	2007	Michael Kitt	male	0	0	0	1
Ireland	2008	Peter Power	male	0	0	0	1
Ireland	2011	Jan O'Sullivan	female	-0.5	0	0	1
Ireland	2012	Joe Costello	male	-0.5	0	0	1
Italy	1967	Giulio Andreotti	male	0.5	0	0	2
Italy	1969	Mario Tanassi	male	-0.5	0	0	1
Italy	1970	Silvio Gava	male	0.5	0	0	1
Italy	1972	Mauro Ferri	male	-0.5	0	0	1
Italy	1974	Ciriaco De Mita	male	0.5	0	0	1
Italy	1975	Carlo Donat-Cattin	male	0.5	0	0	1
Italy	1979	Franco Nicolazzi	male	-0.5	0	0	1
Italy	1980	Antonio Bisaglia	male	0.5	0	0	1
Italy	1981	Giovanni Marcora	male	0.5	0	0	1
Italy	1983	Filippo Maria Pandolfi	male	0.5	0	0	1
Italy	1984	Renato Altissimo	male	0.5	0	0	1
Italy	1987	Adolfo Battaglia	male	0	0	0	1
Italy	1991	Guido Bodrato	male	0.5	0	0	1
Italy	1992	Giuseppe Guarino	male	0.5	0	0	1
Italy	1993	Paolo Savona	male	0	0	1	1
Italy	1994	Vito Gnutti	male	0.5	0	0	1
Italy	1995	Alberto Clo	male	0	0	0	1
Italy	1996	Pierluigi Bersani	male	-0.5	0	0	1
Italy	2000	Enrico Letta	male	-0.5	0	0	1
Italy	2001	Antonio Marzano	male	0.5	0	0	1
Italy	2005	Claudio Scajola	male	0.5	0	0	1
Italy	2006	Pierluigi Bersani	male	0	0	0	5
Italy	2008	Claudio Scajola	male	0.5	0	0	2
Italy	2010	Silvio Berlusconi	male	0.5	0	0	1
Italy	2011	Paolo Romani	male	0.5	0	0	1
Italy	2012	Corrado Passera	male	0.5	0	1	1
Japan	1967	Takeo Miki	male	0.5	0	0	1
Japan	1969	Kiichi Aichi	male	0.5	0	0	1
Japan	1971	Takeo Fukuda	male	0.5	0	0	1
Japan	1973	Masayoshi Ohira	male	0.5	0	1	3
Japan	1975	Kiichi Miyazawa	male	0.5	0	0	1
Japan	1977	Iichiro Hatoyama	male	0.5	0	0	1
Japan	1978	Sunao Sonoda	male	0.5	0	0	1
Japan	1980	Saburo Okita	male	0.5	0	0	1
Japan	1981	Sunao Sonoda	male	0.5	0	0	3
Japan	1982	Yoshio Sakurauchi	male	0.5	1	1	1
Japan	1983	Shintaro Abe	male	0.5	0	0	1
Japan	1987	Tadashi Kuranari	male	0.5	0	0	1
Japan	1988	Sosuke Uno	male	0.5	0	0	1
Japan	1990	Taro Nakayama	male	0.5	0	0	1
Japan	1991	Michio Watanabe	male	0.5	0	1	1
Japan	1993	Tsutomu Hata	male	0	0	1	1
Japan	1994	Yohei Kono	male	0.5	0	1	1
Japan	1996	Yukihiko Ikeda	male	0.5	0	0	1
Japan	1998	Keizo Obushi	male	0.5	1	0	1
Japan	1999	Masahiko Komura	male	0.5	0	0	1
Japan	2000	Yohei Kono	male	0.5	0	1	3

Country	Year	Name	Gender	Ideology	Experience dev. coop.	Economics education	Tenure
Japan	2001	Makiko Tanaka	female	0.5	0	1	1
Japan	2002	Yoriko Kawaguchi	female	0.5	1	1	1
Japan	2005	Nobutaka Machimura	male	0.5	0	1	1
Japan	2006	Taro Aso	male	0.5	0	1	1
Japan	2008	Masahiko Komura	male	0.5	0	0	2
Japan	2009	Hirofumi Nakasone	male	0.5	0	1	1
Japan	2010	Katsuya Okada	male	0	0	0	1
Japan	2011	Takeaki Matsumoto	male	0	0	0	1
Japan	2012	Koichiro Genba	male	0	0	0	1
South Korea	1987	Choe Gwang-su	male	0.5	0	0	2
South Korea	1989	Choe Ho-jung	male	0.5	0	0	1
South Korea	1991	Lee Sang-ok	male	0.5	0	0	1
South Korea	1993	Han Seung-ju	male	0.5	0	0	1
South Korea	1995	Gong Ro-myeong	male	0.5	1	1	1
South Korea	1997	Yu Jong-ha	male	0.5	0	0	1
South Korea	1998	Park Jeong-su	male	0	0	0	1
South Korea	1999	Hong Sun-yeong	male	0	1	0	1
South Korea	2000	Lee Jeong-bin	male	0	1	0	1
South Korea	2001	Han Seung-su	male	0.5	0	1	1
South Korea	2002	Choe Seong-hong	male	0	1	0	1
South Korea	2003	Yoon Young Kwan	male	0.5	0	1	1
South Korea	2004	Ban Ki-moon	male	0.5	1	0	1
South Korea	2007	Song Min-sun	male	0	0	0	1
South Korea	2008	Yu Myung-hwan	male	0.5	1	0	1
South Korea	2011	Kim Sung-hwan	male	0.5	1	1	1
Luxembourg	1980	Gaston Thorn	male	1	0	0	12
Luxembourg	1981	Colette Flesch	female	1	1	1	1
Luxembourg	1985	Jacques Poos	male	0	0	1	1
Luxembourg	2000	Charles Goerens	male	1	0	0	1
Luxembourg	2004	Jean-Louis Schiltz	male	0.5	0	0	1
Luxembourg	2010	Marie-Josée Jacobs	female	0.5	0	0	1
Netherlands	1967	Berend Jan Udink	male	0.5	0	1	1
Netherlands	1972	Kees Bortien	male	0.5	0	0	1
Netherlands	1973	Jan Pronk	male	0	0	1	1
Netherlands	1978	Jan de Koning	male	0.5	1	0	1
Netherlands	1982	Kees van Dijk	male	0.5	1	1	1
Netherlands	1983	Eegje Schoo	female	1	0	0	1
Netherlands	1987	Pieter Bukman	male	0.5	0	0	1
Netherlands	1990	Jan Pronk	male	0	0	1	6
Netherlands	1999	Eva Herfkens	female	0	1	0	1
Netherlands	2003	Agnes van Ardenne	female	0.5	0	0	1
Netherlands	2007	Albert Gerard Koenders	male	0	1	1	1
Netherlands	2010	Maxime Verhagen	male	0.5	1	0	1
Netherlands	2011	Ben Knapen	male	0.5	0	0	1
Norway	1967	John Lyng	male	0.5	0	0	2
Norway	1970	Svenn Stray	male	0.5	0	0	1
Norway	1971	Andreas Zeier Cappelen	male	0	0	0	1
Norway	1973	Dagfinn Varvik	male	1	0	1	1
Norway	1974	Knut Frydenlund	male	0	0	0	1
Norway	1981	Svenn Stray	male	0.5	0	0	2
Norway	1983	Reidun Brusletten	female	0.5	0	0	1
Norway	1986	Vesla Vetlesen	male	0	1	0	1
Norway	1988	Kirsti Kolle Grøndahl	female	0	1	0	1
Norway	1990	Tom Vraalsen	male	1	1	1	1
Norway	1991	Grete Faremo	female	0	1	0	1
Norway	1993	Kari Nordheim-Larsen	female	0	0	0	1
Norway	1998	Hilde Frafjord Johnson	female	0.5	0	0	1
Norway	2000	Anne Kristin Sydnes	female	0	0	0	1
Norway	2002	Hilde Frafjord Johnson	female	0.5	1	0	3
Norway	2006	Erik Solheim	male	-0.5	1	0	1

Country	Year	Name	Gender	Ideology	Experience dev. coop.	Economics education	Tenure
Norway	2012	Heikki Holmås	male	-0.5	0	1	1
New Zealand	1967	Keith Holyoake	male	1	0	0	8
New Zealand	1973	Norman Kirk	male	0	0	0	1
New Zealand	1975	Bill Rowling	male	0	0	1	1
New Zealand	1976	Brian Talboys	male	1	0	0	1
New Zealand	1982	Warren Cooper	male	1	0	0	1
New Zealand	1985	David Lange	male	0	0	0	1
New Zealand	1988	Russell Marshall	male	0	0	0	1
New Zealand	1990	Mike Moore	male	0	0	0	1
New Zealand	1991	Donald Charles McKinnon	male	1	0	0	1
New Zealand	2000	Phil Goff	male	0	0	0	1
New Zealand	2006	Winston Peters	male	0	0	0	1
New Zealand	2009	Murray McCully	male	1	0	0	1
Portugal	1980	Diogo Freitas do Amaral	male	0.5	0	0	1
Portugal	1981	André Gonçalves Pereira	male	0.5	1	0	1
Portugal	1982	Vasco Futscher Pereira	male	0.5	1	0	1
Portugal	1983	Jaime Gama	male	0	0	0	1
Portugal	1986	Pedro Pires de Miranda	male	0.5	0	0	1
Portugal	1988	Joao de Deus Pinheiro	male	0.5	0	0	1
Portugal	1993	José Manuel Barroso	male	0.5	0	1	1
Portugal	1996	Jaime Gama	male	0	0	0	4
Portugal	2002	António Martins da Cruz	male	0.5	0	0	1
Portugal	2004	Teresa Gouveia	female	0.5	0	0	1
Portugal	2005	Diogo Freitas do Amaral	male	0.5	1	0	2
Portugal	2006	Luís Amado	male	0	0	1	1
Portugal	2011	Paulo Portas	male	0.5	0	0	1
Sweden	1967	Alva Myrdal	female	0	1	0	1
Sweden	1974	Gertrud Sigurdsen	female	0	0	0	1
Sweden	1977	Ola Ullsten	male	0.5	0	0	1
Sweden	1983	Lennart Bodström	male	0	0	0	1
Sweden	1986	Lena Hjelm-Wallén	female	0	0	0	1
Sweden	1992	Alf Svensson	male	0.5	0	0	1
Sweden	1995	Pierre Schori	male	0	0	0	1
Sweden	2000	Maj-Inger Klingvall	female	0	0	0	1
Sweden	2002	Jan Olov Karlsson	male	0	0	1	1
Sweden	2004	Carin Jämtin	female	0	0	0	1
Sweden	2007	Gunilla Carlsson	female	1	0	1	1
USA	1967	David Dean Rusk	male	0	0	0	7
USA	1969	William Pierce Rogers	male	1	0	0	1
USA	1974	Henry Kissinger	male	1	1	0	1
USA	1977	Cyrus Vance	male	0	0	0	1
USA	1980	Edmund Muskie	male	0	0	0	1
USA	1981	Alexander Haig	male	1	0	1	1
USA	1983	George Pratt Shultz	male	1	0	1	1
USA	1989	James Baker	male	1	0	0	1
USA	1993	Warren Christopher	male	0	0	0	1
USA	1997	Madeleine Albright	female	0	1	0	1
USA	2001	Colin Powell	male	1	0	1	1
USA	2005	Condoleezza Rice	female	1	0	0	1
USA	2009	Hillary Rodham Clinton	female	0	0	0	1

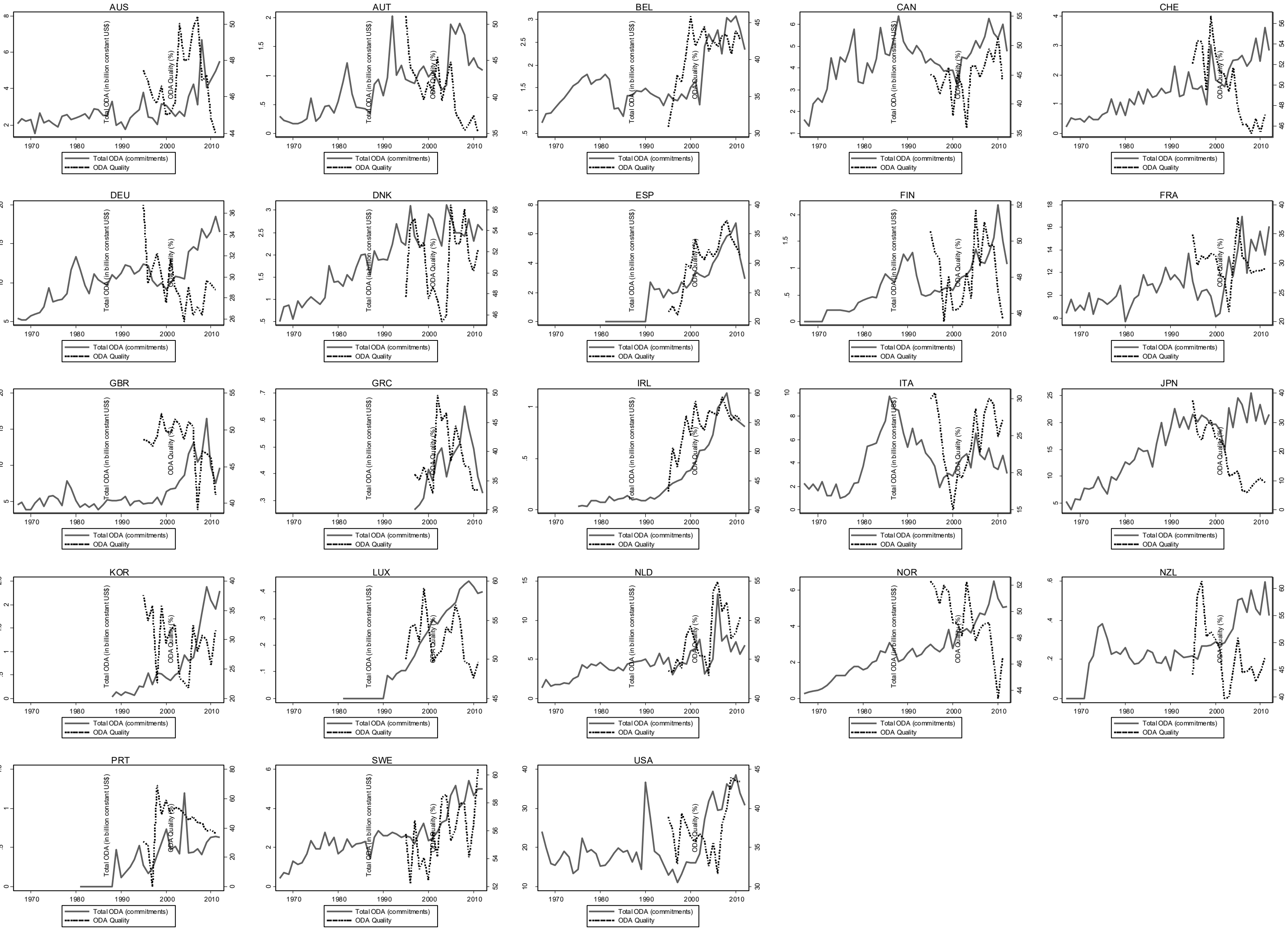
.

Appendix A2.2. List of variables

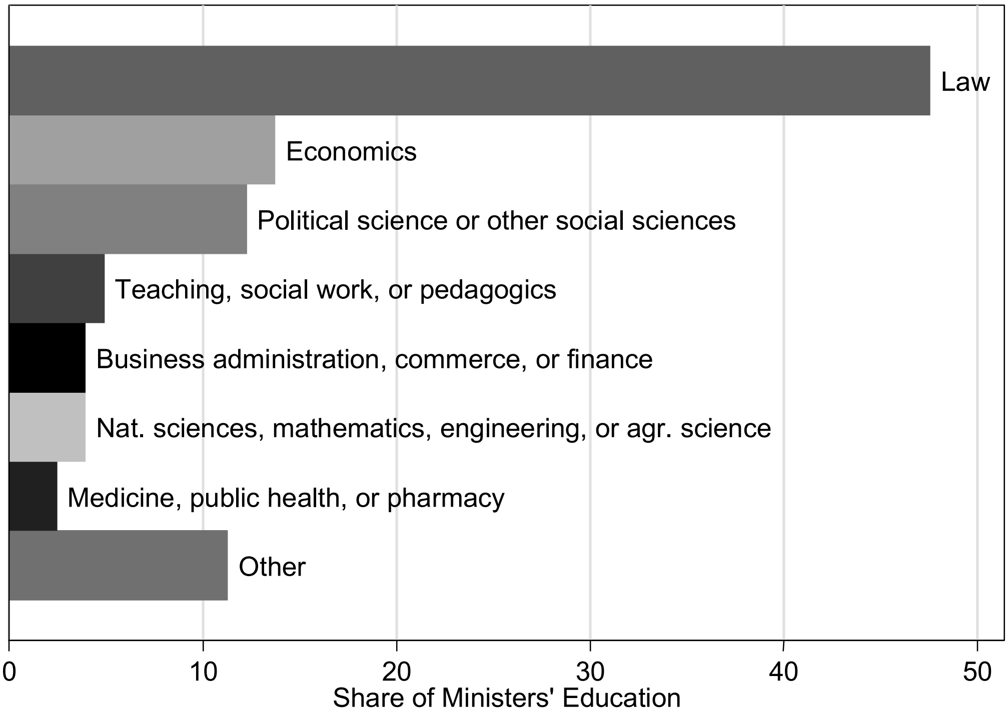
Variable	Definition	Source
Dependent variables		
(log) Total ODA com.	(log) Total ODA commitments (in constant 2012 US$)	OECD (2014)
(log) Total ODA disb.	(log) Total ODA disbursements (in constant 2012 US$)	OECD (2014)
Quality ODA	Roodman's (2012) bilateral quality-adjusted gross aid disbursement (% of bilateral gross aid)	Own construction with data from Roodman (2012)
Quality ODA (no selectivity reward)	Roodman's (2012) bilateral quality-adjusted gross aid disbursement excluding the discount for selectivity with regards to richer and poorly governed countries (% of bilateral gross aid)	Own construction with data from Roodman (2012)
Quality ODA (no proliferation penalty)	Roodman's (2012) bilateral quality-adjusted gross aid disbursement excluding the discount for project proliferation (% of bilateral gross aid)	Own construction with data from Roodman (2012)
Quality ODA (no tied aid penalty)	Roodman's (2012) bilateral quality-adjusted gross aid disbursement excluding the discount for tied aid (% of bilateral gross aid)	Own construction with data from Roodman (2012
Gender		
Female minister	1 if development minister is female	Own construction
Female gov. head	1 if head of government is female	Own construction
Female parliament	Seats held by women in national parliaments (% of total seats)	WDI (World Bank 2014) and Brady et al. (2014)
Political ideology		
Right-wing minister	Political ideology of development minister (liberalist economic policy (1), conservative (0.5), social democratic party (0), modern socialist (-0.5), unreformed socialist and communist (-1))	Own construction following Bjørnskov and Potrafke (2011)
Right-wing gov. head	Political ideology of head of government (liberalist economic policy (1), conservative (0.5), social democratic party (0), modern socialist (-0.5), unreformed socialist and communist (-1))	Own construction following Bjørnskov and Potrafke (2011)
Right-wing parliament	Average political ideology of all parliament members (liberalist economic policy (1), conservative (0.5), social democratic party (0), modern socialist (-0.5), unreformed socialist and communist (-1))	Bjørnskov and Potrafke (2011)
Ideological difference	1 if political ideology of development minister and head of government is different	Own construction based on Bjørnskov and Potrafke (2011)
Ideological difference (more right-wing)	1 if development minister is more right-wing than the head of government	Own construction based on Bjørnskov and Potrafke (2011)
Ideological difference (more left-wing)	1 if development minister is more left-wing than the head of government	Own construction based on Bjørnskov and Potrafke (2011)
Right-wing minister (election manifesto)	Right-left position of development minister's party (or of his head of government if he or she is not affiliated with a party)	Manifesto Project (Volkens et al. 2016)
Experience		
Prof. dev. coop. minister	1 if development minister has a professional background in development cooperation	Own construction
Prof. dev. coop. gov. head	1 if head of government has a professional background in development cooperation	Own construction
Economics & business minister	1 if development minister has a degree of higher education in economics or business administration, commerce or finance	Own construction
Economics & business gov. head	1 if head of government has a degree of higher education in economics or business administration, commerce or finance	Own construction
Economics minister	1 if development minister has a degree of higher education in economics	Own construction
Economics gov. head	1 if head of government has a degree of higher education in economics	Own construction

Business administration, commerce, or finance minister	1 if development minister has a degree of higher education in business administration, commerce or finance	Own construction
Business administration, commerce, or finance minister	1 if head of government has a degree of higher education in business administration, commerce or finance	Own construction
Law minister	1 if development minister has a degree of higher education in law	Own construction
Law gov. head	1 if head of government has a degree of higher education in law	Own construction
Medicine, public health, or pharmacy minister	1 if development minister has a degree of higher education in medicine, public health, or pharmacy	Own construction
Medicine, public health, or pharmacy gov. head	1 if head of government has a degree of higher education in medicine, public health, or pharmacy	Own construction
Nat. sciences, mathematics, engineering, or agr. science minister	1 if development minister has a degree of higher education in natural sciences, mathematics, engineering, or agricultural science	Own construction
Nat. sciences, mathematics, engineering, or agr. science gov. head	1 if head of government has a degree of higher education in natural sciences, mathematics, engineering, or agricultural science	Own construction
Political science or other social sciences minister	1 if development minister has a degree of higher education in political science or other social sciences	Own construction
Political science or other social sciences gov. head	1 if head of government has a degree of higher education in political science or other social sciences	Own construction
Teaching, social work, or pedagogics minister	1 if development minister has a degree of higher education in teaching, social work, or pedagogics	Own construction
Teaching, social work, or pedagogics gov. head	1 if head of government has a degree of higher education in teaching, social work, or pedagogics	Own construction
Tenure minister	Number of years the development minister is in office	Own construction
Tenure gov. head	Number of years the head of government is in office	Own construction
General ministerial experience minister	Number of months the development minister has worked in a ministerial position other than the development minister before assuming office	Own construction
Tenure minister + General ministerial experience minister	Total number of months the development minister has worked as development minister or any other minister prior to assuming the development office	Own construction
Head of government in the future	1 if development minister will become head of government in the future	Own construction
Control variables		
(log) GDP per capita	Log of gross domestic product divided by population (constant 2005 US$)	WDI (World Bank 2014)
Openness	Trade (% of GDP)	WDI (World Bank 2014)
Gov. expenditures	General government final consumption expenditure (% of GDP)	WDI (World Bank 2014)
Debt	Government debt (% of GDP)	Abbas et al. (2010)
Political globalization	KOF Index on Political Globalization (components: embassies (25%), membership in international organizations (28%), participation in UN Security Council missions (22%), international treaties (25%))	Dreher (2006), updated in Dreher, Gaston and Martens (2008)
Aid agency	1 if existence of national aid agencies operating independently from the Ministry of Foreign Affairs	Fuchs et al. (2014), own update
(log) Colonial history	(log) Population of former colonies on DAC list of ODA recipients (0 if no colonial history)	Own calculations based on CEPII data (Mayer and Zignago 2006) and WDI (World Bank 2014)

Appendix A2.3. Quantity and quality of ODA by donor country (1967-2012)



Appendix A2.4. Development ministers' educational background (1971-2012)



Appendix B2.1. Development minister characteristics and total ODA budgets (alternative specification)

	(1)	(2)	(3)	(4)	(5)	(6)
	(log) Total ODA com.	(log) Total ODA disb.	(log) Total ODA com.	(log) Total ODA disb.	(log) Total ODA com.	(log) Total ODA disb.
	1971-2012	1971-2012	<=1990	<=1990	>=1991	>=1991
Lagged DV			0.4898***	0.4885***	0.3603***	0.5393***
			[0.000]	[0.000]	[0.000]	[0.000]
(log) GDP per capita	2.6073***	2.9678***	1.4275***	1.8873***	1.9005***	1.3337***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Openness	0.0035**	0.0056***	-0.0054*	-0.0018	0.0023*	0.0015
	[0.029]	[0.000]	[0.079]	[0.385]	[0.072]	[0.127]
Gov. expenditure	0.0467***	0.0759***	0.0099	0.0322***	0.0155	0.0034
	[0.000]	[0.000]	[0.389]	[0.001]	[0.210]	[0.755]
Debt	-0.0022***	-0.0029***	-0.0013*	-0.0006	-0.0001	-0.0005
	[0.000]	[0.000]	[0.057]	[0.361]	[0.833]	[0.340]
Political globalization	0.0095***	0.0127***	0.0054	0.0052	-0.0008	0.0004
	[0.005]	[0.002]	[0.455]	[0.192]	[0.764]	[0.859]
Aid agency	0.0181	0.1108**	-0.1378	0.0971	0.0150	0.0862**
	[0.706]	[0.030]	[0.320]	[0.548]	[0.739]	[0.017]
(log) Colonial history	-0.3592***	-0.4424***	0.4274**	0.0927	0.2853	0.0701
	[0.002]	[0.000]	[0.043]	[0.541]	[0.131]	[0.597]
Female minister	-0.0513*	-0.0608*	-0.0160	-0.0320	-0.0499*	-0.0387
	[0.086]	[0.055]	[0.761]	[0.404]	[0.054]	[0.164]
Female gov. head	0.0408	0.0440	-0.2970***	-0.1788***	0.0548	0.0668***
	[0.497]	[0.483]	[0.000]	[0.001]	[0.142]	[0.004]
Female parliament	0.0116***	0.0146***	0.0178***	0.0147***	0.0004	-0.0002
	[0.000]	[0.000]	[0.002]	[0.001]	[0.890]	[0.934]
Right-wing minister	0.0044	-0.0531	-0.0213	-0.0890**	-0.0214	-0.0388
	[0.934]	[0.257]	[0.755]	[0.043]	[0.528]	[0.126]
Right-wing gov. head	-0.0170	-0.0013	-0.0403	-0.0073	0.0324	0.0417
	[0.813]	[0.984]	[0.652]	[0.905]	[0.460]	[0.232]
Right-wing parliament	0.0326	0.6919***	0.3957**	0.8388***	-0.0358	0.0713
	[0.813]	[0.002]	[0.012]	[0.000]	[0.736]	[0.411]
Ideological difference	0.1063***	0.0771**	0.1314***	0.1342***	0.0157	-0.0175
	[0.002]	[0.010]	[0.002]	[0.000]	[0.469]	[0.330]
Prof. dev. coop. minister	0.0082	0.0079	0.0388	0.0239	-0.0048	0.0329
	[0.810]	[0.846]	[0.311]	[0.394]	[0.883]	[0.185]
Prof. dev. coop. gov. head	0.0153	0.0210	0.0219	0.0207	0.0195	-0.0221
	[0.756]	[0.592]	[0.725]	[0.624]	[0.649]	[0.467]

Economics & business minister	-0.0223	0.0007	-0.0489	0.0044	0.0169	0.0112
	[0.334]	[0.977]	[0.110]	[0.839]	[0.433]	[0.431]
Economics & business gov. head	0.0760**	0.0902***	0.0481*	0.0437*	-0.0004	0.0206
	[0.018]	[0.005]	[0.077]	[0.079]	[0.989]	[0.243]
Tenure minister	0.0100**	0.0121**	0.0067	0.0049	0.0070*	0.0054**
	[0.042]	[0.022]	[0.394]	[0.369]	[0.083]	[0.034]
Tenure gov. head	-0.0045	-0.0066	-0.0003	-0.0044	-0.0020	-0.0054**
	[0.253]	[0.126]	[0.957]	[0.283]	[0.521]	[0.013]
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Gov. head FE	No	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	812	838	317	345	491	493
R-squared (within)	0.761	0.850	0.720	0.826	0.768	0.848
Number of countries	23	23	18	20	23	23

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.2. Development minister characteristics and ODA (election manifestos)

	(1) (log) Total ODA com.	(2) (log) Total ODA disb.	(3) Quality ODA
	1971-2012	1971-2012	1995-2011
Panel A (baseline specification)			
Right-wing minister	0.0008	-0.0356*	-0.6600
	[0.979]	[0.088]	[0.421]
Panel B			
Right-wing minister (election manifesto)	-0.0006	-0.0003	-0.0191
	[0.365]	[0.471]	[0.263]

Notes: Column 1 is based on column 3 of Table 2.2. Column 2 is based on column 4 of Table 2.2. Column 3 is based on column 3 of Table 2.3. All regressions include the lagged dependent variable, all control variables, country-fixed effects, and year-fixed effects. Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.3. Development minister characteristics and ODA (ideological differences)

	(1) (log) Total ODA com.	(2) (log) Total ODA disb.	(3) Quality ODA
	1971-2012	1971-2012	1995-2011
Panel A (baseline specification)			
Ideological difference	0.0693***	0.0371**	-0.1828
	[0.001]	[0.015]	[0.752]
Panel B			
Ideological difference (more left-wing)	0.0410	0.0189	1.4534
	[0.525]	[0.643]	[0.382]
Ideological difference (more right-wing)	0.0987	0.0559	-1.9682
	[0.179]	[0.219]	[0.140]
Test for joint significance (p-value)	0.006	0.053	0.307
Test for same coefficient size (p-value)	0.660	0.646	0.225

Notes: Column 1 is based on column 3 of Table 2.2. Column 2 is based on column 4 of Table 2.2. Column 3 is based on column 3 of Table 2.3. All regressions include the lagged dependent variable, all control variables, country-fixed effects, and year-fixed effects. Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.4. Development minister characteristics and ODA (various experience variables)

	(1) (log) Total ODA com.	(2) (log) Total ODA disb.	(3) Quality ODA
	1971-2012	1971-2012	1995-2011
Panel A (baseline specification)			
Tenure minister	0.0068*	0.0054**	0.1703**
	[0.059]	[0.047]	[0.040]
Panel B			
Tenure minister	-0.00024	-0.00339	-0.22195
	[0.980]	[0.628]	[0.313]
Tenure minister squared	0.00070	0.00087	0.03514*
	[0.440]	[0.159]	[0.052]
Test for joint significance (p-value)	(0.156)	(0.035)**	(0.018)**
Panel C			
Tenure minister + General ministerial experience minister	0.0003	0.0003*	0.0007
	[0.134]	[0.088]	[0.899]
Panel D			
Tenure minister	0.0071**	0.0057**	0.1613**
	[0.048]	[0.037]	[0.046]
General ministerial experience minister	0.0001	0.0001	-0.0239**
	[0.771]	[0.533]	[0.042]
Panel E			
Tenure minister	0.0235*	0.0151	0.0035
	[0.062]	[0.102]	[0.997]
Minister FE	Yes	Yes	Yes
Panel F			
Tenure minister	0.0064*	0.0054*	0.1677**
	[0.070]	[0.051]	[0.045]
Head of government in the future	-0.0687*	-0.0151	-1.1916
	[0.051]	[0.602]	[0.371]

Notes: Column 1 is based on column 3 of Table 2.2. Column 2 is based on column 4 of Table 2.2. Column 3 is based on column 3 of Table 2.3. All regressions include the lagged dependent variable, all control variables, country-fixed effects, and year-fixed effects. Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.5. Development minister characteristics and ODA (various university degrees)

	(1) (log) Total ODA com.	(2) (log) Total ODA disb.	(3) Quality ODA
	1971-2012	1971-2012	1995-2011
Panel A (baseline specification)			
Economics & business minister	-0.0050	0.0081	-0.5221
	[0.753]	[0.508]	[0.304]
Economics & business gov. head	0.0377*	0.0390***	-0.4718
	[0.066]	[0.005]	[0.472]
Panel B			
Economics minister	-0.0141	0.0007	-0.6310
	[0.428]	[0.960]	[0.268]
Economics gov. head	0.0425*	0.0546***	0.8200
	[0.066]	[0.001]	[0.395]
Panel C			
Business administration, commerce, or finance minister	-0.0128	0.0264	0.4959
	[0.719]	[0.276]	[0.520]
Business administration, commerce, or finance gov. head	0.0291	0.0039	-2.0081**
	[0.336]	[0.860]	[0.011]
Panel D			
Law minister	-0.0025	-0.0169	0.5245
	[0.898]	[0.240]	[0.366]
Law gov. head	-0.0118	-0.0156	1.1910*
	[0.576]	[0.368]	[0.071]
Panel E			
Medicine, public health, or pharmacy minister	0.0062	-0.0455*	-3.0340***
	[0.860]	[0.074]	[0.001]
Medicine, public health, or pharmacy gov. head	-0.1608**	-0.0604	4.6814**
	[0.020]	[0.340]	[0.019]
Panel F			
Nat. sciences, mathematics, engineering, or agr. science minister	-0.0322	0.0380	0.7383
	[0.237]	[0.123]	[0.620]
Nat. sciences, mathematics, engineering, or agr. science gov. head	-0.0209	-0.0987**	-1.2157
	[0.685]	[0.050]	[0.595]
Panel G			
Political science or other social sciences minister	0.0383*	0.0232	-0.2396
	[0.060]	[0.152]	[0.703]
Political science or other social sciences gov. head	0.0148	-0.0062	0.4000
	[0.500]	[0.716]	[0.657]
Panel H			
Teaching, social work, or pedagogics minister	-0.0316	-0.0062	1.2208
	[0.202]	[0.715]	[0.184]
Teaching, social work, or pedagogics gov. head	-0.0631	-0.0481	1.3535
	[0.135]	[0.143]	[0.480]

Notes: Column 1 is based on column 3 of Table 2.2. Column 2 is based on column 4 of Table 2.2. Column 3 is based on column 3 of Table 2.3. All regressions include the lagged dependent variable, all control variables, country-fixed effects, and year-fixed effects. Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.6. Development minister characteristics and total ODA budgets (1971-2012, contemporaneous explanatory variables)

	(1) (log) Total ODA com.	(2) (log) Total ODA disb.	(3) (log) Total ODA com.	(4) (log) Total ODA disb.	(5) (log) Total ODA com.	(6) (log) Total ODA disb.
Lagged DV	0.4553***	0.6765***	0.4608***	0.6492***	0.0909	0.3058***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.110]	[0.000]
(log) GDP per capita	1.3949***	0.9555***	1.5253***	1.0437***	2.3482***	2.1493***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Openness	0.0012	0.0008	0.0013	0.0014**	-0.0013	-0.0012
	[0.245]	[0.106]	[0.259]	[0.021]	[0.239]	[0.220]
Gov. expenditure	0.0371***	0.0244***	0.0379***	0.0228***	0.0752***	0.0628***
	[0.000]	[0.001]	[0.000]	[0.002]	[0.000]	[0.000]
Debt	-0.0018***	-0.0009***	-0.0015***	-0.0009***	-0.0012**	0.0000
	[0.000]	[0.002]	[0.001]	[0.005]	[0.042]	[0.999]
Political globalization	0.0052***	0.0030**	0.0075**	0.0038*	0.0038	0.0063*
	[0.010]	[0.047]	[0.015]	[0.092]	[0.341]	[0.072]
Aid agency	0.0140	0.0447*	0.0077	0.0392	0.0239	0.0856*
	[0.658]	[0.069]	[0.816]	[0.109]	[0.654]	[0.056]
(log) Colonial history	-0.2583***	-0.1469***	-0.1869**	-0.1288**	-0.3636	-0.4715**
	[0.001]	[0.007]	[0.013]	[0.027]	[0.280]	[0.013]
Female minister	-0.0174	-0.0013	-0.0162	-0.0047	-0.0141	-0.0108
	[0.468]	[0.942]	[0.496]	[0.801]	[0.606]	[0.638]
Female gov. head			0.0197	0.0080		
			[0.591]	[0.789]		
Female parliament			0.0061**	0.0051***		
			[0.011]	[0.005]		
Right-wing minister	0.0015	0.0099	0.0100	0.0006	-0.0090	0.0169
	[0.943]	[0.540]	[0.760]	[0.981]	[0.749]	[0.485]
Right-wing gov. head			-0.0163	-0.0038		
			[0.679]	[0.904]		
Right-wing parliament			-0.0835	0.1585*		
			[0.460]	[0.052]		
Ideological difference			0.0345	0.0084		
			[0.160]	[0.630]		
Prof. dev. coop. minister	0.0043	-0.0219	-0.0105	-0.0255	-0.0289	-0.0720**
	[0.875]	[0.395]	[0.700]	[0.320]	[0.371]	[0.024]
Prof. dev. coop. gov. head			0.0350	-0.0097		
			[0.284]	[0.708]		
Economics & business minister	-0.0015	0.0014	-0.0153	0.0011	-0.0015	0.0007
	[0.949]	[0.908]	[0.458]	[0.926]	[0.969]	[0.968]
Economics & business gov. head			0.0477**	0.0210		
			[0.021]	[0.151]		
Tenure minister	0.0068*	0.0059**	0.0062	0.0056**	0.0040	0.0055**
	[0.092]	[0.014]	[0.131]	[0.021]	[0.422]	[0.025]
Tenure gov. head			-0.0006	-0.0012		
			[0.849]	[0.589]		

Country FE	Yes	Yes	Yes	Yes	No	No
Gov. head FE	No	No	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	829	860	802	833	829	860
R-squared (within)	0.818	0.919	0.823	0.917	0.473	0.664
Number of countries	23	23	23	23	23	23
Number of gov. heads	181	186	178	183	181	186
Average number of years	36.0	37.4	34.9	36.2	4.6	4.6

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.7. Development minister characteristics and total ODA budgets as a share of government expenditures (1971-2012)

	(1) (log) Total ODA disb. (% gov. exp.)	(2) (log) Total ODA disb. (% gov. exp.)	(3) (log) Total ODA disb. (% gov. exp.)
Lagged DV	0.8017***	0.7810***	0.3513***
	[0.000]	[0.000]	[0.000]
(log) GDP per capita	0.3821***	0.5199***	1.7588***
	[0.001]	[0.000]	[0.000]
Openness	0.0013*	0.0022***	0.0004
	[0.097]	[0.008]	[0.779]
Gov. expenditure	-0.0135*	-0.0149**	0.0139
	[0.059]	[0.050]	[0.403]
Debt	-0.0005	-0.0007	-0.0004
	[0.187]	[0.103]	[0.565]
Political globalization	-0.0001	0.0009	0.0057
	[0.977]	[0.737]	[0.173]
Aid agency	0.0911**	0.1005**	0.1825*
	[0.017]	[0.014]	[0.067]
(log) Colonial history	-0.2430***	-0.2458**	-1.0213***
	[0.008]	[0.011]	[0.002]
Female minister	-0.0196	-0.0272	-0.0727
	[0.534]	[0.420]	[0.198]
Female gov. head		0.0481	
		[0.210]	
Female parliament		0.0038	
		[0.139]	
Right-wing minister	-0.0140	-0.0269	0.0426
	[0.507]	[0.382]	[0.319]
Right-wing gov. head		-0.0125	
		[0.746]	
Right-wing parliament		0.2056**	
		[0.020]	
Ideological difference		0.0048	
		[0.840]	
Prof. dev. coop. minister	0.0262	0.0306	-0.0106
	[0.326]	[0.235]	[0.800]
Prof. dev. coop. gov. head		-0.0148	
		[0.732]	
Economics & business minister	0.0150	0.0135	0.0058
	[0.502]	[0.544]	[0.862]
Economics & business gov. head		0.0513**	
		[0.022]	
Tenure minister	0.0082*	0.0094**	0.0090*
	[0.059]	[0.041]	[0.086]
Tenure gov. head		-0.0066**	
		[0.028]	
Country FE	Yes	Yes	No
Gov. head FE	No	No	Yes
Year FE	Yes	Yes	Yes
Observations	866	838	866

R-squared (within)	0.829	0.835	0.494
Number of countries	23	23	23
Number of gov. heads	187	185	187
Average number of years	37.7	36.4	4.6

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.8. Development minister characteristics and total ODA budgets (without potential "bad controls")

	(1) (log) Total ODA com.	(2) (log) Total ODA disb.	(3) (log) Total ODA com.	(4) (log) Total ODA disb.	(5) (log) Total ODA com.	(6) (log) Total ODA disb.	(7) Quality ODA	(8) Quality ODA
	1971-2012	1971-2012	1971-2012	1971-2012	1971-2012	1971-2012	1995-2011	1995-2011
(log) GDP per capita	2.6919***	2.9756***	2.6820***	3.1219***	2.4534***	3.1976***	13.2682**	16.7866**
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.049]	[0.015]
Openness	0.0033**	0.0053***	0.0040***	0.0067***	-0.0002	0.0009	-0.0079	-0.0055
	[0.013]	[0.000]	[0.006]	[0.000]	[0.789]	[0.248]	[0.709]	[0.754]
Gov. expenditure	0.0536***	0.1022***	0.0535***	0.0949***	0.0374**	0.0856***	0.9956**	0.8111*
	[0.000]	[0.000]	[0.000]	[0.000]	[0.017]	[0.000]	[0.017]	[0.066]
Debt	-0.0021***	-0.0021***	-0.0018***	-0.0024***	-0.0011	-0.0005	-0.1314***	-0.1106***
	[0.000]	[0.001]	[0.002]	[0.002]	[0.289]	[0.435]	[0.000]	[0.000]
Openness	-0.4351***	-0.4067***	-0.3513***	-0.3406***	-0.5319	-0.5306**	15.7455***	15.1474***
	[0.000]	[0.000]	[0.001]	[0.001]	[0.176]	[0.045]	[0.002]	[0.005]
Female minister	-0.0462	-0.0291	-0.0424	-0.0460	-0.0422	-0.0417	0.7275	0.8525
	[0.118]	[0.362]	[0.163]	[0.161]	[0.193]	[0.175]	[0.218]	[0.191]
Female gov. head			0.0804	0.0871				-2.1743*
			[0.208]	[0.189]				[0.092]
Female parliament			0.0112***	0.0158***				0.1067
			[0.001]	[0.000]				[0.275]
Right-wing minister	0.0158	0.0136	0.0056	-0.0426	0.0209	0.0420	0.6748	-0.7347
	[0.646]	[0.693]	[0.919]	[0.372]	[0.704]	[0.262]	[0.423]	[0.475]
Right-wing gov. head			-0.0248	-0.0266				2.8666**
			[0.734]	[0.682]				[0.027]
Right-wing parliament			-0.0031	0.6662***				-5.4174
			[0.983]	[0.005]				[0.125]
Ideological difference			0.1135***	0.0865***				0.3783
			[0.001]	[0.010]				[0.600]
Prof. dev. coop. minister	0.0378	0.0197	0.0087	0.0026	-0.0539	-0.0317	1.2071**	1.1823*
	[0.307]	[0.712]	[0.804]	[0.953]	[0.163]	[0.394]	[0.049]	[0.055]
Prof. dev. coop. gov. head			0.0264	0.0487				-2.0780
			[0.595]	[0.277]				[0.160]

Economics & business minister	0.0139	0.0276	-0.0129	0.0176	-0.0205	-0.0020	-0.7762	-1.0946*
	[0.637]	[0.373]	[0.586]	[0.516]	[0.531]	[0.938]	[0.147]	[0.087]
Economics & business gov. head			0.0833**	0.1057***				-0.5075
			[0.011]	[0.002]				[0.534]
Tenure minister	0.0110**	0.0131**	0.0094*	0.0101*	0.0084	0.0082**	0.1402	0.1895*
	[0.036]	[0.018]	[0.066]	[0.056]	[0.186]	[0.043]	[0.166]	[0.056]
Tenure gov. head			-0.0039	-0.0055				0.0804
			[0.333]	[0.243]				[0.421]
Country FE	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Gov. head FE	No	No	No	No	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	840	866	812	838	840	866	840	866
R-squared (within)	0.748	0.825	0.755	0.839	0.437	0.589	389	387
Number of countries	23	23	23	23	23	23	0.217	0.252

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Appendix B2.9. Development minister characteristics and ODA quality (1995-2011, lagged explanatory variables)

	(1)	(2)	(3)	(4)	(5)
	Quality ODA	Quality ODA	Quality ODA	Quality ODA	Quality ODA
Lagged DV			0.2100*	0.6967***	0.0286
			[0.059]	[0.000]	[0.723]
(log) GDP per capita	13.9470**	18.5659**	10.6669	-0.0640	17.1457
	[0.031]	[0.017]	[0.179]	[0.954]	[0.215]
Openness	0.0284	0.0369	0.0198	0.0237**	0.0033
	[0.250]	[0.174]	[0.426]	[0.033]	[0.874]
Gov. expenditure	0.7766**	0.8748**	0.5335	0.2632	0.9219
	[0.038]	[0.029]	[0.212]	[0.125]	[0.220]
Debt	-0.1224***	-0.1160***	-0.0721***	-0.0554**	-0.0561*
	[0.000]	[0.000]	[0.004]	[0.025]	[0.100]
Political globalization	-0.2757**	-0.1818	-0.1381	0.1491**	-0.1491**
	[0.018]	[0.126]	[0.152]	[0.035]	[0.039]
Aid agency	-1.5301*	-1.7594**	-1.7850**	-1.2961	-1.7498
	[0.074]	[0.036]	[0.010]	[0.104]	[0.183]
(log) Colonial history	16.7812***	14.7386**	12.9526**	-0.0902*	13.5561
	[0.001]	[0.010]	[0.028]	[0.086]	[0.237]
Female minister	0.4325	0.6984	0.6109	0.3897	-0.0859
	[0.424]	[0.245]	[0.235]	[0.455]	[0.876]
Female gov. head		-2.3204*	-2.5747*	-0.7213	-2.2248*
		[0.056]	[0.059]	[0.343]	[0.078]
Female parliament		0.0826	-0.0004	-0.0020	0.0496
		[0.381]	[0.997]	[0.958]	[0.759]
Right-wing minister	1.0544	-0.0875	0.3154	-0.1010	0.2600
	[0.154]	[0.917]	[0.685]	[0.919]	[0.799]
Right-wing gov. head		1.4348	0.7836	-0.6174	-0.1320
		[0.136]	[0.438]	[0.581]	[0.917]
Right-wing parliament		-0.3676	-0.9763	6.5968**	0.1840
		[0.920]	[0.781]	[0.041]	[0.973]
Ideological difference		1.0091	0.5464	0.2978	0.2564
		[0.106]	[0.314]	[0.635]	[0.702]
Prof. dev. coop. minister	1.3633***	1.2758**	0.8779*	-0.9362	0.0220
	[0.008]	[0.031]	[0.069]	[0.232]	[0.975]
Prof. dev. coop. gov. head		-2.9010***	-2.0997**	0.1754	
		[0.002]	[0.011]	[0.875]	
Economics & business minister	-1.4710**	-1.4336**	-1.3003**	-0.1462	-0.9289
	[0.019]	[0.033]	[0.047]	[0.818]	[0.336]
Economics & business gov. head		-0.4551	-0.3569	0.2854	
		[0.490]	[0.516]	[0.622]	
Tenure minister	0.1975*	0.2587***	0.2131**	0.1411*	0.2130*
	[0.052]	[0.006]	[0.020]	[0.100]	[0.061]
Tenure gov. head		-0.0514	-0.0434	-0.0836	
		[0.446]	[0.464]	[0.212]	

Country FE	Yes	Yes	Yes	No	No
Gov. head FE	No	No	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	388	385	364	364	364
R-squared (within)	0.253	0.280	0.282	0.807	0.145
Number of countries	23	23	23	23	23
Number of gov. heads	88	88	82	82	82
Average number of years	16.9	16.7	15.8	15.9	4.5

Notes: Standard errors are clustered at the government-head level. P-values are in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

CHAPTER 3

The Political Economy of International Finance Corporation Lending[104]

Co-authored with Axel Dreher

3.1. INTRODUCTION

Consider the International Finance Corporation's (IFC) multi-million credit to Ghana approved in 2010. The credit was intended to facilitate the renovation of a five-star Mövenpick Hotel in Accra, the country's capital. According to the IFC's "summary of proposed investment," the project company is a fully owned subsidiary of Kingdom Hotels Investments, a global player with operations in 18 countries, and owned by a Saudi Arabian prince.[105] As the IFC points out, the project's development impact consists of adding important business infrastructure, creating new jobs, and providing demand for local food and non-food supplies. According to the IFC, the project has been a "great success," obtaining most revenue per room in all Accra.[106] Despite the ongoing global financial crisis at the start of the project it is hard to imagine that private capital would not have been available to finance the project, or that other, more obviously developmentally oriented projects, could not have been more worthy of support.

---

[104] An earlier version of Chapter 3 is available as working paper: Dreher, Axel and Katharina Richert (2017). The Political Economy of International Finance Corporation Lending. *CEPR Discussion Paper* No. 12290.

[105] See http://ifcext.ifc.org/ifcext/spiwebsite1.nsf/DocsByUNIDForPrint/D8FCCDDFAFCFB6FF852577E5005E6194?opendocument (last accessed December 12, 2016).

[106] See http://www.ifc.org/wps/wcm/connect/0a5122004c23a313bf56bfd8bd2c3114/The+Movenpick+Ambassador+Hotel,+Ghana.pdf?MOD=AJPERES (last accessed December 12, 2016).

One might expect projects such as the five-star luxury hotel to be an exception in the IFC's portfolio. However, two thirds of the IFC's investments go to companies from the world's richest countries, while only one fifth goes to companies of the poorest countries (Ellmers et al. 2010). What is more, the bulk of the IFC's investments go to projects implemented in middle-income countries, rather than poorer ones (Ellmers et al. 2010). With the majority of projects, obvious developmental benefits are hard to find.[107] It is implausible to assume (at least for the authors of this chapter) that international companies investing in middle-income countries like Brazil and Romania cannot easily access private capital markets. Overall, it seems that the typical IFC investment finances a project in a middle-income country, of dubious developmental impact, and executed by a large conglomerate from an industrialized country that would have enjoyed easy access to private capital.

Still, as the IFC's guidelines explain, its official mandate is to finance poverty-reducing projects for which "sufficient private capital is not otherwise available on reasonable terms" (e.g., IFC 2004: 4). The IFC is part of the World Bank Group, which features the ending of extreme poverty and promotion of shared prosperity as its key mandate. This mandate is explicitly shared by the IFC.[108]

The apparent contrast between the IFC's official mandate and observed reality presents an interesting and important puzzle. Over the last decade or so, the IFC became a major player in development lending, with a 2016 portfolio of new commitments amounting to almost US$ 19 billion. Understanding the drivers of IFC lending is thus of vital importance. In this chapter – as we outline in some detail in Section 3.2. – we argue that the governments of industrialized country companies receiving IFC money form coalitions with governments of middle-income countries where the bulk of investments are executed. Both types of governments can influence IFC lending in various ways. We therefore expect preferential treatment when such countries hold positions of power. Preferential treatments should be most pronounced when

[107] That is, unless one is willing to view every project that has a positive contribution to GDP, does not substantially hurt the environment, and creates some jobs as developmental. In 2016, of the US$ 19 billion in long-term investments, less than one third went to countries eligible for concessional aid from the International Development Association – the so-called IDA countries (IFC 2016). More than half of the IFC's investments go to ten middle-income countries – the BRICs alone receive about one third (Ellmers et al. 2010). According to the World Bank Group's Independent Evaluation Group (IEG, 2007), 43 percent of the IFC's projects that were evaluated during the 2000-2005 period did not receive high development ratings.

[108] See, e.g., http://www.ifc.org/wps/wcm/connect/corp_ext_content/ifc_external_corporate_site/solutions/solutions (last accessed December 12, 2016).

the representatives of both the recipient's and the company's countries are best able to collude in exerting their influence. We argue that this will be the case when both countries' governments are represented among the IFC's Board of Directors[109] – the body that makes the final decision about loan approval – and at times these countries are members of the United Nations Security Council (UNSC), where they have extraordinary clout with major IFC shareholders who control the Board.

As we outline in Section 3.3., we have collected data for more than 3,000 IFC projects over the 1995-2015 period. We test whether the (joint) influence of countries receiving the projects and countries hosting the companies that implement them helps these countries to garner a disproportionate share of IFC funding. To evaluate our hypothesis, we rely on two proxies for shareholder influence in international organizations widely used in the related literature. Kaja and Werker (2010), Morrison (2013), and Malan (2016) have shown that country representation on the Executive Boards of international organizations substantially increases the size of the loans that countries receive from these organizations. What is more, a large number of papers have shown that those countries with influence over an international organization's major shareholders receive more of the organization's loans at more favorable terms (Thacker 1999, Kuziemko and Werker 2006, Dreher and Jensen 2007, Dreher et al. 2009a, 2009b, Vreeland and Dreher 2014). We follow this literature and investigate whether the IFC extends larger loans at times when the recipient government and the government of the country whose company executes the project (i) are members of the IFC's Board of Directors or (ii) are represented as members of the United Nations Security Council (UNSC), which is of paramount importance to major IFC shareholders.

We extend the literature in three dimensions. One, we are the first to investigate the importance of political influences for IFC lending. Two, in addition to investigating the political importance of borrowing countries we also focus on the importance of countries that represent the interests of their private companies, which has received little attention in the international financial institutions literature.[110] And finally, we speak to the recent policy

[109] Throughout the chapter we use "representation" rather than "direct representation" when we refer to countries that nominate an Executive Director. Of course, the Board member for a group is supposed to represent all countries in the group.

[110] More broadly, our chapter thus also relates to the literature on politically connected firms. This literature typically focusses on individual countries and finds that firms with connections to the country's government receive larger loans (e.g., Khwaja and Mian 2005).

debate on leveraging private funding for development (as for example highlighted in the 2015 Addis Ababa Action Agenda). To the extent that the allocation of funding is shaped by political interests rather than need or expected rates of return, an allocation of private funds that gives politicians or international bureaucrats some weight might be less effective in promoting development than commonly thought, just as political considerations in allocating official aid make the aid less effective in raising growth (Dreher et al. 2018).

Our results – presented in Section 3.4. – show that the representation of a country and in particular its project partners on the IFC's Board of Directors or on the UNSC significantly and substantially increases IFC loan size. We find additional effects in cases where both countries hold one of these powerful positions, resulting in more IFC projects and larger loans. Given that these significant effects disappear in the years after countries' joint representation, these estimates seem to represent the causal effect of membership rather than some permanent omitted characteristics shared by these countries. This is most true regarding the two-year spells of temporary membership on the UNSC, which has been shown to be rather idiosyncratic, and not determined by variables that also affect the lending behavior of international organizations (Dreher et al. 2014, Vreeland and Dreher 2014).

Section 3.5. concludes the chapter. We argue that the commercial incentives of the political coalitions involved work contrary to the IFC's goal of poverty reduction and economic development. Given the voting power of the current system's beneficiaries in the IFC, however, reforming the IFC so that it truly fosters development will be all but impossible.

3.2. THE ARGUMENT

The IFC is the private sector arm of the World Bank Group.[111] It has 184 member countries who provide the organization's capital of US$ 2.56 billion. The paid-in capital in turn determines the members' voting weight.[112] In the fiscal year 2016, about one third of the World Bank Group's commitments were through the IFC, amounting to US$ 19 billion committed to 344 projects in 78 recipient countries. Within the World Bank Group, the IFC is becoming an

[111] The Group's original institution is the International Bank for Reconstruction and Development (IBRD). The Group also features the International Development Association (IDA), the Multilateral Investment Guarantee Agency (MIGA), and the International Centre for Settlement of Investment Disputes (ICSID). The IBRD and the IDA are jointly known as the World Bank.

[112] More specifically, each share of the IFC's capital stock comes with one vote. In addition, 5.55 percent of the total votes are shared as basic votes by all members equally.

increasingly important institution.[113] Its current President sees the future of the Group as a broker between private lenders and developing countries, which would substantially strengthen the role of the IFC.[114] According to Ellmers et al. (2010: 8) "private sector finance may even become the new core business of the Bank."

The IFC's highest decision making body – the Board of Governors – has delegated the institution's day-to-day business to the Board of Directors, which consists of the World Bank President and 25 representatives of the IFC's member countries. The six countries with the largest shares in the World Bank Group each appoint one Director;[115] the remaining Directors are elected by groups of countries[116] and have the final say over all the votes of the group they represent. This gives some countries substantially more power over the IFC's decisions than others. For example, the United States – by far the IFC's largest shareholder – is in charge of almost 21 percent of the votes, Japan is in command of six percent, while no other country has a vote share exceeding five percent.[117] The smallest vote share is 0.03 percent – held by Afghanistan, Belize, Benin, and Botswana, among others. Though formal voting over lending decisions is rare, and proposals are typically not voted down, decisions require a majority of the Directors and at least half of the total votes of the Board.[118] It is thus well-known to IFC staff preparing the loans which countries are expected to receive favorable treatment and which loans, if proposed, would likely be rejected. We expect staff to make use of this knowledge when preparing a loan. Consequently, loans that favor countries whose interests are represented on the Board are more likely to be put forward. What we have in mind here is thus a combination of formal and informal influence in the IFC's decision-making. Formal power derives from the voting weight of a country. The importance of informal channels for international organizations' decision-making has been pointed out by Stone (2011, 2013) and Koremenos (2013), among others. As Lang and Presbitero (2017) explain, preemptive

---

[113] Commitments have been as low as US$ 4 billion in the year 2000 only, which might explain why the IFC has received little attention in the academic literature so far.

[114] See President Jim Yong Kim's lecture "Rethinking Development Finance" on 11 April 2017 at the London School of Economics, http://www.lse.ac.uk/website-archive/newsAndMedia/videoAndAudio/channels/publicLectures AndEvents/player.aspx?id=3802.

[115] That is, France, Japan, United Kingdom, Germany, United States, and China.

[116] The exception is Saudi Arabia whose vote share is high enough to "elect" its own Director. The Director nominated by Russia represents the votes of only one additional country, Syria.

[117] The World Bank Group provides details at https://data.worldbank.org/data-catalog/gsdw-avpz (last accessed September 7, 2017).

[118] Voting records are not publically available, so that analysis at the vote-level is impossible.

obedience by World Bank bureaucrats can give rise to political biases even in the absence of any direct major power intervention.

Our argument rests on two pillars. First, we expect governments of countries where IFC projects are typically implemented and those of countries whose companies are interested in applying for IFC funding to hold some sway over the IFC. This is either because they are influential in the IFC themselves or because they have powerful allies who support their interests. Second, we argue that the private companies applying for IFC loans exert pressure on these governments. We outline our expectations regarding these two channels of influence in turn.[119]

The influence of international organizations' major shareholders on the organizations' policies has been investigated before. Among others, international organizations grant their major shareholders and their shareholders' political allies more and larger loans at more generous terms and with shorter preparation times, better growth and inflation forecasts and policy surveillance reports, and better risk ratings in the Debt Sustainability Framework of the World Bank and the International Monetary Fund (IMF).[120] These types of political influences have received particular attention for the IMF, the International Development Association (IDA), and the International Bank for Reconstruction and Development (IBRD). We expect the same mechanisms to hold for the IFC. The governing structure of the IFC is very much in analogy to that of the IMF, and the other World Bank Group institutions. In fact, its Board of Directors is identical to those of the IDA and the IBRD.[121] To the extent that the other institutions of the World Bank Group are receptive to shareholder influence, there is little reason to expect this influence to be absent from IFC lending.

Countries represented on the IFC's Board of Directors have direct control over the loans approved by the IFC. We thus expect the countries that are represented there to receive more favorable treatment. As Kaja and Werker (2010) point out, the boardroom culture created by frequent meetings within the same group of decision-makers leads to the Directors engaging

[119] We do not explore other potentially important explanations for the IFC's lending pattern, such as the institutional pressures for the IFC to grow and to generate funds for other branches of the World Bank Group.

[120] See Aldenhoff (2007), Kuziemko and Werker (2006), Dreher et al. (2008), Kaja and Werker (2010), Fratzscher and Reynaud (2011), Vreeland and Dreher (2014), Dippel (2015), Kilby (2011, 2013, 2015), Kersting and Kilby (2016), Lang and Presbitero (2017). For a broader overview of the political economy of international organizations see Dreher and Lang (2016).

[121] This is even though some members of the IBRD are not members of the IDA or IFC.

in logrolling behavior. The Board is thus likely to support loans that are in the interest of a specific director, who in turn supports loans in the interest of other members of the Board. Indeed, the results of Kaja and Werker (2010) show that countries represented on the Board receive substantially more loans from the IBRD compared to what these same countries receive at other times, including the years directly before joining the Board and after leaving the Board. Until the World Bank's Country Policy and Institutional Assessment (CPIA) gained substantial weight in the decision of which countries receive IDA support in recent years, the same was true for the IDA (Morrison 2013).[122] Given the similarity between these organizations and the IFC we expect that countries represented on the IFC's Board of Directors as well as their allies will receive larger IFC loans at times they request them.

Previous research has indeed shown that countries of importance to the World Bank's major shareholders receive more and larger loans from the IDA and the IBRD. A widely used proxy for political importance is membership on the UNSC. This follows the seminal work by Kuziemko and Werker (2006) who show that temporary UNSC members receive a surge in foreign aid from the United States during the two years of their membership. The United States apparently cares a great deal about the UNSC,[123] and the same holds for other important shareholders of the World Bank (Vreeland and Dreher 2014). With respect to the World Bank, Dreher et al. (2009b) find that temporary members receive substantially more IBRD and IDA projects during their time on the UNSC.[124] They attribute this to the interests of the major shareholders in doling out favors to countries of importance to them and the resulting shareholders' pressure on the Bank for more favorable terms. Given that the major shareholders of the World Bank care about whether or not a country is a member of the UNSC, we expect them to be attentive to these countries' interests at the IFC as well.

The second pillar in our theory – the influence of private companies on international organizations' lending behavior – is key in deriving our hypotheses, as it is these companies that apply for IFC loans. The motives of private companies have received much less attention than the direct interests of governments. Notable exceptions are Broz and Hawes (2006), Malik and Stone (2016) and McLean (2017).

---

[122] See Malan (2016) for similar results on the IMF.

[123] Also see Lai and Lefler (2016).

[124] Kersting and Kilby (2017) show that UNSC membership is an important determinant of the Bank's supplemental loans.

Broz and Hawes (2006) focus on banks. They show that countries with larger exposure to U.S. banks are more likely to receive IMF programs and larger loans. They attribute this to the banks' influence over U.S. politicians and the politicians' power over the IMF. McLean (2017) investigates contract allocation in the IDA and the IBRD. As she explains (2017: 257): "donor governments pay significant attention to the inflow of contracts funded by multilateral aid to their economies due to domestic political and economic considerations."[125] Malik and Stone (2016) point out that large multinational companies are important political actors, with some influence over both their home country and the country where they invest. They show that IDA and IBRD projects involving multinational U.S. companies receive larger disbursements and better evaluations relative to their performance, compared to other companies. Malik and Stone investigate political influences at the disbursement – rather than the commitment – level. They consequently attribute their result to lobbying with World Bank employees who are in charge of disbursing funds and evaluating projects.

Malik and Stone describe a number of channels by which private companies can influence the World Bank's decisions. Among them is lobbying with the local congressional office, which then passes on the company's request to the country's authorities in charge of the World Bank – the Treasury in case of the United States. A Treasury representative could then pass on the request to the United States representative on the Bank's Board of Directors, in charge of approving loans, or to the director of a specific department preparing the loan.[126] For everyone involved complying is easier than ignoring the request. As Malik and Stone (2016: 8) put it "Lobbying is effective because the chain of delegation ensures that there is no one holding the door shut." As they explain (p. 8), the congressional office routinely forwards requests like these, because compliance comes at almost no cost, while non-compliance might easily become costly in the future. The Treasury to some extent depends on congressional votes to pass appropriations bills, while costs to put pressure on the World Bank are hardly significant.[127]

---

[125] Broz and Hawes (2006) and Broz (2008, 2011) show that this holds true for U.S.-policies on international organizations as well. They show that U.S. commercial banks hold some sway over how U.S. congress legislates the IMF.

[126] Parízek (2017) shows that powerful states are over-represented among the staff of international organizations' secretariats, which facilitates the exercise of power over them. Also see Novosad and Werker (2018).

[127] Oatley and Yackee (2004) argue that commercial banks' influence in part runs through domestic interest groups within the United States. U.S. policymakers are receptive to U.S. commercial bank

Overall, our argument is thus an indirect one. Private companies have a clear interest in receiving the IFC's loans. This is because, even though the interest rates on loans are comparable to market rates, IFC loans are subsidized. The IFC raises its funds in the international debt markets. Due to its high paid-in capital it has a AAA rating and can borrow at prime conditions. The IFC does not pay dividends to its shareholders and is exempt from corporation taxes. These subsidies are to some extent passed on to borrowers, in the form of longer maturing loans or longer grace periods compared to market conditions (Te Welde and Warner 2007). As Te Welde and Warner (2007) point out, technical assistance – underpriced compared to market conditions – is a key subsidy in support of IFC projects. What is more, complementary private money might be more easily available for projects with IFC support, for example because private creditors rely on the IFC's screening of projects or expect preferential political treatment for these projects in the recipient country.[128]

Private companies lobby their national government which, in turn, either holds some influence at the IFC itself, or uses its political influence with other governments that are powerful in the IFC (for example because they hold positions on the UNSC). We expect companies to lobby the government of the country where the projects are implemented as well. Large multinational corporations can have substantial influence on low- and middle-income governments (e.g., Jensen 2008). These governments arguably have an interest in promoting business in their country, in particular if it comes at little cost to them. Using their seat on the IFC Board or their clout with important shareholders of the IFC, we expect recipient country governments to give in to companies' demands, and extend their influence at times they have some.

In summary, we expect country representation on the UNSC and the IFC's Board to result in larger loans to companies from that country. We also expect that the larger the number of actors interested in a certain Board decision, the easier it will be to build an effective coalition in the Board that favors this outcome. The effect of these countries' presence on projects in their shared interest could easily be larger than the sum of their individual influences. We therefore investigate the joint influence of the government in the country

---

pressure to use their privileged access to IMF decision-making to represent the banks' interests in the Fund.

[128] This might work across similar lines as the IMF's "seal of approval" – the catalyzing impact attributed to IMF programs with respect to private capital flows (e.g., Biglaiser and DeRouen 2010).

receiving the project and the government of the company implementing it as well. We expect that loan size will be maximal during times of joint representation. This should hold most with respect to the low- or middle-income countries receiving the project, which are arguably less well politically connected compared to rich country governments and will find it more difficult to muster effective support from the IFC's shareholders at times they do not form such coalitions.

3.3. DATA AND METHOD

Our key dependent variable is new IFC loan amounts committed. These data are available from the World Bank Group's website.[129] Crucially, we require data on the country of investment ("recipient country") and on the country of the project company ("sponsor"). While the former can easily be extracted from the World Bank Database, the latter had to be hand-coded for each individual project based on the information about the project company's headquarters given on the website.[130] For example, the project sheet for the Mövenpick project in Ghana given in the introduction details that the project company is KHI Ghana Limited, which is a fully owned subsidiary of Kingdom Hotels Investments, owned by Prince Walid of Saudi Arabia. We therefore coded Ghana as the country receiving the project and Saudi Arabia as the sponsor.

Appendix A3. shows the shares of projects and loan volumes committed to the 30 largest sponsor and recipient countries. As can be seen, the highest percentage of loan volumes has been committed to companies from the United States with 7.8 percent of overall commitments and 4.8 percent of all projects over our sample period. In addition to the United States, the ten sponsor countries receiving the bulk of IFC funding are Brazil, India, Turkey, China, Russia, France, Mexico, the United Kingdom, and Indonesia. Overall, companies from these ten countries received almost half (49 percent) of the total investment volume in our sample period.[131] The largest recipient country is India, with a share of 8.5 percent of total IFC commitments over the sample period.

---

[129] See http://ifcextapps.ifc.org/ifcext/spiwebsite1.nsf/$$Search?openform (last accessed April 29, 2017).

[130] As an interesting extension, future research might consider coding the individual shareholders of multinational companies as well.

[131] Note that a substantial number of projects in a specific host country are implemented by a company from the same country (69 percent of the loans; 67 percent of the projects). However, in 27 percent of the observations with such projects at least one additional project is implemented by a company from another country.

Our variables of interest are membership on the IFC's Board of Directors and membership on the UNSC. We code a binary indicator showing whether or not a country was represented on the IFC's Board of Directors relying on information from the World Bank's Annual Reports for our sample years. We updated data on temporary membership on the UNSC from Dreher et al. (2009a) and added the five permanent members – United States, China, Russia, United Kingdom, and France.[132]

Our first set of regressions focus on the benefits to the recipient country.[133] They are consequently at the recipient-year level:

$$y_{it} = \beta_1 Member_{it} + \beta_2 Partner\ Member_{it} + \beta_3 Member_{it} * Partner\ Member_{it} + \beta_4 \#Partners_{it} + \beta_5 GDPpc_{it-1} + \beta_6 Population_{it-1} + \gamma_i + \tau_t + \varepsilon_{it}, \quad (1)$$

where y_{it} are (logged) IFC loan amounts committed to recipient country i in year t, in constant 2010 US$.[134] $Member_{it}$ represents positions of power that country i holds in a year t – either resulting from membership on the UNSC or from membership on the Board of the IFC.[135] $Partner\ Member_{it}$ tests whether a country's IFC loans turn out larger at times where it receives greater support from other influential governments. More precisely, it measures the share of a country's project partners that are represented on the Bank's Board or hold positions as members of the UNSC. A country is coded as partner when it sponsors at least one project in a specific recipient country in a given year. While we include projects where the sponsor and recipient countries are identical in most regressions, we do not code recipients to be their own partners (i.e., the share variable is zero for countries that only receive projects which are implemented by companies from the same country). The interaction between $Member_{it}$ and $Partner\ Member_{it}$ tests the effect of their joint presence over and in addition to them being present themselves.

---

132 The previous literature excludes the permanent members given that they are multicollinear with country fixed effects included in most studies. Given that we focus on interactions between members with positions on the UNSC these permanent members add variation even in the presence of country fixed effects.

133 We exclude high-income economies from the set of recipient countries, as they are not eligible for IFC financing. However, there are some exceptions of high-income countries receiving projects anyway. We keep the high-income countries with at least one project approved in our recipient sample. This makes a total number of 155 potential recipient countries, of which we lose two in the regressions below, due to missing data for our control variables.

134 We added one dollar to all values before taking the logarithm to avoid losing zero observations.

135 In what follows, we refer to positions on the UNSC or the IFC Board as positions of power, or influence.

In most regressions, we also control for the total number of a country's project partners ($\#Partners_{it}$). We thus test whether countries that are equal in terms of the overall number of their partners receive larger loans when these partners are more influential. This is an important variable to control for in our setting. At any point in time, a country's influence in the IFC is likely to increase with the number of its project partners – whether or not these partners are represented on the Board. This influence arguably gives them easier access to loans. At the same time, the number of partners is highly correlated with the share of partners in positions of influence.[136] Controlling for the number of project partners is thus essential for our identification strategy, as we would otherwise not be able to separate the influence of the share of partners in positions of influence from that of the number of partners itself. We also include an interaction between a country's influence and the influence of its total number of partners to test whether their joint influence gives them larger sway, above the sum of their individual powers. We include this interaction for the same reason we include the number of project partners per se: At any time a country enters positions of influence it is likely to be more successful in achieving its goals with a larger number of project partners, independent of whether these partners hold extraordinary positions of influence at the same point in time.

We control for (lagged and logged constant 2010 US$) GDP per capita of the recipient country as well as its (lagged and logged) population. All regressions include dummies for each recipient country (γ_i) and year (τ_t); finally, ε_{it} represents the error term. The inclusion of country fixed effects is of particular importance. Members of the UNSC and the IFC's Board of Directors systematically differ from non-members, on average (Vreeland and Dreher 2014). We therefore compare the amount of loans a country receives at the time it holds extraordinary power compared to what the same country receives from the IFC at other times. This does not fully rule out endogeneity of IFC Board membership, where countries tend to hold positions for protracted periods of time, and thus might be different in other respects compared to the same country at other times. With regards to UNSC membership however the inclusion of country-fixed effects makes it unlikely that omitted variables bias our estimates. While larger and more developed countries enter the UNSC more frequently than others, these variables do not predict the timing of temporary membership (Dreher et al. 2014). Temporary membership on the UNSC is limited to two years, with no immediate re-election possible.

[136] Most obviously, a country with more partners has a higher probability that the share of its partners holding positions of influence exceeds zero in any year.

Given that most countries announce their intention to compete for a seat on the UNSC many years in advance, and 80 percent of the elections are uncontested, it is unlikely that a country receives higher IFC loans for exactly the two years of membership for any reason unrelated to membership itself (Dreher et al. 2014, Mikulaschek 2017). This is even more true for the share of a country's project partners that hold a seat on the UNSC at any particular point in time.

To further increase confidence that membership on the UNSC and the Board of Directors are not driven by time-varying country-specific variables that are correlated with the amount of IFC loans, we also estimate specifications that exclude the interaction between recipient and sponsor influence (thus reducing complexity), but instead include leads and lags of our membership variables. Some countries might be more important at certain points in time than at others. As a consequence, they receive larger IFC loans while, at the same time, they might be more likely to enter the UNSC or the Board of the IFC's Directors. A more important country might have more important partners, so that the share of a country's project partners on the UNSC and the Board might increase as well. To test whether time-varying country-specific events bias our estimates we rely on the following specification:

$$y_{it} = \sum_h \beta_{I,h}\, Member_{ih} + \sum_h \beta_{PI,h}\, Partner\ Member_{ih} + \beta_1 \#Partners_{it} + \beta_2 Member_{it} * \#Partners_{it} + \beta_3 GDPpc_{it-1} + \beta_4 Population_{it-1} + \gamma_i + \tau_t + \varepsilon_{it}, \qquad (2)$$

with *h* indicating the years *t-2* to *t*+2. Any trend in country-specific time-varying variables that affects the probability that a country or its project partners enter positions of influence should be reflected in the years before or after a country holds such a position. It is often known years in advance who will run for election and, as in the bulk of cases election is uncontested, who will eventually be elected. For example, European countries announce their decision to run for election between five and fifteen years ahead of the actual election (Mikulaschek 2017). For these countries it is, more often than not, known at least five years in advance – when no other country decided to run – who will be on the UNSC in future (Mikulaschek 2017).

The logic of diffuse reciprocity discussed in Vreeland and Dreher (2014) for the IMF and the World Bank leads us to expect significant increases in loans in the years immediately prior to membership. Countries in positions of power are not necessarily bribed or rewarded for any particular vote they cast in line with the interests of powerful allies on the UNSC (or the IFC Board). Instead, these allies hold them in their debt by supporting their requests as soon as it becomes known that these countries will hold positions of power, relying on these countries' goodwill when casting votes in decisions to come (also see Thorvaldsdottir 2015, Mikulaschek 2017). Given that in many cases it is well known which country will be on the UNSC or hold a seat on the IFC's Board in one or two years, it might well be that loans increase in the immediate years before a country enters the UNSC or the Board. We take this as an advantage rather than a disadvantage of our estimation strategy. While the timing of UNSC membership is idiosyncratic,[137] the consequences of membership might not be, and might potentially be correlated with IFC loan size. As one example, joint membership in international fora might lead to enhanced opportunities for cooperation between countries, so that more IFC projects result from such cooperation at the time of membership. Given the manifold possibilities of interaction between countries outside these fora, we consider this unlikely.[138] In any case, the effects of cooperation can hardly explain an increase in IFC loans *prior to entering* the IFC Board or the UNSC.

Some projects agreed on during a country's time on the UNSC or Board might also formally be committed only in the following years, so that larger loans occur in the first or second year following the end of a country's term in these positions of influence. However, the significance of longer lags would threaten the plausibility of our identification strategy.

In a second step we change perspectives, focusing on sponsor- rather than recipient-countries.[139] We repeat the regressions shown in specifications (1) and (2), explaining (logged) loan commitments to sponsor country *j*:

$$y_{jt} = \beta_1 Member_{jt} + \beta_2 Partner\ Member_{jt} + \beta_3 Member_{jt} * Partner\ Member_{jt} + \beta_4 \#Partners_{jt} + \beta_5 GDPpc_{jt-1} + \beta_6 Population_{jt-1} + \gamma_j + \tau_t + \varepsilon_{jt}. \tag{3}$$

While the regression is similar to those discussed above, note that $Partner\ Member_{jt}$ now reflects the membership of countries receiving the IFC's loans, rather than other sponsor countries. We also test an event-time specification, in analogy to that discussed above.

137 We are not aware of studies that investigate the determinants of membership on the IFC's Board across countries. For a study of Switzerland's election to the Boards of the IMF and the World Bank – and thus, implicitly, of the IFC – see Vreeland (2011).

138 To the extent that shared positions on the Board and the UNSC improve information about IFC projects, we would expect countries with previous experience there to receive more projects compared to those without. Membership should then become less important as determinant of IFC projects over time, given the easier spread of information via modern technology and due to the increase in IFC activity over time. We find no evidence for this.

139 We include all 184 IFC member countries as potential sponsor countries (of investing companies), but exclude two of them from our regressions due to missing data on the control variables. 163 of these countries actually are "sponsor countries" in our sample period.

Our third set of regressions investigates the interaction among recipient- and sponsor-power in more detail. We aim to test potential mechanisms behind the expected increase in IFC loans at times where countries enjoy (joint) political power. While focusing on the country-year level is crucial to test whether and to what extent temporarily influential countries benefit from their positions overall, regressions at the recipient-sponsor-year level can help shedding light on the mechanisms driving such increases. Specifically, we test whether countries on the UNSC or the Board of Directors receive a larger number of projects or larger projects, on average, when they operate in tandem with countries that also hold positions of influence. We therefore proceed at the dyadic level and estimate:

$$y_{ijt} = \beta_1 Member_{it} * Member_{jt} + \sum_k \beta_k Control_{kijt-1} + \gamma_{it} + \delta_{jt} + \tau_{ij} + \varepsilon_{ijt}, \quad (4)$$

with y_{ijt} representing one of our dependent variables (number of projects or (logged) average project volumes committed) involving projects to recipient country i and a company from sponsor country j in year t. $Member_{it}$ reflects the recipient country's influence at the IFC in year t, either indicated by membership on its Board of Directors or on the UNSC. $Member_{jt}$ measures the influence of the sponsor country. The levels of these variables are captured by fixed effects for recipient-years (γ_{it}) and sponsor-years (δ_{jt}). Our variable of interest is the interaction between the two, $Member_{it} * Member_{jt}$, which measures the joint influence of the recipient country and the sponsor country.

We include a number of control variables that vary at the recipient-sponsor-year level. Our most conservative regressions also include dyadic sponsor-recipient-fixed effects (τ_{ij}). At the dyad-year level, we control for whether, in year *t-1*, the recipient and sponsor country shared a common currency. We include the (lagged and logged) amount of foreign aid from the sponsor to the recipient country, the (lagged and logged) recipient country's imports from the sponsor, and a binary variable indicating whether both countries were part of the same trade agreement in the previous year.[140]

We also estimate less conservative regressions, excluding dyad-fixed effects. These regressions in addition include a binary variable indicating that for a specific project the

[140] The indicators for a common currency and a joint regional trade agreement are taken from Head et al. (2010), data on gross aid disbursements from the OECD (2017), and data on imports from the sponsor to the recipient from the IMF (2016). All monetary values are logged constant 2010 US$. See Appendix B3. for details. Note that the number of countries is slightly lower compared to the monadic regressions, due to missing data for control variables.

sponsor is identical to the recipient. We control for whether the sponsor and recipient share a border ("Neighbors"), a common official language or minority language, ethnicity, colonizer, or legal origin. We include a variable indicating a colonial relationship between the two (after 1945) and control for the population-weighted distance between them as well as their time difference, and for whether one of the two countries is a current or former hegemon of the other ("Recipient Hegemon" and "Sponsor Hegemon").[141] Finally, ε_{ijt} is the error term.

3.4. RESULTS

Table 3.1. shows the results for specification (1), at the recipient-country-year level. Columns 1-4 start with simple regressions that exclude the number of partners and the two interactions. While these regressions ignore the potential bias arising from countries with more partners also having a higher share of partners in positions of influence, they are appealing due to their comparably straightforward interpretation.

[141] Appendix B3. shows the definitions and sources for all variables, while Appendix C3. provides descriptive statistics. One might also think of controlling for recipient and sponsor country voting power at the IFC, their (log) real per capita GDP and (log) population size. These variables however are captured by the set of fixed effects we include.

Table 3.1. IFC Loans to Recipient Countries, 1995-2015, OLS

	(1) IFCEB	(2) UNSC	(3) IFCEB	(4) UNSC	(5) IFCEB	(6) UNSC	(7) IFCEB	(8) UNSC
IFCEB/UNSC	0.714	-0.339	0.364	0.581	1.260**	0.544	1.342**	0.626
	(1.60)	(0.61)	(0.68)	(1.33)	(2.28)	(0.86)	(2.57)	(1.01)
IFCEB/UNSC partner (share)	7.459***	7.224***	12.675***	11.881***	2.044***	1.406**	2.364***	1.665***
	(14.55)	(11.95)	(30.71)	(20.02)	(4.08)	(2.45)	(4.71)	(3.26)
Member*Share of partners					1.356	1.591*		
					(1.07)	(1.66)		
Member*Number of partners					-3.459***	-3.814***	-3.074***	-3.547***
					(4.64)	(5.68)	(4.03)	(5.11)
Number of partners					5.232***	5.690***	5.087***	5.623***
					(9.16)	(10.96)	(9.35)	(11.03)
GDP p.c. (log, t-1)	2.771**	2.396**	0.576	-0.136	2.688**	2.600**	2.672**	2.586**
	(2.36)	(2.11)	(1.05)	(0.25)	(2.39)	(2.33)	(2.38)	(2.32)
Population (log, t-1)	0.883	1.448	0.090	1.012	1.050	1.388	1.036	1.375
	(0.42)	(0.70)	(0.08)	(0.78)	(0.54)	(0.72)	(0.53)	(0.71)
Own loans included?	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Number of observations	3,145	3,145	3,145	3,145	3,145	3,145	3,145	3,145
Number of recipient countries	153	153	153	153	153	153	153	153
R-squared (within)	0.13	0.08	0.35	0.20	0.18	0.18	0.18	0.18

Notes: Robust t-values (clustered at the recipient-level) in parentheses. All regressions include country fixed effects. *** p<0.01, ** p<0.05, * p<0.1

According to the results, IFC loan size increases significantly with the share of a country's partners on the IFC's Board (column 1) or the UNSC (column 2), while there is no significant effect of a country's membership itself (though the coefficient in column 1 is only marginally insignificant).[142] Columns 3 and 4 exclude loans that are executed by a country's "own" companies. The results are similar, though the coefficients for the share of a country's partners in positions of influence increase substantially.

Columns 5 and 6 report the full specification (1). The results show that loan size increases with the number of partners a country has projects with and its (logged) per capita GDP, at least at the five-percent level of significance, but not with its population. Column 5 again focuses on the IFC's Board of Directors; column 6 shows the results for membership on the UNSC. As can be seen, the interactions between a country's own positions of influence and the share of influential partners are positive and sizable. While the coefficient is marginally insignificant for IFC Board membership, it is significant at the ten-percent level for UNSC membership.[143] In both regressions, holding a seat, the share of partners holding a seat, and the interaction between the two is jointly significant at the one-percent level. The coefficients of column 6 imply that an increase in the share of a country's partners that hold a seat on the UNSC by ten percentage points increases loan size by 35 percent at times the recipient country is represented on the UNSC itself, compared to an increase of 15 percent at times it is not (the average loan in the sample being US$ 27.4 million). According to column 5, the corresponding

142 Results are similar when we control for a country's number of partners, though the coefficients of a country's share of partners in positions of influence are half in size, as could be expected. Results are also similar in regressions with a binary dependent variable that indicates the presence of at least one project, as well as in regressions explaining the amount of loans conditional on receiving at least one project (i.e., in the first and second stage of a two-step model).

143 The extent to which countries vote the same as major shareholders of international organizations in the United Nations General Assembly (UNGA) is another widely used proxy for political influence (e.g., Humphrey and Michaelowa 2013). There is evidence that the United States government cares about UNGA voting and uses its foreign aid and power over international organizations to influence how countries vote. Andersen et al. (2006) show that countries voting more frequently in line with the United States receive substantially larger loans from the IDA. Kersting and Kilby (2016) provide evidence that World Bank disbursements accelerate during the run up to a competitive election if the recipient country government voted in line with the U.S. in the UNGA but decelerate if that government voted against the United States. Given that voting in the UNGA could be determined by omitted variables that also affect IFC lending we do not use it here. Note however that the average voting coincidence of a country's project partners in line with the United States (the major shareholder of the IFC) is significant at the one-percent level when it is included in the regressions of columns 5 and 6 instead of membership on the UNSC and the IFC's Board. A country's own voting pattern as well as an interaction between the two is not significant at conventional levels.

increases resulting from partner country representation on the IFC's Board are 40 percent and 23 percent.

According to the coefficients of both column 5 and column 6, loan size decreases with the total number of partners at times a country holds a position of influence, at the one-percent level of significance. It thus seems that countries that already have a large number of projects (proxied by the number of project partners) are less able to benefit from positions of power. The coefficients imply that membership on the UNSC or the Board of Directors only benefits those 17 percent of our sample countries that enter these positions of power while they exclusively receive projects they themselves sponsor (i.e., the company implementing the project is from the same country). Given that this holds true for the economically and politically more powerful countries in our sample, our results seem to indicate that more powerful countries benefit from entering positions of power, while less powerful countries lose out.[144] Potentially, these results can be explained with the possibility of weaker countries receiving substantial increases in loans from another part of the World Bank Group – the International Bank for Reconstruction and Development (Dreher et al. 2009b, Kaja and Werker 2010).[145]

In columns 7 and 8 we exclude the interaction between a country's influence and those of their partners. Column 7 focuses on the IFC's Board. The results show that loan size increases with the share of a country's partners that hold a seat at the Board, at the one-percent level of significance. The coefficient indicates that a one percentage point increase in the share of a country's partners that hold a seat on the Board of Directors increases loan size by 2.4 percent (the average share of a country's partners on the Board being 0.1 and the average loan size US$ 27.4 million). At the five-percent level, the point estimate indicates that a seat at the Board increases loan size by more than 280 percent (as long as a Board member only receives

[144] When we split the sample according to a country's vote share in the IFC's Board we indeed find a significant and positive effect of membership on the IFC's Board and the share of partners on the Board (but not the interaction between the two) for countries with a vote share above the median. Focusing on those below, the impact of a Board seat turns negative and marginally insignificant, while the share of partners on the Board and the interaction between the two are significantly positive. We have also run similar regressions for the UNSC using a country's share of senior positions in the United Nations Secretariat to proxy for countries' influence in the United Nations (Novosad and Werker 2018). We find however no significant interaction between the share of senior positions and UNSC membership.

[145] We intend to investigate the nexus between IFC and IBRD financing for countries holding positions of power in future work.

projects that are implemented by its own companies).[146] Again, the effect of membership on the Board on loan size decreases with the total number of a country's partners. A seat on the Board can increase loan size when the number of partners is zero or one, depending on the share of partners that also hold a position on the Board. For example, a seat amounts to an increase in loan size of almost 90 percent for a recipient country with one partner that also holds a seat on the Board. The effect turns negative when the number of partners exceeds one (which holds for less than one percent of the recipient-years in our sample).

Column 8 shows that members of the UNSC do not receive larger loans (though the coefficient is positive, the t-value indicates that it is not precisely estimated). The results also show that a larger share of a country's project partners on the UNSC increase loan size at the one-percent level of significance. According to the estimate, a one percentage point increase in the share of a country's partners that hold a seat on the Board of Directors increases loan size by 1.68 percent.

Overall, we conclude that positions of power matter. We find robust evidence that the share of a country's partners holding positions of influence increases loan size. We find weaker evidence that a country's own position of power increases loans; the same holds for the interaction between recipient and sponsor power.

Table 3.2. investigates the timing of membership (specification (2) above). Neither IFC Board membership nor UNSC membership are unanticipated.[147] We therefore test whether substantial changes in influence begin to occur in the years before a country enters a board or council position. As a placebo test, we also investigate loan size after membership has ended. When structural changes make a country more likely to enter influential positions in international fora and this influence is correlated with the number of IFC projects implemented in one country and executed by companies from another, it is unlikely that this change disappears at the time the country leaves its board or council position. To the extent that loans increase at times a country holds positions of influence but not thereafter and UNSC membership is uncorrelated with variables that typically determine lending by international organizations (as shown in Dreher et al. 2014), the increase during membership is arguably

[146] For comparison, note that Kaja and Werker (2010) show that a seat at the Board increases IBRD loans by roughly 300 percent.

[147] See Dreher et al. (2014), Vreeland and Dreher (2014), and Lai and Lefler (2017) for evidence on UNSC elections.

the consequence of membership itself, rather than those of structural changes more broadly. We follow Kuziemko and Werker (2006) and Kaja and Werker (2010) and include binary indicators for the two years before and after a country assumes membership on the UNSC and the IFC Board.[148]

Table 3.2. IFC Loans to Recipient Countries, Event-Time Specification, 1995-2015, OLS

	(1) IFCEB	(2) UNSC	(3) UNSC	(4) UNSC
IFCEB/UNSC (t-2)	0.826	-0.258	-0.266	-0.627
	(0.93)	(0.36)	(0.37)	(0.85)
IFCEB/UNSC (t-1)	0.790	0.067	0.050	0.099
	(0.99)	(0.10)	(0.07)	(0.13)
IFCEB/UNSC (t)	1.395*	0.750		0.802
	(1.97)	(1.22)		(1.27)
UNSC first year (t)			1.335	
			(1.62)	
UNSC second year (t)			0.112	
			(0.14)	
IFCEB/UNSC (t+1)	0.282	0.340	0.338	0.370
	(0.28)	(0.47)	(0.47)	(0.47)
IFCEB/UNSC (t+2)	-0.675	0.100	0.086	-0.347
	(0.50)	(0.11)	(0.10)	(0.38)
UNSC (t+3)				1.202
				(1.52)
IFCEB/UNSC partner (share, t-2)	-0.436	0.428	0.393	0.572
	(0.64)	(0.56)	(0.51)	(0.72)
IFCEB/UNSC partner (share, t-1)	0.869	1.436**	1.424**	1.253*
	(1.24)	(2.30)	(2.27)	(1.92)
IFCEB/UNSC partner (share, t)	2.390***	2.301***		2.260***
	(3.31)	(3.77)		(3.41)
UNSC partner first year (share, t)			2.583**	
			(2.42)	
UNSC partner second year (share,			-0.451	
			(0.43)	
IFCEB/UNSC partner (share, t+1)	0.325	1.533**	1.534**	1.333*
	(0.44)	(2.19)	(2.21)	(1.75)
IFCEB/UNSC partner (share, t+2)	-0.065	1.425	1.371	1.546
	(0.09)	(1.53)	(1.47)	(1.54)
UNSC partner (share, t+3)				0.214
				(0.25)
Number of partners (t-3)				0.206
				(0.75)
Number of partners (t-2)	0.501*	0.231	0.240	0.170
	(1.78)	(0.82)	(0.85)	(0.61)
Number of partners (t-1)	1.108***	0.983***	0.988***	1.025***
	(3.71)	(3.41)	(3.44)	(3.23)
Number of partners (t)	5.077***	5.517***	5.552***	5.516***
	(9.26)	(11.58)	(11.48)	(11.95)
Number of partners (t+1)	-0.017	0.037	0.024	0.084
	(0.07)	(0.16)	(0.11)	(0.35)
Number of partners (t+2)	0.206	-0.050	-0.053	-0.086
	(0.78)	(0.19)	(0.20)	(0.33)

[148] We continue to include the interaction of the number of a country's partners with positions of influence, but do not introduce additional interactions between the number of partners in past and future years. Our results are overall unchanged when we (i) exclude the interaction or (ii) include interactions with the number of partners in previous and future years.

Number of partners (t+3)				-0.043
				(0.16)
IFCEB/UNSC * Number of	-3.341***	-3.666***	-3.570***	-3.875***
	(4.19)	(5.42)	(4.88)	(5.51)
GDP p.c. (log, t-1)	2.259**	2.056*	2.056*	2.174*
	(2.00)	(1.85)	(1.85)	(1.90)
Population (log, t-1)	0.081	0.341	0.333	-0.667
	(0.04)	(0.16)	(0.16)	(0.30)
Number of observations	2,843	2,843	2,843	2,554
Number of recipient countries	153	153	153	153
R-squared (within)	0.19	0.19	0.19	0.18

Notes: Robust t-values (clustered at the recipient-level) in parentheses. All regressions include country fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Column 1 shows that neither membership on the IFC's Board nor the share of project partners on the Board affects the size of previous or future loans, at conventional levels of significance. Compared to column 5 of Table 3.1. – excluding the leads and lags – results for our variables of interest are basically unchanged. Column 2 shows similar results for the UNSC. Note however that loan size increases with the share of project partners on the UNSC in *t*, and both in the year before and the year after a country's temporary membership. Given that most members announce their intention to run for election to the UNSC years in advance and the election takes place in September of the year before tenure begins, we are not surprised by the significance of the share of a country's partners on the UNSC on loan sizes in the election year. In fact, the logic of diffuse reciprocity outlined in Vreeland and Dreher (2014) for the IMF and the World Bank makes us expect the effect of political importance to be visible as soon as the election to the UNSC becomes likely (or even, in September, certain). To the extent that formal commitments are given with some delay, political power in one year could also explain increases in loans in the year after holding the position of influence. Note that loan size is unaffected two years before and two years after membership. The coefficients associated with the share of a country's partners on the UNSC increases from two years before to one year before; is largest in the year of membership, and then declines. Overall, this pattern is in line with our expectations.

Column 3 separates the first from the second year of membership on the UNSC.[149] Results are basically unchanged. Note however that we observe the increase in loans resulting from a higher share of a country's partners being on the UNSC during the first year of

[149] We do not run this regression for the IFC's Board, given that tenure is not limited to a fixed number of years (among the non-permanent members the average number of years on the Board is 4.6 with a maximum of 21 years).

membership, a result that is again in line with the logic of diffuse reciprocity. In column 4 we add an additional lag, given that a skeptical reader might remain unconvinced by previous specifications, showing the second year of the share of a country's partners on the UNSC to be just marginally insignificant, and positive in sign. As can be seen, the coefficient of the third lag is completely insignificant, and its introduction does not change any of the results. We thus conclude that the timing of effects is overall in line with our hypotheses.

We next turn to the sponsor- rather than recipient-country (specification (3) above). Columns 1 and 2 of Table 3.3. show the full specifications. We then subsequently exclude the interaction between the sponsor country's position on the IFC or the UNSC with the share of its partners holding these positions (columns 3 and 4) and the number of the country's partners and its interaction with the share of partners in positions of influence (columns 5 and 6).

The results are overall similar to those at the recipient level. There is a strong and highly significant effect of the number of a sponsor-country's partners on the IFC Board and the UNSC on the volume of IFC lending the country receives. A country's presence on the Board also results in larger loans (at the one percent level), while the interaction between the country's own presence and those of a larger number of partners on the Board is again only marginally significant. A country's own presence on the UNSC and its interaction with the number of partners on the UNSC do not affect loan size, at conventional levels of significance. According to column 3, a one percentage point increase in the share of a country's partners that hold a seat on the Board of Directors increases loan size by 4.12 percent for countries that only receive projects that are implemented by its own companies, while a seat on the Board is rewarded with an increase of 267 percent. A one percentage point increase in the share of a country's partners on the UNSC increases IFC loans by 4.16 percent (column 4).

Columns 7 and 8 show that the increase in loans starts two years before membership, and holds in the first year a country's partners have left their positions of influence. There is no significant effect in the second year after the end of membership. Again, this pattern is broadly in line with our expectations.

Table 3.3. IFC Loans to Sponsor Countries, 1995-2015, OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IFCEB	UNSC	IFCEB	UNSC	IFCEB	UNSC	IFCEB	UNSC
IFCEB/UNSC (t-2)							1.451*	1.153*
							(1.80)	(1.73)
IFCEB/UNSC (t-1)							1.467	0.425
							(1.65)	(0.57)
IFCEB/UNSC (t)	1.242***	0.235	1.299***	0.228	0.797*	-0.539	2.119***	0.492
	(2.61)	(0.43)	(2.70)	(0.41)	(1.80)	(1.05)	(2.82)	(0.78)
IFCEB/UNSC (t+1)							0.787	0.417
							(0.86)	(0.51)
IFCEB/UNSC (t+2)							-0.768	0.688
							(0.71)	(0.93)
IFCEB/UNSC partner (share, t-2)							3.130***	3.589***
							(3.43)	(4.31)
IFCEB/UNSC partner (share, t-1)							1.876*	1.161
							(1.92)	(1.15)
IFCEB/UNSC partner (share)	2.736**	4.196***	4.037***	4.073***	8.720***	8.075***	5.525***	5.250***
	(2.46)	(4.81)	(5.73)	(5.89)	(11.19)	(8.84)	(7.02)	(6.64)
IFCEB/UNSC partner (share,							1.824**	2.333**
							(1.98)	(2.37)
IFCEB/UNSC partner (share,							0.996	0.801
							(0.82)	(0.71)
Member*Share of partners	2.518*	-0.488						
	(1.89)	(0.30)						
Number of partners (t-2)							-0.231	-0.308
							(0.88)	(1.29)
Number of partners (t-1)							-0.150	-0.260
							(0.71)	(1.40)
Number of partners (t)	7.447***	5.297***	7.181***	5.311***			7.172***	5.366***
	(5.68)	(4.88)	(5.79)	(4.98)			(6.04)	(5.15)
Number of partners (t+1)							0.243	0.325
							(1.01)	(1.34)
Number of partners (t+2)							-0.132	-0.130
							(0.45)	(0.50)
Member*Number of partners (t)	-4.702***	-2.974**	-4.367***	-3.010**			-4.302***	-2.965**
	(3.17)	(2.26)	(3.10)	(2.41)			(3.11)	(2.43)

GDP p.c. (log, t-1)	3.039***	3.156***	3.045***	3.159***	3.427***	3.556***	2.284**	2.401**
	(2.64)	(2.68)	(2.64)	(2.68)	(2.70)	(2.77)	(2.06)	(2.10)
Population (log, t-1)	2.077	1.871	2.117	1.871	2.768	2.652	0.217	-0.053
	(0.99)	(0.86)	(1.01)	(0.86)	(1.16)	(1.10)	(0.10)	(0.02)
Number of observations	3,746	3,746	3,746	3,746	3,746	3,746	3,386	3,386
Number of sponsor countries	182	182	182	182	182	182	182	182
R-squared (within)	0.153	0.141	0.153	0.141	0.08	0.07	0.161	0.148

Notes: Robust t-values (clustered at the sponsor-level) in parentheses. All regressions include country fixed effects. *** p<0.01, ** p<0.05, * p<0.

Table 3.4. IFC Projects and Mean Investments at the Sponsor-Recipient-Dyad, 1995-2015, OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Projects				Mean investment (log)			
	IFCEB	UNSC	IFCEB	UNSC	IFCEB	UNSC	IFCEB	UNSC
Sponsor and recipient on	0.001	0.011	0.069***	0.114**	0.026	0.025	0.280***	0.460***
	(0.42)	(1.06)	(2.80)	(2.56)	(1.18)	(0.60)	(4.77)	(4.06)
Aid disbursements (log, t-1)	0.000**	0.000**	0.000	0.000	0.004***	0.004***	0.004***	0.005***
	(2.32)	(2.29)	(0.35)	(0.54)	(3.75)	(3.75)	(5.53)	(5.68)
Imports from sponsor (log, t-1)	-0.000**	-0.000**	-0.002***	-0.002***	-0.001***	-0.001***	-0.006***	-0.006***
	(2.23)	(2.23)	(3.56)	(3.55)	(3.10)	(3.12)	(5.82)	(5.83)
Regional trade agreement (t-1)	-0.001	-0.001	0.005*	0.005*	-0.017	-0.017	0.023**	0.024**
	(0.87)	(0.87)	(1.85)	(1.87)	(0.95)	(0.94)	(2.26)	(2.31)
Common currency (t-1)	0.002*	0.002	0.005**	0.006**	0.014	0.011	0.006	0.008
	(1.75)	(1.26)	(2.00)	(2.12)	(1.24)	(1.02)	(0.42)	(0.60)
Recipient and sponsor identical			0.621***	0.618***			4.460***	4.451***
			(6.63)	(6.63)			(12.36)	(12.36)
Neighbours			0.002	0.002			0.052***	0.054***
			(0.79)	(1.07)			(2.84)	(2.98)
Common language			0.007***	0.007***			0.040***	0.039***
			(3.71)	(3.71)			(3.76)	(3.74)
Common ethnicity			-0.001	-0.001			0.003	0.003
			(0.33)	(0.34)			(0.26)	(0.25)
Common colonizer			0.005*	0.005*			0.035***	0.035***
			(1.84)	(1.88)			(3.51)	(3.53)
Colonial relation			0.004	0.005			-0.011	-0.005
			(0.28)	(0.42)			(0.12)	(0.06)
Common legal origin			-0.001	-0.001			-0.014***	-0.014***
			(0.91)	(0.85)			(3.28)	(3.22)
Distance			-0.000	-0.000			-0.000***	-0.000***
			(0.42)	(0.34)			(2.69)	(2.63)
Time difference			-0.001	-0.001			0.002	0.002

			(1.02)	(1.08)			(0.76)	(0.69)
Recipient hegemon			-0.017	-0.019			0.002	-0.005
			(0.96)	(1.05)			(0.01)	(0.04)
Sponsor hegemon			0.005	0.004			0.165**	0.159**
			(0.58)	(0.39)			(2.32)	(2.21)
Number of observations	570,706	570,706	570,664	570,664	570,706	570,706	570,664	570,664
Number of sponsor countries	180	180	180	180	180	180	180	180
Number of recipient countries	151	151	151	151	151	151	151	151
Dyad-fixed effects	yes	yes	no	no	yes	yes	no	no
R-squared (within)	0.00	0.00	0.14	0.14	0.00	0.00	0.18	0.18

Notes: Robust t-values (clustered at the dyad-level) in parentheses. All regressions include sponsor-year and recipient-year fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table 3.4. investigates the interaction between recipient and sponsor countries in more detail, focusing on potential mechanisms behind the (marginally significant) results in Tables 3.1. and 3.3. We test whether the increase in loan size resulting from joint positions of power can be explained by an increase in the number of projects or larger mean loan sizes between countries holding joint positions of influence. We therefore turn to the dyadic recipient-sponsor-year level (specification 4 above).

We focus on the presence of a recipient and sponsor country on the UNSC or the Board of Directors at the same point in time, netting out as much variation as possible. Controlled for dyad-specific fixed effects, recipient-year and sponsor-year fixed effects, and the variables varying at the recipient-sponsor-year level, we do not find a significant effect of joint sponsor- and recipient-country presence on either the Board of the IFC or the UNSC on the number of new IFC projects (columns 1 and 2).

Columns 3 and 4 exclude the dyad-fixed effects (and include the additional control variables that do not vary at the recipient-sponsor-level). At the one-percent level, the number of projects increases between countries in years they both hold a seat at the IFC's Board of Directors (column 3). According to column 4, the same holds in years the recipient and sponsor countries share membership on the UNSC. The coefficients show that these effects are of moderate size. In years of joint membership on the Board (UNSC), 0.07 (0.11) additional projects are initiated between the two countries, according to our estimates.[150]

Columns 5-8 replicate the regressions focusing on (log) mean loan size rather than the number of projects, with similar results. With the inclusion of the dyad-fixed effects, coefficients are not significant at conventional levels (columns 5 and 6). When we exclude them, we find that the average loan between two countries with joint membership on the Board (UNSC) is 32 percent (58 percent) larger compared to loans for countries without joint representation.

Overall, these results are in line with our hypotheses. The strictest specifications including dyad-fixed effects do not allow us to identify significant effects. This might potentially be because countries that share positions of power are different from those that do

[150] We also estimated these regressions with sponsor- and recipient-country fixed effects and year fixed effects instead, including $Member_{it}$ and $Member_{jt}$ rather than their interaction. Both coefficients are completely insignificant for UNSC membership and sponsors being IFC Board members. IFC Board membership of recipient countries is marginally significant with the expected positive coefficient.

not at any point in time. Given that at least for the UNSC the timing of a country holding power – let alone the timing of two countries holding power at the same time – is random (Dreher et al. 2014, Mikulaschek 2017) we consider this unlikely. However, in order to increase the readers' confidence that dyad-specific, time varying factors that might be correlated with the number and mean size of IFC loans and joint membership on the Board do not bias our results, we further investigate these findings in Table 3.5.

Column 1 of Table 3.5. again focuses on the timing of IFC Board membership. Specifically, we include binary indicator variables for the four years before and the four years after joint membership on the Board. When IFC-supported countries are different at any point in time we would expect these leads and lags to be significant as well. To the extent that the coefficients shown in Table 3.4. represent the causal effect of joint membership, we expect the deeper lags to be insignificant. Increases directly before assuming positions of power could be attributed to diffuse reciprocity; increases directly after to delays between a company's application of a project and the actual commitment.[151]

The results show significant coefficients for the years of membership as well as the year directly before and after leaving the Board. We interpret the insignificance of the deeper leads and lags as evidence in support of our hypothesis.

Column 2 turns to the UNSC. Rather than focusing on the full sample, we restrict our regressions to countries from the African continent. This is because the African seats on the UNSC follow a pattern closest to rotation, so that the timing of membership is most plausibly exogenous (Dreher et al. 2014, Vreeland and Dreher 2014).[152] Our results are similar to those for the full sample shown in column 4 of Table 3.4. Given the exogenous timing of UNSC membership of any specific country in Africa, the joint membership of any two countries is exogenous as well.

---

[151] The IFC's project cycle includes seven stages prior to commitment: business development, early review, appraisal, investment review, negotiations, public notification, and Board review and approval. See http://www.ifc.org/wps/wcm/connect/corp_ext_content/ifc_external_corporate_site/solutions/ifc-project-
cycle (accessed July 10, 2017).

[152] Note that Africa has three seats on the UNSC. There are 114 country-year observations of joint membership in our sample. Aid disbursements are excluded in this regression as no aid is disbursed from African sponsors to African recipient countries.

Table 3.5. IFC Projects and Mean Investment at the Sponsor-Recipient-Dyad, Event-Time-Specification, 1995-2015, OLS

	(1)	(2)	(3)	(4)	(5)	(6)
		Projects			Mean investment (log)	
	IFCEB	UNSC (exogenous)	UNSC (exogenous)	IFCEB	UNSC (exogenous)	UNSC (exogenous)
Sponsor and recipient on IFCEB/UNSC	0.069***	0.107**	0.119**	0.280***	0.817**	0.911**
	(2.80)	(2.52)	(2.56)	(4.75)	(2.24)	(2.34)
Recipient and sponsor identical	0.614***	0.306***	0.268***	4.423***	2.978***	2.687***
	(6.60)	(5.33)	(5.79)	(12.44)	(6.80)	(6.89)
Aid disbursements (log, t-1)	0.000			0.004***		
	(0.29)			(5.46)		
Imports from sponsor (log, t-1)	-0.002***	-0.002***	-0.002***	-0.006***	-0.015***	-0.015***
	(3.54)	(3.31)	(3.46)	(5.84)	(3.55)	(3.70)
Regional trade agreement (t-1)	0.005*	-0.003	-0.001	0.023**	-0.025	-0.011
	(1.81)	(0.74)	(0.29)	(2.23)	(0.68)	(0.32)
Common currency (t-1)	0.005**	0.010	0.011*	0.006	0.049	0.063
	(1.98)	(1.56)	(1.75)	(0.41)	(1.04)	(1.35)
Neighbours	0.002	0.010**	0.010**	0.052***	0.092***	0.092***
	(0.80)	(2.50)	(2.52)	(2.84)	(2.75)	(2.75)
Common language	0.007***	0.002	0.002	0.040***	0.003	0.005
	(3.73)	(0.41)	(0.50)	(3.77)	(0.07)	(0.15)
Common ethnicity	-0.000	-0.007	-0.006	0.003	-0.055	-0.050
	(0.29)	(1.42)	(1.35)	(0.29)	(1.34)	(1.28)
Common colonizer	0.004*	0.002	0.001	0.035***	0.038	0.030
	(1.84)	(0.33)	(0.15)	(3.53)	(0.88)	(0.75)
Colonial relation	0.003	-0.039	-0.036	-0.012	-0.218	-0.193
	(0.28)	(1.03)	(1.01)	(0.13)	(0.69)	(0.65)
Common legal origin	-0.001	0.007	0.006	-0.015***	0.044	0.043
	(1.00)	(1.57)	(1.59)	(3.36)	(1.43)	(1.42)
Distance	-0.000	-0.000	-0.000	-0.000***	-0.000	-0.000
	(0.41)	(1.14)	(0.94)	(2.68)	(1.46)	(1.33)

Time difference	-0.001	0.001	0.001	0.002	0.016	0.013
	(1.05)	(0.47)	(0.36)	(0.72)	(0.64)	(0.55)
Recipient hegemon	-0.017	-0.008	-0.010	0.001	-0.111	-0.129
	(0.96)	(0.56)	(0.74)	(0.01)	(0.84)	(0.98)
Sponsor hegemon	0.005	-0.002	-0.005	0.166**	-0.010	-0.027
	(0.60)	(0.16)	(0.31)	(2.32)	(0.07)	(0.19)
Sponsor and recipient on IFCEB/UNSC (t-1)	0.118**		0.201**	0.471**		1.643**
	(1.99)		(2.27)	(2.51)		(2.24)
Sponsor and recipient on IFCEB/UNSC (t-2)	0.094		0.144	0.462*		1.335
	(1.32)		(1.49)	(1.76)		(1.61)
Sponsor and recipient on IFCEB/UNSC (t-3)	0.021		0.275	0.222		2.020**
	(0.35)		(1.62)	(0.75)		(2.10)
Sponsor and recipient on IFCEB/UNSC (t-4)	-0.054		0.305**	0.013		2.627**
	(1.05)		(2.22)	(0.04)		(2.45)
Sponsor and recipient on IFCEB/UNSC	0.134*		0.177*	0.613***		1.508*
	(1.83)		(1.79)	(2.64)		(1.90)
Sponsor and recipient on IFCEB/UNSC	0.067		0.253*	0.308		1.367*
	(0.72)		(1.71)	(0.96)		(1.85)
Sponsor and recipient on IFCEB/UNSC	0.046		0.143	0.082		1.014
	(0.73)		(1.35)	(0.29)		(1.43)
Sponsor and recipient on IFCEB/UNSC	0.077		-0.015	0.697		-0.184
	(0.86)		(0.25)	(1.25)		(0.35)
Number of observations	570,664	58,989	58,989	570,664	58,989	58,989
Number of sponsor countries	180	53	53	180	53	53
Number of recipient countries	151	53	53	151	53	53
Dyad-fixed effects	no	no	no	no	no	no
R-squared (within)	0.150	0.188	0.203	0.190	0.220	0.231

Notes: Robust t-values (clustered at the dyad-level) in parentheses. All regressions include sponsor-year and recipient-year fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Column 3 adds the leads and lags, in analogy to those of column 1.[153] The results turn out to be similar to those of column 1. Joint membership on the UNSC increases the number of projects, at the five-percent level of significance. The same holds for the year of election to the UNSC, and the first and second year after UNSC membership (at the ten-percent level). However, the number of projects increases four years prior to membership as well, which might be due to the logic of diffuse reciprocity, omitted variables that we fail to control for, or random chance. Columns 4-6 replicate the regressions for (log) mean IFC loan size. Results are similar, though project size increases in the third year prior to membership as well.

Overall, our dyadic results are in line with the hypothesis that recipient and sponsor countries can exploit positions of joint influence to their benefit. At times they share a seat on the IFC's Board or the UNSC they receive more and larger projects. This increase occurs in the years prior to joining these organizations, but fades out afterwards. This pattern does not fit a potential alternative explanation for our main results, according to which the countries that are represented on the Board of the IFC or the UNSC might increase collaboration between themselves more broadly. If collaboration is the prime explanation, we would expect a different pattern: the number of IFC projects would only increase at the time of membership and would be likely to continue thereafter. We find this not to be the case. To the contrary, the patterns shown by the data are well in line with the logic of diffuse reciprocity: As soon as it is known that countries will assume positions of influence in future, influential countries hold them in their debt by granting favors they would not receive at other times. Importantly, such favors could be granted years ahead of actual membership on the UNSC or the IFC's Board of Directors. On average, we find a pattern that fits such a mechanism in the two years preceding membership.

3.5. CONCLUSIONS

Two thirds of the IFC's investments go to companies from the world's richest countries, while only one fifth go to companies of the poorest ones (Ellmers et al. 2010). What is more, the bulk of the IFC's investments benefits projects implemented in middle-income countries, rather

[153] We also estimated this regression for the full sample. While the interaction between recipient and sponsor membership on the UNSC is significant at the one-percent level, the regressions excluding dyad-fixed effects do not seem to sufficiently control for omitted variables, so that leads and lags turn out to be significant as well.

than poorer ones (Ellmers et al. 2010). The majority of its projects do not serve obvious development goals.

In this chapter, we argued that coalitions between middle-income countries receiving the bulk of the IFC's projects and countries representing the companies that receive the loans influence IFC decisions in their favor. Our results based on more than 3,000 IFC projects over the 1995-2015 period show that during the years that countries are jointly represented at the IFC Board or the UNSC they indeed attract significantly more and larger IFC projects. Recipient and sponsor countries receive larger overall lending as well, at times they and – most robustly – a larger share of their project partners are in positions of influence. These results are in line with the pattern of IFC lending at large.[154]

IFC lending distorts markets in two ways. First, the IFC competes with alternative types of financing. Due to exemptions on dividends and corporation tax, it can offer better deals compared to private competitors. Coming back to the Mövenpick example from the introduction, we find it hard to believe that no private lender would have been willing to finance the project, even with the ongoing global financial crisis. Second, IFC-funded investors receive an unfair advantage over their competitors. The subsidized loan to Mövenpick, for example, gives it an edge over other hotels in Accra, who have access to capital on inferior terms.

Additionally, the IFC's lending is unlikely to result in the most efficient allocation of funds to achieve its mandate of "ending of extreme poverty." According to the results in Dreher et al. (2018) political motives in granting foreign aid reduce the effectiveness of the aid in promoting economic growth. To the extent that IFC lending is driven by the interests of powerful companies rather than the interests of the poor, we expect these loans to be less successful in reducing poverty.[155]

[154] Note that we do not claim to explain the IFC's lending pattern in total. Like other international bureaucracies, IFC management and staff face incentives to expand their mandate if they can (Vaubel et al. 2007). What is more, the IFC's profits generate funds for other branches of the World Bank Group. IFC staff thus has incentives to increase the volume of lending whenever there is demand for their loans. Loans to companies from richer countries are typically larger and repayment is more secure, arguably resulting in larger loans to richer countries.

[155] We would like to test this proposition making use of the World Bank's Independent Evaluation Group data. Unlike for the IBRD and IDA these data are however not available to the public at the project level. Preliminary tests with country-level data show no significant effect of lagged membership on the Board or UNSC on the share of projects with a positive evaluation. Interestingly, there is some evidence that evaluations improve with countries' contemporaneous positions of power, however. The benefits of these positions could thus extend to the project evaluation stage as well.

Our results also speak to the policy debate on leveraging private funding for development. The 2015 Addis Ababa Action Agenda and Germany's recent "Compact with Africa" are just two of the recent examples for development initiatives highlighting the need to combine private and official efforts in financing sustainable development. The World Bank's President envisions the Bank to become a broker that leverages private capital rather than mainly providing its own funds.[156] To the extent that the allocation of these funds can be shaped by the politically powerful, they might instead be captured to finance projects that benefit powerful middle-income countries and the companies that implement the projects. The additional funds might then be ineffective in achieving development goals, just as political considerations in allocating official aid make the aid less effective in raising growth.

Fundamental reforms would be required. However, the interests represented by the IFC's major shareholders hardly focus on poverty reduction. The voting power of the world's poorest countries, where companies' access to private capital is scarcest, is minuscule. So is the political influence of companies that implement projects with mainly developmental aims in mind compared to largely commercially oriented multinational conglomerates. We therefore expect that – given the incentives of the political coalitions involved and these countries' voting power in the IFC – reforming the IFC will be all but impossible. Those who would gain from a more development-oriented policy are politically weak.

The limited influence of emerging powers in the IFC and the reluctance of major shareholders to agree to substantial reforms have contributed to China initiating the Asian Infrastructure Investment Bank (AIIB), which also has a mandate to lend to the private sector. Its governing structure is however similar to the IFC's, with mainly rich countries being represented on its Board of Directors. Investigating the influence of private companies and coalitions between countries representing the investor and those receiving the investment projects funded by other international organizations is an important question that we leave to future research.

[156] Again, see the reference given in footnote 114.

APPENDIX FOR CHAPTER 3

Appendix A3. Loan Commitments and Projects Over Sponsor and Recipient Countries, 1995-2015

Sponsor country	Share in total investment	Share in total projects	Recipient country	Share in total investments	Share in total projects
United States	7.80	4.78	India	8.50	7.88
Brazil	6.97	4.69	Brazil	8.17	5.04
India	6.76	7.08	Russia	7.28	5.83
Turkey	6.38	5.12	Turkey	6.76	4.76
China	5.40	4.75	China	6.68	6.02
Russia	3.81	3.60	Mexico	4.76	3.53
France	3.47	2.08	Indonesia	3.35	2.36
Mexico	3.08	2.86	Argentina	2.77	2.49
United Kingdom	2.93	1.99	Colombia	2.66	1.80
Indonesia	2.75	2.17	Philippines	2.55	1.73
Argentina	2.25	2.05	Ukraine	2.16	1.89
Colombia	2.20	1.65	Egypt	2.01	1.95
Philippines	1.87	1.37	Peru	1.97	1.70
South Africa	1.70	1.77	Nigeria	1.76	2.17
Germany	1.68	2.05	Romania	1.51	1.42
Nigeria	1.50	1.96	South Africa	1.45	1.80
Chile	1.48	0.96	Thailand	1.44	0.72
Ukraine	1.39	1.21	Pakistan	1.36	1.48
Netherlands	1.33	0.87	Chile	1.33	0.85
Peru	1.26	1.46	Ghana	1.26	1.17
Pakistan	1.09	1.18	Kazakhstan	1.11	0.91
Italy	1.08	0.62	Kenya	1.10	1.70
Thailand	1.08	0.59	Bangladesh	1.02	1.04
Austria	1.05	0.65	Jordan	0.99	1.26
Singapore	1.04	0.71	Korea, Rep.	0.91	0.54
Romania	1.02	0.93	Vietnam	0.84	1.17
Kenya	1.01	1.55	Croatia	0.81	0.66
Lebanon	0.94	1.06	Bulgaria	0.79	0.82
Egypt	0.88	1.15	Morocco	0.76	0.60
Switzerland	0.88	0.78	Panama	0.76	0.54

Notes: 30 largest recipient and sponsor countries in terms of their share in total IFC commitments.

Appendix B3. Definitions and Sources

Variable	Description	Data Source
Total investment (log)	Log of total IFC investments approved by the IFC Executive Board per country or country-pair (constant 2010 US$)	Own calculations based on IFC (2017)
Mean investment (log)	Log of mean IFC investments per project approved by the IFC Executive Board per country or country-pair (constant 2010 US$)	Own calculations based on IFC (2017)
Projects	Number of IFC projects approved per country or country-pair	Own calculations based on IFC (2017)
IFCEB	1 for IFC Executive Board (IFCEB) membership	Own construction based on World Bank Annual Reports 1995-2015
UNSC	1 for UNSC membership	Dreher et al. (2009a)
IFCEB partner (share)	Share of a country's project partners that are represented on the IFC Executive Board (IFCEB)	Own calculations based on IFC (2017)
UNSC partner (share)	Share of a country's project partners that hold positions as members of the UNSC	Own calculations based on IFC (2017)
Number of partners	Total number of a country's IFC project partner countries	Own calculations based on IFC (2017)
Recipient and sponsor identical	1 if the approved IFC project is sponsored and received by the same country	Own calculations based on IFC (2017)
Aid disbursements (log)	Log of total gross ODA disbursements (constant 2010 US$).	OECD (2017), Table DAC2a
Colonial relation	1 for pairs in colonial relationship post 1945	Head et al. (2010)
Common colonizer	1 for common colonizer post 1945	Head et al. (2010)
Common currency	1 for common currency	Head et al. (2010)
Common ethnicity	1 if a language is spoken by at least 9 percent of the population in both countries	Head et al. (2010)
Common language	1 for common official primary language	Head et al. (2010)
Common legal origin	1 for common legal origin	Head et al. (2010)
Distance	Weighted distance (population-weight, km)	Head et al. (2010)
GDP p.c. (log)	Log of GDP per capita (constant 2010 US$)	WDI (World Bank 2016)
Imports from sponsor (log)	Log of value of imports from sponsor to recipient country, defined as goods, cost, insurance, freight (CIF) (constant 2010 US$)	IMF (2016)
Neighbours	1 for contiguity	Head et al. (2010)
Population (log)	Log of population	WDI (World Bank 2016)
Recipient hegemon	Recipient is current or former hegemon of sponsor	Head et al. (2010)

Regional trade agreement	1 for joint regional trade agreement in force	Head et al. (2010)
Sponsor hegemon	Sponsor is current or former hegemon of recipient	Head et al. (2010)
Time difference	Number of hours difference between sponsor and recipient country	Head et al. (2010)

Appendix C3. Descriptive Statistics

Variables Tables 3.1. and 3.2.	Mean	Sd	Min	Max
Total investment (log)	6.49	8.32	0	20.71
IFCEB	0.10	0.30	0	1
IFCEB partner (share)	0.10	0.29	0	1
UNSC	0.07	0.25	0	1
UNSC partner (share)	0.05	0.22	0	1
Number of partners	0.19	0.53	0	6
GDP p.c. (log)	7.88	1.22	4.75	11.10
Population (log)	15.60	2.07	9.76	21.04

Variables Table 3.3.	Mean	Sd	Min	Max
Total investment (log)	6.05	8.22	0	21.59
IFCEB	0.16	0.36	0	1
IFCEB partner (share)	0.03	0.15	0	1
UNSC	0.08	0.27	0	1
UNSC partner (share)	0.02	0.13	0	1
Number of partners	0.16	0.59	0	8
GDP p.c. (log)	8.31	1.51	4.75	11.62
Population (log)	15.63	2.01	9.76	21.04

Variables Table 3.4.	Mean	Sd	Min	Max
Projects	0.004	0.13	0	22
Mean investment (log)	0.04	0.84	0	19.56
IFCEB Sponsor	0.16	0.36	0	1
IFCEB Recipient	0.10	0.30	0	1
UNSC Sponsor	0.08	0.27	0	1
UNSC Recipient	0.07	0.25	0	1
Recipient and sponsor identical	0.01	0.07	0	1
Aid disbursements (log)	1.51	4.44	0	23.24
Imports from sponsor (log)	7.60	7.91	0	26.83
Neighbours	0.02	0.15	0	1
Common official language	0.17	0.37	0	1
Common ethnology	0.15	0.36	0	1
Common colonizer	0.12	0.32	0	1
Colonial relation	0.00	0.07	0	1
Distance	7946.36	4503.73	0	19650.13
Time difference	4.60	3.38	0	12
Recipient hegemon	0.00	0.04	0	1
Sponsor hegemon	0.01	0.08	0	1
Regional trade agreement	0.06	0.23	0	1
Common legal origin	0.33	0.47	0	1
Common currency	0.01	0.12	0	1

Variables Table 3.5.	Mean	Sd	Min	Max
Projects	0.01	0.12	0	7
Mean investment (log)	0.07	1.06	0	18.86
UNSC Sponsor	0.05	0.23	0	1
UNSC Recipient	0.05	0.23	0	1
Recipient and sponsor identical	0.02	0.14	0	1
Imports from sponsor (log)	5.85	6.95	0	22.26
Neighbours	0.09	0.29	0	1
Common official language	0.45	0.50	0	1
Common ethnology	0.33	0.47	0	1
Common colonizer	0.27	0.45	0	1
Colonial relation	0.00	0.03	0	1
Distance	3587.52	1945.68	0	9772.06
Time difference	1.36	1.05	0	5
Recipient hegemon	0.00	0.03	0	1
Sponsor hegemon	0.00	0.03	0	1
Regional trade agreement	0.16	0.36	0	1
Common legal origin	0.55	0.50	0	1
Common currency	0.09	0.28	0	1

CHAPTER 4

What Makes a Successful Development Intervention?

The Theory of Planned Behavior – An Application to Implementation Research

Co-authored with Lennart Kaplan, Jana Kuhnt and Sebastian Vollmer

4.1. INTRODUCTION

"Wouldn't economics make a lot more sense
if it were based on how people actually behave,
instead of how they should behave?"
— Dan Ariely,
Predictably Irrational: The Hidden Forces That Shape Our Decisions

A large focus in the literature studying development co-operation naturally lies on its effectiveness. On the macroeconomic cross-country level, the effectiveness of aid is studied to an impressive extent, while results are still inconclusive (Burnside and Dollar 2000; Easterly, Levine, and Roodman 2004). In focus of the literature typically stand donor characteristics (Berthélemy 2006; Minasyan, Nunnenkamp, and Richert 2017), recipient characteristics (e.g., Rajan and Subramanian 2008; The World Bank 1998), or certain types of development assistance (e.g, Clemens, Radelet, and Bhavnani 2012; Dreher, Nunnenkamp, and Thiele 2008; Roodman 2015). Much less attention is drawn to the specific implementation features of development interventions, which might likewise and very likely predict success or failure of interventions. Take for instance two very similar interventions on HIV/Aids education for young people in Uganda from Kinsman et al. (2001) and Karim et al. (2009). While Karim et al. (2009) show quite positive effects of the intervention on female participants with regard to increased condom use, Kinsman et al. (2001) see almost no effect of their large-scale intervention. Can we accordingly assume that HIV/Aids education works in all evaluated eight districts, but Masaka, where Kinsman et al. (2001) conducted their study? Alternatively in 2009, but not in 2001? Possible, but unlikely. The probability is higher that the implementation strategy, which Karim et al. (2009) tested, led more successfully to behavioral change than the approach evaluated by Kinsman et al. (2001) in the given setting. However, what makes a successful development intervention?

At the heart of development interventions is regularly the aim to change human behavior – generally as a mediator to reach a certain goal (e.g., increased use of condoms to reduce sexually transmitted diseases). Limited participation or support from the respective study population challenges these interventions (e.g., Banerjee et al. 2010; Cole et al. 2013). In this chapter, we want to address the puzzle of success and failure of interventions and examine incentivizing factors for intervention uptake. What we have in mind here, is a framework guiding researchers and practitioners in designing successful interventions. A systematic and deep understanding of what drives behavioral change in response to development activities is in high demand and studies partly acknowledge this by building a theory of change (UNICEF 2014; Nayiga et al. 2014). However, the application of a general framework is missing (Duflo, Glennerster, and Kremer 2010; World Bank 2015). Instead, most interventions in development economics still predominantly rely on monetary incentives to increase uptake. Other important drivers of human behavior have attracted limited attention (e.g., Kettle et al., 2016). This is the case, despite insights from behavioral economics stressing the importance of elements besides monetary incentives that shape human motivation and behavior (e.g., Bowles and Polania-Reyes 2012; Gneezy, Meier, and Rey-Biel 2011), and scholarly work showing that these factors play a role in the successful design of interventions (e.g., Banerjee et al. 2010; Cole et al. 2013; Ashraf, Field, and Lee 2014).[157]

[157] These factors 'disturbing' the rational decision-making are increasingly acknowledged by economists (here often-called psychological biases and cognitive limitations) and insights from behavioral economics are applied to public policy (e.g., Behavioral Insights Team in the UK; Mind, Behavior and Development Unit at the World Bank; Madrian 2014).

We make use of a psychological theory called the "Theory of Planned Behavior" (TPB), which provides a straightforward framework to identify and respond to facilitating and hindering factors related to human behavior. The framework rests upon three determining factors that influence a person's behavior (Ajzen 1985; Fishbein and Ajzen 1980). The first determinant is the personal *attitude towards the behavior*, which refers to the degree to which a person has a favorable or unfavorable evaluation of performing the behavior in question. A certain attitude (e.g., dis-/trust) is mostly acquired through knowledge or learning, which can be influenced by various factors, including information or previous experience (Perugini and Bagozzi 2001; Vogel and Wanke 2016). The second predictor termed *'subjective norm'* reflects the social influence felt by the individual. It refers to the perceived social pressure to perform or not to perform the behavior. The third behavioral determinant is the degree of *'perceived behavioral control'*, which refers to the perceived own control over the behavior, i.e., ease or difficulty in its performance. Individuals, in general, rather perform tasks, if they believe they are able to (Armitage and Conner 2001). Generally speaking, individuals are more likely to intend a certain behavior if they judge it beneficial (*attitude toward behavior*), if they think important others want them to do it (subjective norm), and if they feel, they are able to do it (perceived behavioral control). Importantly, the TPB links its three predictors to intended behavior, which is the immediate antecedent and, thus, a close predictor of an individual's actual behavior (Ajzen 1991; Bilic 2005).

The TPB is currently the most widely used and accepted social cognition model across disciplines and researchers (e.g., Hobbis and Sutton 2005; McEachan et al. 2011; Ogden 2003) and seems particularly suitable to development economics. This is the case as there is a substantial body of literature which shows the applicability of the TPB to a wide variety of behaviors in different cultural and geographical settings including high- and low-income countries (e.g., Hsu, Chang, and Yansritakul 2017; Kassim et al. 2017; Kiene et al. 2014; Walrave, Heirman, and Hallam 2014; Protogerou et al. 2012). The TPB's predictive power was for instance shown in different settings with regard to technology, health-care, consumption choices, voting or education (Appleby, Roskell, and Daly 2016; Armitage and Conner 2001; Barnard-Brak, Burley, and Crooks 2010; Bilic 2005; Blue 1995; Cheon et al. 2012; Cooke et al.

2014; Landmann 2017).[158] To the best of our knowledge, however, the framework has not yet been used in implementation research to guide interventions in the field of development economics.

We apply the TPB to a real-world intervention, which we conducted ourselves, specifically, the introduction of the World Health Organization (WHO)'s Safe Childbirth Checklist (SCC) within two randomized controlled trials (RCTs) in Pakistan (Kuhnt and Vollmer 2017) and Indonesia (Kaplan, Richert and Vollmer, 2017). Evidently, the checklist can only be effective if health personnel complies with the intervention and actually uses the SCC. Hence, the behavior in question is the uptake (use) of the checklist during deliveries. Based on the TPB determinants, we analyze incentivizing factors. In addition, we will strengthen the analysis by looking into one specific parameter that is likely to influence the behavioral reaction towards development programs. Recently, studies have started to shed light on softer preconditions for the support of interventions: the implementer's characteristics (e.g., Cilliers, Dube, and Siddiqi 2015; Milner, Nielson, and Findley 2016). These fall into our determinant *attitude towards the behavior*, because they influence trust levels. As this determinant is particularly well in control of implementers, the realization of potential incentives should be comparably easy and promising. Accordingly, we deepened our analysis of the determinant *attitude towards checklist use* by conducting a framed field experiment. Within the context of the Indonesian SCC trial, we assess whether health personnel's attitude and support towards checklist use changes conditional on whether the participation of local or international agents in the study is highlighted.[159]

Our results show that the TPB can indeed help in disentangling the puzzle about intervention success and failure and consequently serve as a guideline in determining and shaping factors affecting intervention uptake. In both country settings, all three proposed TPB determinants are positively related to the uptake of the intervention. A focus on the implementation design on stimulating these factors is thus likely to increase the success of interventions through increased support and consequently higher participation rates among the targeted population. Furthermore, our framed field experiment indicates that the change

[158] Studies also looked into long-term predictions of the TPB. While the predictive power oftentimes drops with time, the TPB is still able to predict behavior for time periods as long as 15 years (e.g., Fichten et al. 2016b; McEachan et al. 2011).

[159] For a visualization of our study design, see Figure A4.1.

in attitudes due to the salience of international involvement in projects seems to have advantages over solely locally organized programs in the Indonesian context. The population under study shows higher trust and support for interventions with international involvement. Previous exposure to international research projects drives those positive behavioral reactions towards international agents.

The chapter is structured as follows: Section 4.2. links the "Theory of Planned Behavior" to our intervention and describes our research design and data. Section 4.3. elaborates on the methods used, and the results are described in Section 4.4. In Section 4.5., the generalizability and policy relevance of the results is discussed, and the study concluded.

4.2. Research Design and Data

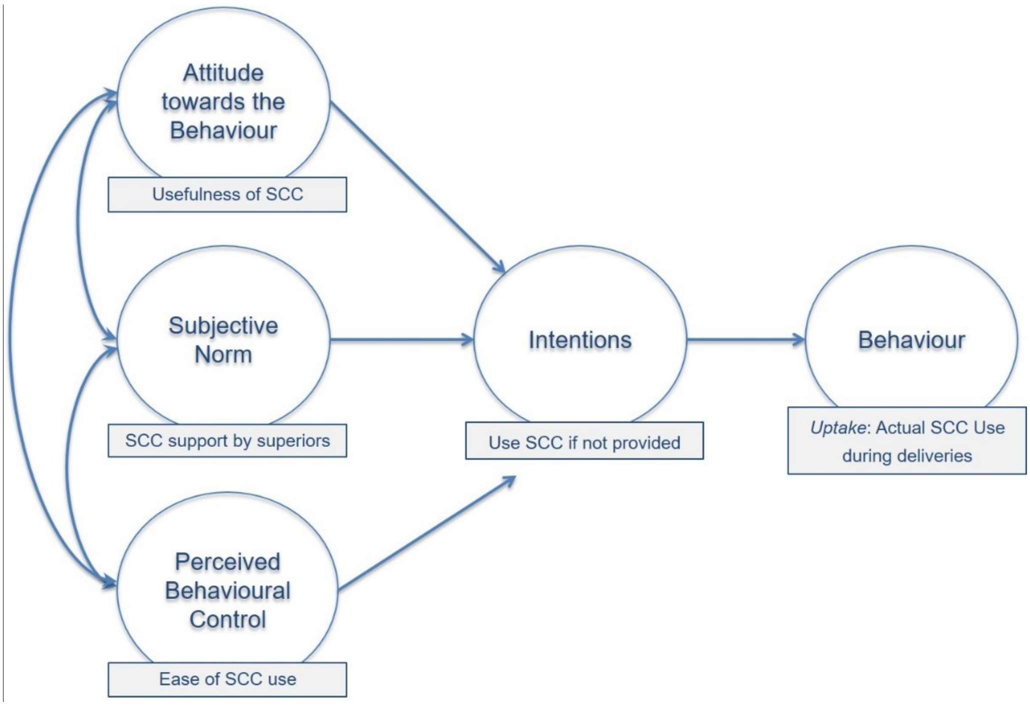
The interventions used in this study address safe childbirth. For a detailed description of the interventions, see the evaluation articles of the main RCTs (Kaplan, Richert, and Vollmer 2017; Kuhnt and Vollmer 2017). Two-thirds of mother and newborn deaths globally occur due to causes which could largely be prevented if well-established essential practices were followed (WHO 2017). However, the gap between the knowledge about what should be done to ensure safe deliveries and what is actually done is large. Following the ideas of the rational choice theory that describes independent agents thriving to maximize their utility (Simon and Feldman, 1959), the deviation should be a matter of information or knowledge availability, assuming that incentives to ensure the well-being of the patient are functioning (e.g., humanity; prestige or punishment and investigation in case of death of mother or child). The WHO Safe Childbirth Checklist (SCC) initiative aims at providing health personnel with a checklist to be used around the delivery process entailing the essential practices addressing the major risk factors for mothers and children in low- and middle-income countries. Experience from other medical fields suggests that checklists could be a promising tool to motivate health personnel to follow essential practices and tackle the know-do gap. Checklists compress and bundle the necessary information into easy-to-use actionable items and, herewith, reduce a possible 'information overload' (e.g., Borchard et al. 2012; Haugen et al. 2015; Workman, Lesser, and Kim 2007). Insights from behavioral economics suggest that human behavior is bounded by limitations of the working memory. In situations characterized by high levels of cognitive load – the amount of mental activity imposed – the successful execution of certain tasks might be interrupted or impaired (e.g., Burgess 2009; Deck and Jahedi 2015; Hoffman et al. 2011; Croskerry 2002; Lichand and Mani 2016). Checklists can be especially helpful to reduce additional cognitive load and allow a reduction of complexity of the task at hand by reminding the user of the essential steps to follow.

Using cluster randomized controlled trials, we evaluated the SCC in 32 health facilities in Indonesia, as well as in 17 health facilities and among 149 individual health providers in Pakistan. In both countries, the intervention we conducted was very similar. The treatment (SCC) was randomly introduced to approximately half of the health providers to causally identify the effect of the intervention on studied outcomes. The randomization took place at the facility level. Hence, all staff working in the same facility were jointly allocated to either treatment or control group.

4.2.1 The TPB in the Setting of our Intervention

In this section, we transform our SCC intervention into the logic of the Theory of Planned Behavior. This identifies the TPB determinants as illustrated in Figure 4.1. In the logic of Ajzen (1991) the *attitude towards the checklist*, the *subjective norm* of health personnel and the *perceived behavioral control* about checklist use will jointly determine whether health staff *intents* to use the checklist, which finally leads to whether the checklist is actually used during deliveries. We will go into more detail in the following.

Figure 4.1. Theory of Planned Behavior in the SCC Setting



Source: Own illustration based upon Ajzen (1991)

The puzzle of this study is as follows: If health personnel *know* that the checklist entails necessary essential practices supporting the safety of deliveries, why would they decide not to *use* the checklist. This is where we apply the TPB to carve out how the perception about the checklist's usefulness and relevance ('Attitude towards the Behavior'), support, and peer-pressure among staff members ('Subjective Norm'), as well as perceived ability to use the checklist ('Perceived Behavioral Control'), shape intended ('Intentions') and actual uptake ('Behavior'). Specifically, the know-do gap can be translated into the TPB determinants: The easiest explanation of why people would not *use* the checklist is because they do not *know* its benefits. The research design assured all health personnel to be *informed* about the checklist's benefits. *Knowing* the benefits, however, presumes that health personnel also *believed* in the information attained. Trusting in the checklist would therefore be a first important precondition for checklist uptake, which is subsumed under *attitude towards the behavior* within the TPB. On the perspective of the *do*-side from the know-do gap, people might still not use the checklist as they feel unable to use it (*perceived behavioral control*) or not obliged to do so (*subjective norm*). Using the real-world setting of the SCC interventions in Indonesia and Pakistan, we are able to empirically test the influence of the TPB determinants on intended and actual use of the SCC.[160] Of all TPB determinants, the *attitude towards the behavior* building on how trust-worthy the intervention is perceived seems to be particularly well in control of the intervention implementer. We therefore elaborate additionally on this determinant within our field experiment.

Data: Measuring TPB Determinants and Outcomes

We measured our data through surveys with health personnel and clinical observations of the delivery process. Our TPB determinants were collected through survey questions and serve as explanatory variables in our analysis. We conducted baseline and endline surveys at the health facilities in Indonesia and Pakistan at the beginning and the end of the interventions. Importantly, the data for the TPB analysis was only collected for the respondents working in treatment facilities, as at the time of the endline survey health staff in control facilities had not been in contact with the SCC. Hence, asking about the perceptions of the SCC would not have been possible and limits our sample to those interviewed at treatment facilities. This leaves us with 79 respondents in Pakistan and 163 health workers in Indonesia.[161] Including only the treatment facilities, gives us a non-random sample limiting causal inference, which is discussed below.

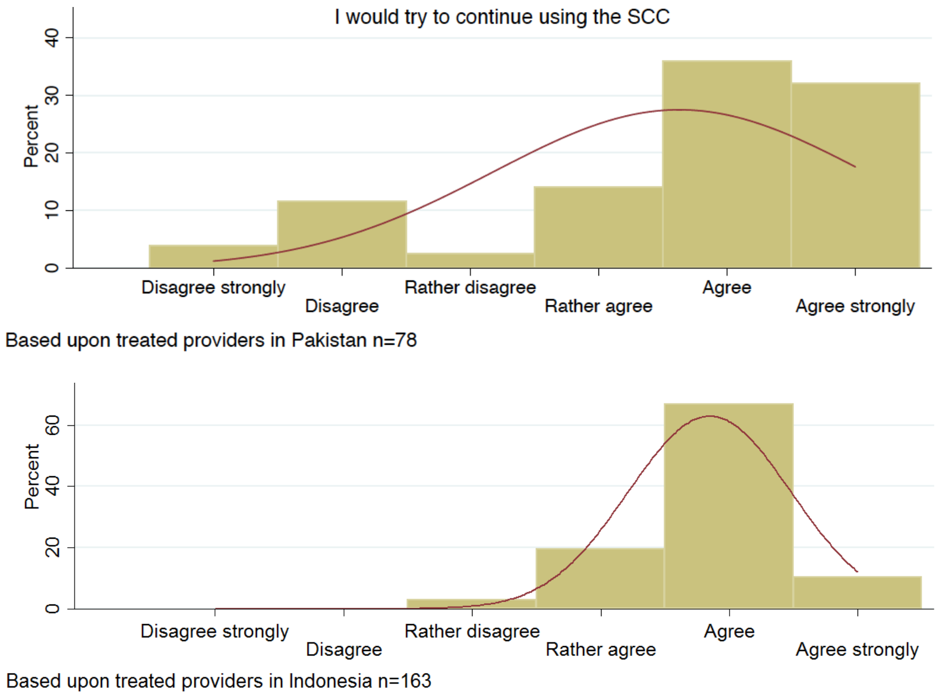
The numerous applications of the TPB to a wide array of contexts ease the measurement of TPB determinants (e.g., French and Hankins 2003; McEachan et al. 2011).[162] We were thus able to follow the respective literature when formulating survey questions. The first determinant *attitude towards the behavior,* here towards the use of the SCC, we prompt by asking the respondents to judge the usefulness of the SCC in their professional context (based upon Kam et al. 2012). *Subjective norm* would translate into the degree of support by health

[160] Theoretically, opportunity costs of using the SCC might be an impeding factor. However, monetary costs are very low and non-monetary components are implicitly part of *attitudes* and *subjective norms*.
[161] The Pakistani health staff worked at 70 different providers (including individual providers but also larger health facilities). While we surveyed every individual provider, we increased the number of interviews at health facilities proportionally with their number of delivery staff to get a more nuanced picture within larger teams. The Indonesian trial involved interviews at 16 health facilities.
[162] It has to be noted that the TPB can be applied in various ways, which is likely to influence its effects (Lugoe and Rise 1999). In order to increase the TPB's explanatory power and flexibility to address also varying intentions and behavior, several studies extended the original framework by further constructs and components (e.g., Armitage and Conner 2001; Bilic 2005; Cheon et al. 2012; Conner and Armitage 1998; Perugini and Bagozzi 2001; Klöckner and Blöbaum 2010). We will stick to the original theory when applying it to development economics, while we acknowledge the propositions made to deepen or broaden the TPB.

practitioners' superiors. *Perceived behavioral control* takes into account how easy the health practitioners judge the checklist to be applicable in their daily work routine. Additionally, surveys included demographic background information, which serves as control variables.

Following the TPB, the three components then influence whether health staff intends to use the checklist and, ultimately, if they actually use it during deliveries conducted (see Figure 4.1.). Intentions to use the checklist and actual checklist use represent our outcome measures. We investigated respondents' intended behavior towards the SCC use, by asking whether they intend to continue using the SCC after termination of the study applying a 6-point Likert scale.[163] Descriptive statistics show that the SCC is generally valued by the practitioners in Indonesia and Pakistan (Figure 4.2.). Yet, there is some distinct variation within and across the settings.

Figure 4.2. Intended behavior in Pakistan and Indonesia

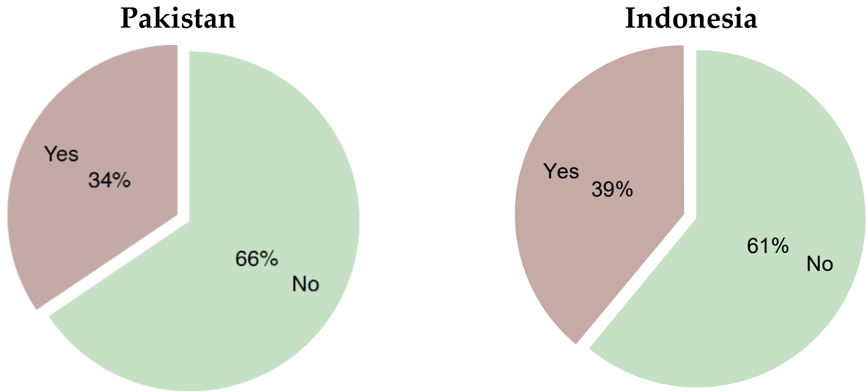


Additionally, Figure 4.3. describes the actual SCC use by health practitioners in Indonesia and Pakistan. It indicates a limited uptake and, hence, a potential gap between intended and actual use. Therefore, it is important to examine the factors that possibly constrain the behavior more carefully.[164]

[163]As an additional outcome measure we asked participants whether they would recommend the SCC to colleagues. Results are displayed in the Appendix.

[164] More detail on the data collected can be found in Kuhnt and Vollmer (2017) and Kaplan, Richert, and Vollmer (2017).

Figure 4.3. Actual behavior in Pakistan and Indonesia – fraction of use ("Yes") and non-use ("No")



Notes: Based upon 212 treated providers in Pakistan and 233 in Indonesia

In line with the previously reviewed literature, we suggest that *attitudes* towards interventions – one of the three TPB determinants – is linked to *attitudes* towards the implementing agent. These, in turn, are likely to be influenced by previous exposure to and herewith experience with program implementers. In the Acehnese health sector, 10 percent (17.5 percent) of the surveyed providers have previously participated in research projects by international (local) actors. Those interactions date back significantly before our intervention as only 2.5 percent of the respondents faced international research projects in their facility during the previous two years.

To also assess the actual use of the SCC, we additionally conducted standardized clinical observations in a subsample of the health facilities. Trained observers documented the delivery processes and marked whether the attending health staff had used the checklist.[165] This information was collected for 212 deliveries at 9 treatment providers in Pakistan and 233 deliveries at 15 treatment facilities in Indonesia.[166]

All measures (except for the actual behavior measure through clinical observations) are perception-based and, hence, subjective indicators. While this sheds light on subjective experiences, these questions are more difficult to compare across individuals and are subject

[165] Checklist use was either defined by whether the checklist was picked up during or directly after care, or whether the checklist poster was observed during the delivery process. To hang up a checklist poster in the delivery room for simultaneous consultation formed part of our intervention.
[166] In Pakistan, our observations capture 50 percent of all monthly conducted deliveries at the observed health facilities as well as 94 percent of all monthly conducted deliveries at observed individual providers. In Indonesia, the fraction relates to 64 percent of all monthly conducted deliveries at observed health facilities.

to social desirability bias.[167] However, evidence from TPB studies suggests that self-reported behavior can have higher explanatory power for intended behavior than objective measures as the latter can hardly reflect intentions, which are by nature subjective (e.g., Armitage and Conner 2001; McEachan et al. 2011).

4.2.2. What Shapes Attitudes Towards Interventions? A Framing Experiment

We investigate further on the TPB determinant *attitudes towards the behavior* in a framed field experiment, as this is particularly in control of intervention implementers. If we can identify positive incentives with our analysis, those should be comparably easy to implement and therefore promising to actually materialize in improved uptake. Precisely, the experiment aims at shedding more light on what influences people's trust in the intervention. For practical reasons, we conducted the experiment within the Indonesian trial only.

Experimental evidence within the context of the SCC intervention strengthens the real-world applicability and external validity. It has been prominently voiced that these types of experiments are a valuable and important tool to generate policy-relevant insights, e.g., by better understanding structural parameters obtained from experimental interventions like RCTs (Duflo, Glennerster, and Kremer 2010; Viceisza 2016). The experiment is, hence, not only designed to inform our specific intervention but to generate insights for international development interventions in more general terms.

Recent literature suggests that our channel in focus – *attitudes towards the behavior* – in implemented interventions is influenced by characteristics of the implementers themselves. International development interventions are mostly implemented in collaboration between international and local actors. These might include non-governmental organizations (NGOs), governmental agencies, or profit-oriented service providers. Also, the growing number of impact evaluations in the domain of development economics are often implemented by a research team working at an institution of higher education in a high-income country that collaborates with varying intensity with local partners of low- and middle-income countries to evaluate development policies or programs (Cameron, Mishra, and Brown 2016).[168] Based

[167] The social desirability bias describes the bias respondents can have in their responses due to the desire to act in a socially acceptable manner (Kemper et al. 2014).

[168] Cameron, Mishra, & Brown (2016) find that in a random sample of development evaluation studies more than 50 percent of first authors were affiliated to an institution in North America or Europe. More

on insights from previous studies, we propose that the implementer's local or international background might influence the participants' *attitude* towards the intervention.

Scholarly work has developed driving factors explaining this phenomenon. Cilliers, Dube, and Siddiqi (2015) show that the presence of a foreigner versus a local as a third-party bystander positively affects the contributions of participants in a dictator game in Sierra Leone and identify two potential channels: First, an increase in contributions to impress the foreigner and, second, reduced contributions in areas that were previously exposed to the aid-industry. In the latter locations, they show that participants more frequently believed that the game tested their need for aid, and subsequently contributed less. Milner, Nielson, and Findley (2016) find that the support of Ugandans for foreign-funded as compared to national government-funded programs is substantially larger. They stress the importance of general levels of confidence and trust towards the implementing agents for the support of projects. Dietrich, Mahmud, and Winters (2018), as well as Winters, Dietrich, and Mahmud (2017) show more specifically that donors are linked to higher quality perceptions vis-à-vis the national government. This relates to the general debate on how aid can be delivered most successfully, and whether foreign funding undermines state legitimacy (e.g., Dietrich and Winters 2015). Previous involvement and experiences with the respective agents might play a substantial role in shaping those attitudes and support vis-à-vis an implementer's project. In this vein, Dietrich, Mahmud, and Winters (2018) condition their experimental effect on previous political participation and Milner, Nielson, and Findley (2016) find that the support for foreign-funded as compared to national government funded programs is substantially larger, if participants are in favor of opposition parties, and had negative experiences with the government in the past.[169] Here, the authors, especially, stress the role of corruption and clientelism (e.g., Findley et al. 2017; Milner, Nielson, and Findley 2016).[170] In contrast, the 'home bias'-phenomenon suggests that participants have more trust in locals than in

specifically in our Indonesian context, seven out of nine RCTs in Indonesia registered with the 'American Economic Association: RCT Registry', had an US-based principal investigator and only one out of nine was led by an Indonesian investigator (American Economic Association 2018).

[169] Milner, Nielson, and Findley (2016) also assess sub-group effects with regard to gender, education, poverty, media exposure, geographic region, experience with aid, type of donor and political connections, but find mainly insignificant results.

[170] Although not testing it explicitly, Findley et al. (2017) name perceptions on accountability, capacities, and level of control as further potential channels.

internationals as cultural proximity could increase people's trust (e.g., Fuchs and Gehring 2017).

To the best of our knowledge, the described strand of the literature is currently limited to state versus non-state actors. However, against the background of the numerous international development cooperation projects and in light of the increasing number of large research projects as outlined above, it is important to understand whether the origin of the program implementer also matters, irrespective of an affiliation to the state. To this question, we dedicate our framed field experiment.[171]

Experimental Design

In the aggregate, our experiment compares whether the salience of international versus local program implementers affects support for the respective project. Stressing certain aspects of a particular situation and otherwise equivalent descriptions can lead to very different perceptions and behavioral reactions (Kahneman 2003; Tversky and Kahneman 1981; Payne et al. 2013; Hossain and List 2012; Johnson and Goldstein 2003). The result is what is called the *framing effect.*[172] Stressing certain aspects invokes different associations and leads to different evaluations by the decision maker.[173] Framing effects have been incorporated into theories on human behavior to explain deviations from rational choices (e.g., prospect theory). Their application to real-world decision-making can have important practical implications. Based upon their own intervention, Bertrand, Mullainathan, and Shafir (2006) specifically point out that framing might be a particularly cost-effective way to increase interventions' uptake, which we aim to test here.

[171] We follow the classification of experiments proposed by Harrison and List (2007).

[172] The framing effect became popular through its essential role in Kahneman and Tversky's prospect theory (Kahneman and Tversky 1979) in which they describe gambles either by their loss or gain probability. There are three different types of framing approaches that have been described and used in the literature: Most prominently and widely researched is the *risky choice framing* (risk of losing vs. risk of winning) as introduced by Kahneman and Tversky (1979). *Attribute framing* makes certain characteristics of a choice or good more salient (ground beef that is 75 percent lean vs. 25 percent fat). Lastly, *goal framing* where either punishment or reward is emphasized (behavioraleconomics.com 2017).

[173] Since then, framing experiments have been extensively applied in medical sciences both in hypothetical (Wilson, Kaplan, and Schneiderman 1987) and real contexts, often related to message framing experiments, e.g., with regard to smoking cessation, HIV screening as well as skin and breast cancer prevention (Detweiler et al. 1999; Kalichman & Coley 1995; Schneider et al. 2001; Toll et al. 2007).

We make use of the randomized phase-in design of the SCC intervention in Indonesia. Within the endline survey of the larger RCT project, we performed the experiment with midwives at control facilities that neither have received the SCC nor were in contact with the implementation team. Within this group of midwives, we used a between-subject design and randomly assigned the study participants to two different framing information on the actually conducted SCC intervention: The first framing information stressed the involvement of international actors in the SCC program, while the second made the participation of local counterparts more salient (see Figure A4.1. in the Appendix for an overview over the study design).[174] We use the fact that the SCC evaluation has been implemented jointly by both – international and local – actors and therefore, highlight different attributes of the project. We then investigated the participants' respective behavior towards the intervention by assessing the support for the SCC project. Since we randomized participants into different treatment groups, we can make causal inference on how the origin of implementers affects behavioral reactions (i.e., different levels of support for the SCC intervention).

In a short pre-experimental survey, we collected background information, including socio-economic and contextual work characteristics, of each participant.[175] In appreciation of participants' survey participation, each respondent received a voucher for a phone credit top-up worth 25.000 IDR (approx. 1.75 US$). Afterwards, the respondents were offered to participate in the experiment.[176] Lastly, we conducted a short post-experimental survey, including questions capturing potential framing mechanisms and additional control variables, like the experience of current financial distress.

The 'experimental commodity' was derived from the on-going RCT intervention on the SCC. First, the idea and structure of the SCC was explained to the participants. Afterwards, they were presented with one of the two framings that selectively either stressed the involvement of 'local' or 'international' actors respectively, in the SCC intervention.[177] A

[174] The framing experiment does not include a control group as development programs are always either conducted exclusively locally or have an international component. We believe that it is very unlikely that the implementer's identity is unknown to program participants.
[175] This survey was included in the endline survey of the larger SCC intervention.
[176] All respondents chose to continue the survey and participated in the following framing experiment.
[177] As it is likely that respondents equate an international actor as a donor, we specifically addressed the relevant actors as researchers and professionals in our framing component.

qualitative investigation was conducted prior to the experiment to ensure that the correct terms were used to describe 'local' versus 'international' agents.[178]

Our framing information reads as follows:

"Among other researchers, [INTERNATIONAL / LOCAL] researchers took an active role in introducing the checklist to 17 facilities in Aceh province. The research team received approval from the provincial health office of Aceh. However, no funding was provided by the provincial health office. [LOCAL / INTERNATIONAL] research assistants and [INTERNATIONAL / LOCAL] health professionals with a lot of experience in delivery services were important partners and greatly supported the project."

In order to be able to draw broader conclusions and to generalize the findings to different types of interventions, we named different actors (e.g., researchers, practitioners). To prevent potential effects through assumptions on political involvement, we specifically address the role of the provincial health office in the information given to the study participants. Further, to counter potential bias through speculations on the financial capabilities of different actors, we stress that funding of the intervention is ensured irrespective of the framing given to the participant. For the detailed experimental protocol see Appendix A1.

We hypothesize that the level of support would significantly differ between the local and the international framing. Following the literature, there are arguments for directive effects on both sides, which leads us to handle the issue as an empirical question.

[178] For this purpose, we talked to health-care providers from different facilities, which were not part of the sampled institutions. In the Acehnese setting 'local' is understood as 'Acehnese' identity, whereby 'Indonesian' would be an external concept. Certainly, it would have been of large interest to examine the difference between local and Indonesian implementers. However, due to power constraints, we decided to focus on this more specific framing without splitting the group and reducing the sample. The distinctness of 'Acehnese' and 'Indonesian' is also underlined by the fact that a small set of respondents named Indonesia and certain provinces as international countries. To deepen our understanding of the term 'international' in the Acehnese context, we asked respondents to name the three countries, they first think of when hearing this term (see Figure A4.2. in the Appendix). There is a large consensus among respondents regarding the main countries associated with 'international', namely Germany (24 percent), Malaysia (19 percent), USA (13 percent), Australia (8 percent). The high prominence of Germany among the foreign countries named, could first – of course – be attributed to the fact that parts of the implementing researchers, were German. Second, it is nevertheless likely, that Germany is indeed particularly present the Acehnese people as it was the largest European donor after 2004's Tsunami (BBC 2005). Moreover, Germany's reconstruction efforts were characterized by a strong focus on health interventions (Federal Ministry for Economic Cooperation and Development (BMZ) 2005).

Experimental data

In total, the experiment was conducted with 236 female midwives from the SCC intervention's control group. The average study participant was 33 years old (minimum: 21 years, maximum 50 years), had 10 years of work experience (minimum: 0 years; maximum 28 years) and 15 years of education (minimum: 12 years; maximum 17 years) (see Table A4.1. in the Appendix). Participants in the experiment were comparable in their characteristics to health workers of the main RCT study (see Appendix Table A4.2.).[179] Individual characteristics and further contextual variables are balanced across treatments indicating that the randomization was successful (Table A4.1.). In our main analysis, we focus on those participants that have not been in prior contact with the SCC as 27.92 percent of the respondents state that they were previously exposed to the SCC.[180] As we cannot infer how much these respondents know about the SCC intervention and how intense the exposure was, excluding them is the more conservative choice.[181] This reduces our sample to 173 participants.[182] Balance on important covariates is still given in this reduced sample (see Appendix Table A4.3.). Previous SCC exposure was equally distributed across the framing treatments, ruling out selection concerns and enabling us to interpret the estimates causally.

We proxy SCC support by asking the respondents whether they would contribute parts of the money they had received through the voucher for phone credit top-up in appreciation of their survey participation to buy checklist copies, which would then support the implementation of the SCC in other anonymous health facilities within the province.[183] The contribution was made anonymously. After the experiment, all participants received a

---

[179] Health workers in the treatment group seem to have experienced on average five more months of education.

[180] Reasons for previous exposure might be a second job at another (treatment) facility (11.11 percent of respondents have a second job) or communication with other health practitioners within the district. Contact to midwives from other facilities is also significantly correlated with prior checklist contact.

[181] As a robustness check, we also report the full sample results including a prior contact binary variable in the regression model in Appendix Table A4.11. However, as we assume a large heterogeneity of exposure – health practitioners with a job at another facility might have worked with the SCC, others might have just heard the name of the SCC from colleagues - we prefer the reduced sample for our main results.

[182] Due to two outcome measures that could not be matched to respondents and 4 respondents that refrained from answering on control questions, the sample is reduced to n=165 in our main specifications.

[183] If they wanted to contribute, we offered them five options from 5,000 to 25,000 IDR (equivalent to 0.4 - 1.86 US$) due to pragmatic reasons of specific top-up values.

debriefing.[184] To create transparency on the use of the collected funds, we publicly made information on total amounts available after the end of the study and informed the participant about this procedure. In addition to this traditional monetary outcome, we also collected measures suggested by other disciplines. Psychologists commonly assess the respondent's behavior through time investments (Wildschut et al. 2014).[185] Hence, we asked the participant's willingness to invest additional time to practice checklist use during regular working weeks. Further, in order to counter potential social desirability bias, we asked the participants to estimate the average monetary contribution of colleagues in other health facilities in the province.[186] Those elicitation exercises based on introspection have been shown to reduce potential conformity bias in the experimental literature (Trautmann and van de Kuilen 2015). We focus on the traditionally employed monetary outcome as due to the costs incurred by the respondent (hypothetical or real) this is likely to be the strongest measure, while the additional outcomes are presented in the Appendix. Summary statistics for all measures employed can be found in Appendix Table A4.4. for Indonesia and A4.5. for Pakistan.

In the post-experimental survey, we asked several questions on potential mechanisms to explain differential preferences towards implementers. Following Milner, Nielson, and Findley (2016), we measured participants' level of trust towards different actors (international/local actors) and towards the previously named countries that they understood by the term 'international'. We used 4-point Likert scales. In addition, we asked participants

---

[184] After the debriefing, we offered participants to change their monetary contribution. 39 (16.5 percent) participants made use of this option. Generally, this led to an increase in contributions by on average one category (about 4200 IDR), but the amount is not contingent on the framing applied. The main analysis focuses on the pre-debriefing contribution, as we are interested in the framing effect.

[185] Actual behavior measured by contributing money may be strongly influenced by general or situational economic living conditions of respondents (Wildschut et al. 2014). In case respondents face strong economic constraints, small or zero contributions might reflect a high neediness rather than lack of support for the intervention.

[186] We incentivized an honest answer through an additional payment if the respondent's estimate was close to the real average within our study sample. The underlying idea is that respondents might be biased to not perfectly reveal their preferences in order to keep their face. However, when being financially incentivized to assess the potential answer of an anonymous third person, opportunity costs of not revealing the own true assessment increase. Moreover, in a resource constrained setting the beliefs about the willingness of others to contribute could provide more accurate information about preferences as they are less subject to the idiosyncratic financial situation of the respondent.

whether they have previously participated in interventions by international or local experts or researchers, respectively.[187]

4.3. METHOD

In the first part of our regression analysis we address the role of the TPB determinants for intended behavior with regard to checklist use. Our regression line for intended behavior reads as follows:

$$y_i = \alpha + \beta_1\, TPB\ determinant_i + \beta_k \Sigma_k X_i + \varepsilon_i \qquad (1)$$

As throughout the study, we estimate models for Indonesia and Pakistan separately using ordinary least squares (OLS) regressions. Our level of analysis is the individual health worker i (79 respondents for Pakistan and 163 individuals for Indonesia). y_i determines our outcome variable, which measures intended behavior employing 6-point Likert scales. α is a constant, and *TPB determinant*$_i$ capture our variables of interest (also using 6-point Likert scales)[188] via our three perception measures for the three TPB pillars: *Attitudes, subjective norms,* and *perceived behavioral control*. In adjusted regressions we add $\Sigma_k X_i$, which represents our set of k control variables. These include a variable indicating the type of facility,[189] a binary variable indicating the location of the facility (rural versus urban), a variable capturing the district where the provider is located and we include the level of service provision, which is proxied by a dummy for 24/7 opening hours. The idea is that those time-invariant facility characteristics might affect the drivers of the TPB. Perceived behavioral control could be affected by staffing and equipment, which is captured by facility type and geographical remoteness (district dummies and rural/urban distinction) as well as the 24/7 service

[187] To investigate additional potential mechanisms, we also collected information on perceived corruption, skills and control to implement interventions, sufficient funding capabilities, and accountability. All these data were collected after the experiment was conducted in order to not affect our main outcome measures. However, this procedure comes with the trade-off of potential justification bias, where individuals would adapt their answers ex-post to justify the previously indicated support. We indeed find that the framing statistically significantly affects some of these variables. Hence, in our analysis, we focus on those variables not significantly affected to avoid bad control issues. Not statistically significantly affected were: Participation in international or local projects, trust in internationals, trust in named foreign countries and trust in locals. We use these later in our regression analysis.

[188] Further, we also estimated regressions with an alternative coding for robustness, where we defined a dummy variable with the value one for the highest category and zero otherwise.

[189] Our sample included a wide heterogeneity of facilities from primary to tertiary health providers.

provision. Provider's *attitudes* and the *subjective norms* of superiors towards the SCC are subject to the facilities' safety culture, which is captured again by facility type, service provision, and geographical remoteness (district dummies and rural/urban distinction).

Our second part of regressions is the equivalent to the first but changes the outcome variable to birth observations j measuring the actual behavior. Here, y_j, is a binary variable equalling one if the checklist was used by the health worker during the delivery. As we cannot link each delivery to the specific health workers' responses, we take averages of *attitudes, subjective norms* and *perceived behavioral control* per health facility. This would provide us with an intuition of more supportive environments being associated with more or less take-up.[190] The control variables X_j stay the same as in regression *(1).*

Following the clustered setup of the intervention, in all specifications, the error terms are clustered at the facility level to account for joint correlation within the clusters.[191] We employed Likert scales to all perception-based survey questions, which are relatively continuous measures. Hence, we consider them as continuous variables in the estimations, which is the preferred method of analysis proposed in the literature (Pasta 2009).[192] As our sample is restricted to our treatment group and includes, thus, a non-random set of individuals, estimations are not derived within the randomization framework and do not allow a causal interpretation. Nonetheless, controlling for several potentially confounding variables, we will receive informative correlations about how behavioral processes are associated with intervention uptake.

The third part of our regression analysis concerns the experimental data. Our analysis of the experiment aims to identify the existence of a systematic difference in the support for

[190] As our analysis, thus, involves different aggregation levels and our measure of intention and actual behavior capture slightly different concepts, we do not estimate a model on the direct link between intentions and behavior.
[191] Due to a limited number of clusters we also present results with wild bootstrapped standard errors following Cameron, Gelbach, and Miller (2008) for all our baseline models in the Appendix. However, this is only possible for the unadjusted regressions (without controls). In the regressions which include control variables and bootstrapped SE, we face problems of overfitting. This is the case as our controls consist mainly of dummy or categorical variables, which reduce variation among our relatively small number of observations too strongly to calculate meaningful adjusted standard errors. Accordingly, we prefer to present regressions without bootstrapped standard errors in our main models.
[192] We also assessed the feasibility of continuous items with a scale from 0 to 100, but learned that those were harder to comprehend for respondents in the field.

our intervention by health practitioners, conditional on whether the local or international implementation was more salient. Our results are based on the following regression equation:

$$y_i = \alpha + \beta_1 framing_i + \beta_2 framing_i * c_i + \beta_3 c_i + \beta_m \Sigma_m C_i + v_i \qquad (2)$$

In our most parsimonious model, y_i is the outcome variable, indicating the support of the SCC by health worker i. α is a constant, and $framing_i$ is a binary variable, which equals one if the respondent was exposed to an international, and zero for a local framing. Moreover, heterogeneous effects are assessed by the inclusion of an interaction between the framing and channel c_i, which is prior participation in international or local projects. In adjusted regressions we add $\Sigma_m C_i$, which is our set of control variables. The controls include a variable indicating the respective facility type, where the participant is employed, as this is likely to influence the *attitude* of the respondent towards the tool, and a binary variable marking whether the respondent experienced financial problems within the past days as this might affect monetary contributions.[193] Further, to control for a potential social desirability bias, we measured social conformity following the social desirability scale developed by Kemper et al. (2014). This measure was adopted to the Acehnese context and we transformed its five items into a composite index.[194] We control also for the subjective perception regarding the amount of paperwork during deliveries, which was motivated by an often-experienced perception during implementation that the new tool adds to the already existing paperwork. Finally, v_i describes the residual. Errors are clustered at the facility level to take into account similarities within teams. We are, thus, mainly interested in the effect sizes of β_1 and β_2.[195]

[193] Previous research on the SCC has shown differential effects of the checklist across different healthcare facility settings. Applicability to the respective work environment is likely to be influenced by factors like e.g., team size, resource access, or delivery load. Related research has similarly controlled for a constructed wealth index (e.g., Cilliers, Dube, and Siddiqi 2015).
[194] We adapted the social desirability measures to the respective context in cooperation with Indonesian counterparts. For instance, one of the items reads "I have occasionally thrown litter away in the countryside or on to the road." As environmental concerns are less salient in the Acehnese context than religious concerns, we changed the item to "When I had the chance to donate for religious purposes, I always contributed a lot." The full set of questions we used for the construction of the social desirability index are displayed in Appendix A4.2.
[195] Estimates using ordered probit regressions are shown for robustness in the Appendix. For the ease of interpretation, we prefer to present OLS results in the main part.

4.4. RESULTS

Main results: TPB determinants and SCC support

For all three TPB determinants, *attitudes*, s*ubjective norms,* and *perceived behavioral control*, in both study sites Pakistan and Indonesia, we find that coefficients point towards a consistently similar direction. Table 4.1. displays the regression results of the intended and actual SCC uptake for the data from Pakistan and Indonesia. While the first row always presents the unadjusted coefficients, the second displays results adjusted for control variables as described in Section 4.3. Results show that respondents who express a strongly positive *attitude* towards the SCC are also more likely to intend to use the new tool even if it is not freely provided to them anymore (columns (1a) to (2b)). In Pakistan and Indonesia the coefficients are positive and statistically significant (ranging from the 1-percent to 5-percent level). This is also supported by the actual SCC use (columns (3a) to (4b)). The stronger the positive stance towards the checklist, the more often health staff actively uses the SCC during the delivery process. If the SCC is perceived to be more useful (*attitude*), its actual use among Indonesian health workers increases by 39.4 percentage points and among Pakistani practitioners by 47.1 percentage points. Further, we find consistently positive coefficients in both countries with respect to the support of superiors for the new tool (*social norms*). While it seems to play an important role for intended and actual SCC uptake in Indonesia, it is less important for intended behavior as compared to the actual SCC use in the Pakistani setting. This can be explained by different samples across our outcome measures. While the actual behavioral outcome was mainly collected for health practitioners working in facilities, the sample measuring the intended SCC uptake is dominated by individual health workers (like community midwives). Hence, for them the opinion of superiors is less of a concern but rather the perceived usability (*perceived behavioral control*). In this regard, we see that the ease of use is a statistically significant predictor of intended SCC use in Pakistan (at the 5 percent level in the adjusted regression), while it is positive but not statistically significant in the Indonesian context or for actual SCC uptake in both countries.[196]

[196] As outlined above, we use wild cluster bootstrapped standard errors as robustness tests in samples with small number of clusters (9 in Pakistan and 15 in Indonesia). Results are displayed in Appendix Table A4.6. showing that results are by and large robust to this standard error correction. When we generate a dummy variable as an outcome, equalling one for the highest category only ("fully agree" to "Would try to use SCC even if copies are not provided") results are qualitatively unchanged (see Appendix Table A4.7.).

These results – though not allowing the establishment of a causal pathway – give a consistent indication: Influencing the TPB determinants into the respective positive direction is associated with increased intended and actual uptake of the SCC. The regressions without controls in the (a)-columns indicate that the TPB determinants capture 5 to 13 percent of the variation in intentions among Indonesian respondents, and 0.3 to 4 percent of the variation in actual behavior (measured by the adjusted R-Squareds between 0.048 and 0.132). Adjusted R-Squareds for the Pakistani case are exceeding those from Indonesia and range between explaining 0.8 to 19 percent of the variation in intensions with the TPB determinants. The explanatory power for actual behavior lies between 10 and 29 percent. Hence, the three TPB determinants are important predictors for intended and actual behavioral outcomes here the use of the SCC.

Differences in the adjusted R-Squareds across TPB determinants are well in line with qualitative evidence. Indonesian coaches, who assisted health personnel in using the checklist, were seldomly asked for help regarding the content of the SCC, which corresponds to the ease of use of this intervention. In contrast, the assessment of the supervisor seems to play an important role in the hierarchically structured Indonesian society. This is also borne out by inter-facility staff meetings and midwives' correspondence with coaches in Indonesia, stressing the salience of supervisors and colleagues reminding each other to use the checklist regularly. In the Pakistani case, we see the strongest explanatory power for the determinants *attitudes* and *control* and far behind for *norms* (12 to 19 percentage points difference). In line with explanations from above, the effect is likely to be driven by the sample of community midwives, who work rather self-employed and do not depend on superiors' norms, accordingly.

Both sets of results imply that in both countries, specifically, *attitudes* are crucial in shaping intentions and actual behavior. As indicated in the previous literature review, perceptions about the implementer can be strong predictors in shaping intentions and behavior. This is assessed in the subsequent section.

Table 4.1. Main regressions – Intended and Actual SCC uptake (1 "disagree strongly" - 6 "agree strongly")

	Intended SCC Uptake: Would try to use SCC even if copies are not provided				**Actual SCC Uptake:** Was Checklist actively used or looked at during delivery process?			
	Pakistan		**Indonesia**		**Pakistan**		**Indonesia**	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Attitudes: Using the SCC in my professional role is: 1 "completely useless" - 6 "completely useful"	0.984***	0.818***	0.4535***	0.3088**	0.655***	0.471**	-0.356	0.394***
p-value	(0.00001)	(0.00024)	(0.00422)	(0.01181)	(0.00285)	(0.01948)	(0.24536)	(0.00000)
Adjusted R-Squared:	0.187	0.254	0.114	0.272	0.288	0.346	0.017	0.061
N	79	79	163	163	212	212	219	219
Subjective Norms: SCC is supported by superiors: 1 "not at all" - 6 "completely"	0.143	0.164*	0.5356***	0.3162***	0.207*	0.0781**	0.654*	0.279***
p-value	(0.115)	(0.0597)	(0.00740)	(0.00082)	(0.09708)	(0.02679)	(0.09143)	(0.00002)
Adjusted R-Squared:	0.008	0.304	0.132	0.261	0.095	0.325	0.041	0.062
N	58	58	163	163	212	212	219	219
Perceived Behavioral Control: Ease to use SCC in work environment: *1 "very difficult" - 6 "very easy"*	0.439***	0.366**	0.2614*	0.0226	0.306***	0.112	0.0589	0.0154
p-value	(0.00319)	(0.0290)	(0.08960)	(0.86269)	(0.00026)	(0.16877)	(0.42346)	(0.97900)
Adjusted R-Squared:	0.128	0.211	0.048	0.222	0.253	0.318	0.003	0.057
N	78	78	163	163	212	212	219	219
Control variables	no	yes	no	yes	no	yes	no	yes
Mean of dep. var.	4.628	4.628	4.847	4.847	0.344	0.344	0.389	0.389
Median of dep. var.	5	5	5	5	-	-	-	-
Standard Deviation of dep. var.	1.452	1.452	0.634	0.634	0.476	0.476	0.489	0.489

Asterisks indicate p-values according to: *** p<0.01, **p<0.05, * p<0.1. All regressions are based upon the treated providers and standard errors (SE) are clustered at the facility level. Adjusted regressions (b) additionally control for a variable indicating the facility type, a binary variable indicating rural/urban location, a variable indicating the district and a binary variable indicating whether the facility is open 24/7.

Main results: framing experiment

Table 4.2. displays the main results of the framing experiment conducted in Indonesia. We only include our main outcome measure (monetary investment) here, while results of the alternative outcomes are presented in the Appendix (Table A4.8.).[197] The first column presents the unadjusted results, whereas the second column gives the results adjusted for additional control variables.[198] We limit our sample to those respondents who were not exposed to the SCC prior to this experiment. Full sample regression results controlling for prior contact, are shown in the Appendix (Table A4.11.) and are comparable to the findings presented here.[199] As conservative robustness check, we also present random inference based p-values.[200]

Table 4.2. Main regression results of the framing experiment

	Financial Contribution in support of SCC project (in IDR)	
	(a)	(b)
Framing (=1 if 'international')	557.6236	1,283.7717**
p-value	(0.39637)	(0.02119)
RI p-value	(0.4500)	(0.0570)
N	165	165
Control variables	no	yes
Mean of dep. var.	4757.576	4757.576
Std. dev. of dep.	4711.366	4711.366

[197] Similarly, we present estimates using ordered probit regressions in the Appendix Table A4.9. Results are qualitatively unchanged to OLS regressions.
[198] In line with the randomized setup of the study, results are robust to the inclusion of further covariates, which increases the precision of estimates. The full specification including all control variables is presented in the Appendix Table A4.10.
[199] As a further robustness check we estimate a regression which controls for an interaction of the framing with the indicator for past contact. First, this is more conservative as the framing should have a lower effect on the persons that are acquainted to the SCC and induce, thus, a downward bias. Second, individuals with prior contact to the checklist might react heterogeneously due to more comprehensive information.
Table A4.12. depicts the corresponding results. While the framing indicator decreases slightly in size, but stays significant in the adjusted regressions, there is no significantly different treatment effect for those respondents with past contact.
[200] Randomization inference takes the randomization explicitly into account and follows R.A. Fisher's idea of statistical inference via permutation tests of treatment allocation (Young 2016). The idea is to assume uncertainty about the treatment allocation and compare the actual treatment allocation to possible alternative allocations.

Asterisks indicate p-values based on standard errors clustered at the facility level according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level. RI p-values are computed with a permutation test based on Hess (2017). Specifications (b) include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork. The same regression with wild cluster bootstrapped SE can be found in Appendix Table A4.13., for which significance levels hold.

In the unadjusted regressions, international framing has a positive but at conventional levels insignificant effect on financial contributions of respondents. Once adjusting for control variables, this coefficient turns significant at the 5-percent level. Respondents facing an international framing contribute on average more money in support of the SCC project than their counterparts being confronted with the local framing. In the adjusted specification, their contribution is 1,284 IDR higher. These results are supported by the alternative outcome measures presented in Appendix Table A4.8.[201] Hence, our results suggest that the intervention is increasingly supported by the respondents, if it is perceived as an internationally-led endeavor. The representativeness of the experiment is supported by the balance of important individual and contextual characteristics between the experimental sample and the larger sample of the SCC intervention.

[201] Our alternative outcome measures are first, whether respondents would recommend the SCC to fellow colleagues, second, whether they would be willing to invest additional time for the SCC project, third, how high they estimate the average contribution by others and fourth an index of all four outcome measures, using principal component analysis (PCA). Estimates in Table A4.8. show robustly positive coefficients, when controls are included and reach statistical significance for recommending the SCC to others and for the PCA-index. Here, however, the financial contribution is the variable that explains the major part of the variance in the index. Additionally, one's willingness to support an intervention might also be strongly determined by the beliefs about others' contribution. However, reporting one's perception about others might be subject to conformity bias, especially, in the Indonesian society, where a large focus is put on keeping one's face. Elicitation exercises based on introspection have been shown to reduce potential conformity bias in the experimental literature (Trautmann and van de Kuilen 2015). We, thus, incentivized respondents with an additional pay-off of 5,000 IDR to estimate the average contribution category of respondents at other facilities. Therefore, we use the outcome variable elicitation as control in our main specification as robustness (see Table A4.14.). As expected, elicitation shows to be highly significant and positive, while the framing effect holds.

Channels: previous exposure

In order to understand in more detail why respondents show stronger support towards projects implemented by international actors as compared to local implementers, we investigate a mechanism that could influence the *attitude* of respondents. Previous exposure is one prominent factor determining *attitudes*. Hence, it might play a role whether respondents have been in contact with locally or internationally-led research projects in the past. Their respective experiences are likely to influence their present attitudes and reactions to the intervention.

Descriptive correlations (see Appendix Table A4.15.) indicate that first-hand experiences – both with local and international research programs – are associated with positive perceptions towards the corresponding implementer – though no claims regarding the causal direction can be made here. Hence, it seems that those positive experiences affect not only the respective implementer but also the support for other actors.[202] It is, therefore, of particular interest to examine the interaction of the international or local framing with previous exposure vis-à-vis the implementing agent. Table 4.3. displays the results for the interaction of our experimental framing with the binary variables indicating if respondents already participated in international or local research projects. While the randomization ensured that the framing could be considered as exogenous, project participation is potentially endogenous regarding other traits of the surveyed respondent. However, as recent research by Bun and Harrison (2018) and Nizalova and Murtazashvili (2016) indicates, the interaction of an exogenous and an endogenous variable can be considered as exogenous, when controlling for the endogenous variable.[203] Moreover, balancing tests provided in Table A4.2. and A4.3. underscore that previous participation is balanced across both framing treatments. The results in columns (1a-b) are structured to compare respondents with similar previous experience (e.g., participation in international/local projects) across framings. The corresponding comparison group are locally framed respondents, who did neither participate in a local nor in an international project. Rows I and II show that if a person had been exposed

[202] The results in Table A4.15. also hold if including as control variable local or international participation, respectively, and if standard errors are bootstrapped. The majority of respondents, who participated in international projects also participated in local projects, but not vice versa.

[203] Nonetheless, one needs to be aware that, especially, with a limited sample size omitted variables might not be homogenously distributed and, hence, it is not inherently clear, which other factors are correlated with our interaction variable of interest.

both to an international and local research project in the past, their contribution is approx. 6,500-8,500 IDR (e.g., 0.45-0.65 US$) higher if framed international.[204] Thus, the effect of the *attitude* towards the intervention in the unadjusted and adjusted specification is significantly higher if respondents knowing both implementers are framed internationally (p-value: 0.025 and 0.000, respectively). Respondents who previously participated in local projects do not contribute different amounts of money when faced with an international framing. However, if respondents were only exposed to international projects in the past, they do contribute significantly less if locally framed, both significant with and without adjusting for controls (p-value: 0.012 and 0.052, respectively). Finally, row VII does not depict any significant framing effects, if respondents did not have any prior experience. Those estimates suggest that the positive effects of the international framing are driven by previous experience with the respective implementer. The reduced willingness to contribute to local projects is most pronounced if respondents have participated both in local and international projects.

Table 4.3. Main regression results – Previous participation in international and local projects

	Financial Contribution in support of SCC project (in IDR)	
	(1a)	(1b)
I. International Framing (1) x Int Partipation (1) x Loc Participation (1)	2,708.3333	4,202.8921**
p-value	(0.2369)	(0.0186)
II. International Framing (0) x Int Partipation (1) x Loc Participation (1)	-3,791.6667***	-4,313.2256***
p-value	(0.0067)	(0.0001)
Coefficient Equality Row (I) & (II)	0.0247	0.0000
III. International Framing (1) x Int Partipation (0) x Loc Participation (1)	-2,291.6667*	-1,196.6307
p-value	(0.0684)	(0.2871)
IV. International Framing (0) x Int Partipation (0) x Loc Participation (1)	-148.8095	-537.1762
p-value	(0.9183)	(0.7621)

[204] Although this amount seems small, it corresponds to one meal or half an hour of work of a midwife in the local context.

Coefficient Equality Row (III) & (IV)	0.1864	0.6602
V. International Framing (1) x Int Partipation (1) x	-625.0000	1,433.0596
p-value	(0.7101)	(0.5073)
IV. International Framing (0) x Int Partipation (1) x Loc Participation (0)	-4,791.6665***	-4,184.6089
p-value	(0.00001)	(0.1275)
Coefficient Equality Row (V) & (VI)	0.0124	0.0520
VII. International Framing (1) x Int Partipation (0) x Loc Participation (0)	646.9298	1,009.8640
p-value	(0.4626)	(0.2006)
N	165	165
Control variables	No	Yes

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level. Specifications (b) include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork.

The positive *attitudes* towards international projects might, however, depend on the local context as every country will have its specifics in experiences with and attitudes towards the local and international community. The Acehnese context is a very interesting case to study as to its large exposure to various international as well as local actors in the aftermath of the Tsunami 2004 causing more than 130 000 deaths in the country.[205] Due to previous experiences with both local and international implementers, the assessment of *attitudes* towards the different implementers is facilitated. However, this context of ultimate human emergency, might have induced a more positive *attitude* towards the international assistance and could make the interpretation specific to the context.[206] Qualitative data based on 66 surveys with health practitioners suggest that positive *attitudes* towards internationals are mostly linked to

[205] Despite the individual tragedies, the natural disaster was perceived by parts of the population as a chance to restart, as it coincided with the cessation of the Aceh insurgency after almost 30 years of combat and successful reconstruction efforts.

[206] Moreover, Aceh might be specific due to its strong Muslim heritage and introduction of Islamic law in 2006.

perceptions of better knowledge and more structured implementation approaches.[207] This is in line with the positive and significant correlation of the international framing with positive perceptions of international control capabilities and skills vis-à-vis local implementers (confer Table A4.16.) and corresponds to higher trust levels after previous project participation (see Table A4.15.).[208]

Taken together, those results, first, suggest to consider the previous experience of the targeted population when aiming to achieve high project uptake and accordingly frame development policies. Second, they call for caution when thinking about scalability of projects by the local government if piloted by internationals. Third, they call for caution, both from internationals and locals, when implementing development policies. Their past actions might affect subsequent take-up and success of other projects.

4.5. Discussion and Conclusion

Evidence from behavioral economics supports the importance of non-monetary incentives, trust, or peer effects to explain human behavior. These insights are also of utmost importance to the design of interventions in development economics. This chapter makes use of the *Theory of Planned Behavior (TPB)* – a well-established theory originating from social psychology – offering a systematic approach to explain and influence supportive human behavior by considering three determinants: A positive *attitude* towards the behavior or intervention, supporting *subjective norms*, and a high degree of *perceived behavioral control*. We provide evidence of the positive association of these mechanisms with the uptake of a program by study participants in two different cultural contexts. Using the settings of two randomized controlled trials in Pakistan and Indonesia, we show that a more positive *attitude* towards the new tool (here the Safe Childbirth Checklist (SCC)), more salient *subjective norms* in favor of the intervention, and greater *perceived behavioral control* to actively use and implement the checklist were associated with increased intended and actual use of the checklist. Applying

---

[207] These were the most often named categories in an open question reading "Please describe your experience working with international teams. What did you find surprising?"

[208] We asked midwives if they would attribute certain attributes rather to local or international researchers (e.g., skills, corruption, financial capabilities) in order to carve out how those channels might affect support for the intervention. Those questions were asked intentionally after collecting the outcomes in order to not confound the results. However, this comes with the risk of justification bias, indicated by the significant framing effects in Table A4.16. Hence, we did not use those channels for further analysis. Yet, they might be still informative in terms of general attribute ascription.

the TPB in two diverse study contexts strengthens the claim of generalizability of the results. Previous studies on the TPB also support its broad applicability to explain and influence human behavior. However, it is important to note that we left the random setting for the TPB analysis and, hence, our study does not allow us to infer causal effects of the TPB on intended and actual behavioral reactions.

Recent evidence shows the importance of implementers' characteristics in shaping behavior towards an intervention and it is likely that this affects the TPB determinant *attitude towards the behavior*. Hence, we further investigate how the salience of the implementer's background, in particular, whether a project is led by an international or local agent, influences the participants' support for the project. The implementer's background is particularly interesting with regard to increasing experimental research in low- and middle-income countries, which is often a collaboration between international and local researchers and practitioners. The results of the framed field experiment in Indonesia indicate that respondents are more supportive towards interventions (measured through monetary support) implemented by international actors as compared to solely locally led projects. This finding is in line with previous research on behavioral reactions towards international and multilateral donor agencies (e.g., Milner, Nielson, and Findley 2016; Winters, Dietrich, and Mahmud 2017). Even though research projects might be characterized by different conditions than practical development cooperation, our findings could be informative for development cooperation in practice and development research likewise. Our results might be important for potential replication or scaling of interventions by local actors that were previously implemented by international agents. Extra effort might be needed to generate a positive, supportive behavior towards the intervention if solely implemented by local agents (or probably vice-versa in countries with higher trust in local than international implementers). Generally, trust towards both groups is high in the Indonesian case. Interestingly, those respondents that have already been exposed to previous internationally-led research interventions take a more positive stance towards future international projects. This relationship cannot be established for those who already participated in local research projects. Overall, the results suggest that previous experience with the respective agents influences the attitude and support for future interventions, underscoring the importance of responsible conduction of interventions.

By investigating potential pathways of the framing effects, particularly previous project participation towards the respective agents, we aim at opening this 'black box' to foster the explanation of the effects found. The chapter also stresses the effect of the salience of project implementers to influence support and contribution towards an intervention in case trust levels towards the implementer are high. However, experiences with local and international actors might differ across contexts. For this reason, our results can be considered as one of the first steps of evaluating the TPB and, more specifically, *attitudes* towards implementers, experimentally.

This study provides evidence in favor of an active consideration of the TPB determinants in the design and implementation of interventions to increase uptake, cooperative behavior, and general support by the targeted population. Certainly, researchers and practitioners will already have intuitively taken determinants of the TPB into account when designing their intervention. In our study, however, we argue for a systematic application of the TPB to increase interventions' success. A qualitative investigation prior to the project implementation and close cooperation with people knowing the local context to identify behavioral, normative, and control beliefs (that underlie the TPB determinants) within the study sample is recommended (Protogerou et al. 2012). Following the logic of the TPB, changing the respective beliefs in the appropriate direction will increase supportive behavior towards the intervention (Hobbis and Sutton 2005). Further research needs to contribute to a clearer understanding by randomly altering the other TPB determinants or replicating results in different settings. This way, important knowledge can be gained to improve not only research interventions, but also practical development cooperation in more general terms.

APPENDIX
FOR CHAPTER 4

Figure A4.1. Study design flow chart

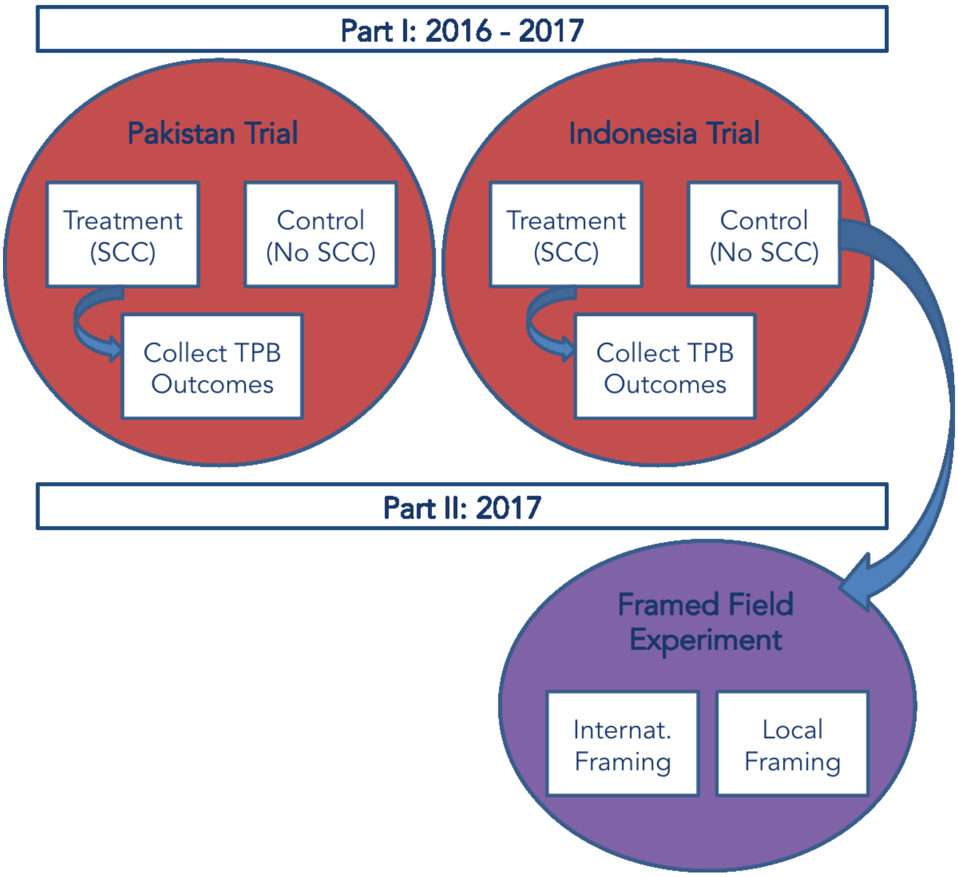
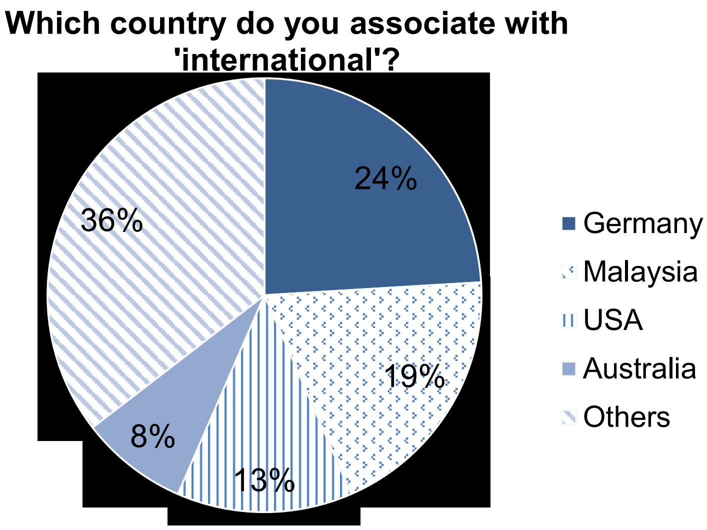


Figure A4.2. 'International' country concept



Appendix A4.1. Experimental Protocol[209]

General Remarks

If respondent asks you something, kindly answer by mentioning that you are only involved as an enumerator in the project and that you do not have any information on the Safe Childbirth Checklist. Furthermore, please connect the respondent with the contact number, which has been stated before

Of course if there are misunderstandings, you should repeat the provided information. However, please do not explain the information in different words.

Part A

"Now, we would like to present you a new tool and would like to learn about your opinion towards it."

[Before the start of the experiment (after the completed Survey); give 25.000 IDR voucher to the respondent]

"This is in appreciation of your time. Thank you very much. Subsequently, we will provide you with some information on a new tool for health-care in Aceh province. After this, you can decide whether you want to take the money for yourself or if you want to contribute some for the implementation of this tool."

Part B

[Enumerator: Please, read this introduction out aloud and clear.])

"During complex events, like performing a surgery or a delivery, people can be forgetful or might be distracted by other emergencies or duties. This can potentially have terrible consequences, in the worst case losing the patient. Research proofs that checklists can save lives and prevent these mistakes. Like a surgeon is responsible for patients' lives in the operation theater, the delivery team can have great impact on the safety of mothers and babies. We would like to present you a new tool, which was developed especially for your everyday work: The Safe Childbirth Checklist. It comprises 30 easy to use items. The checklist begins with the admission of the patient and ends with the discharge of mother and baby from the hospital. In each delivery, the doctor or midwife fills in one checklist for every patient. You will fill-in the checklist step by step and the checklist will remind you to

---

[209] The Indonesian version of the experimental protocol is available upon request.

perform the important steps during delivery. If you would like to know more about the checklist, here it is."

[Enumerator: Please hand a checklist copy over to the doctor or midwife.]

"For example, the checklist reminds you to perform easy things, which are nevertheless very important like hand washing."

[Enumerator: Show item "Confirm supplies are available to clean hands and wear gloves for each vaginal exam." on checklist]

"The checklist also reminds you to share important information with patients, including danger signs."

[Enumerator: Show item "Danger Signs" on checklist to the midwife or doctor]

"All these steps are already part of the study curriculum. Hence, every checklist item is easy to understand. Generally, most of the health workers already practice these important steps in the delivery process. The checklist just has the purpose to remind you of all the important steps during the delivery process. Especially, when health practitioners are under a lot of pressure, e.g., during night shifts or if complications arise, it can be very helpful. For instance, a research study has proven that during surgeries simple checklists can help to reduce death rates even by almost half."

Part C

"Among other researchers, [INTERNATIONAL / LOCAL] researchers took an active role in introducing the checklist to 17 facilities in Aceh province. The research team received approval from the provincial health office of Aceh. However, no funding was provided by the provincial health office. [LOCAL / INTERNATIONAL] research assistants and [INTERNATIONAL / LOCAL] health professionals with a lot of experience in delivery services were important partners and greatly supported the project."

Part D

"I will now read to you information about the funding of the Safe Childbirth study conducted by the [INTERNATIONAL / LOCAL] researchers. The following is a page of paper containing information on the checklist."

[Enumerator: Please hand over the SCC leaflet to the participant]

Introducing the Safe Childbirth Checklist in Aceh, Indonesia

In 2013, 6.5 million children under the age of five died worldwide of mainly preventable causes. Neonatal and maternal death can be prevented in many cases. Aceh province has reported neonatal death rates above the Indonesian level, with 28 per 1,000 live births (Statistics Indonesia, 2013).

Research has proven that checklists can help remember essential steps while complex procedures like flying an airplane or childbirth and thus minimize the risk of death.

A new tool has been developed, the Safe Childbirth Checklist (SCC), since 2008. The checklist targets (aims to reduce) the major causes of maternal and neonatal deaths. It comprises 28/29 items, which lead (function as guidance) through the childbirth procedure step-by-step. The first item is checked at admission of the patient and is finalised at the discharge of the mother. The checklist is designed (made) to provide assistance in these important and demanding (smth like crucial) situations.

For instance, it reminds of wearing gloves or the communication (to inform about any existing danger signs) of maternal and neonatal danger signs. These steps are part of the study curriculum and comprehensive standards of procedures. However, when health practitioners are under a lot of pressure, e.g. during night shifts or if complications arise, it can be very helpful. However, in a stressful situation one can be forgetful and the easy use of the checklist enables the pursuit of a safe birth.

Although this research is independent from the Indonesian government, it received approval by the provincial health office. The checklist was adapted to the Acehnese context in order to meet the local requirements.

The checklist was introduced to 17 health facilities in Banda Aceh, Bireuen and Aceh Besar. To provide all necessary information, the introduction of the checklist took place within the scope of a workshop and was complemented by continuous coaching visits over six months. After these six months, the researchers provide feedback. It is planned to introduce the checklist with the same procedure to your facility in 2017.

All information that is provided (by the health facilities) will be treated with highest (strictest) confidentiality and serves only for the purpose of improving neonatal and maternal health. Although the research may not change things in the short run, the hope is that this report will be helpful to governments when planning childbirth services in the future.

"The funds for the study have been used to implement the Safe Childbirth Checklist in 17 health facilities in Aceh province during October 2016. Funds are still available to introduce the checklist to 16 further facilities. The budget is enough to provide the 17 health facilities over six months with checklist copies. Therefore, every delivery during these six months can be conducted with the checklist. After this survey ends, the first six months of the checklist implementation are also over. There will be no funds remaining to provide additional checklists to those 17 health facilities, where the checklist was already introduced before."

Part E

"The researchers are collecting funds to be able to provide checklist copies at those health facilities. Are you willing to support the activity? Remember that the money collected will exclusively be used to provide Checklist copies to the health facilities. The total amount of money that was contributed by all donors together will be made transparent. After finalizing the data collection, the amount of money collected will be published openly in every participating facility of this research."

"If you would like to support the activity, please decide on the amount of money you would like to contribute and note it down on the voucher. You can choose to not contribute at all, or you can give 5,000, 10,000, 15,000, 20,000 or 25,000 IDR. Every contribution can help to conduct more deliveries with a Safe Childbirth Checklist. When you are done, please put the voucher in the envelope and seal it. If you do not wish to contribute anything, please put the number 0 on the voucher. In the end, only the aggregate amount of contributions from all participating facilities will be announced. Your individual contribution will be treated confidentially."

Part F

[Enumerator: Read this introduction out aloud to the participant]

"During the following task you have to estimate the most chosen answer, which neither refers to the total amount nor the average."

"We have asked also other health practitioners/workers in the district how much is their willingness to contribute to the provision of checklist copies. Which amount do you think was contributed to the checklist copies by your colleagues per person at other facilities? This estimation is not at all related to your personal opinion. Instead, we would like you to estimate which amount of contribution that was given by most of the other health practitioners per person."

"For this question, if you assessed the most chosen amount per person correctly, you will be given an additional 10,000 IDR. If you estimated the right amount, the 10.000 IDR will be topped up to your phone credit together with the voucher within the next few days."

"The other health practitioners also had to choose to contribute 0; 5,000 ; 10,000 ; 15,000 ; 20,000 or 25,000 IDR. Which category do you think was the most frequently chosen by the health workers?/ Which amount do you think most other health workers chose to contribute per person?"

Part G

"Your facility is one of the other 16 facilities, where the research team would like to implement the Safe Childbirth Checklist. Experience shows that checklist use needs to be practiced with coaches regularly in order to make deliveries safer. How committed are you in investing your time to practice the use of the checklist in every week?"

Debriefing

"Thank you very much for your participation. We asked you previously several questions. The aim is to find out what is your opinion about [local/international] researchers and how this opinion influences your motivation to use the Safe Childbirth Checklist. The checklist was previously pilot tested in other countries around the world. This way the most crucial practices during child delivery were identified. The research collaboration was led by the Harvard School of Public Health and the World Health Organization. Local researchers from Syiah Kuala University worked together with international researchers to adapt the checklist to the local context. Both parties hope that the Safe Childbirth Checklist can be implemented sustainably to serve as a tool for safe deliveries in Aceh province."

"Thank you very much for your participation. We asked you previously several questions. The aim is to find out what is your opinion about [local/international] researchers and how this opinion influences your motivation to use the Safe Childbirth Checklist. The checklist was previously pilot tested in other countries around the world. This way the most crucial practices during child delivery were identified. The research collaboration was led by the Harvard School of Public Health and the World Health Organization. Local researchers from Syiah Kuala University worked together with international researchers to adapt the checklist to the local context. Both parties hope that the Safe Childbirth Checklist can be implemented sustainably to serve as a tool for safe deliveries in Aceh province."

"If these information change your attitude towards contributing to the checklist copies in any way, you are free to to change your indicated contribution."

[Enumerator: If the respondent decides to change his/her contribution, please hand the envelope back.]

Table A4.1. Balancing table for the experimental sample – comparison between treatment (international framing) and control (local framing)

Variable	N Full Sample	Mean Full sample	SD Full sample	Mean Control	SD Control	Mean Treatment	SD Treatment	p-value of difference in means t-test
Facility Type	236	1.538		1.690		1.433		0.021**
Respondent's Gender (1=m, 2=f)	236	2.000		2.000		2.000		
Respondent's Age	236	33.314	7.493	33.650	7.806	33.112	7.316	0.593
Respondent's Education (Yrs)	236	15.051	0.527	15.020	0.603	15.067	0.462	0.619
Experience (Yrs)	236	9.576	7.271	9.690	7.736	9.537	6.979	0.886
Sufficient Income	236	3.208	1.008	3.160	1.012	3.246	1.014	0.526
Financial problems	236	1.678		1.720		1.642		0.081*
Donate Less if others donate	236	4.657	1.264	4.710	1.225	4.627	1.296	0.564
Social Acceptability Index	236	3.411	0.838	3.450	0.821	3.381	0.857	0.513
Social Acceptability Item 1	236	4.966	0.690	5.000	0.778	4.940	0.622	0.480
Social Acceptability Item 2	236	4.568	1.027	4.600	0.932	4.545	1.101	0.650
Social Acceptability Item 3	236	5.343	0.558	5.310	0.506	5.366	0.595	0.172
Social Acceptability Item 4	233	4.644	1.074	4.694	1.069	4.602	1.087	0.475
Social Acceptability Item 5	236	2.229	1.254	2.250	1.298	2.216	1.235	0.784
Paperwork takes too much time	236	2.814	1.343	3.000	1.497	2.664	1.195	0.173
Routines make work easier	236	5.153	0.734	5.150	0.626	5.179	0.764	0.660
Previous checklist experience	236	2.564	1.831	2.500	1.795	2.627	1.871	0.536
Previous checklist use	236	0.547		0.540		0.560		0.772
Access to essential resources	236	3.470	0.517	3.530	0.502	3.425	0.526	0.080*
Team efficacy indicator	236	5.246	0.513	5.220	0.462	5.261	0.547	0.570
Participation in local projects	236	1.831		1.870		1.806		0.235
Participation in int projects	236	1.898		1.880		1.910		0.511
Participation in donor projects	236	1.907		1.920		1.896		0.511

* Notes: Proportions in the two groups are significantly different from each other, $p \leq 0.05$ (t test). Based upon the full sample with N denoting the number of observations, SD gives the standard deviation. Standard errors were clustered at the facility level.

Table A4.2. Balancing table – Experiment and respective comparison group of SCC intervention

Variable	N Full Sample	Mean Full Sample	SD Full Sample	Mean SCC Intervention	SD SCC Intervention	Mean Experiment	SD Experiment	p-value of difference in means t-test
Facility Type	335	1.676		1.859		1.503		0.002
Respondent's Gender (1=m, 2=f)	335	1.994		1.988		2.000		0.150
Respondent's Age	335	32.529	0.403	32.706	0.606	32.379	0.539	0.687
Respondent's Education (Yrs)	335	15.195	0.064	15.405	0.121	14.994	0.040	0.001
Experience (Yrs)	335	8.928	0.404	8.969	0.600	8.905	0.547	0.937
Resource Access	335	3.486	0.027	3.534	0.039	3.444	0.038	0.102
Team Efficacy	335	5.240	0.025	5.282	0.036	5.195	0.034	0.081

* Notes: Proportions in the two groups are significantly different from each other, $p \leq 0.05$ (t test). 'Full Sample' refers to the pooled Indonesian SCC intervention (treatment and control group), 'SCC Intervention' to the treatment group of the SCC intervention, and 'Experiment' to the SCC intervention's control group where framing experiment was conducted. N denotes the number of observations, SD gives the standard deviation. Standard errors are clustered at the facility level.

Table A4.3. Balancing table – Reduced sample used for empirical analysis (excluding those with prior SCC contact)

Variable	N Full Sample	Mean Full sample	SD Full sample	Mean Control	SD Control	Mean Treatment	SD Treatment	p-value of difference in means t-test
Facility Type	170	1.500		1.618		1.409		0.050
Respondent's Gender (1=m, 2=f)	170	2.000		2.000		2.000		
Respondent's Age	170	32.359	6.997	33.118	7.680	31.774	6.395	0.232
Respondent's Education (Yrs)	170	14.994	0.516	14.974	0.565	15.011	0.478	0.742
Experience (Yrs)	170	8.888	7.094	8.974	7.494	8.849	6.824	0.908
Sufficient Income	170	3.200	1.069	3.118	1.083	3.269	1.065	0.348
Financial problems	170	1.741		1.763		1.720		0.396
Donate Less if others donate	170	4.606	1.411	4.658	1.381	4.581	1.440	0.613
Social Acceptability Index	170	3.329	0.827	3.316	0.852	3.344	0.814	0.808
Social Acceptability Item 1	170	5.000	0.738	4.987	0.887	5.011	0.599	0.834
Social Acceptability Item 2	170	4.459	1.142	4.461	1.026	4.462	1.239	0.991
Social Acceptability Item 3	170	5.429	0.584	5.408	0.521	5.452	0.634	0.436

Social Acceptability Item 4	167	4.545	1.063	4.649	1.065	4.457	1.063	0.239
Social Acceptability Item 5	170	2.118	1.286	2.184	1.334	2.065	1.258	0.375
Paperwork takes too much time	170	2.906	1.364	3.145	1.547	2.720	1.174	0.150
Routines make work easier	170	5.100	0.727	5.079	0.648	5.151	0.722	0.471
Previous checklist experience	170	2.765	1.983	2.632	1.945	2.882	2.026	0.298
Previous checklist use	170	0.541		0.553		0.538		0.854
Access to essential resources	170	3.441	0.498	3.513	0.503	3.387	0.490	0.060
Team efficacy indicator	170	5.200	0.443	5.158	0.434	5.226	0.445	0.459
Participation in local projects	170	1.829		1.868		1.796		0.131
Participation in int projects	170	1.918		1.895		1.935		0.272
Participation in donor projects	170	1.935		1.934		1.935		0.959

* Notes: Proportions in the two groups are significantly different from each other, $p \leq 0.05$ (t test). Based upon the reduced sample excluding observations with prior contact to the checklist. N denotes the number of observations, SD gives the standard deviation. Standard errors were clustered at the facility level.

Table A4.4. Summary statistics for Indonesian data

Variable	N Full Sample	Max Full Sample	Min Full Sample	Mean Full sample	SD Full sample
Actual Behavior					
Active SCC Use	219	1	0	0.389	0.489
Intended Behavior					
Would try to use SCC even if copies are not provided	163	6	3	4.847	0.634
Would recommend the SCC to fellow colleagues	163	6	2	5.092	0.495
Using the SCC in my professional role is	163	6	4	5.325	0.483
Ease to use SCC in work environment	163	6	4	5.141	0.565
SCC is supported by superiors	163	6	4	5.828	0.439
Urban (1) – Rural (2) Dummy	163	2	1	1.515	0.501
CEmONC Service Provision 24/7	163	1	0	0.178	0.384
Facility Type: Community Health Centre	163	1	0	0.589	0.494
Facility Type: Public Hospital	163	1	0	0.135	0.343
Facility Type: Private Hospital	163	1	0	0.190	0.394
Facility Type: Private Midwife Clinic	163	1	0	0.086	0.281
District: Aceh Besar	163	1	0	0.276	0.448
District: Banda Aceh	163	1	0	0.331	0.472
District: Bireuen	163	1	0	0.393	0.490

Table A4.5. Summary statistics for Pakistani data

Variable	N Full Sample	Max Full Sample	Min Full Sample	Mean Full sample	SD Full sample
Actual Behavior					
Active SCC Use	212	1	0	0.344	0.476
Intended Behavior					
Would try to use SCC even if copies are not provided	78	6	1	4.628	1.452
Would recommend the SCC to fellow colleagues	78	6	1	5.141	1.090
Using the SCC in my professional role is	79	6	1	5.380	0.821
Ease to use SCC in work environment	79	6	1	4.962	1.305
SCC is supported by superiors	58	6	1	5.155	1.508
Urban (1) – Rural (2) Dummy	80	1	0	0.813	0.393
Open 24/7	80	1	0	0.150	0.359
Facility Type: Health Facility	80	1	0	0.2125	0.412
Facility Type: Community Midwife	80	1	0	0.5625	0.500
Facility Type: Lady Health Visitor	80	1	0	0.225	0.420
District: Haripur	80	1	0	0.45	0.501
District: Nowshera	80	1	0	0.55	0.501

Appendix A4.2. Social desirability index

We modify social desirability questions developed by Kemper et al. (2014) to reflect social desirability norms in the Acehnese context. The social desirability index was constructed by adding up the top categories (5 and 6) indicated in the subsequent questions.

1. "In an argument, I always remain objective and not become emotional." 2. "Even if I am sad, I always smile when talking to others." 3. "When talking to someone older, I always listen carefully to what she/he says." 4. "When I had the chance to donate for religious purposes, I always contributed a lot." 5. "Sometimes I only help people if I hope to get something in return."	1. Disagree strongly 2. Disagree 3. Rather disagree 4. Rather agree 5. Agree 6. Agree strongly 7. Not applicable

Table A4.6. Main Results Intended and Actual SCC Use with Wild Bootstrapped SE

	Intended SCC Use: Would try to use SCC even if copies are not provided (1 "disagree strongly" - 6 "agree strongly")	Actual SCC Use: Was Checklist actively used or looked at during delivery process?	
	Indonesia	**Pakistan**	**Indonesia**
	(2a)	(3a)	(4a)
Attitudes: Using the SCC in my professional role is: 1 "completely useless" - 6 "completely useful"	0.4535***	0.655***	-0.364
WB p-value	(0.00400)	(0.00000)	(0.50450)
Subjective Norms: SCC is supported by superiors: 1 "not at all" - 6 "completely"	0.5356*	0.207	0.642
WB p-value	(0.0721)	(0.32032)	(0.50250)
Perceived Behavioral Control: Ease to use SCC in work environment: *1 "very difficult" - 6 "very easy"*	0,2614	0.306***	0.0381
WB p-value	(0.10210)	(0.00000)	(0.43243)
N	163	212	218
Control variables	no	no	no
Mean of dep. var.	4.847	0.344	0.389
Median of dep. var.	5	-	-
Standard Deviation of dep. var.	0.634	0.476	0.489

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, * p<0.1. All regressions are based upon the treated providers and standard errors (SE) are clustered at the facility level and wild cluster bootstrapped due to the small number of clusters (15), following Cameron, Gelbach and Miller (2008). No bootstrapping is provided for intended SCC use in Pakistan as a sufficient number of clusters (70) was sampled.

Table A4.7. Binary definition of outcome variable

	Would try to use SCC even if copies are not provided (Dummy variable=1 if "fully agree")			
	Pakistan		**Indonesia**	
	(1a)	(1b)	(2a)	(2b)
Attitudes: Using the SCC in my professional role is: 1 "completely useless" - 6 "completely useful"	0.930***	0.704**	0.4509***	0.3173**
p-value	(0.00717)	(0.02494)	(0.00564)	(0.01273)
Subjective Norms: SCC is supported by superiors: 1 "not at all" - 6 "completely"	0.508	0.244	0.6996***	0.4444***
p-value	(0.11763)	(0.47477)	(0.00864)	(0.00257)
Perceived Behavioral Control: Ease to use SCC in work environment: *1 "very difficult" - 6 "very easy"*	0.763**	0.675**	0.3027	-0.0573
p-value	(0.01099)	(0.04077)	(0.16626)	(0.74646)
N	78	78	163	163
Control variables	no	yes	No	yes
Mean of dep. var.	4.628	4.628	4.847	4.847
Median of dep. var.	5	5	5	5
Standard Deviation of dep. var.	1.452	1.452	0.634	0.634

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, * p<0.1. All regressions are based upon the treated providers and standard errors (SE) are clustered at the facility level. Adjusted regressions (b) additionally control for a variable indicating the facility type, a binary variable indicating rural/urban location, a variable indicating the district and for the Pakistan data a binary variable indicating whether the facility is open 24/7.

Table A4.8. Main Results – Alternative Outcome Measures

	Would recommend the SCC to fellow colleagues (1 "disagree strongly" – 6 "agree strongly")		Willingness to invest additional time for SCC project (in 5 min categories)		Estimated average contribution by fellow health staff (in IDR)		PCA index (all outcome variables)	
	(1a)	(1b)	(2a)	(2b)	(4a)	(4b)	(5a)	(5b)
Framing (=1 if 'international')	0.0491	0.1259*	-0.1505	0.0949	605.9285	769.9561	0.1082	0.3167**
p-value	(0.53505)	(0.05828)	(0.40361)	(0.62377)	(0.44686)	(0.30365)	(0.52512)	(0.01225)
RI p-value	(0.6000)	(0.1220)	(0.6680)	(0.7460)	(0.3420)	(0.2390)	(0.5840)	(0.0530)
WB p-value	(0.53053)	(0.07608)	(0.37037)	(0.57057)	(0.45245)	(0.28228)	(0.52452)	(0.01001)
N	167	167	167	167	167	167	167	167
Control variables	No	yes	no	Yes	no	yes	No	Yes
Mean of dep. var.	5.108	5.108	5.084	5.084	7365.269	7365.269	-0.117	-0.117
Std. dev. of dep. var.	0.581	0.581	2.237	2.237	3950.536	3950.536	1.289	1.289

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level. We present results based on clustered SE indicated as "p-values" and wild bootstrapped due to limited cluster number (13) for the specifications indicated as "WB p-values", following Cameron, Gelbach and Miller (2008). RI p-values are computed with a permutation test based on Hess (2017). Specifications (b) include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork.

Table A4.9. Ordered probit results

	Would recommend the SCC to fellow colleagues (1 "disagree strongly" – 6 "agree strongly")		Willingness to invest additional time for SCC project (in 5 min categories)		Financial Contribution in support of SCC project (in IDR)		Estimated average contribution by fellow health staff (in IDR)	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Framing (=1 if 'international')	0.1911	0.1259*	-0.0871	0.0949	0.0813	1,283.7717**	0.1291	769.9561
p-value	(0.31609)	(0.05828)	(0.23911)	(0.62377)	(0.59964)	(0.02119)	(0.53468)	(0.30365)
N	167	167	167	167	165	165	167	167
Control variables	no	yes	no	yes	no	yes	no	yes

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. Reported coefficients are not transformed and represent ordered probit coefficients. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level. Specifications (b) include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork.

Table A4.10. Covariates across outcome variables

	Would recommend the SCC to fellow colleagues (1 "disagree strongly" 6 "agree strongly")	Willingness to invest additional time for SCC project (in 5 min categories)	Financial Contribution in support of SCC project (in IDR)	Estimated average contribution by fellow health staff (in IDR)	PCA index (all outcome variables)
Facility Type (Base: Com. Health Clinics)					
Public Hospital	-0.0629	-1.0438	-3,444.5249***	415.6412	-0.7102*
p-value	(0.59493)	(0.07348)	(0.00001)	(0.81588)	(0.06426)
WB p-value	(0.65065)	(0.13413)	(0.00200)	(0.69469)	(0.20020)
Private Hospital	-0.2171	0.8256	-1,093.5731	1,162.3579	0.0420
p-value	(0.29571)	(0.26514)	(0.66748)	(0.33678)	(0.92297)
WB p-value	(0.30230)	(0.34434)	(0.54054)	(0.45445)	(0.87287)
Social Acceptability Index (5 components: 1'low' - 6 'high')	0.1318*	0.9340***	825.2196*	-81.4618	0.4459***
p-value	(0.07054)	(0.00000)	(0.09055)	(0.70392)	(0.00185)
WB p-value	(0.08208)	(0.00000)	(0.11411)	(0.63664)	(0.00000)
Paperwork takes too much time (1 "disagree strongly" - 6 "agree strongly")	-0.1485***	-0.6369***	-978.2253***	-599.9692**	-0.4430***
p-value	(0.00266)	(0.00010)	(0.00203)	(0.01911)	(0.00008)
WB p-value	(0.00400)	(0.00200)	(0.00200)	(0.01201)	(0.00400)

Financial problems	-0.0078	0.2482	1,266.2008	1,169.0571*	0.2960
p-value	(0.96621)	(0.55271)	(0.19761)	(0.06300)	(0.16188)
WB p-value	(0.98098)	(0.54254)	(0.24224)	(0.06206)	(0.13213)
N	167	167	165	167	167

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact (refer to Table A4.3.). Standard errors (SE) are clustered at the facility level. We present results based on clustered SE indicated as "p-values" and wild bootstrapped due to limited cluster number (13) for the specifications indicated as "WB p-values", following Cameron, Gelbach and Miller (2008).

Table A4.11. Framing – Full sample with prior contact control

	Would recommend the SCC to fellow colleagues (1 "disagree strongly" – 6 "agree strongly")		Willingness to invest additional time for SCC project (in 5 min categories)		Financial Contribution in support of SCC project (in IDR)		Estimated average contribution by fellow health staff (in IDR)		PCA index (all outcome variables)	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
Framing (=1 if 'international')	0.0582	0.1275**	-0.0478	0.1774	537.5565	1,206.2985*	458.1032	789.4081	0.1151	0.3230***
p-value	(0.29056)	(0.03919)	(0.79610)	(0.25000)	(0.44476)	(0.06153)	(0.59195	(0.24813)	(0.50173)	(0.00812)
WB p-value	(0.28629)	(0.04004)	(0.78478)	(0.24024)	(0.45045)	(0.05005)	(0.61862	(0.26026)	(0.46046)	(0.00801)
N	230	230	230	230	226	226	230	230	226	226
Control variables	no	yes	no	yes	no	yes	no	Yes	no	yes

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the full sample. Standard errors (SE) are clustered at the facility level. We present results based on clustered SE indicated as "p-values" and wild bootstrapped due to limited cluster number (13) for the specifications indicated as "WB p-values", following Cameron, Gelbach and Miller (2008). Specifications (b) include a binary variable indicating whether respondent was in prior contact with the checklist, a variable capturing the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork.

Table A4.12. Main Results – Full sample using interaction with prior contact

	Financial Contribution in support of SCC project (in IDR)	
	(a)	(b)
No Prior Contact X International Framing	557.6236	1,164.8298**
p-value	(0.39538)	(0.03264)
Prior Contact X Local Framing	225.9725	627.9609
p-value	(0.83548)	(0.54703)
Prior Contact X International Framing	706.5217	1,955.2291
p-value	(0.54728)	(0.10483)
N	226	226
Control variables	no	yes
Mean of dep. var.	4757.576	4757.576
Std. dev. of dep. var.	4711.366	4711.366

Notes: Asterisks indicate p-values based on standard errors clustered at the facility level according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level. Specifications (b) include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork. The base category is No Prior Contact and Local Framing.

Table A4.13. Main Results: Wild Bootstrapped SE

	Financial Contribution in support of SCC project (in IDR)	
	(a)	(b)
Framing (=1 if 'international')	557.6236	1,283.7717**
WB p-value	(0.40440)	(0.03203)
N	165	165
Control variables	no	yes
Mean of dep. var.	4757.576	4757.576
Std. dev. of dep. var.	4711.366	4711.366

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level and wild boostrapped due to limited cluster number (13) for the specifications indicated as "WB p-values", following Cameron, Gelbach and Miller (2008).). Specifications (b) include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork.

Table A4.14. Main Results – Controlling for elicitation

	Financial Contribution in support of SCC project (in IDR)
Framing (=1 if 'international')	852.6100*
p-value	(0.06413)
Elicited Contribution of Others as control	0.5000***
p-value	(0.00163)
N	165
Control variables	yes
Mean of dep. var.	4757.576
Std. dev. of dep.	4711.366

Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level. The specification include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork. Moreover, the elicited contribution of health practitioners from other facilities is added as a control variable.

Table A4.15. Association between previous project participation and trust

	Trust in local actors (1"not at all" - 4"great deal")		Trust in international actors (1"not at all" – 4 "great deal")		Trust in foreign countries (average trust level)	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Participation international project	0.6039***	0.4738***	0.2857	0.3676**	0.2522*	0.3782***
p-value	(0.00731)	(0.00775)	(0.11531)	(0.01980)	(0.08280)	(0.00202)
WB p-value	(0.00000)	(0.00000)	(0.09610)	(0.00000)	(0.11211)	(0.01201)
Participation local project	0.0647	0.1395	0.3017	0.3115*	0.3985***	0.3699***
p-value	(0.81084)	(0.56675)	(0.13285)	(0.06467)	(0.00251)	(0.00021)
WB p-value	(0.79079)	(0.54454)	(0.10210)	(0.04004)	(0.00200)	(0.00000)
N	168	168	168	168	168	168
Control variables	No	yes	No	Yes	no	Yes

Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the sample limited to those respondents without prior SCC contact. Standard errors (SE) are clustered at the facility level. . We present results based on clustered SE indicated as "p-values" and wild bootstrapped due to limited cluster number (13) for the specifications indicated as "WB p-values", following Cameron, Gelbach and Miller (2008). Specifications (b) include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork.

Table A4.16. Association between framing and potential channel variables

	Control Capabilities	Implementation Skills	Funding Capabilities	Accountability	Trust Foreign Countries	Participation International Project	Participation Local Project
Framing (=1 if 'international')	0.8015***	0.7738***	0.6040***	0.4445*	0.0447	0.0225	-0.0652
SE	(0.214)	(0.210)	(0.188)	(0.243)	(0.051)	(0.047)	(0.055)
p-value	(0.00247)	(0.00275)	(0.00670)	(0.08993)	(0.39321)	(0.63772)	(0.25714)
WB p-value	(0.00400)	(0.00801)	(0.00801)	(0.11812)	(0.37437)	(0.71872)	(0.22422)
N	230	230	230	230	230	230	230

Notes: Asterisks indicate p-values according to: *** p<0.01, **p<0.05, *p<0.1. All specifications are based upon the full sample. Standard errors (SE) are clustered at the facility level. All specifications include a variable indicating the facility type, a binary variable indicating if the respondent had financial problems, a composite index of social desirability variables and a variable indicating the subjective perception of the amount of paperwork.

CHAPTER 5

The Impact of Impact Evaluations:

Cross-Country and Sub-National Evidence

Single-authored

5.1. INTRODUCTION

"Success depends on knowing what works."
– Bill Gates, Co-Chair
Bill & Melinda Gates Foundation

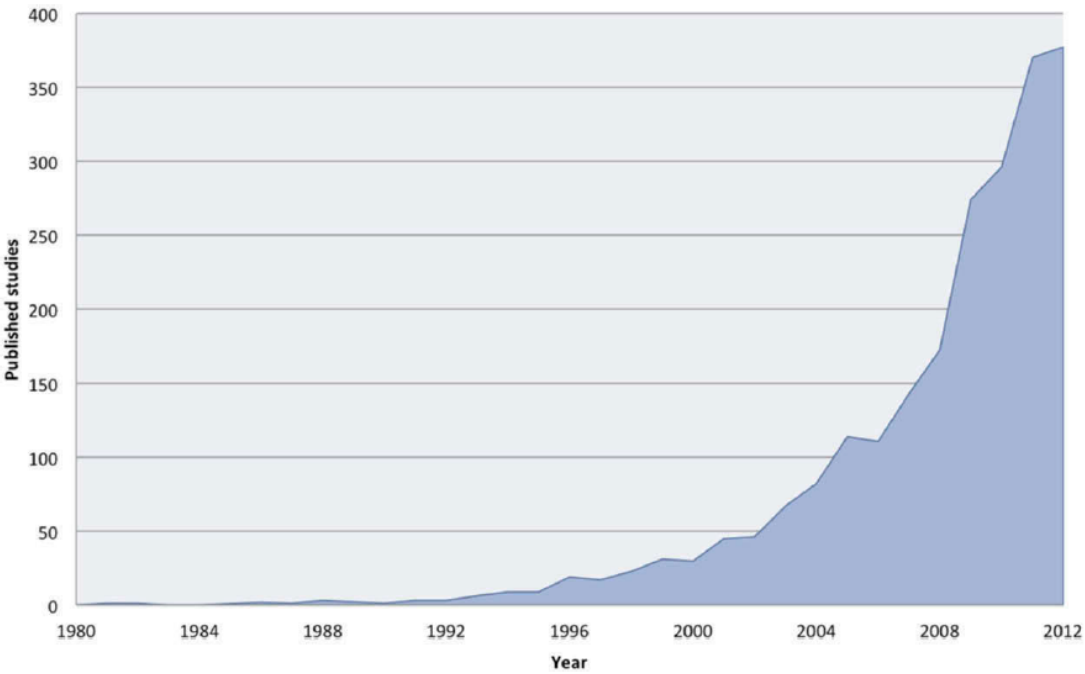
The likewise simple and wise statement of Bill Gates illustrates exactly the idea behind impact evaluations (IEs) in development economics. With impact evaluations, we try to find out, what in interventions enhances development and what not. At the same time, evidence is scarce on whether this strategy of employing impact evaluations actually works successfully in leading to improved outcomes – their final purpose and meaning of existence. This evidence gap remains, in spite of the wealth of experience coming from a history of almost 40 years with impact evaluations in development economics, which should be enough to draw conclusions. Challenges for evaluations to be effective, are large. Even most rigorously conducted evaluations that find significant evidence cannot have an impact by themselves, but only through policy practitioners that employ this evidence to design, improve, up-scale or discontinue their projects (Shah et al. 2015).[210] Similarly important to evidence in aid effectiveness is therefore rigorous evidence in impact evaluation effectiveness,

[210] Beyond reaching the policy level, Banerjee et al. (2017) describe six further challenges which might impede real-world benefits.

which allows a judgement of whether the money spent is well invested or whether something in their mechanics needs refinement.

Another argument why impact evaluations' effectiveness needs evidence is based on their opportunity costs. Resources which are increasingly channeled into one area are likely to be withdrawn from another in development economics. The rise in impact evaluations of international development interventions is particularly well documented by Cameron et al. (2016). From the year 2000 onwards, impact evaluations were growing at a rapid pace, as illustrated in Figure 5.1. (Cameron et al. 2016).

Figure 5.1. Increase in impact evaluations published as journal articles, working papers, reports and book chapters per year (1981-2012)



Source: Cameron et al. (2016, 7)

Overall, 66 percent (1491/2259) of all listed evaluations employ randomized control trials (RCTs), which are likely to be conducted by researchers. The increase in research output goes hand in hand with increased resources channeled into evaluating impacts. As a proxy for this shift in resources to RCT studies, Banerjee, Duflo and Kremer (2016) refer to the composition of the association of development economists, Bureau for Research and Economic Analysis of Development (BREAD). While around 34 percent of the fellows and associates who obtained their PhD between 1981 and 1990 conducted at least one RCT, the fraction doubles for the ones with PhD years between 2006 and 2015. For current PhDs, the absolute and relative number increases further. Accordingly, the share of RCT-papers presented at the annual conference of BREAD increased from 8 percent in 2005 to 63 percent in 2010 and consolidated between 40 to 50 percent afterwards. With regard to publications in top journals,[211] Banerjee et al. (2016) document a quite significant number of 10 RCT-studies out of 32 development papers in total in the year 2015, while none of the 21 development papers published in top journals were RCTs in the year 2000.

Critics fear a crowding out of other strands in development economics (Ravallion 2012; Deaton 2010), while advocates see rather beneficial effects of RCT-studies for the discipline as a whole (Banerjee et al. 2016). Banerjee et al. (2016) link the general advancement in clean identification strategies in development economics not least to the rise of randomized interventions, which could stimulate finding creative ways for causal identification strategies beyond random settings. The research question of this chapter, however, is not so much what the rise of impact evaluations imply for the scientific side, but whether there are measurable real-world impacts. Nonetheless, this chapter is no meta-study, assessing what makes similar evaluations work. It rather aims at statistically quantifying a potential direct effect of impact evaluations on the outcomes their interventions hint at – via improved international development or national projects.

A growing, mainly qualitative literature, addresses exactly this question of how much influence impact evaluations have on real-world policy and social outcomes. They unite in finding it generally difficult to judge whether impact evaluations successfully influence policy. Shah et al. (2015) measure the effectiveness of IEs in influencing policy by the number of IEs that are up-scaled.[212] Turning to 626 evaluations conducted through the Abdul Latif Jameel Poverty Action Lab (JPAL) for instance, only around 15 were up-scaled. Proportions are similar for evaluations from the Innovations for Poverty Action (IPA). Accordingly, the authors see a rather weak effect of impact evaluations leading to direct action. However, looking at mere scale-ups overlooks the potential influence of impact evaluations leading to improved designs or termination of poorly functioning programs (Buddlemeyer and Skoufias 2003). Baanante and Valdivia (2015) draw a slightly more positive picture. They ask 19 impact

[211] Banerjee et al. (2016) include the American Economic Review (AER), Quarterly Journal of Economics (QJE), Journal of Political Economy (JPE), Review of Economic Studies (Restud) and Econometrica in their list of top journals.

[212] With "upscaling" I refer to the extension of projects to further project sites (e.g., to other facilities or provinces).

evaluation experts to rate the level of policy influence of the most policy relevant impact evaluations they had conducted. Out of 37 impact evaluations, 68 percent (25/37) were rated to have influenced policy decisions, while 32 percent (12/37) had not. Banerjee et al. (2016) choose another approach to measure evaluations' real-world influence and include 43 evaluations conducted through USAID's Development Innovation Ventures (DIV) between 2010-2012.[213] They measure real-world influence by the estimated number of people reached through the innovations' evaluations and their adaptions. They find that 14 percent (6/43) of the innovations have reached more than one million people after 3-5 years and 30 percent (13/43) have reached more than 100,000 people within 3-5 years. At the same time, this implies that a relatively high fraction of 44 percent (19/43) of the evaluations were successfully up-scaled. The authors identify cost-effectiveness and the involvement of researchers in the evaluation design as determinants for success. Legovini et al. (2016) take a quantitative stance and evaluate the influence of the World Bank's impact evaluations on project characteristics for 100 evaluations and 1135 World Bank projects between 2005 and 2011. They find that projects with IEs are disbursed in a timelier manner and have decreased differences between commitments and disbursements. Overall, evidence on the success of impact evaluations to directly inform and influence policy decisions is thus limited, so far, and mixed. However, none of the authors doubts the important influence of impact evaluations in generating general knowledge on development effectiveness. Shah et al. (2015) further differentiate the success probability of impact evaluations between being supply and demand driven. Naturally, demand driven IEs, i.e., having been asked for by project operators or policy have a much greater chance to influence policy levels. However, following Jones et al. (2009), IEs tend to be rather supply-driven in all but the health and social development sector. Nevertheless, the impact of impact evaluations could still be limited with regard to improving outcomes, if they are primarily used by policymakers to legitimize actions (which would have been undertaken anyway) and defend budgets, as found by work of Jones et al. (2009) and the World Bank (2009).

The existing literature is accordingly mainly characterized by qualitative evidence based on small samples, to which this chapter adds a quantitative analysis of the real-world

[213] USAID awards each year innovative solutions to development challenges and provides funding for their rigorous evaluation and potential up-scaling (https://www.usaid.gov/div). Accordingly, these evaluations can be assumed to be disproportionately successful.

effects of impact evaluations, in the sense of analyzing the effect on social outcomes based on a comprehensive sample. The current approach is feasible thanks to the newly built database of the International Initiative for Impact Evaluation (3ie), which includes all impact evaluations with development focus in developing countries from 1981-2012. To the best of my knowledge, this is the first time that this rich database is explored beyond descriptive relationships (as in Cameron et al. 2016). I focus the analysis on impact evaluations in the health sector only, which accounts for around 50 percent of the interventions being conducted so far (Cameron et al. 2016). In an analysis on three different levels, I first explore cross-country, macroeconomic average effects, second, sub-national district-level panel-analysis for Uganda, which allows a more thorough analysis of the mechanisms behind, and third, birth-level quasi-panel evidence. The cross-country regression analysis with 81 countries in the main specification explores the effect of the number and stocks of impact evaluations in health per country on changes in Disability-Adjusted Life Years (DALYs) two years later. At the sub-national level, I link the number and stocks of specific impact evaluations on maternal and newborn health to projects within this array and examine their interacted effects on district-level as well as disaggregated birth-level infant mortality rates below the age of 12 months.

The obvious challenge in identifying causal effects of impact evaluations is that IEs themselves are not randomly allocated (Banerjee et al. 2016). The analysis thus needs to carefully address endogeneity. Generally, the lag structure that I apply diminishes concerns. Following my line of argument outlined in the chapter, evaluators collect endline data in t-3, one year later, they inform the policy about their results (t-2). Policymakers or practitioners would then be able to improve, up-scale or down-scale their projects in t-1, which would lead to improved outcomes in t if there is a measurable direct effect of IEs. The decision to conduct an impact evaluation could therefore not be taken later than four years before I measure potential effects on outcomes, rather earlier. The actual timing of impact evaluations depends on many exogenous factors (i.e., research permit, researchers' availability, shifts in the implementation schedule due to unforeseen barriers in the field, etc.). Therefore, concerns about factors that are systematically linked to the announcement of evaluation results, are reduced. Nevertheless, I discuss several potential biases through, for instance, more governmental effort leading to more health IEs or more research interest after the cessation of a conflict, which could both coincide with generally improving health outcomes. Consequently, the correlation with the outcome measures applied in this paper (DALYs or infant mortality) would be negative, such that the omission of these variables would downward bias my results.

The three levels of analyses offer different opportunities to tackle the biases. On the cross-country-level, I control for the obvious, measurable confounders by a large set of control variables, as well as country-specific linear time trends to capture the general improvement of health outcomes through increasing living standards and technological or medical advancements. Year-fixed effects additionally control for non-linear events like global climate shocks or crises affecting health.

The analysis of the sub-national level for Uganda and its large advantage to work with more precisely measured data further reduces endogeneity concerns. I code district-level development projects from international development cooperation and country-level health IEs into eight narrow health categories, such as maternal and newborn health or HIV/Aids. I then focus the analysis on impact evaluations and development projects in maternal and newborn health and their interacted effect on infant mortality. Only if there is a project in maternal and newborn health in t, and a corresponding evaluation from the same health category disseminates evaluation results in t-1, the interaction is larger than zero. I call these incidences a "match". The identification strategy mainly builds on accurately identifying the matches and comparing them to incidences when development action through projects take place, but cannot rely on information provided through IEs.

My final estimation at the birth-level employs fixed effects for mothers and, thus, exploits variation among siblings, depending on whether their mother could benefit from an impact evaluation-project match or not. Many confounders are automatically controlled for through the application of mother-fixed effects like time-invariant characteristics of mothers' surroundings (including district characteristics or vegetation) and mothers' socio-economic status. Year-fixed effects control additionally for events out of mothers' control, like epidemics or nation-wide changes in governmental regulations.

Cross-country level results show a significant correlation between impact evaluations in health and improved health outcomes. One additional impact evaluation in health reduces the average number of DALYs by 0.1 percent. Sub-national evidence supports the findings found at an aggregated level. Different specifications show robustly a reduced infant mortality rate after matching impact evaluations and projects in maternal and newborn health. Mean health aid disbursements of 38,601 US$ lead to a decreasing IMR by 0.5 in 1,000 live births if an impact evaluation takes place, compared to disbursements without an impact evaluation. The robustness of the findings in different specifications and levels of analyses raises confidence that health impact evaluations conditional on corresponding international development projects lead indeed to reduced infant mortality.

"Nearly five trillion dollars seems a high price to pay for uncertainty and misunderstanding[.]" (Tierney et al. 2011, 1891). The statement of Tierney et al. (2011) underlines the objective of this chapter in adding to the literature that tackles aid effectiveness, which is still characterized by *"uncertainty and misunderstanding"* about what makes aid effective. First, the chapter contributes to filling the knowledge gap in aid effectiveness on sub-national levels (Kotsadam et al. 2018; Dreher and Lohmann 2015) in utilizing newly available fine-grained data from AidData (Tierney et al. 2011). Second, the presented evidence adds to understandings about the importance of project implementation processes, where evidence is generally scarce (e.g., Marchesi and Masi 2018; Kaplan et al. 2018/Chapter 4). Third, the chapter contributes to scholarly work on the role of differing donor quality (e.g., Minasyan et al. 2017/Chapter 1; Roodman 2012; Bermeo 2011; Minoiu and Reddy 2010) and ineffectiveness of politically motivated aid (e.g., Dreher et al. 2018; Kilby and Dreher 2010), where the ineffectiveness of poor-quality donors or politically motivated aid could be partly explained by their unlikeliness to take insights from impact evaluations into account.

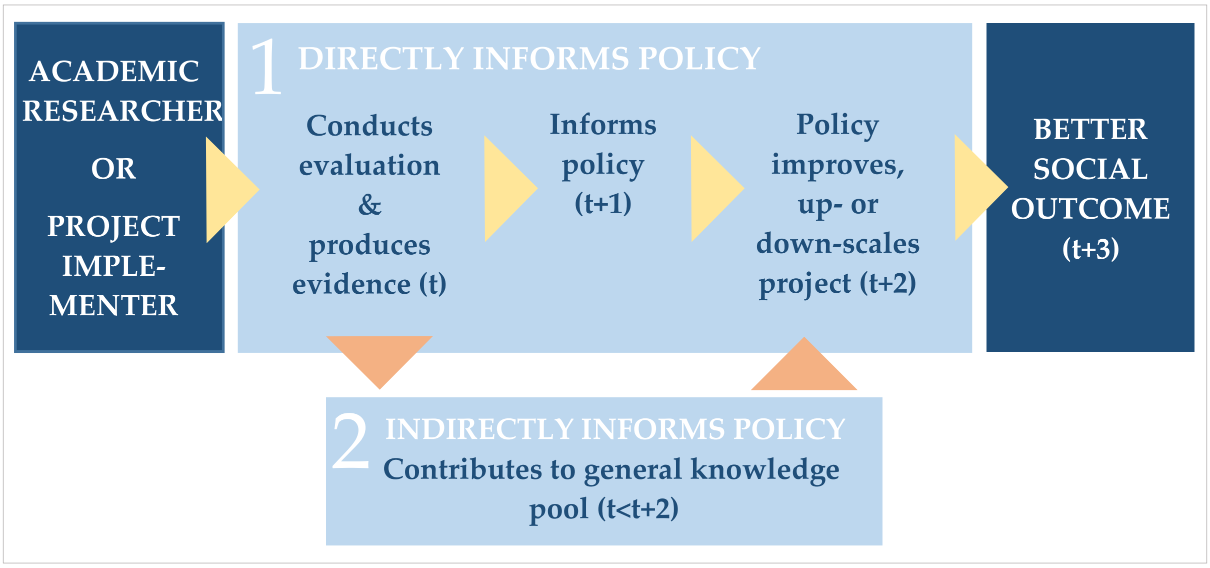
The chapter is structured as follows: Section 5.2. explains the mechanisms through which impact evaluations are assumed to lead to improved outcomes and builds the Hypothesis. Section 5.3. describes the datasets and gives some descriptive information. Section 5.4. establishes the empirical analysis and Section 5.5. presents the quantitative results. The chapter concludes with Section 5.6.

5.2. THE MECHANISM

The question on the impact of impact evaluations is directly linked to the question on how policymakers update information. Policymakers can be persons belonging to national governments or international development cooperation that conduct development interventions. Vivalt and Coville (2017) look at the mechanisms behind policy effectiveness of impact evaluations and state three necessary conditions for policymakers to update: 1. Evidence, 2. The accurate update of the beliefs, based on the evidence, 3. Policymakers' capability and willingness to change their decisions according to the evidence.[214]

Under the assumption that policymakers are willing and able to learn from impact evaluations, scholarly work identifies three major channels through which IEs can affect policy (Baanante and Valdivia 2015). First, the use can be classified as instrumental or direct, i.e., employing IEs to improve, scale-up or scale-down projects (e.g., Buddlemeyer and Skoufias 2003), second, indirect through general knowledge building on development (e.g., Banerjee et al., 2016) and third, legitimizing, to justify continuation of desired policies or budgets (e.g., Jones et al. 2009; Baanante and Valdivia 2015). The theory of change for impact evaluations in Figure 5.2. is built on Shah et al. (2015) – with modifications relevant to this chapter – and illustrates the mechanism:

Figure 5.2. Theory of change for impact evaluations



[214] One of the major obstacles in making impact evaluations effective lies probably in bridging step one and step two, i.e., reaching policymakers with IE evidence. Initiatives like Eva Vivalt's AidGrade (www.aidgrad.org) try to tackle this important challenge.

Source: Own design based on IDinsight (Shah et al. 2015, p. 4, 21)

The yellow (horizontal) arrows in Figure 5.2. illustrate direct ways to influence policy for impact evaluations and the orange (vertical) arrows indirect ones. The third channel, via the legitimization of policy is not included as this way of using impact evaluations does not add actual value.[215] The theory of change for the direct channel 1 starts with an academic researcher or project implementer, who identify an evidence gap and conduct an impact evaluation accordingly. In year t, they collect outcomes, then conduct the analysis and are 1 year later able to inform project operators (subsumed under policy) about the results in t+1. Policy then needs another year to implement the evidence in their projects by improving certain aspects, up-scaling or down-scaling them for negative or insignificant results (t+2). The modified projects can have an immediate effect (at least on health outcomes) and should lead to measurable improved social outcomes in the following year (t+3). Importantly, the theory of change highlights that impact evaluations cannot have an effect by themselves, but only through development projects.

The second, indirect channel starts after impact evaluations have produced evidence. This evidence is then available via reports and studies and adds to the pool of global knowledge about the specific policy area.[216] Policy can always learn from this pool of global knowledge, which is produced any time before t+2. Hence, also the global knowledge, if used by policymakers, can contribute to improved social outcomes in t+3.

Three mechanisms follow from the theory of change about potential measurable effects of impact evaluations. First, impact evaluations can and are frequently used as pilot studies, the Progresa success story builds a prominent example (e.g., Baanante and Valdivia 2015). If IEs show the intervention to be effective, it should be up-scaled.[217] Accordingly, the number of project sites could increase after the incidence of an IE in a certain year, if the intervention was positively evaluated.

---

[215] However, Buddlemeyer and Skoufias (2003) argue convincingly that also the third, legitimizing channel of impact evaluation has an important function as proof of effectiveness helps projects to survive changes in government.
[216] The pool of general knowledge consists of evidence created through national and internationally done impact evaluations. To the extent in which work by Vivalt (2015) or Duflo et al. (2012), for instance, cautions against generalizability of impact evaluations, IE evidence certainly increases the general global understanding of human behavior (Banerjee et al. 2016).
[217] Banerjee et al. (2016) indicate that up-scaling also depend on intervention costs per capita, rigorous design and likewise simplicity of the intervention.

Second, if IE results are disseminated while a project is being implemented, project implementers might still be able to learn from the presented evidence and adjust their ongoing project designs. A change in design should lead to increasing differences between the planned (committed) and actual (disbursed) project costs.

Third, the up-scaled projects following a pilot impact evaluation and the projects benefitting from the dissemination of IE results during ongoing project cycles should be disproportionately effective and therefore lead to improved social outcomes. Also, projects can be implemented from scratch, without a pilot testing the interventions' effectiveness. In this case, project implementers can learn from the global knowledge pool created by all impact evaluations conducted in the relevant policy area before. The relevant unit of analysis would be the stocks of impact evaluations done before, accordingly. Arguably, evidence from the most recent impact evaluation done could have the strongest influence, even if it was not the corresponding pilot – results might be particularly present and easy to access for policymakers. Following the outlined mechanisms, I hypothesize that IE stocks and the incidence of an IE conducted directly before project implementation should lead to improved outcomes, in combination with projects. Vivalt and Coville (2017) show that policymakers are more likely to learn from positive results. Accordingly, also the overall influence of impact evaluations should depend on the direction of evaluation results.

5.3. Data and Descriptives

This section describes the different datasets on a macro- and microeconomic level, which is created for the subsequent analysis of the research question and gives some descriptive statistics.

Macroeconomic data

Conducting the current study is only feasible thanks to a newly built comprehensive database of impact evaluations with a development focus by the non-governmental organization International Initiative for Impact Evaluation (3ie) (2018).[218] In a large data collection effort, the 3ie included all impact evaluation studies and reports in their repository that meet the criteria of building a counterfactual-based evaluation, evaluating development interventions in developing countries and being published as journal articles, book chapters, reports or working papers. The repository spans the time from 1981-2016. However, as the 3ie considers data collection completed only until 2012, I will use their data accordingly. The dataset includes 2259 studies, of which 77 percent are journal articles, 18.9 percent working papers, 3.9 percent reports and 0.3 percent books or book chapters (see Cameron et al. 2016 for a comprehensive description of the database).

Due to two pragmatic reasons, I focus my analysis on impact evaluations in the health sector only. First, the large majority of impact evaluations is being conducted in the health sector. Within the 3ie sample studies, 54.2 percent belong to interventions in health and offer thus the largest data variation. Second, health interventions can have immediate effects on health outcomes (De and Becker 2015), which makes the linkage between interventions and outcomes comparably easy. My final dataset covers the 1995-2012 period and entails in total 703 health impact evaluations. Unsurprisingly, the maximum of impact evaluations per year is found in India with 36 evaluations in the year 2006 alone.

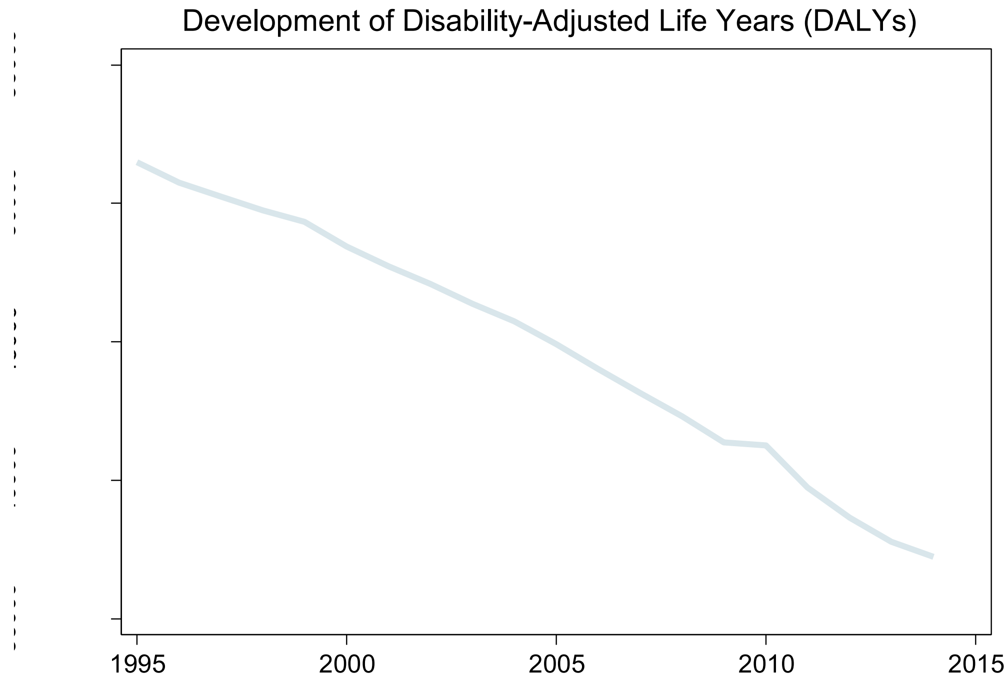
I follow USAID's Development Innovation Ventures (DIV) in measuring the social return of evaluations with Disability-Adjusted Life Years (DALYs) (Banerjee et al. 2016). DALYs present a comprehensive measure of the health status of societies because they

---

[218] I am grateful to the 3ie for providing me with the data in a ready-to-use format, including information on the evaluation method, publication type and source.

calculate the number of years, which are lost to every person due to any kind of experienced disability per life year. That means, if a person is blind, for instance, his or her year of living experienced in blindness is discounted by 0.187, a moderate measles infection would be discounted by 0.051 (Global Burden of Disease Collaborative Network 2017). Weight factors range from 0 (perfect health) to 1 (dead).[219] The full sum of DALYs for a society are the Years of Life Lost (YLLs) (i.e., number of deaths*remaining life expectancy) + Years Lived with Disability (YLD) (i.e., number of cases*disability weights) (WHO 2018). A reduction in DALYs thus corresponds to a healthier society. The measure is commonly applied in public health and builds for instance, a major indicator in the WHO country health profiles (WHO 2015). DALYs are constantly decreasing in line with increasing global living standards. See Figure 5.3. for the development of DALYs aggregated over all sample countries from 1995-2014:

Figure 5.3. Development of DALYs from 1995-2014



[219] Accordingly, also premature deaths are assigned 1 full DALY per every life year between the experienced deaths and the maximal healthy life expectancy. The same maximal life expectancy is applied to every person in every country (WHO 2018).

Sub-national data

The country of choice for the sub-national analysis is Uganda. This choice is pragmatically data driven. For my analysis, I need disbursement data on aid projects in health on the district level. Such georeferenced data are available from AidData for 13 countries. Of these countries, Uganda has received most impact evaluations in health (106) and offers, thus, the largest variation in data.[220] Moreover, as being part of one of the first coding waves, the Ugandan dataset[221] belongs to the most reliable AidData-coded project databases (Civelli et al. 2017). I aggregate the project level data to 56 Ugandan districts as of the year 2000. This is a reasonably large number with regard to data variation for regression analyses and the level on which the GADM Global Administrative Areas allocation of latitudinal and longitudinal data to districts (ADM2 level units) is available (http://www.gadm.org/).[222]

The AidData database contains 1709 projects for the period between 1978 and 2014 of which 565 were geocoded. According to AidData geocodes were attributed to all projects for which sufficient information was available.[223] Out of all coded projects, 214 are classified into the health sector and 69 of them geocoded and hence adequate for my approach. This number further reduces, as I only keep projects with precision codes from 1 to 3 (larger than 3 relates to too broad administrative levels of regions and above) for which the transaction year is available.[224] I am left with 50 projects disbursed through 421 individual transactions (one project is disbursed in different years to different project sites) over the time period 1995-2013. Disbursements per project site and the number of project sites are aggregated per district per year. All of the 56 districts receive a health project site over the time period studied (see Appendix A5.1., column 1).

---

[220] The 12 countries besides Uganda are Afghanistan with 5 impact evaluations, Burundi with 4, Colombia with 25, Honduras with 8, Iraq with 2, Nigeria with 29, Nepal with 39, Philippines with 27, Senegal with 14, Sierra Leone with 5, Somalia with 0 and Timor-Leste with 2 (https://www.aiddata.org/datasets accessed June 3rd, 2018).

[221] Released in April 2016 as version "Release Level 1 v 1.4.1" (https://www.aiddata.org/data/uganda-aims-geocoded-research-release-level-1-v1-4-1).

[222] The Ugandan government further divided districts and reaches a number of 111 as of July 2018 (http://www.statoids.com/uug.html) (see Green, 2008 for potential explanations). All data points after 2000 were accordingly allocated to the 56 districts.

[223] Personal e-mail contact from April 2018.

[224] I lose 125 transaction year-site observations due to missing transaction years, which I could not identify via online research.

In order to be able to link projects to impact evaluations as precisely as possible, I categorized them further according to their project title available in the dataset (and if ambiguous further online research on the project) into more narrow categories. The categories I built were General health (115 transactions), HIV/Aids (161), Access to medicine (54), Maternal and newborn health (49), Health infrastructure (29), Malaria (11), and Garbage management (2) (see Appendix Table A5.1.).

Subsequently, I conducted the same coding process for all Ugandan impact evaluations in health with a view to subsume them into the same categories, allowing a (potential) direct link between projects and impact evaluations. Additionally, I coded the district where the impact evaluation was conducted, the outcomes intended to affect (like "reduction in neonatal mortality" for maternal and newborn health interventions or "increased condom use" for HIV/Aids interventions), success of affecting the outcome with the intervention and time until measurement of outcomes. The coding was based on the information available in the corresponding articles or reports. I coded the success of affecting the intended outcome with the evaluation on a scale from 1 to 5. The value 1 referred to a negative impact of the intervention, 2 to insignificant results, 3 to moderate positive results, 4 to positive and 5 to very positive results.

There are in total 96 health impact evaluations for the sample period of 1995-2012 in Uganda.[225] The year refers to the estimated year that policy was informed about the evaluation results (under the assumption that results are disseminated one year after the endline data collection as outlined in Section 5.2.).[226] The majority of impact evaluations was done in HIV/Aids with 43 in total. Second comes general health with 13 interventions in topics like community-based monitoring of health providers, for instance (Björkman and Svensson 2007). Twelve interventions were done in mental health, mainly addressing mental support for war-

---

[225] Out of the 106 impact evaluations originally included in the 3ie repository for the whole available period from 1995 to 2014, I dropped six because they were either duplicates (five times) or cross-country studies (one time). Four more studies were dropped for the final sample, as they were from beyond the sample period 1995-2012.

[226] From the repository database, the publication date of studies or reports is available. However, I need more precise timing data on the year policy information, i.e. the year when policymakers were informed about evaluation results and hence able to include new insights in their project design. I make use of the estimated time until publication created by Cameron et al. (2016). Based on a random subset of all repository studies, they estimate the time between the endline data collection and the publication in academic health journals to take 3.75 years, for book chapters 4.8 years, for working papers 3.63 years and for policy reports 1 year. I use these estimates to calculate the estimated year of policy information.

affected youth (e.g., Ertl et al. 2011) or HIV/Aids-affected families (e.g., Nabunya et al. 2014). Malaria interventions (11) come next, then maternal and newborn health interventions (9), six interventions in nutrition, one intervention in health infrastructure and one in sanitation. The large majority of interventions (90.6 percent) is published in academic journals, while the remaining are working papers or policy reports. A similarly high fraction with 86.5 percent of the interventions was conducted as RCT. The evaluation results referring to the success of the intervention in positively influencing the respective outcome is on average "positive," the second highest category. Results for HIV/Aids, Malaria and nutrition evaluations are below average.

From the available categories and outcome measures, I chose evaluations in maternal and newborn health for deeper investigation due to two reasons. First, these interventions can have a relatively proximate effect on the outcome measure infant mortality according to Kotsadam et al. (2018), for instance. Second, the outcome measure infant mortality is arguably objective and reliable data for its construction are available in a georeferenced format through Demographic Health Surveys (DHS) for Uganda.

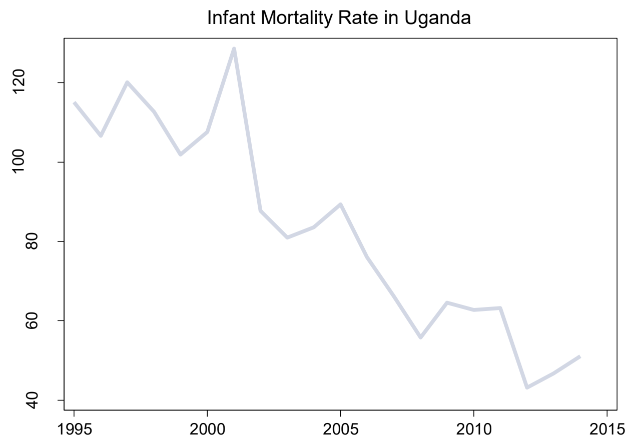
Five DHS surveys for Uganda from 1995-2016 build the base for the sub-national dataset. The DHS data are collected in a nationally representative way and include surveys with women aged 15-49 (DHS 2018). For my sample period from 1995-2014, the dataset consists of 88,348 children and 25,780 mothers. I construct infant deaths below 12 months as a binary outcome variable for the birth-level dataset. Each child born alive is coded 1 if the infant died before 12 months of age. Within the aggregated dataset on district levels, I build the infant mortality rate (IMR) based on the calculation method used for the World Development Indicators (UN 2018). For every year, I divide the number of infants dying below the age of 12 months by the number of children born alive and multiply with 1,000:[227]

$$IMR = \left(\frac{Number\ of\ deaths\ among\ infants < 12\ moths\ of\ age}{Number\ of\ live - born\ children\ per\ year}\right) * 1{,}000$$

Figure 5.4. displays Uganda's average IMR over the sample period 1995-2014. Uganda's decrease in the IMR is in line with a steadily decreasing infant mortality rate worldwide. While the global IMR was at 61 children in 1,000 live births in the year 1995, it

[227] For robustness, I construct the below-5 mortality ratio equivalently.

dropped to 32 in 2014 (UN 2018). Numbers for Uganda are expectedly above the global average with an IMR of 115 in the year 1995 and 51 in 2014 (own dataset based on DHS data). Figure 5.4. Infant Mortality Rate for Uganda over the sample period (1995-2014)



Uganda belongs to the low-income country group classified by the World Bank with a Gross National Income (GNI) per capita of 1370 US$ (int PPP) and a life expectancy at birth of 57 years. Uganda's national health expenditure per capita is below the average expenditures among its African WHO region. Among the highest causes of deaths for children under five, are acute respiratory infections (15 percent), Malaria (13 percent), Prematurity (12 percent), Birth asphyxia (11 percent), Diarrhea (8 percent) and HIV/Aids (7 percent).[228]

Descriptive statistics

The first and second mechanism described in Section 5.2. with regard to the influence of impact evaluations can only be addressed descriptively with the data at hand. For testing the first mechanism, I assess whether project characteristics change, if evaluation results are disseminated during corresponding projects in the same category (i.e., I code a binary variable to 1, if during the duration of an HIV/Aids project, an evaluation in HIV/Aids informed policymakers about its results). More than half of all projects (51 percent) had a corresponding impact evaluation during their project duration. Based on the information available in the project dataset from AidData, I construct the share of disbursements over commitments and the number of project sites per project. The expectation is that disbursements are larger than commitments and the number of project sites increases if evaluation results are published

[228] Data are based on the latest available WHO statistical profile for Uganda from 2015 and mainly refers to the year 2013 (http://www.who.int/gho/countries/uga.pdf?ua=1).

during project cycles. Running simple correlations, I do not find support for either indicator. The correlation between disbursements as share of commitments and the existence of an evaluation is insignificant and negative (-0.183, p-value 0.309) and between the number of project sites and simultaneous evaluation incidence is positive and insignificant (+0.048, p-value 0.762). The relationship does also not change if I drop all 11 evaluations with negative results (negative or insignificant) from the sample, as hypothesized following Vivalt and Coville (2017). Accordingly, there is no descriptive support for the first potential mechanism, through which policymakers might improve or abandon their projects if they have the possibility to learn from simultaneously disseminated evaluation results.

Following the second mechanism, impact evaluations' influence in corresponding health projects could also materialize through project up-scaling. I test this by checking whether projects have more project sites if a corresponding impact evaluation was done directly, i.e., one year before project start. This was the case for 39.5 percent of the projects in the sample. Indeed, there is a significantly positive correlation between the number of project sites and an impact evaluation immediately before project start (+0.369 p-value 0.015). According to these descriptive statistics, a potential mechanism through which impact evaluations could affect projects is hence through up-scaling of a priori conducted pilot studies.

5.4. Empirical Method

A rigorous empirical analysis needs to identify whether impact evaluations were indeed able to materialize in improved outcomes – the main research question of this chapter formulated in the Hypothesis in Section 5.2. On the cross-country level, I will employ Ordinary Least Squares (OLS) with fixed effects. In a second approach, I will turn to sub-national, district-level and birth-level analyses for Uganda. This allows me to include high-dimensional fixed effects and a more precise linkage of the interventions and outcomes, which both reduces endogeneity concerns (Civelli et al. 2017). The analyses on all three levels have to deal with the obvious causality problem of unambiguously linking the number of IEs, which is not random, to changes in outcomes (Banerjee et al. 2016). There are several potential biases, which I will discuss in the following along with their potential direction. When outlining the empirical strategy below, I will show how the analysis addresses the biases.

The first concern which comes to mind is a simultaneity bias in the sense that a larger number of impact evaluations could well coincide with governments who generally increase efforts in national health and their demand for impact evaluations at the same time. This would downward bias my results. This means that the coefficient of the number of impact evaluations in health captures parts of the health improving effect actually originating from increased government effort. Health improvement translates to reductions in my outcome variables DALYs and the IMR. Accordingly, if I cannot control for governmental effort in my analysis, my estimate for impact evaluations is likely to be downward biased, i.e., the effect of impact evaluations with regard to improving outcomes appears to be stronger (more negative) than it is. The cessation of conflicts or wars could have a similar effect. Health outcomes can be expected to improve after conflicts and likewise the interest of researchers increases, who could be specifically interested in the effectiveness of interventions improving the health of the war-affected population, or, have always had a research interest in the specific country or region but were not able to enter before, due to the conflict. In line with explanations for increasing government effort, also this potential confounder will be downward biasing the effect of impact evaluations. A third bias can be associated to the fact that foreign evaluators of development interventions who want to study their effectiveness in improving health outcomes will typically go to places which have problems with these health outcomes, or, they will go to places they like. More accurate is probably to assume a combination of both. For instance, the evaluation of interventions addressing infant mortality is more likely to take place in areas with high infant mortality rates, which could be stagnant, increasing or also decreasing, but at a comparably low pace. This reverse causality, would then associate increasing numbers of health evaluations to also increasing infant mortality rates and therefore upward biasing the results. Whether researchers like to do research in a place will depend on research regulations like the ease to obtain research permits and the obligations linked to them. These are unlikely to be correlated with health outcomes. However, researchers will also prefer to go to places where living standards are acceptable and not harmful to their own health, like through easily communicable diseases. These aspects are obviously related to the health outcomes, researchers might study, and might again downward bias the estimated effect of impact evaluations. Overall, are the discussed biases likely to downward bias the results, and need to be tackled in the analysis in order to produce meaningful findings.

Macroeconomic level

On the macroeconomic level, I will look at a potential average effect of impact evaluations in health on improved health outcomes applying Ordinary Least Squares (OLS) with fixed effects. The dependent variable will employ the logged number of DALYs. Alternative specifications use the DALYs rate in 100,000 inhabitants as robustness. In my baseline specification, I regress the logged number of DALYs on the number of health evaluations conducted per country per year, including a set of controls. I focus the analysis on this direct effect of the incidence of impact evaluations, as this is most controversially discussed in the literature. A contribution of impact evaluations on general knowledge building is generally acknowledged (see Figure 5.2. for direct versus indirect influences of policy). However, I will still address the effect of a general knowledge increase from nationally conducted IEs by including stocks of IEs per country, in an alternative specification.

$$(1) \quad log(DALYs_{it}) = \beta_1 \#HIE_{it-2} + \beta_2\, (log)Health\ disb._{it-1} + \beta_3\, Health\ gov.\ exp._{it-1} + \Sigma_m \beta_4\, Controls_{imt-1} + \beta_5\, lintt_{it} + \eta_i + \mu_t + u_{it}$$

On a cross-country, aggregated level, more impact evaluations in health – measured directly as incidence ($\#HIE_{it-2}$) or indirectly as stock ($Stock\ HIE_{it-2}$) – should increase the effectiveness of every dollar spent in health, via international development disbursements ($(log)Health\ disb._{it-1}$) or national government expenditures ($Health\ gov.\ exp._{it-1}$), if the policymakers designing the projects learn from the evaluations (see Section 5.2.). Accordingly, I will include aid disbursements and government expenditure on health as controls, to observe the ceteris paribus effect of impact evaluations. Controlling for government expenditure in health also diminishes the first source of endogeneity bias identified above. However, the effect of health IEs can also materialize through up-scaling as described in Section 5.2., which would lead to increased health aid disbursements or government expenditures.[229] The two types of health expenditures would represent bad controls in this case. Therefore, I will accordingly show regressions which exclude them.

Following suggestions from the literature, I include the following control variables ($Controls_{imt-1}$) in my model: Logged GDP in constant 2011 US$ controls for the positive relationship between living standards and health and reduces endogeneity bias from researchers' potential preference to conduct impact evaluations rather in countries with higher

[229] If the money is not cut elsewhere, which would again lead to greater effectiveness per dollar spent.

living standards. Logged population is an important control, as health outcomes might be easier or more difficult to handle conditional on the size of the population. Also, Cameron et al. (2016) show a positive link between higher population numbers and increased incidence of impact evaluations. In addition, when using the total number of DALYs as an outcome, controlling for population size is essential. A dummy for regime change is included as priorities change when politicians change, which might affect the use of IEs and affect outcomes (Buddlemeyer and Skofias 2003). This also captures parts of the endogeneity originating from potentially changing governmental preferences for health, which could at the same time increase demand in impact evaluations. A conflict dummy, indicating an internal armed conflict in a given year, eliminates health effects stemming from conflict violence. Additionally, the control captures endogeneity from conflict cessation leading to increased impact evaluations and improved health outcomes at the same time. A political corruption index is included as corruption should slow down learning from IEs. Presumably, more corrupt governments would focus less on effective policies but rather on policies beneficial to government elites or bureaucracy and therefore disregard IE results with a higher likelihood. Following a similar line of argument, I include a measure for government effectiveness, the polity IV score to categorize democratic vs. autocratic regimes, freedom of academic expression and a dummy for the existence of a local government[230], which might influence whether the results reach policy. With more government fractionalization, i.e., the existence of local governments it might be more difficult to spread knowledge gained through IEs (see Appendix 5.3. for a detailed description of the variables as well as summary statistics). All time-varying control variables enter the regression with a lag of one year. The inclusion of control variables hence addresses three of the four endogeneity concerns, government effort, conflicts, and living standards in the destination country. The potential upward bias through worse outcomes attracting researchers' scientific interest remains, but is less worrisome as it would render my effects weaker than they actually are.

Considering the steadily decreasing shape of the outcome DALYs over time as illustrated in Figure 5.3., the inclusion of country-specific linear time trends $lintt_{it}$ is important (Herzer and Nagel 2015). They account for general advancements in health due to

[230] It might first come at a surprise, that the dummy for existence of local governments is time-varying, but the degree of decentralization can change over time within countries. For instance the dummy switches from non-existence to existence of local government for Azerbaijan in 1999.

technological or medical improvements and increasing living standards. These general factors are likely to be overall responsible for the observed linear improvements in health or life expectancy. However, as we still see large differences in technological change and health status improvements in different countries, the change is likely to come at a different pace for every country, however linear, which is why country-specific linear trends are included. Additionally, there may be non-linear shocks affecting all countries at the same time like the global financial crisis or certain seminal health inventions. I include year-fixed effects μ_t, which capture these non-linear health shocks equal to all countries. Finally, I include country-fixed effects η_i, which capture time-invariant factors like countries' geography and the error term u_{it}. Robust standard errors are clustered at the country-level. The choice of control variables and fixed effects should capture a large fraction of endogeneity in my model. Note that linear time trends also account for the global knowledge increase through impact evaluations in health. My analysis thus aims at isolating the direct and indirect country-specific effects of locally conducted impact evaluations – as described in the Hypothesis. Any effect found will, therefore, be a lower bound of impact evaluations' impact, which could additionally materialize through general global knowledge building (see Section 5.2.).

The timing of the specification is built on insights about the mechanism outlined in Section 5.2. of how impact evaluations work. Accordingly, I include health interventions with a two-year lag and all other control variables with a one-year lag, including health aid and national health expenditures. The lag-structure I apply reduces further endogeneity concerns. Impact evaluations, which are mostly conducted as a cooperation between foreign researchers and local specialists, need time to be set up. Research permits need to be applied for, field work abroad planned, and the like. As described in the *Introduction* and *Mechanisms* Sections, evaluations will need to be planned at least four years before their outcomes are measured in my analysis. All biases outlined above could confound the decision to conduct an impact evaluation (government demand, conflict cession, country attractiveness with regard to research and living standards, or health challenges). The actual timing of results' dissemination, however, depends on many exogenous factors. Such factors are the availability of researchers, length of the administrative process until research permits are forwarded, unforeseen barriers in the field or delays in the data analyses. The list could be further extended. It is therefore hard to imagine that confounding factors systematically coincide with the dissemination of impact evaluation results and would therefore be the actual drivers of effects, while they are likely to confound decisions to undertake impact evaluations, which is, however less of a threat to my identification. Moreover, the careful selection of adequate DVs, control variables and lag structures reduces the measurement error, omitted variable and simultaneity bias and can therefore lead to meaningful correlations. However, the results do not claim causality.[231]

Sub-national level

Similar to studies on the aggregate effect of aid on growth (e.g., Dreher and Langlotz 2017), also the aggregate effect of health evaluations on overall health is informative from a general perspective using the full variation of all possible country contexts. Nevertheless, the analysis might suffer from an "over-aggregation" problem as outlined in Civelli et al. (2017). They consequently turned to the sub-national level of – coincidentally – Uganda as well to evaluate the effect of aid on growth. I will make use of Uganda's district level data (and below) to assess my research question.

Due to the fact that impact evaluations can arguably only show an effect through aid projects or national government programs and the sub-nationally coded data provide high precision to reasonably link specific evaluations to such projects, the sub-national analysis will focus on the effects of impact evaluations conditional on aid disbursements. Only the

[231] One way of addressing causality could go through an instrumental variable approach with an exogenous instrument for the endogenous number of impact evaluations. One possibility for such an instrument would be the ease of doing research in a country, which is arguably exogenous, as it depends on governmental regulations, requirements and administrative barriers for research permits and research visa, but should not be linked to health outcomes in any way. As there is no time varying indicator for the ease of doing research in different countries available, I chose an approach building on Google hits. I was inspired by Rose (2007), who used the "desirability of residing in a particular country as Foreign Service Officer" as an instrument for the number of Foreign Missions. He measured this desirability by the number of Google hits for the search '+"Travel Destination" + "city" + "x"' with x being the capital city. Similarly, I construct a year and country varying measure for the ease of doing research in a specific country in a specific year. To this end, I count the Google hits for the phrase "[country] difficult foreign research permit [year]", where the [country] varied over all sample countries and the [year] for all years from 1997, when Google was founded, to 2012. Anecdotal evidence among researchers or websites suggests that research permit requirements can be an underestimated hurdle that may well lead to a change in the destination country if the barrier becomes too high (e.g., https://clinregs.niaid.nih.gov/country/india#_top). However, the first stage shows a positive correlation between the number of Google hits and the number of impact evaluations, rather than the expected negative. The Google hits therefore are more likely to capture general interest in doing research than difficulty to do research and are accordingly not a valid instrument. Potentially, future research can however build on the idea in order to address causal findings.

disaggregated, narrow categorization of impact evaluations and development projects into the eight categories described in Section 5.3. allows the linkage between IEs and projects. Again, I argue that there should only be a direct effect measurable if the policy recommendations from an impact evaluation are announced one year before a development project is disbursed. I call these incidences, when HIEs inform policy one year before projects in the same category take place, a match. In the sub-national analysis, I interact the number of impact evaluations informing policy in t-2 with project disbursements in the same health category in t-1 and investigate the effect of these matches on the specific outcome in time t. The interaction effect thus compares observations with matches to observations in which development projects take place without being directly informed by impact evaluations. Additionally, I evaluate an indirect effect of HIEs contributing to an increased pool of knowledge by examining the stocks of HIEs.

Specifically, I link evaluations and aid projects addressing maternal and newborn health and observe their conditional influence on the infant mortality rate below the age of 12 months in 1,000 live births.[232] The ability to precisely link IEs to disbursements and specific outcomes reduces endogeneity concerns based on omitted variable bias and measurement error. An additional advantage over aggregated cross-country data is that I have information on specific monetary transactions in the eight health categories per districts and years (Civelli et al. 2017).

As data on governmental health expenditure are unfortunately not available, I will only be able to assess the link between impact evaluations and international development projects or disbursements on the IMR. On the one hand, every effect I may find will consequently represent a lower bound of the impact evaluations' influence as their impact may also materialize through national health expenditures. On the other hand, the disability to control for governmental efforts could increase the simultaneity bias described above and downward bias my results.

[232] The timing assumptions for these specific interventions in Maternal and newborn health are based on corresponding impact evaluations conducted in Uganda. I follow the studies from Fuentes et al. (2006) and Björkman and Svensson (2007), which form part of the impact evaluation repository for Uganda. Their interventions lasted 11 (Fuentes et al. 2006) and 12 months (Björkman and Svensson 2007) before they measured infant mortality and under-five mortality, respectively. Accordingly, I lag aid projects or disbursements by one year.

In the first set of regressions, I exploit district-level variation and the conditional effect of impact evaluations concerning mother and newborn health on the IMR aggregated per district (IMR_{jt}).

(2) $IMR_{jt} = \beta_1 \#Mat\&NeoIE_{t-2} * (log)Mat\&Neo\ disb._{jt-1} + \beta_2 (log)Mat\&Neo\ disb._{jt-1} + \Sigma_m \beta_3\ Controls_{jmt-1} + \eta_j + \mu_t + e_{jt}$

As explained above, the number of maternal and newborn health evaluations ($\#Mat\&NeoIE_{t-2}$) enters with a two-year lag and is interacted with one-year lagged maternal and newborn health disbursements ($(log)Mat\&Neo\ disb._{jt-1}$). Even though I coded information on the exact location where the impact evaluations were conducted, I assume their results can influence policy nationwide based on Jones et al. (2009).[233] The same set of health evaluations is therefore attributed to every district *j*. The interaction's constitutive term of Maternal and newborn health evaluations ($\#Mat\&NeoIE_{t-2}$) is hence captured in the year-fixed effects included in the model (μ_t). This theoretically justified strategy comes with the advantage of reducing many endogeneity concerns. Since it does not matter in which district exactly the evaluation was conducted, but results are assumed to be able to inform policy in all districts likewise, it is also irrelevant whether evaluators picked specific evaluation locations due to better or worse infant mortality outcomes, better living standards, or after cessation of conflicts. What matters is only the match of information spread all over Uganda and specific projects in maternal and infant health per district. Moreover, a national increase in governmental health efforts is captured by the year-fixed effects μ_t. The year-fixed effects account again for all health shocks equal to all districts at the same time, like epidemics. Lastly, district-fixed effects η_j control for time-invariant district characteristics like being landlocked. The error term is represented by e_{jt} and robust standard errors are clustered at the district level. In a robustness specification, I will additionally employ district-specific linear time trends to capture linear advancements at a different pace in different districts.

I largely follow Cruzatti et al. (2018) in specifying my set of control variables. As time-varying geographic controls, I include mean precipitation (*"Rainfall"*), mean temperature (*"Temperature"*) and the mean Normalized Difference Vegetation Index (NDVI) (*"Vegetation"*). As population controls, I include the population density and urban land coverage (Goodman

[233] Jones et al. (2009) find that the communication of evaluation results at the national level is relatively robust for the health sector.

et al. 2016). In addition to controls used by Cruzatti et al. (2018), I employ logged population numbers due to their relevance for impact evaluation incidences, the number of deaths due to conflicts ("Conflict deaths"), which can increase infant mortality as well as the average age of mothers at their children's births (see Appendix A5.3. for detailed sources and descriptions).[234] All control variables are included without a time lag as they simultaneously affect children's survival during their first year of life. The estimation period ranges from 1995-2014.

In alternative specifications, I use stocks of maternal and newborn health evaluations (*Stock Mat&NeoIE*$_{t-2}$) instead of incidences and project numbers (*#Mat&Neo pr.*$_{jt-1}$) instead of disbursements. In robustness specifications, I will employ IMR growth rates as dependent variable or effects on the under-5 mortality rate.

Specific interest is put in gaining more insights on mechanisms. I will investigate differential effects for different project operators. Over the sample period maternal and neonatal health projects are realized by the following donors: The World Bank in 18 project-years with 206,986 US$ mean disbursements, Japan in 16 district-years with an average disbursement of 40,474 US$, Sweden in 8 project-years with 231,310 US$ mean disbursements and the European Union (EU) in 5 district-years with an average disbursement of 715,018 US$.

An even higher degree of data precision is reached with the third level of analysis using birth-level data and the application of mother-fixed effects following Kotsadam et al. (2018). The dataset is structured as quasi-panel, comparing birth cohorts, but not individuals over time (Guillerm 2017). The regression line reads as follows:

(3) $$\textit{Infant death}_{kt} = \beta_1\, \textit{Active Mat\&Neo Match}_t + \beta_2\, \textit{Order}_{kt} + \beta_3\, \textit{Twin}_{kt} + m_m + \mu_t + e_{kjt}$$

The dependent variable "*Infant death*$_{kt}$" is defined as a binary indicator, which equals one, if the child *k* died below the age of 12 months. The major advantage of the analysis based on the quasi-panel is the possibility to apply mother-fixed effects (m_m). With mother-fixed effects, all potential confounders which do not change for the same mother over time are controlled for. This strategy should capture most heterogeneity, when it comes to survival probabilities of infants (like mother's education and socio-economic factors and all time-invariant-characteristics of the village, town or district, she lives in). However, necessary

[234] I do not include the index of aridity and travel to next larger settlement of 50,000 people like Cruzatti et al. (2018) do. I expect aridity to be captured by the combination of "Rainfall" and "Vegetation". Distance to next larger settlement is not applicable to me as I have district-level data.

controls, which are correlated with infant mortality are the birth order of the child (*Order*$_k$) and a binary variable indicating whether the child is part of a multiple birth (*Twin*$_k$). The strategy thus looks at the within-sibling variation. In a robustness specification, I will include a district-specific linear time trend as well, which controls for technological advancements (however, less important for differences in births of the same mother) and experience gains for the mother.

I build on the approach of Kotsadam et al. (2018) in defining a binary variable "*Active Mat&Neo Match*$_t$" which is one as soon as an impact evaluation matches a corresponding project per mother. For instance, if there is an impact evaluation in maternal and newborn health in the year 2000 and a development project in maternal and newborn health takes place in a specific district in 2001, this is a match for this district. Assume a mother living in this district has three children, the first born in 1999, the second in 2002 and the third in 2004. The "*Active Mat&Neo Match*$_t$" will turn one for the mother (and her children) in 2002 and 2004, but zero in 1999. The assumption is that the mother's delivery for the third child can still benefit from the intervention she experienced with her second child, while there was no intervention match during the birth of her first child.

In an additional specification, I add birth order-fixed effects, like Kotsadam et al. (2018) do. Birth order-fixed effects in regressions across mothers follow a similar logic as year-fixed effects in cross-country estimations. We observe variation within siblings per mother, while controlling for all factors, which are equal to children of the same birth order. Controlling additionally for year-fixed effects takes out all confounders, which are equal to children born in a specific year all over Uganda. This could be country-wide climatic phenomena, epidemics or government regulations. It is difficult to come up with an additional source of endogeneity, which would systematically relate a project-IE match to a confounder actually driving the effect, other than the project-IE match. The fact that IEs can be conducted in any district and not necessarily near the mother works additionally in our favor as explained above. Robust standard errors are clustered at the district level *j* and sample weights are included.

5.5. RESULTS

In the results section, I will first address the cross-country evidence on the influence of impact evaluations and subsequently turn to more narrow sub-national country evidence for Uganda.

Table 5.1. shows OLS evidence for the link between the number of impact evaluations in health and the logged number of DALYs per country. The results show a strongly negative and statistically significant correlation, which is robust to the in- or exclusion of all control variables, importantly excluding the potential bad controls health disbursements, government expenditure and GDP in column 4.[235] Table A5.3. in the Appendix shows the full set of controls including different model specifications. Control variables show the expected correlations. My preferred specification with regard to model fit measured by adjusted R-squared and the number of observations is displayed in column 2 and will serve as the baseline. The effect size can be interpreted as one additional health intervention reducing the Disability Adjusted Life Years by 0.1 percent. With an average of 23.6 million life years lost due to disabilities per country per year, 0.1 percent would translate into 23,600 life years less lost to disabilities, if one more impact evaluation in health is conducted.[236]

Table 5.1. OLS Regressions with Fixed-Effects

	(1)	(2)	(3)	(4)
	LOGGED NUMBER OF DALYS			
	w/o controls	preferred	all controls	w/o bad controls
Number of Health IEs (t-2(=t_{policy}))	-0.003***	-0.001**	-0.002**	-0.001**
	(0.002)	(0.035)	(0.023)	(0.035)
Logged int. health disbursements,		-0.000	0.001	
const (t-1)		(0.851)	(0.483)	
Gov. health exp. (% gov exp.) (t-1)		-0.001	-0.001	
		(0.543)	(0.497)	
Country- and year FE	√	√	√	√
Country-specific linear time trends	√	√	√	√
Number of observations	2,880	1,314	1,005	1,174
Number of countries	160	81	81	93
R-squared (within)	0.618	0.585	0.520	0.515
Adjusted R-squared (within)	0.593	0.548	0.464	0.464

[235] Only the exclusion of the country-specific linear time trend reduces the significance level and leaves us with insignificant, but negative estimates for the number of health IEs. The theoretically chosen timing seems to be empirically adequate, as the effect is only significant with a one- and two-year lag and loses significance before and after (see Appendix Table A5.4.).

[236] Overall, we have 483 observations with a positive number of health evaluations in the sample, which is 36.8 percent of all country-years.

Notes: Robust p-values in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The Appendix entails robustness checks for an alternative dependent variable (DALY rate in 1,000 inhabitants), stocks of evaluations and periodic data (six times three year periods from 1995-2014). It is reassuring that the effects of impact evaluations show robustly in the same negative direction at similar coefficient sizes for all robustness checks. However, significance cannot be established (see Appendices A6, A7 and A8).

Overall, the findings presented by OLS estimates with fixed effects hint at a positive correlation between impact evaluations in health and improved health outcomes, meaning that more impact evaluations correlate with fewer life years of the population whose quality is discounted due to any kind of disability experienced. Significance, however, is not robust.

Arguably, impact evaluations cannot have direct impacts by themselves, but function through improved health interventions. As the data are not specific enough on a cross-country level to meaningfully link impact evaluations to health projects in the same specific health category through interacting the variables. I turn to the sub-national level for a more fine-grain analysis, where this is possible.

Uganda sub-national analysis

As discussed before, an analysis on the sub-national level solves many of the estimation challenges encountered on the cross-country level. I will first present district level OLS results for the effects of IEs in maternal and newborn health conditional on corresponding international health disbursements. Second, I will turn to the more rigorous birth-level estimations counting on mother-fixed effects.

District-level evidence is displayed in Table 5.2. for the interaction term presenting the variable of interest only. Column 1 shows the specification without any control variable, column 2 includes geographic controls only, column 3 includes population controls only and column 4 includes both sets of controls and mothers' age at birth additionally. As model 4 produces the highest adjusted R-Squared, explaining 29 percent of the variation in the infant mortality rate, this specification is preferred and serves as the baseline model for the remaining tables presented in the following. Table A5.9. in the Appendix presents full specifications and displays that the control variables for infant mortality show the expected effects. [237]

Table 5.2. District-Level Conditional Effect of Impact Evaluations on IMR

	(1)	(2)	(3)	(4)
	INFANT MORTALITY RATE (IN 1,000 LIVE BIRTHS)			
	w/o controls	with geo. controls	with pop. controls	preferred (full)
Logged disb (t-1) * IEs (t-2) in mat & neo health	-1.091*	-1.106*	-1.222*	-1.234**
	(0.072)	(0.061)	(0.053)	(0.044)
Logged mat & neo health disb (t-1)	1.062	1.012	1.183	1.182
	(0.202)	(0.203)	(0.157)	(0.161)
Constant	196.507***	24.055	47.175	-339.302
	(0.000)	(0.848)	(0.877)	(0.259)
District- and year-fixed effects	√	√	√	√
Observations	1,008	1,008	1,008	1,008
R-squared (within)	0.291	0.293	0.297	0.304
Adjusted R-squared (within)	0.277	0.277	0.280	0.285
Number of districts	56	56	56	56

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1. Note that the levels of IEs in mat & neo health (t-2) are kept in the year fixed effects as they apply on the country level.

Findings in Table 5.2. show a robust, significant, and negative effect of increasing health interventions in maternal and neonatal health, conditional on increasing corresponding disbursements for different models. Coefficients are negative and significant at the 10 percent-level in all but the preferred specification in column 4, where significance levels reach 5 percent. The margins plot in Figure 5.5. allows a more meaningful interpretation of the marginal effect of impact evaluations presented in the preferred model.[238]

[237] All control variables enter the regression without time lag. However, the size and significance levels of the coefficients of our variables of interest are robust to the inclusion of all control variables with a one-year lag, if readers are concerned about potential bad control problems. Results are available upon request.

[238] Estimates are robust to the inclusion of district-specific time trends, which further reduce endogeneity concerns. Appendix Table A5.10. presents full results. The interaction in column 4 materializes 30 times in the sample, which is 3 percent of the observation-years. As this constitutes a relatively small fraction driving the effect, I conduct robustness regressions including interactions with all health impact evaluations. Results show that apart from interactions between maternal and neonatal health impact evaluations and disbursements, also interactions for HIV/Aids interventions have a robustly significant negative effect on neonatal mortality rates and under-five mortality rates. Considering that Uganda belongs to one of the most HIV/Aids-affected countries, this finding seems plausible. HIV/Aids

Figure 5.5. Marginal effect of impact evaluations

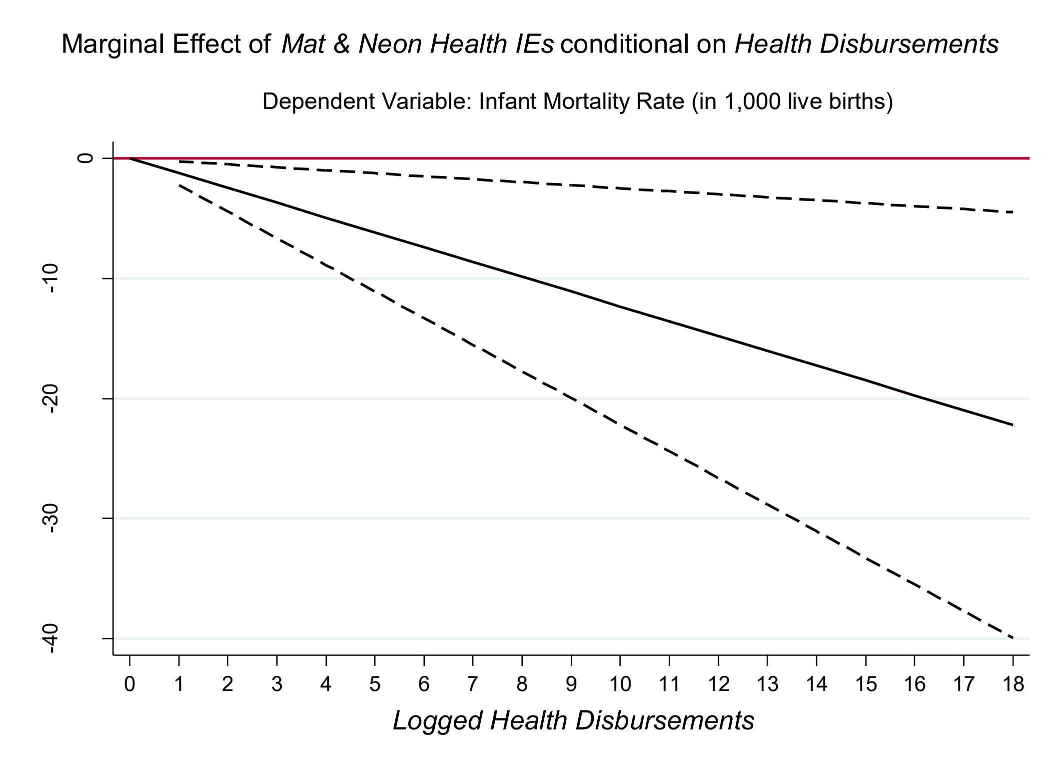


Figure 5.5. shows that as the number of impact evaluations increases, the IMR reduces if logged disbursements in maternal and neonatal health projects increase. Precisely, the interaction term indicates that the marginal effect of impact evaluations on IMR varies between -1.23 and -22.21 depending on the level of health disbursements. With mean disbursements of 38,601 US$ (in logs 0.41), the IMR decreases by 0.5 in 1,000 live births if there is one additional impact evaluation from which projects were able to learn (the average infant mortality rate is 80 infants in 1,000 live births).[239]

The effect is robust and significant when growth rates of the IMR are applied as the DV, but loses significance for the number of projects instead of disbursements. Similar to regressions on a cross-country level, the effect turns insignificant when stocks are applied,

---

disbursements and impact evaluations match more frequently in 60 observation-years, which is in 6 percent of all observations. Interactions for general health or malaria interventions do not correlate with either measure. Also, a joint significance of all intervention interactions cannot be established. Results are available upon request.

[239] Please note that the estimations presented do not give insights about the individual effect of impact evaluations. As outlined above, however, impact evaluations should not have measurable individual effects other than through health projects.

even though the negative direction is robust – like throughout all robustness checks employed.[240]

The finding on insignificant effects of impact evaluation stocks is unintuitive. The picture on learning mechanisms through impact evaluations gets more nuanced when looking at individual donors. Results are shown in Table 5.3. for different interactions of impact evaluation numbers with disbursements (column 1) or project numbers (column 2) and interactions of impact evaluation stocks and disbursements (column 3) or project numbers (column 4).

Table 5.3. Influence by donor country

	(1) HIE numbers and disburse-ments	(2) HIE numbers and project numbers	(3) Stocks and disburse-ments	(4) Stocks and project number
EU disbursements / projects (t-1)	-1.282	3.466	-0.299	85.255***
	(0.156)	(0.735)	(0.873)	(0.000)
EU effort (t-1) * HIEs (t-2)	0.490	1.884	-0.127	-14.797***
	(0.288)	(0.757)	(0.677)	(0.000)
Japan disbursements / projects (t-1)	0.535	4.317	0.835	30.169
	(0.514)	(0.734)	(0.700)	(0.276)
Japan effort (t-1) * HIEs (t-2)	-0.826	-9.809	-0.145	-4.336
	(0.219)	(0.253)	(0.588)	(0.216)
Sweden disbursements / projects (t-1)	-0.494	-4.751	26.122**	466.181***
	(0.583)	(0.523)	(0.021)	(0.000)
Sweden effort (t-1) * HIEs (t-2)			-3.005**	-52.493***
			(0.025)	(0.000)
WB disbursements / projects (t-1)	4.040***	53.459***	20.300***	57.968*
	(0.000)	(0.000)	(0.001)	(0.074)
WB effort (t-1) * HIEs (t-2)	-3.624***	-53.951***	-2.260***	-6.429*
	(0.002)	(0.000)	(0.005)	(0.071)
District- and year-fixed effects	√	√	√	√
Observations	1,040	1,040	1,040	1,040
R-squared (within)	0.242	0.242	0.242	0.242
Adjusted R-squared (within)	0.218	0.218	0.218	0.218
Number of districts	56	56	56	56

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1. Note that the levels of IEs in mat & neo health (t-2) are kept in the year fixed effects as they apply on the country level.

[240] Results for these robustness checks are available upon request.

Columns 1 and 2 in Table 5.3. show that only the World Bank projects seem to receive enough direct input from corresponding impact evaluations to lead to significantly more efficient disbursements and overall projects. The significantly negative effects of impact evaluations conditional on projects from the EU (column 4) and projects and disbursements from Sweden and the World Bank (columns 3 and 4) indicate that these donors benefit from the general past knowledge creation in maternal and newborn health. Japan's aid, however, never seems to benefit from impact evaluation insights. These findings are well in line with general rankings of donor quality. The World Bank has a long history of using impact evaluations, Sweden and the EU are generally ranked as good quality donors and Japan is lacking behind regarding quality (Roodman 2012).

The final setting is built on birth-level data and employs the strict identification strategy using mother- and year-fixed effects. Table 5.4. shows the results for the effects of an active match between impact evaluations in maternal and newborn health and corresponding projects on infant deaths. Column 1 includes mother-fixed effects and year-fixed effects.

Table 5.4. Regressions with Mother Fixed Effects

	(1)	(2)	(3)	(4)
	INFANT DEATHS			
Active Mat&NeoIE Match (t)	-0.025**	-0.023**	-0.019*	-0.018*
	(0.017)	(0.028)	(0.062)	(0.074)
Birth order (t)	-0.030***		-0.031***	
	(0.000)		(0.000)	
Twin status (t)	0.112***	0.113***	0.112***	0.113***
	(0.000)	(0.000)	(0.000)	(0.000)
Mother-fixed effects	√	√	√	√
Year-fixed effects	√	√	√	√
Birth order-fixed effects		√		√
District-specific linear time trend			√	√
Observations	74,484	74,483	74,484	74,483
Number of districts	56	56	56	56
R-squared	0.305	0.307	0.307	0.308
R-squared (within)	0.0154	0.0130	0.0156	0.0130
Adjusted R-squared (within)	0.0154	0.0130	0.0155	0.0129

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1. Regressions include sample weights. Robust standard errors are clustered at the district-level.

The dummy indicating an active Health IE – project match shows a negative effect significant at the 5 percent-level. The effect shows that the probability for infants dying is reduced by 2.5 percent compared to their siblings who were born before the existence of such a match. Importantly, the birth order is controlled for, as first born children would generally have a higher mortality risk. The effects are also robust and significant at the 5 percent-level, to the inclusion of birth order-fixed effects in column 2, in addition to mother-fixed effects and year-fixed effects. This specification captures all factors equal to all first born, second born, etc. – comparable to the logic of year-fixed effects. Columns 3 and 4 show that results are also robust to the inclusion of district-specific linear time trends, which capture general linear improvements over the time period each mother gives birth to all her children. As expected, coefficients decrease if another source of endogeneity bias with downwards direction is eliminated. The robustly significant impact of impact evaluations on outcomes established with this conservative estimation strategy establishes confidence in the results obtained throughout the analysis, which indicate the existence of a measurable positive impact of impact evaluations on improved outcomes, which works through corresponding development projects.

5.6. CONCLUSION

This chapter presents several models and estimation strategies at three levels of analyses including the cross-country, district-level and birth-level to test a potential impact of impact evaluations. In particular, the chapter focuses on impact evaluations in the health sector, as these can arguably have relative immediate effects on health-related outcome measures and therefore simplify timing assumptions. The challenge of the analysis is to establish a meaningful link between impact evaluations and outcomes, because it cannot be direct but only work via development projects (from the government, international donors or NGOs). Importantly and the more remarkable, all presented models tell the same story – that there is a positive effect of impact evaluations on final outcomes. Endogeneity concerns are largely reduced by the lag-structure employed as well as a large set of control variables and different fixed effects.

Cross-country evidence leads to the conclusion that the number of impact evaluations done per country significantly correlates with improved health outcomes measured through Disability-Adjusted Life Years (DALYs). One additional evaluation of health interventions reduces the average DALYs per country by 0.1 percent. On average, this corresponds to 23,600 less life years lost due to disabilities. For testing the existence of an effect of impact evaluations conditional on corresponding health projects on a global scale, more fine-grain data would be needed that allow a precise linkage between evaluations and development projects. For the example of one country – Uganda – this chapter provides such data and a more rigorous analysis. Results show that increasing impact evaluations in maternal and newborn health lead to significantly reduced infant mortality rates, if aid disbursements in the same category increase. Interestingly, there are heterogenous effects for different donors. With the limitation of the small sample of four different donors, suggestive evidence for the influence of impact evaluations reproduce general donor rankings: the World Bank with a long and deep history of using impact evaluations seems to make the best use of them, leading to more efficient projects. Positive effects are also shown for Sweden and the EU, generally classified as very good (Sweden) and well enough (EU) donors. Japan, however, which is generally low ranked with regard to its aid quality, does not seem to be influenced at all by impact evaluations. These findings, however, need deeper analyses to be conclusive.

The effects found for the impact of impact evaluations and their robustness is surprisingly strong in light of the available qualitative evidence, which judges the direct policy influence of impact evaluations rather moderate. One reason for the discrepancy could be that effects can be expected to be larger for demand-driven evaluations, which are particularly prominent in health and social development sectors (Jones et al. 2009). Effects are therefore likely to be smaller in other policy areas. However, the provided evidence can only hint at lower bounds of the impact of impact evaluations in Uganda, as sub-national data on health expenditures from the national government, through which impact evaluations are similarly likely to work, is not available.

This chapter is the first to give a quantitative response to the policy debate on whether the rise in impact evaluations and increased resources channeled into it, is reasonable. Naturally, this question is linked to aid effectiveness, in general. As we still know generally very little about what makes development assistance effective, Tierney et al. (2011, 1891) state right to the point that *"[n]early five trillion dollars seems a high price to pay for uncertainty and misunderstanding[.]"* in their article *"More dollars than sense?"*. In light of the huge amounts of money potentially spent without much "sense" through ineffective aid disbursements, every dollar invested in rigorous impact evaluations that have the potential to find out what works and what does not, seems sensibly invested – based on the findings of this chapter.

The deep analysis of mechanisms in how evaluations work conditional on ODA, as presented in this chapter is restricted to a small sample of sub-national data for only one country. Pooling sub-national IEs and ODA project data for more countries would allow deeper analyses of the mechanisms and is open to future research along with identifying channels to bridge academic impact evaluations and policy.

APPENDIX FOR CHAPTER 5

Table A5.1. Number of health transactions per district and category

District	Number of transactions in: All health projects	General health	Maternal and newborn health	HIV/Aids	Health infrastructure	Malaria	Access to medicine	Garbage management
Adjumani	1	1	0	0	0	0	0	0
Apac	23	2	8	8	0	3	2	0
Arua	6	3	0	2	0	0	1	0
Bugiri	15	0	0	14	0	0	1	0
Bundibugyo	4	1	2	0	0	0	1	0
Bushenyi	12	2	8	2	0	0	0	0
Busia	1	1	0	0	0	0	0	0
Gulu	8	3	1	1	0	3	0	0
Hoima	6	1	2	2	0	0	1	0
Iganga	36	7	0	26	2	0	1	0
Jinja	13	0	0	3	9	0	1	0
Kabale	14	4	1	5	3	0	1	0
Kabarole	7	1	1	3	0	0	1	1
Kaberamaido	5	1	0	3	0	0	1	0
Kalangala	1	0	0	1	0	0	0	0
Kampala	26	5	3	16	1	1	0	0
Kamuli	19	0	0	19	0	0	0	0
Kamwenge	4	1	1	1	0	0	1	0
Kanungu	6	1	3	2	0	0	0	0
Kapchorwa	4	2	0	0	0	0	2	0
Kasese	7	1	4	1	0	0	1	0
Katakwi	2	1	0	0	0	0	1	0
Kayunga	4	0	0	3	0	0	1	0
Kibale	3	1	1	0	0	0	1	0
Kiboga	1	1	0	0	0	0	0	0
Kisoro	20	18	1	0	0	0	1	0
Kitgum	4	1	0	1	0	2	0	0
Kotido	5	3	0	1	0	0	1	0
Kumi	5	2	0	0	0	0	3	0
Kyenjojo	5	1	2	1	0	0	1	0
Lira	12	2	3	5	2	0	0	0
Luwero	1	0	0	0	0	0	1	0
Masaka	9	1	0	4	0	0	4	0
Masindi	5	3	0	1	0	0	1	0
Mayuge	7	0	1	5	0	0	1	0
Mbale	9	2	0	4	0	0	3	0
Mbarara	8	3	1	3	0	0	1	0
Moroto	4	1	1	2	0	0	0	0

Moyo	4	1	0	2	0	0	1	0
Mpigi	5	0	0	2	0	0	3	0
Mubende	5	1	2	0	0	0	1	1
Mukono	6	0	0	4	0	0	2	0
Nakapiripirit	3	1	0	2	0	0	0	0
Nakasongola	1	0	0	0	0	0	1	0
Nebbi	8	1	0	0	6	0	1	0
Ntungamo	17	14	1	0	2	0	0	0
Pader	5	1	0	0	1	2	1	0
Pallisa	2	1	0	0	0	0	1	0
Rakai	7	1	0	1	3	0	2	0
Rukungiri	4	2	1	0	0	0	1	0
Sembabule	2	1	0	0	0	0	1	0
Sironko	2	1	0	1	0	0	0	0
Soroti	14	9	1	2	0	0	2	0
Tororo	6	3	0	1	0	0	2	0
Wakiso	7	0	0	7	0	0	0	0
Yumbe	1	1	0	0	0	0	0	0
Total	421	115	49	161	29	11	54	2

Table A5.2. Number and characteristics of impact evaluations in health per category

	Number of impact evaluations in:								
	All health sectors	General health	Maternal and newborn health	HIV/Aids	Health infrastructure	Malaria	Nutrition	Mental health	Sanitation
Total number of which	96	13	9	43	1	11	6	12	1
Journal articles	87	11	7	42	1	10	4	12	0
RCTS	83	10	4	41	0	11	5	12	0
Mean result (scale 1-5)	3.79	4	4	3.3	4	3.21	3.7	4.5	4

Appendix A5.3. Definitions of variables and summary statistics

Variables for Tables 5.1., A5.4., A5.5., A5.6., A5.7.	Description of Variable	Source
Logged DALYs number	Logged disability adjusted life years (DALYs). DALYs calculation: Years of Life Lost (number of deaths*remaining life expectancy) + Years Lived with Disability (number of cases*disability weights).	Global Burden of Disease Collaborative Network (2017)
DALYs number	Total number of DALYs. DALYs calculation: Years of Life Lost (number of deaths*remaining life expectancy) + Years Lived with Disability (number of cases*disability weights).	Global Burden of Disease Collaborative Network (2017)
DALYs rate	Number of DALYs in 100,000 inhabitants. DALYs calculation: Years of Life Lost (number of deaths*remaining life expectancy) + Years Lived with Disability (number of cases*disability weights).	Global Burden of Disease Collaborative Network (2017)
Stock of health IEs	Stock of impact evaluations in health per country (adding the ones, of which the results are communicated to the policy in that year).	International Initiative for Impact Evaluation (2018)
Number of health IEs	Number of impact evaluations in health per country, of which the results are communicated to the policy in that year.	International Initiative for Impact Evaluation (2018)
Logged int. health disbursements, const US$	Logged health sector ODA gross disbursements in constant 2011 US$. 1 US$ was added to all monetary values before logging.	OECD (2018)
Int. health disbursements, const US$	Total health sector ODA gross disbursements in constant 2011 US$.	OECD (2018)
Gov. health exp. (% gov exp.)	Health expenditure, public as percentage of total government expenditure.	WDI (World Bank 2018)

Logged population	Logged population numbers.	WDI (World Bank 2018) and Coppedge et al. (2017) (V-Dem)
Logged GDP, const US$	Logged Gross Domestic Product (GDP) in constant 2011 US$. 1 US$ was added to all monetary values before logging.	WDI (World Bank 2018)
Polity IV	Democraty measure from fully autocratic (-10) to fully democratic (+10).	Coppedge et al. (2017) (V-Dem)
Political corruption	Political corruption index combining the executive corruption index and public sector corruption index referring to public sector employees. Ranges between zero and one from less corrupt to more corrupt systems.	Coppedge et al. (2017) (V-Dem)
Election dummy	Binary variable equal to one if executive election takes place that year.	Coppedge et al. (2017) (V-Dem)
Internal conflict	Binary variable equal to one if country suffered internal armed conflict in given year.	Coppedge et al. (2017) (V-Dem) based on Brecke (2001)
Local government	Binary variable equal to one if local government exists.	Coppedge et al. (2017) (V-Dem)
Government effectiveness	Government effectiveness measures the quality of public service provision, bureacracy, competence of civil servants, independence of civil service and credibility of the government's commitment to policies.	Coppedge et al. (2017) (V-Dem) based on Kaufmann et al. (2016)
Free acad. Expression	Measure based on survey question: "Is there academic freedom and freedom of cultural expression related to political issues?" Responses: 0 - not respected, 1 - weakly respected, 2 - somewhat respected, 3 - mostly respected, 4 - fully respected.	Coppedge et al. (2017) (V-Dem) based on Pemstein et al. (2015)

Variables for Tables 5.2., 5.3., A5.9.	Description of Variable	Source
IMR	Infant mortality rate in 1,000 live births per district in Uganda, constructed as follows: (Number of deaths among infants<12 months of age per year / Number of live-born children per year)*1,000.	Own construction based on DHS (2018) Birth Recode (UGBR)
Under-five-mortality rate	Below-five mortality rate in 1,000 live births per district in Uganda, constructed as follows: (Number of deaths among infants<59 months of age per year / Number of live-born children per year)*1,000.	Own construction based on DHS (2018) Birth Recode (UGBR)
Stocks of IEs in mat & neo health	Stock of impact evaluations in maternal and newborn health (adding the ones, of which the results are communicated to the policy in that year).	International Initiative for Impact Evaluation (2018) data coded into own categories
Number of IEs in mat & neo health	Number of impact evaluations in maternal and newborn health, of which the results are communicated to the policy in that year.	International Initiative for Impact Evaluation (2018) data coded into own categories
Logged mat & neo health disb	Logged total aggregated disbursements in maternal and newborn health projects in constant 2011 US$. 1 US$ was added to all monetary values before logging.	AidData (2018) data coded into own categories
Mat & neo health disb	Total aggregated disbursements in maternal and newborn health projects in constant 2011 US$.	AidData (2018) data coded into own categories
Conflict deaths	Number of total fatalities resulting from conflict event per Ugandan districts.	Goodman et al. (2016), GeoQuery AidData, created using UCDP Georeferenced Event Dataset (GED) global version 17.1.
Logged population	Logged total population count per Ugandan districts, linearly interpolated.	Goodman et al. (2016), GeoQuery AidData, World v4.GPWv4, UN Adjusted
Population density	Population density per Ugandan districts, persons per square km, linearly interpolated.	Goodman et al. (2016), GeoQuery AidData, World v4.GPWv4, UN Adjusted
Urban landcover	Land coverage with urban regions of districts per year.	Goodman et al. (2016), GeoQuery AidData, Global Land Cover Facility (MCD12Q1)

Rainfall	Mean precipitation per district per year.	Goodman et al. (2016), GeoQuery AidData, created using UDel Precipitation dataset (v4.01)
Temperature	Average air temperature per district per year.	Goodman et al. (2016), GeoQuery AidData, created using UDel Air Temperature dataset (v4.01)
Vegetation index	Yearly value for Normalized Difference Vegetation Index (NDVI), mean per Ugandan districts.	Goodman et al. (2016), GeoQuery AidData, created using the NASA Long Term Data Recode (v4) AVHRR data.
Mother's age at birth	Mother's age at birth.	DHS (2018) Birth Recode (UGBR)

Variables Table 5.4.	**Description of Variable**	**Source**
Infant deaths	Binary variable equal to one, if infant died below 12 months of age, zero otherwise (excluding stillbirths).	Own construction based on DHS (2018) Birth Recode (UGBR)
Active Mat&NeoIE Match	Binary variable equal to one, as soon as there was a health impact evaluation in t-2 in maternal and newborn health and a project in maternal and newborn health in t-1, calculated within the period, the mother is giving birth to all her children. The variable is zero, as long as there is no such match.	Own construction based on DHS (2018) Birth Recode (UGBR)
Twin status	Binary variable equal to one, if infant is part of a multiple birth, zero if single birth.	Own construction based on DHS (2018) Birth Recode (UGBR)
Birth order	Birth order of child.	DHS (2018) Birth Recode (UGBR)

Summary Statistics Variables Tables 5.1., A5.4., A5.5., A5.6., A5.7.	Mean	Sd	Min	Max
Logged DALYs number	15.59	1.52	12.11	20.08
DALYs number	23,600,000	71,400,00	183362.7	529,000,000
DALYs rate	46825.63	25030.57	18068.65	188268.1
Stock of health IEs	6.36	18.16	0.00	206.00
Number of health IEs	1.33	3.20	0	36
Logged int. health disbursements, const US$	15.58	3.79	0	20.32
Int. health disbursements, const US$	38,300,000	68,500,000	0.00	670,000,000
Gov. health exp. (% gov exp.)	10.01	4.62	0.76	30.60
Logged population	16.49	1.52	13.07	21.01
Logged GDP, const US$	24.13	1.87	19.72	29.53
Polity IV	2.60	5.87	-10	10
Political corruption	0.63	0.21	0.03	0.96
Election dummy	0.14	0.35	0	1
Internal conflict	0.05	0.23	0	1
Local government	0.97	0.16	0	1
Government effectiveness	-0.46	0.64	-1.98	1.31
Free acad. Expression	2.48	0.93	0.16	4

Summary Statistics Variables Tables 5.2., 5.3., A5.9.	Mean	Sd	Min	Max
IMR	79.94	48.93	0	287.12
Under-five-mortality rate	105.25	59.19	0	352.94
Stocks of IEs in mat & neo health	4.11	3.57	0	9
Number of IEs in mat & neo health	0.50	0.68	0	2
Logged mat & neo health disb	0.41	2.13	0	17.15
Mat & neo health disb	38,601.67	922,524.3	0	28,300,000
Conflict deaths	11.51	54.28	0	662
Logged population	12.96	0.65	9.63	14.41
Population density	322.28	998.92	15.77	10023.75
Urban landcover	74.80	259.80	0	1853
Rainfall	97.97	19.28	48.74	235.68
Temperature	22.97	1.78	17.15	27.58
Vegetation index	5951.48	517.69	3957.63	7379.76
Mother's age at birth	25.75	1.31	21.15	33.23

Summary Statistics Variables Table 5.4.	Mean	Sd	Min	Max
Infant deaths	0.07	0.25	0.00	1
Active Mat&NeoIE Match	0.03	0.17	0.00	1
Twin status	0.03	0.17	0.00	1
Birth order	3.92	2.45	1.00	18

Appendix Table A5.4. Timing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				LOGGED NUMBER OF DALYS			
Number of health IEs (t)	-0.000						
	(0.966)						
Number of health IEs (t-1)		-0.002**					
		(0.021)					
Number of health IEs (t-2(= t_{policy}))			-0.001**				
			(0.035)				
Number of health IEs (t-3(= $t_{endline}$))				-0.001			
				(0.329)			
Number of health IEs (t-4)					-0.002		
					(0.410)		
Number of health IEs (t-5)						-0.000	
						(0.976)	
Number of health IEs (t-6)							-0.000
							(0.914)
Number of observations	1,314	1,314	1,314	1,314	1,314	1,314	1,314
R-squared (within)	0.584	0.586	0.585	0.585	0.585	0.584	0.584
Number of countries	81	81	81	81	81	81	81

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1. Controls from the preferred specification as well as country- and year-fixed effects and country-specific linear time trends are included. Full results are available upon request.

Table A5.5. Influence of the number of health IEs on logged DALYs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				LOGGED NUMBER OF DALYS			
Number of health IEs (t-2(= t_{policy}))	-0.003***	-0.002***	-0.001**	-0.001**	-0.002**	-0.001**	-0.001**
	(0.002)	(0.008)	(0.042)	(0.035)	(0.023)	(0.035)	(0.029)
Logged int. health disbursements,		-0.001	-0.002	-0.000	0.001		0.001
const (t-1)		(0.135)	(0.116)	(0.851)	(0.483)		(0.495)
Gov. health exp. (% gov exp.) (t-1)		-0.000	-0.000	-0.001	-0.001		-0.001
		(0.878)	(0.733)	(0.543)	(0.497)		(0.376)
Logged population (t-1)			0.818***	0.990***	0.922***	0.652	0.696*
			(0.004)	(0.001)	(0.007)	(0.135)	(0.085)
Logged GDP, const (t-1)			-0.001	-0.065*	-0.040		
			(0.991)	(0.062)	(0.314)		
Polity IV (t-1)			0.001	0.001	0.003	0.003*	0.003*
			(0.556)	(0.367)	(0.103)	(0.073)	(0.099)
Political corruption (t-1)				-0.005	-0.057	-0.043	-0.070
				(0.910)	(0.266)	(0.412)	(0.208)
Election dummy (t-1)				-0.003	-0.006	-0.006	-0.006
				(0.568)	(0.247)	(0.193)	(0.256)
Internal conflict (t-1)				0.002	-0.005	-0.006	-0.008
				(0.741)	(0.610)	(0.610)	(0.447)
Local government (t-1)				0.042***	0.037***	0.037***	0.031***
				(0.000)	(0.000)	(0.000)	(0.004)
Government effectiveness (t-1)					-0.002	-0.001	0.009
					(0.883)	(0.947)	(0.645)
Free acad. expression (t-1)					-0.012	-0.022*	-0.014
					(0.296)	(0.073)	(0.203)
Constant	14.814***	14.112***	1.695	0.348	0.876	5.004	3.652

	(0.000)	(0.000)	(0.688)	(0.943)	(0.872)	(0.482)	(0.579)
Country- and year FE	√	√	√	√	√	√	√
Country-specific linear time trends	√	√	√	√	√	√	√
Number of observations	2,880	2,332	1,942	1,314	1,005	1,174	1,018
Number of countries	160	142	120	81	81	93	82
R-squared (within)	0.618	0.558	0.552	0.585	0.520	0.515	0.505
Adjusted R-squared (within)	0.593	0.526	0.517	0.548	0.464	0.464	0.449

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A5.6. Influence of the number of health IEs on the DALY rate (in 100,000 inhabitants)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				DALY RATE (IN 100,000 INHABITANTS)			
Number of health IEs (t-2(=t_{policy}))	-100.781*	-54.379	-37.844	-49.186	-55.964	-44.705	-56.337
	(0.082)	(0.203)	(0.363)	(0.230)	(0.186)	(0.279)	(0.187)
Logged int. health disbursements,		-40.028	-77.637	8.949	120.267		118.512
const (t-1)		(0.326)	(0.280)	(0.930)	(0.436)		(0.443)
Gov. health exp. (% gov exp.) (t-1)		-38.335	-48.906	-70.109	-103.244		-109.263
		(0.552)	(0.517)	(0.384)	(0.267)		(0.247)
Logged population (t-1)			2,529.198	8,228.272	4,416.795	-15,258.691	-4,973.650
			(0.873)	(0.645)	(0.823)	(0.435)	(0.804)
Logged GDP, const US$ (t-1)			-2,575.653	-6,944.330*	-4,223.171		
			(0.555)	(0.085)	(0.324)		
Polity IV (t-1)			107.105	130.626	239.987	252.618	242.651
			(0.309)	(0.344)	(0.153)	(0.136)	(0.151)
Political corruption (t-1)				-1,663.464	-4,820.974	-3,450.053	-5,187.435
				(0.597)	(0.245)	(0.386)	(0.226)
Election dummy (t-1)				-376.712	-640.287	-552.585	-620.008
				(0.382)	(0.246)	(0.220)	(0.267)
Internal conflict (t-1)				432.983	-118.164	-166.491	-213.882
				(0.306)	(0.886)	(0.823)	(0.794)
Local government (t-1)				1,927.303***	1,306.822	1,538.963**	838.414
				(0.006)	(0.228)	(0.032)	(0.443)
Government effectiveness (t-1)					174.756	-522.317	306.962
					(0.857)	(0.629)	(0.762)
Free acad. expression (t-1)					-445.697	-1,075.037	-454.027
					(0.539)	(0.157)	(0.517)
Constant	66,967.503***	4,943.817***	26,490.229	31,011.859	28,237.132	307,248.598	83,948.022

	(0.000)	(0.001)	(0.909)	(0.912)	(0.929)	(0.337)	(0.798)
Country- and year FE	√	√	√	√	√	√	√
Country-specific linear time trends	√	√	√	√	√	√	√
Number of observations	2,880	2,332	1,942	1,314	1,005	1,174	1,018
Number of countries	160	142	120	81	81	93	82
R-squared (within)	0.778	0.765	0.755	0.746	0.682	0.672	0.678
Adjusted R-squared (within)	0.764	0.747	0.736	0.724	0.645	0.638	0.641

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A5.7. The influence of stocks of impact evaluations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				LOGGED NUMBER OF DALYS			
Stock of health IEs (t-2(= t_{policy}))	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001
	(0.190)	(0.288)	(0.535)	(0.450)	(0.390)	(0.222)	(0.299)
Logged int. health disbursements,		-0.001	-0.002	-0.000	0.002		0.002
const (t-1)		(0.168)	(0.122)	(0.891)	(0.410)		(0.408)
Gov. health exp. (% gov exp.) (t-1)		-0.000	-0.000	-0.001	-0.001		-0.001
		(0.938)	(0.783)	(0.605)	(0.556)		(0.437)
Logged population (t-1)			0.792***	0.946***	0.873**	0.594	0.644
			(0.007)	(0.005)	(0.014)	(0.180)	(0.125)
Logged GDP, const US$ (t-1)			0.002	-0.062*	-0.035		
			(0.974)	(0.080)	(0.386)		
Polity IV (t-1)			0.001	0.001	0.003*	0.003*	0.003*
			(0.538)	(0.332)	(0.084)	(0.052)	(0.077)
Political corruption (t-1)				0.004	-0.047	-0.032	-0.057
				(0.926)	(0.335)	(0.519)	(0.265)
Election dummy (t-1)				-0.003	-0.006	-0.006	-0.006
				(0.585)	(0.252)	(0.186)	(0.261)
Internal conflict (t-1)				0.002	-0.004	-0.003	-0.006
				(0.689)	(0.737)	(0.809)	(0.589)
Local government (t-1)				0.040***	0.036***	0.035***	0.029***
				(0.000)	(0.001)	(0.000)	(0.007)
Government effectiveness (t-1)					-0.004	-0.003	0.007
					(0.758)	(0.864)	(0.727)
Free acad. expression (1-1)					-0.013	-0.024*	-0.016

					(0.265)	(0.057)	(0.175)
Constant	16.514***	15.502***	2.903	2.097	2.792	7.691	5.998
	(0.000)	(0.000)	(0.550)	(0.717)	(0.654)	(0.318)	(0.420)
Observations	2,880	2,332	1,942	1,314	1,005	1,174	1,018
R-squared (within)	0.621	0.561	0.553	0.587	0.523	0.522	0.510
Number of countries	160	142	120	81	81	93	82
Adjusted R-squared (within)	0.596	0.529	0.518	0.551	0.468	0.471	0.455

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A5.8. The influence of the number of health evaluations on logged DALYs in periodic data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				LOGGED NUMBER OF DALYS			
Number of health IEs (t-1)	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.544)	(0.546)	(0.463)	(0.453)	(0.463)	(0.581)	(0.503)
Logged int. health disbursements,		0.002	0.003	0.003	0.003		0.004
const (t-1)		(0.623)	(0.336)	(0.329)	(0.350)		(0.393)
Gov. health exp. (% gov exp.) (t-1)		-0.000	-0.001	-0.001	-0.001		-0.001
		(0.799)	(0.365)	(0.338)	(0.345)		(0.293)
Logged population (t-1)			0.909**	0.889**	0.903**	0.674	0.754
			(0.029)	(0.031)	(0.033)	(0.185)	(0.115)
Logged GDP, const US$ (t-1)			0.108	0.108	0.108		
			(0.229)	(0.232)	(0.225)		
Polity IV (t-1)			0.007	0.007	0.006	0.005	0.006
			(0.452)	(0.442)	(0.425)	(0.561)	(0.452)
Political corruption (t-1)				0.032	0.037	0.031	0.033
				(0.438)	(0.439)	(0.492)	(0.457)
Election dummy (t-1)				-0.002	-0.002	-0.002	-0.002
				(0.751)	(0.757)	(0.751)	(0.715)
Local government (t-1)				-0.017	-0.018	0.003	-0.013
				(0.358)	(0.363)	(0.591)	(0.442)
Government effectiveness (t-1)					0.018	0.016	0.024
					(0.596)	(0.616)	(0.476)
Free acad. expression (t-1)					0.008	-0.002	0.000
					(0.686)	(0.912)	(0.989)
Constant	14.275***	14.107***	-2.561	-2.234	-2.449	3.991	2.480

	(0.000)	(0.000)	(0.749)	(0.779)	(0.761)	(0.623)	(0.746)
Number of observations	640	616	523	523	523	540	531
Number of countries	160	155	133	133	133	137	135
R-squared (within)	0.787	0.782	0.765	0.766	0.766	0.775	0.784
Adjusted R-squared (within)	0.714	0.706	0.679	0.677	0.676	0.692	0.701

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A5.9. Full specification of Table 5.2. from the main results part

	(1)	(2)	(3)	(4)
	INFANT MORTALITY RATE (IN 1,000 LIVE BIRTHS)			
	w/o controls	with geo. controls	with pop. controls	preferred (full)
Logged disb (t-1) * IEs (t-2) in mat & neo health	-1.091*	-1.106*	-1.222*	-1.234**
	(0.072)	(0.061)	(0.053)	(0.044)
Logged mat & neo health disb (t-1)	1.062	1.012	1.183	1.182
	(0.202)	(0.203)	(0.157)	(0.161)
Conflict deaths (t)			0.018	0.016
			(0.424)	(0.491)
Logged population (t)			12.410	12.332
			(0.624)	(0.629)
Population density (t)			0.011***	0.010***
			(0.000)	(0.000)
Urban landcover (t)			0.028***	0.026***
			(0.001)	(0.008)
Rainfall (t)		-0.142		-0.081
		(0.306)		(0.605)
Temperature (t)		7.735		8.169
		(0.143)		(0.131)
Vegetation index (t)		0.002		0.005
		(0.791)		(0.589)
Mother's age at birth (t)				4.021**
				(0.025)
Constant	196.507***	24.055	47.175	-339.302
	(0.000)	(0.848)	(0.877)	(0.259)
District- and year-fixed effects	√	√	√	√
Observations	1,008	1,008	1,008	1,008
R-squared (within)	0.291	0.293	0.297	0.304
Adjusted R-squared (within)	0.277	0.277	0.280	0.285
Number of districts	56	56	56	56

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1. Note that the levels of IEs in mat & neo health (t-2) are kept in the year fixed effects as they apply on the country-level.

Table A5.10. Main district-level results of interacted effects including district-specific linear time trends

	(1)	(2)	(3)	(4)
	INFANT MORTALITY RATE (IN 1,000 LIVE BIRTHS)			
	w/o controls	with geo. controls	with pop. controls	preferred (full)
Logged disb (t-1) * IEs (t-2) in mat & neo health	-1.439**	-1.391**	-1.457**	-1.388**
	(0.042)	(0.039)	(0.041)	(0.038)
Logged mat & neo health disb (t-1)	1.889*	1.873*	1.886*	1.879*
	(0.056)	(0.051)	(0.057)	(0.054)
Conflict deaths (t)			0.003	0.002
			(0.915)	(0.933)
Logged population (t)			7.796	8.512
			(0.955)	(0.955)
Population density (t)			-0.029	-0.012
			(0.541)	(0.804)
Urban landcover (t)			-0.012	0.005
			(0.830)	(0.930)
Rainfall (t)		0.057		0.064
		(0.712)		(0.700)
Temperature (t)		6.652		7.115
		(0.211)		(0.217)
Vegetation index (t)		0.007		0.006
		(0.503)		(0.602)
Mother's age at birth (t)				4.354***
				(0.010)
Constant	3.517	-193.359	-94.221	-412.974
	(0.535)	(0.144)	(0.959)	(0.835)
District- and year-fixed effects	√	√	√	√
District-specific linear time trends	√	√	√	√
Observations	1,008	1,008	1,008	1,008
R-squared (within)	0.384	0.385	0.384	0.391
Adjusted R-squared (within)	0.335	0.335	0.332	0.337
Number of districts	56	56	56	56

Notes: Robust p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1. Note that the levels of IEs in mat & neo health (t-2) are kept in the year-fixed effects as they apply on the country-level.

References

Abadie, Alberto (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies* 72(1): 1-19.

Abbas, S.M. Ali, Nazim Belhocine, Asmaa El-Ganainy and Mark Horton (2010). A Historical Public Debt Database. *IMF Working Paper WP/10/245.* Washington, DC: International Monetary Fund.

Achen, Christopher H. (2001). Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables. *Mimeo.* Available at https://www.princeton.edu/csdp/events/Achen121201/achen.pdf (accessed: 30 July 2017).

AidData (2018). Development Projects in Uganda. Uganda AIMS Geocoded Research Release, Version 1.4.1. Available at https://www.aiddata.org/data/uganda-aims-geocoded-research-release-level-1-v1-4-1 (accessed: 03 June 2018).

Aidt, Toke and Bianca Dallal (2008). Female Voting Power: the Contribution of Women's Suffrage to the Growth of Social Spending in Western Europe (1869–1960). *Public Choice 134*, 3-4: 391-417.

Ajzen, Icek (1985). From Intentions to Actions: A Theory of Planned Behavior. In Action-Control: From Cognition to Behavior. eds. J Kuhl and J Beckmann. Heidelberg, Berlin: *SSSP Springer Series in Social Psychology*, 11–39.

Ajzen, Icek (1991). The Theory of Planned Behavior. *Organisational Behaviour and Human Decision Processes* 50: 179–211.

Ajzen, Icek and Martin Fishbein (2005). *The Influence of Attitudes on Behavior.* In The Handbook of Attitudes, eds. D. Albarracín, B.T. Johnson, and M.P. Zanna. Mahwah, NJ: Erlbaum, 173–221.

Aldenhoff, Frank-Oliver (2007). Are Economic Forecasts of the International Monetary Fund Politically Biased? A Public Choice Analysis. *Review of International Organizations* 2, 3: 239-260.

Alesina, Alberto and David Dollar (2000). Who Gives Foreign Aid to Whom and Why? *Journal of Economic Growth 5*, 1: 33-63.

American Economic Association (2018). AEA RCT Registry. https://www.socialscienceregistry.org/ (accessed: 23 January, 2018).

Andersen, Thomas B., Henrik Hansen and Thomas Markussen (2006). US Politics and World Bank IDA-lending. *Journal of Development Studies* 42, 5: 772-794.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press.

Anwar, Mumtaz and Katharina Michaelowa (2006). The Political Economy of US Aid to Pakistan. *Review of Development Economics* 10, 2: 195-209.

Appleby, Ben, Carolyn Roskell and William Daly (2016). What Are Health Professionals' Intentions toward Using Research and Products of Research in Clinical Practice? A Systematic Review and Narrative Synthesis. *Nursing open* 3, 3: 125–39.

Armitage, Christopher J. and Mark Conner (2001). Efficacy of the Theory of Planned Behaviour : A Meta-Analytic Review. *British Journal of Social Psychology* 40: 471–99.

Ashraf, Nava, Erica Field and Jean Lee (2014). Household Bargaining and Excess Fertility: An Experimental Study in Zambia. *American Economic Review* 104, 7: 2210–37.

Athey, Susan and Guido Imbens (2017). *The Econometrics of Randomized Experiments.* Handbook of Economic Field Experiments 1: 73-140.

Baanante, Miguel J. and Lorena A. Valdivia (2015). Impact evaluation and policy decisions: Where are we? A Latin American think tank perspective. *3ie Working Paper* 24. New Delhi: International Initiative for Impact Evaluation (3ie).

Banerjee, Abhijit Vinayak, Esther Duflo and Michael Kremer (2016). The influence of randomized controlled trials on development economics research and on development policy. *Mimeo* MIT.

Banerjee, Abhijit Vinayak, Esther Duflo, Rachel Glennerster and Dhruva Kothari (2010). Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns with and without Incentives. *BMJ* 340.

Barnard-Brak, Lucy, Hansel Burley and Steven M. Crooks (2010). Explaining Youth Mentoring Behavior Using a Theory of Planned Behavior Perspective. *International Journal of Adolescence and Youth* 15, 4: 365–79.

BBC (2005). Germany Takes Lead in Aid Effort. http://news.bbc.co.uk/2/hi/europe/4148719.stm (January 17, 2018).

Bearce, David H. and Daniel C. Tirone (2010). Foreign aid effectiveness and the strategic goals of donor governments. *Journal of Politics* 72, 3: 837-851.

Behrman, Jere R. and Emmanuel Skoufias (2006). Mitigating myths about policy effectiveness: Evaluation of Mexico's antipoverty and human resource investment program. *The ANNALS of the American Academy of Political and Social Science* 606, 1: 244-275.

Benjamin A. Toll, Stephanie S. O'Malley, Nicole A. Katulak, Ran Wu, Joel A. Dubin, Amy Latimer, Boris Meandzija, Tony P. George, Peter Jatlow, Judith L. Cooney and Peter Salovey (2007). Comparing Gain- and Loss-Framed Messages for Smoking Cessation With Sustained-Release Bupropion: A Randomized Controlled Trial. *Psychol Addict Behav* 21, 4: 534–44.

Bermeo, Sarah Blodgett (2011). Foreign aid and regime change: A role for donor intent. *World Development* 39, 11: 2021-2031.

Bermeo, Sarah Blodgett (2016). Aid is not oil: Donor utility, heterogeneous aid, and the aid-democratization relationship. *International Organization* 70, 1: 1-32.

Berthélemy, Jean-Claude (2006). Bilateral donors' interest vs. recipients' development motives in aid allocation: Do all donors behave the same? *Review of Development Economics* 10, 2: 179-194.

Bertoli, Simone, Giovanni Andrea Cornia and Francesco Manaresi (2008). Aid Effort and Its Determinants: A Comparison of the Italian Performance with other OECD Donors. *Dipartimento di Scienze Economiche Working Paper N. 11/2008*. Florence, Italy: Universita' degli Studi di Firenze.

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119, 1: 249-275.

Bertrand, Marianne, Sendhil Mullainathan and Eldar Shafir (2006). Behavioral Economics and Marketing in Aid of Decision Making Among the Poor. *Journal of Public Policy & Marketing* 25, 1: 8–23.

Besley, Timothy, Jose G. Montalvo and Marta Reynal-Querol (2011). Do Educated Leaders Matter? *Economic Journal* 121, 554: F205-F227.

Biglaiser, Glen and Karl DeRouen (2010). The Effects of IMF Programs on US Foreign Direct Investment in the Developing World. *Review of International Organizations* 5, 1: 73-95.

Bilic, Bojan (2005). The Theory of Planned Behaviour and Health Behaviours: Critical Analysis of Methodological and Theoretical Issues. *Hellenic Journal of Psychology* 2: 243–59.

Birdsall, Nancy and Homi Kharas (2010). *Quality of Official Development Assistance Assessment.* Washington, DC: Center for Global Development and Brookings Institution.

Björkman, Martina and Jakob Svensson (2007). Power to the people: Evidence from a randomized field experiment of a community-based monitoring project in Uganda. *Quarterly Journal of Economics.* 124, 2: 735-769.

Bjørnskov, Christian (2010). Do Elites Benefit from Democracy and Foreign Aid in Developing Countries? *Journal of Development Economics* 92, 2: 115-124.

Bjørnskov, Christian (2013). Types of Foreign Aid. *Department of Economics and Business Working Paper 2013-08*. Aarhus, Denmark: Aarhus University.

Bjørnskov, Christian and Niklas Potrafke (2011). Strikes, Ideology and Inequality: Is There Any Connection? Paper presented at the Annual Conference of the Southern Economic Association, Washington DC, November 19-21.

Blue, Carolyn (1995). The Predictive Capacity of the Theory of Reasoned Action and the Theory of Planned Behavior in Exercise Research: An Integrated Literature Review. *Research in Nursing and Health* 18: 105–21.

Borchard, Annegret, David L. B. Schwappach, Aline Barbir and Paula Bezzola (2012). A Systematic Review of the Effectiveness, Compliance, and Critical Factors for Implementation of Safety Checklists in Surgery. *Annals of Surgery* 256, 6: 925–33.

Bowles, Samuel and Sandra Polania-Reyes (2012). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature* 50, 2: 368–425.

Brady, David, Evelyne Huber and John D. Stephens (2014). *Comparative Welfare States Data Set*. University of North Carolina and WZB Berlin Social Science Center.

Brech, Viktor and Niklas Potrafke (2014). Donor Ideology and Types of Foreign Aid. *Journal of Comparative Economics* 42, 1: 61-75.

Breuning, Marijke (1999). Ethnopolitical Parties and Development Cooperation: the Case of Belgium. *Comparative Political Studies* 32, 6: 725-752.

Breuning, Marijke (2001). Women's Representation and Development Assistance: A Cross-National Study. *Women and Politics* 23, 3: 35-54

Broz, J. Lawrence (2008). Congressional Voting on Funding the International Financial Institutions. *Review of International Organizations* 3, 4: 351-74.

Broz, J. Lawrence (2011). The United States Congress and IMF Financing, 1944–2009. *Review of International Organizations* 6, 3-4: 341-368.

Broz, J. Lawrence and Michael Brewster Hawes (2006). Congressional Politics of Financing the International Monetary Fund. *International Organization* 60, 2: 367-99.

Burgess, Diana J. (2009). Are Providers More Likely to Contribute to Healthcare Disparities under High Levels of Cognitive Load? How Features of the Healthcare Setting May Lead to Biases in Medical Decision Making. *Medical Decision Making* 30, 2: 246–57.

Burnside, Craig and David Dollar (2000). Aid, policies, and growth. *American Economic Review* 90, 4: 847-868.

Cameron, A. Colin, Jonah B. Gelbach and Douglas L. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics* 90, 3: 414–27.

Cameron, Drew B., Anjini Mishra and Annette N. Brown (2016). The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness*. 8, 1: 1-21.

Chatagny, Florian (2015). Incentive Effects of Fiscal Rules on the Finance Minister's Behavior: Evidence from Revenue Projections in Swiss Cantons. *European Journal of Political Economy* 39: 184-200.

Chattopadhyay, Raghabendra and Esther Duflo (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica* 72, 5: 1409-1443.

Cheon, Jongpil, Sangno Lee, Steven M. Crooks and Jaeki Song (2012). An Investigation of Mobile Learning Readiness in Higher Education Based on the Theory of Planned Behavior. *Computers and Education* 59, 3: 1054–64.

Cilliers, Jacobus, Oeindrila Dube and Bilal Siddiqi (2015). The White-Man Effect: How Foreigner Presence Affects Behavior in Experiments. *Journal of Economic Behavior and Organization* 118: 397–414.

Clemens, Michael A., Steven Radelet, Rikhil R. Bhavnani and Samuel Bazzi (2012). Counting chickens when they hatch: Timing and the effects of aid on growth. *Economic Journal* 122, 561: 590-617.

Cole, Shawn, Xavier Giné, Jeremy Tobacman, Petia Topavola, Robert Townsend and James Vickery (2013). Barriers to Household Risk Management: Evidence from India. *American Economic Journal: Applied Economics* 5, 1: 104–35.

Commission for Africa (2005). *Our common interest.* Report of the Commission for Africa. Available at http://www.commissionforafrica.info/2005-report (accessed: March 2016).

Conner, Mark and Christopher J. Armitage (1998). Extending the Theory of Planned Behavior: A Review and Avenues for Further Research. *Journal of Applied Social Psychology* 28, 15: 1429–64.

Cooke, Richard and Paschal Sheeran (2004). Moderation of Cognition – Intention and Cognition – Behaviour Relations : A Meta-Analysis of Properties of Variables from the Theory of Planned Behaviour. *Journal of Social Psychology* 43: 159–86.

Cooke, Richard, Mary Dahdah, Paul Norman and David P. French (2014). How Well Does the Theory of Planned Behaviour Predict Alcohol Consumption? A Systematic Review and Meta-Analysis. *Health Psychology Review* 10, 2: 148–67.

Copelovitch, Mark S. (2010). Master or Servant? Common Agency and the Political Economy of IMF Lending. *International Studies Quarterly* 54, 1: 49-77.

Coppedge, Michael ,John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Joshua Krusell, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Moa Olin, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Constanza Sanhueza Petrarca, Johannes von Römer, Laura Saxer, Brigitte Seim, Rachel Sigman, Jeffrey Staton, Natalia Stepanova, and Steven Wilson (2017). V-Dem [Country-Year/Country-Date] Dataset v7.1 *Varieties of Democracy (V-Dem) Project*. Available at https://www.v-dem.net/en/data/data-version-7-1/ (accessed: 13 March 2018).

Courtenay-Hall, Pamela and Larson Rogers (2002). Gaps in Mind: Problems in Envionmental Knowledge-Behaviour Modelling Research. *Environmental Education Research* 8, 3.

Croskerry, Pat (2002). Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. *Acad Emerg Med* 9, 11: 1184–1204.

Croson, Uri and Rachel Gneezy (2009). Gender Differences in Preferences. *Journal of Economic Literature* 47, 2: 1-27.

Cruzatti, John C., Axel Dreher and Sarah Langlotz (2018). World Bank aid and health at the regional level. *Mimeo.*

Dalgaard, Carl-Johan, Henrik Hansen and Finn Tarp (2004). On the empirics of foreign aid and growth. *Economic Journal* 114, 496: F191-F216.

Day, Benjamin (2016). *Aid Salience Spikes' as Drivers of Aid Policy Change.* Paper presented at the Workshop "The Domestic Dimensions of Development Cooperation," University of Antwerp, Belgium (October 24-25).

De Luca, Giacomo, Roland Hodler, Paul A. Raschky and Michele Valsecchi (2015). Ethnic Favoritism: An Axiom of Politics? *CESifo Working Paper 5209.* Munich, Germany: CESifo.

De, Rajlakshmi and Charles Becker (2015). The foreign aid effectiveness debate: Evidence from Malawi. *AidData Working Paper* No. 6. Williamsburg, VA: AidData.

Deaton, Angus (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*: 424-455.

Deck, Cary and Salar Jahedi (2015). The Effect of Cognitive Load on Economic Decision Making: A Survey and New Experiments. *European Economic Review* 78: 97–119.

Detweiler, Jerusha, Brian Bedell, Peter Salovey and Emily Pronin (1999). Message Framing and Sunscreen Use: Gain-Framed Messages Motivate Beach-Goer. *Health Psy* 18, 2: 189–96.

DHS (2018). *The DHS Program. Demographic and Health Surveys.* Data Collection. Available at https://dhsprogram.com/data/data-collection.cfm (accessed: 05 August 2018).

Dietrich, Simone and Matthew S. Winters (2015). Foreign Aid and Government Legitimacy. *Journal of Experimental Political Science* 2, 2: 164–71.

Dietrich, Simone, Minhaj Mahmud and Matthew Winters (2018). Foreign Aid, Foreign Policy, and Domestic Government Legitimacy: Experimental Evidence from Bangladesh. *The Journal of Politics* 80, 1: 133–48.

Dippel, Christian (2015). Foreign Aid and Voting in International Organizations: Evidence from the IWC. *Journal of Public Economics* 132: 1-12.

Dollar, David and Lant Pritchett (1998). *Assessing Aid. What Works, What Doesn't, and Why*. Washington DC.

Doucouliagos, Hristos and Elizabeth Manning (2009). The Allocation of Development Aid Assistance: Do New Donors Have Old Motives? *School of Accounting, Economics and Finance Working Paper 2009/17.* Burwood, Australia: Deakin University.

Doucouliagos, Hristos and Martin Paldam (2008). Aid Effectiveness on Growth: A Meta Study. *European Journal of Political Economy* 24, 1: 1-24.

Doucouliagos, Hristos and Martin Paldam (2010). Conditional Aid Effectiveness: A Meta Study. *Journal of International Development* 22, 4: 391–410.

Doucouliagos, Hristos and Martin Paldam (2011). The Ineffectiveness of Development Aid on Growth: An Update. *European Journal of Political Economy* 27, 2: 399-404.

Dreher, Axel (2006). Does Globalization Affect Growth? Evidence from a New Index of Globalization. *Applied Economics* 38, 10: 1091-1110.

Dreher, Axel and Andreas Fuchs (2011). Does Terror Increase Aid? *Public Choice* 149, 3-4: 337-363.

Dreher, Axel and Nathan M. Jensen (2007). Independent Actor or Agent? An Empirical Analysis of the Impact of US Interests on IMF Conditions. *Journal of Law and Economics* 50, 1: 105-124.

Dreher, Axel and Nathan M. Jensen (2013). Country or Leader? Political Change and UN General Assembly Voting. *European Journal of Political Economy* 29: 183-196.

Dreher, Axel and Sarah Langlotz (2017). Aid and Growth. New Evidence Using an Excludable Instrument. *Heidelberg University Discussion Paper* No. 635.

Dreher, Axel and Steffen Lohmann (2015). Aid and growth at the regional level. *Oxford Review of Economic Policy* 31, 3-4: 420-446.

Dreher, Axel and Valentin F. Lang (2016). The Political Economy of International Organizations. *CESifo Working Paper* 6077.

Dreher, Axel, Andreas Fuchs and Peter Nunnenkamp (2013). New Donors. *International Interactions* 39, 3: 402-415.

Dreher, Axel, Anna Minasyan and Peter Nunnenkamp (2015). Government ideology in donor and recipient countries: Does ideological proximity matter for the effectiveness of aid? *European Economic Review* 79, October: 80-92.

Dreher, Axel, Jan-Egbert Sturm and James Vreeland (2009a). Global Horse Trading: IMF Loans for Votes in the United Nations Security Council. *European Economic Review* 53: 742-757.

Dreher, Axel, Jan-Egbert Sturm and James Vreeland (2009b). Development Aid and International Politics: Does Membership on the UN Security Council Influence World Bank Decisions? *Journal of Development Economics* 88: 1-18.

Dreher, Axel, Kai Gehring and Stephan Klasen (2015a). Gesture Politics or Real Commitment? Gender Inequality and the Allocation of Aid. *World Development* 70: 464-480.

Dreher, Axel, Matthew Gould, Matthew Rablen and James Raymond Vreeland (2014). The Determinants of Election to the United Nations Security Council. *Public Choice* 158: 51-83.

Dreher, Axel, Michael J. Lamla, Sarah M. Lein and Frank Somogyi (2009). The Impact of Political Leaders‘ Profession and Education on Reforms. *Journal of Comparative Economics* 37, 1: 169-193.

Dreher, Axel, Noel Gaston and Pim Martens (2008). *Measuring Globalisation – Gauging its Consequences.* New York, NY: Springer.

Dreher, Axel, Peter Nunnenkamp and Maya Schmaljohann (2015b). The Allocation of German Aid: Self-interest and Government Ideology. *Economics & Politics* 27, 1: 160-184.

Dreher, Axel, Rainer Thiele and Peter Nunnenkamp (2008). Does Aid for Education Educate Children? Evidence from Panel Data. *World Bank Economic Review* 22, 2: 291-314.

Dreher, Axel, Sarah Langlotz and Silvia Marchesi (2017). Information transmission and ownership consolidation in aid programs. *Economic Inquiry* 55: 1671-1688.

Dreher, Axel, Silvia Marchesi and James Raymond Vreeland (2008). The Political Economy of IMF Forecasts. *Public Choice* 137, 1-2: 145-171.

Dreher, Axel, Stephan Klasen, James Raymond Vreeland and Eric Werker (2013). The costs of favoritism: Is politically driven aid less effective? *Economic Development and Cultural Change* 62, 1: 157-191.

Dreher, Axel, Vera Eichenauer and Kai Gehring (2018). Geopolitics, aid and growth. *World Bank Economic Review.* 31, 2:268-28

Dudley, Leonard (1979). Foreign Aid and the Theory of Alliances. *Review of Economics and Statistics* 61, 4: 564-571.

Duflo, Esther, Pascaline Dupas and Michael Kremer (2012). School governance, teacher incentives and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. *National Bureau of Economic Research Working Paper* 17939.

Duflo, Esther, Rachel Glennerster and Michael Kremer (2010). *Using Randomization in Development Economics Research: A Toolkit.* In Handbook of Development Economics. eds. Dani Rodrik and Marc R Rosenzweig. Amsterdam, Netherlands: Elsevier B.V., 3895–3962.

Eagly, Alice H., Steven J. Karau and Mona G. Makhijani (1995). Gender and the Effectiveness of Leaders: A Meta-analysis. *Psychological Bulletin* 117: 125-145.

Easterly, William (2002a). The Cartel of Good Intentions: The Problem of Bureaucracy in Foreign Aid. *Journal of Policy Reform* 5, 4: 223-250.

Easterly, William (2002b). *Evaluating Aid Performance of Donors.* Washington, DC: Center for Global Development.

Easterly, William and Tobias Pfutze (2008). Where Does the Money Go? Best and Worst Practices in Foreign Aid. *Journal of Economic Perspectives* 22, 2: 29-52.

Easterly, William, Ross Levine and David Roodman (2004). Aid, policies, and growth: Comment. *American Economic Review* 94, 3: 774-780.

Ellmers, Bodo, Nuria Molina and Visa Tuominen (2010). Development Diverted: How the International Finance Corporation Fails to Reach the Poor. *Eurodad*, Brussels.

Ertl, Verena, Anett Pfeiffer, Elisabeth Schauer, Thomas Elbert and Frank Neuner (2011). Community-implemented trauma therapy for former child soldiers in northern Uganda: A randomized controlled trial. *Journal of the American Medical Association* 306, 5: 503-512.

Escobar-Lemmon, Maria and Michelle M. Taylor-Robinson (2005). Women Ministers in Latin American Government: When, Where, and Why? *American Journal of Political Science* 49, 4: 829-844.

Farvaque, Étienne, Hakim Hammadou and Piotr Stanek (2009). Select Your Committee: The Impact of Central Bankers' Background on Inflation. *Économie internationale* 117, 99-129.

Federal Ministry for Economic Cooperation and Development (BMZ) (2005). *Review of German Government Funding for the Tsunami Relief Operations*. https://www.oecd.org/derec/germany/36472888.pdf (January 17, 2018).

Feld, Lars P. and Christoph A. Schaltegger (2010). Political Stability and Fiscal Policy: Time Series Evidence for the Swiss Federal Level since 1849. *Public Choice* 144, 3: 505-534.

Ferreira, Fernando and Joseph Gyourko (2014). Does Gender Matter for Political Leadership? The Case of U.S. Mayors. *Journal of Public Economics* 112: 24-39.

Fichten, Catherine, Rhonda Amsel, Mary Jorgensen, Mai N. Nguyen, Jillian Budd, Alice Havel, Laura King, Shirley Jorgensen and Jennison Asuncion (2016). Theory of Planned Behavior : Sensitivity and Specificity in Predicting Graduation and Drop-out among College and University Students. *International Journal of Learning, Teaching and Educational Research* 15, October: 38–52.

Fichten, Catherine, Tali Heiman, Mary Jorgensen, Mai Nhu Nguyen, Alice Havel, Laura King, Jillian Budd, Rhonda Amsel (2016). Theory of Planned Behavior Predicts Graduation Intentions of Canadian and Israeli Postsecondary Students with and without Learning Disabilities/Attention Deficit Hyperactivity Disorder. *International Journal of Higher Education* 5, 1: 208–19.

Findley, Michael G., Adam S. Harris, Helen V. Milner and Daniel L. Nielson (2017). Who Controls Foreign Aid? Elite versus Public Perceptions of Donor Influence in Aid-Dependent Uganda. *International Organization* 71: 1–31.

Fishbein, Martin and Icek Ajzen (1980). *Understanding Attitudes and Predicting Social Behavior. Revised.* Prentice-Hall.

Fleck, Robert K. and Christopher Kilby (2006). How Do Political Changes Influence US Bilateral Aid Allocations? Evidence from Panel Data. *Review of Development Economics* 10, 2: 210-223.

Fratzscher, Marcel and Julien Reynaud (2011). IMF Surveillance and Financial Markets – A Political Economy Analysis. *European Journal of Political Economy* 27, 3: 405-22.

Frederiks, Elisha R., Karen Stenner and Elizabeth V. Hobman (2015). Household Energy Use: Applying Behavioural Economics to Understand Consumer Decision-Making and Behaviour. *Renewable and Sustainable Energy Reviews* 41: 1385–94.

Freier, Ronny and Sebastian Thomasius (2016). Voters Prefer More Qualified Mayors, but Does It Matter for Public Finances? *International Tax and Public Finance* 23, 5: 875-910.

French, David P. and Matthew Hankins (2003). The Expectancy-Value Muddle in the Theory of Planned Behaviour - and Some Proposed Solutions. *British Journal of Health Psychology* 8: 37–55.

Frey, Bruno S. and Stephan Meier (2003). Are Political Economists Selfish and Indoctrinated? Evidence from a Natural Experiment. *Economic Inquiry* 41, 3: 448-462.

Fuchs, Andreas, Axel Dreher and Peter Nunnenkamp (2014). Determinants of Donor Generosity: A Survey of the Aid Budget Literature. *World Development* 56: 172-199.

Funk, Patricia and Christina Gathmann (2015). Gender Gaps in Policy Making: Evidence from Direct Democracy in Switzerland. *Economic Policy* 30, 81: 141-181.

Gary, Aurore and Mathilde Maurel (2015). Donors' policy consistency and economic growth. *Kyklos* 68, 4: 511-551.

Gawande, Atul (2009). *The Checklist Manifesto.* New York: Metropolitan Books. Henry Hold and Company.

Gerber, Alan S. and Donald P. Green (2012). *Field Experiments. Design, Analysis, and Interpretation.* Gloucestershire, UK: Paperbackshop UK Import.

Gerhart, Barry and Sara Rynes (1991). Determinants and Consequences of Salary Negotiations by Male and Female MBA Graduates. *Journal of Applied Psychology* 76, 2: 256-262.

Global Burden of Disease Collaborative Network (2017). *Global Burden of Disease Study 2016 (GBD 2016) Results.* Seattle, United States: Institute for Health Metrics and Evaluation (IHME). Available at http://ghdx.healthdata.org/gbd-results-tool (accessed: 05 August 2018).

Gneezy, Uri, Stephan Meier and Pedro Rey-Biel (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives* 25, 4: 191–210.

Göhlmann, Silja and Roland Vaubel (2007). The Educational and Occupational Background of Central Bankers and its Effect on Inflation: An Empirical Analysis. *European Economic Review* 51, 4: 925-941.

Goodman, Seth, Ariel BenYishay and Daniel Runfola (2016). *Overview of the geo Framework.* AidData. Available at geo.aiddata.org. DOI: 10.13140/RG.2.2.28363.59686 (accessed: 09 May 2018).

Green, Elliot (2008). District creation and decentralisation in Uganda. *Crisis States Working Paper* No. 24. Development Studies Institute. London School of Economics and Political Science.

Guillerm, Marine (2017). Pseudo-panel methods and an example of application to household wealth data. *Economics and Statistics* 491-492: 109-130.

Hallerberg, Mark and Joachim Wehner (2012). The Educational Competence of Economic Policymakers in the EU. *Global Policy* 3, Suppl.1: 9-15.

Hansen, Henrik and Finn Tarp (2001). Aid and growth regressions. *Journal of Development Economics* 64, 2: 547-570.

Hardeman, Wendy, Marie Johnston, Derek Johnston, Debbie Bonetti, Nicholas Wareham and Ann Louise Kinmonth (2002). Application of the Theory of Planned Behaviour in Behaviour Change Interventions: A Systematic Review. *Psychology and Health* 17, 2: 123–58.

Haugen, Arvid S., Eirik Söfteland, Stian K. Almeland, Nick Sevdalis, Barthold Vonen, Geir Eide, Monica Nortvedt, Stig Harthug (2015). Effect of the World Health Organization Checklist on Patient Outcomes. *Annals of Surgery* 261, 5: 821–28.

Hayo, Bernd and Florian Neumeier (2012). Leaders' Impact on Public Spending Priorities: The Case of the German Laender. *Kyklos* 65, 4: 480-511.

Hayo, Bernd and Florian Neumeier (2014). Political Leaders' Socioeconomic Background and Fiscal Performance in Germany. *European Journal of Political Economy* 34: 184-205.

Hayo, Bernd and Florian Neumeier (2016). Political Leaders' Socioeconomic Background and Public Budget Deficits: Evidence from OECD Countries. *Economics & Politics* 28, 1: 55-78.

Head, Keith, Thierry Mayer and John Ries (2010). The Erosion of Colonial Trade Linkages After Independence. *Journal of International Economics* 81, 1: 1-14.

Headey, Derek (2008). Geopolitics and the effect of foreign aid on economic growth: 1970-2001. *Journal of International Development* 20, 2: 161-180.

Herzer, Dierk and Korbinian Nagel (2015). The long-run and short-run effects of health aid on life expectancy. *Applied Economic Letters* 22, 17: 1430-1434.

Hess, Simon (2017). Randomization Inference with Stata: A Guide and Software. *Stata Journal* 17, 3: 630–651.

Hicks, Daniel L., Joan Hamory Hicks and Beatriz Maldonado (2016). Women as Policy Makers and Donors: Female Legislators and Foreign Aid. *European Journal of Political Economy* 41: 46-60.

Hicks, Robert L., Bradley C. Parks, J. Timmons Roberts, Michael J. Tierney (2010). *Greening Aid?: Understanding the Environmental Impact of Development Assistance*. Oxford, UK: Oxford University Press.

Hobbis, Imogen C. A. and Stephen Sutton (2005). Are Techniques Used in Cognitive Behaviour Therapy Applicable to Behaviour Change Interventions Based on the Theory of Planned Behaviour? *Journal of Health Psychology* 10, 1: 7–18.

Hodler, Roland and Paul A. Raschky (2014) Regional Favoritism. *Quarterly Journal of Economics* 129, 2: 995-1033.

Hoffman, Marie, Valerie Mukete, Julia Szmanda and Chelsea Witt (2011). *The Effect of Increased Stress Levels on Memory Test Performance.*

Hossain, Tanjim and John List (2012). The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations. *Management Science* 58, 12: 2151–67.

Hsu, Chia Lin, Chi Ya Chang and Chutinart Yansritakul (2017). Exploring Purchase Intention of Green Skincare Products Using the Theory of Planned Behavior: Testing the Moderating Effects of Country of Origin and Price Sensitivity. *Journal of Retailing and Consumer Services* 34: 145–52.

Humphrey, Chris and Katharina Michaelowa (2013). Shopping for Development: Multilateral Lending, Shareholder Composition and Borrower Preferences. *World Development* 44, 4: 142-155.

Imbens, Guido W. and Jeffrey M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47, 1: 5-86.

Independent Evaluation Group (IEG) (2007). *Independent Evaluation of IFC's Development Results 2007: Lessons and Implications from 10 Years of Experience.* The World Bank, Washington, DC.

International Finance Corporation (IFC) (2004). *Annual Report 2004, Management's Discussion and Analysis, Financial Statements, and Investment Portfolio, Volume 2*. The World Bank, Washington, DC.

International Finance Corporation (IFC) (2016). *Annual Report 2016, Experience Matters.* The World Bank, Washington, DC.

International Initiative for Impact Evaluation (2018). *Impact Evaluation Repository.* Available at http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/ (accessed: 20 March 2018).

International Monetary Fund (IMF) (2016). *IMF Data Access to Macroeconomic and Financial Data, Direction of Trade Statistics (DOTS).* Available at http://data.imf.org/ (accessed: 01 September 2016).

Jacqmin, Julien and Mathieu Lefebvre (2016). Does Sector-specific Experience Matter? The Case of European Higher Education Ministers. *Research Policy* 45, 5: 987-998.

Jensen, Nathan M (2008). *Nation-states and the Multinational Corporation: A Political Economy of Foreign Direct Investment.* Princeton University Press.

Jochimsen, Beate and Sebastian Thomasius (2014). The Perfect Finance Minister: Whom to Appoint as Finance Minister to Balance the Budget. *European Journal of Political Economy* 34: 390-408.

Johnson, Eric J. and Daniel Goldstein (2003). Do defaults save lives? *Science* 302, 1338-1339.

Jones, Benjamin F. and Benjamin A. Olken (2005). Do Leaders Matter? National Leadership and Growth since World War II. *Quarterly Journal of Economics* 120, 3: 835-864.

Jones, Nicola, Nicola Jones, Harry Jones, Liesbet Steer, and Ajoy Datta (2009). Improving impact evaluation production and use. *Working Paper* 300. Overseas Development Institute

Kahneman, Daniel (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review* 93, 5: 1449–75.

Kahneman, Daniel and Amos Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2: 263–91.

Kaja, Ashwin and Eric Werker (2010). Corporate Governance at the World Bank and the Dilemma of Global Governance. *World Bank Economic Review* 24, 2: 171-198.

Kalichman, Seth C. and Brenda Coley (1995). Context Framing to Enhance HIV Antibody Testing Messages Targeted to African American Women. *Health Psychology* 14, 3: 247–54.

Kam, Ling, Vikki E. Knott, Carlene Wilson and Suzanne K. Chambers (2012). Using the Theory of Planned Behavior to Understand Health Professionals' Attitudes and Intentions to Refer Cancer Patients for Psychosocial Support. *Psycho-Oncology* 21: 316–23.

Kaplan, Lennart, Jana Kuhnt, Sebastian Vollmer and Katharina Richert (2018). What makes a successful development intervention? The theory of planned behavior – an application to implementation research. *Mimeo.*

Kaplan, Lennart, Katharina Richert and Sebastian Vollmer (2017). The Safe Childbirth Checklist - Evidence from a Randomized Controlled Trial in Indonesia. *Mimeo.*

Karim, Ali Mehryar, Timothy Williams, Leslie Patykewich, Disha Ali, Charlotte E. Colvin, Jessica Posner and Gideon Rutaremwa (2009). The Impact of the African Youth Alliance Program on the Sexual Behavior of Young People in Uganda. *Studies in Family Planning* 40(4): 289-306.

Kassim, Khairil Anwar, Lawrence Arokiasamy, Mohd Hafzi Md Isa and Chieng Heng Ping (2017). Intention to Purchase Safer Car: An Application of Theory of Planned Behavior. *Global Business & Management Research* 9, 1: 188–97.

Katherine M White, Nerina L Jimmieson3, Patricia L Obst, Nicholas Graves, Adrian Barnett, Wendell Cockshaw, Phillip Gee, Lara Haneman, Katie Page, Megan Campbell, Elizabeth Martin and David Paterson5 (2015). Using a Theory of Planned Behaviour Framework to Explore Hand Hygiene Beliefs at the '5 Critical Moments' among Australian Hospital-Based Nurses 59. *BMC Health Services Research* 15, 1: 1–9.

Kemper, Christoph J., Constanze Beierlein, Doreen Bensch, Anastassiya Kovaleva and Beatrics Rammstedt (2014). *Soziale Erwünschtheit-Gamma (KSE-G). Zusammenstellung sozialwissenschaftlicher Items und Skalen.* doi.org/10.6102/zis186

Kersting, Erasmus and Christopher Kilby (2016). With a Little Help from My Friends: Global Electioneering and World Bank Lending. *Journal of Development Economics* 121: 153-165.

Kersting, Erasmus and Christopher Kilby (2017). The Rise of Supplemental Lending at the World Bank. *Villanova School of Business Department of Economics and Statistics Working Paper* 30.

Kettle, Stewart, Mateo Hernandez, Simon Ruda and Michael Sanders (2016). Behavioral Interventions in Tax Compliance - Evidence from Guatemala. *World Bank Policy Research* June: 1–40.

Kiene, Susan M., Sarah Hopwood, Haruna Lule and Rhoda K. Wanyenze (2014). An Empirical Test of the Theory of Planned Behaviour Applied to Contraceptive Use in Rural Uganda. *Journal of Health Psychology* 19, 12: 1564–75.

Kilby Christopher and Axel Dreher (2010). The impact of aid on growth revisited: Do donor motives matter? *Economics Letters* 107, 3: 338-340.

Kilby, Christopher (2011). Informal Influence in the Asian Development Bank. *Review of International Organizations* 6, 3-4: 223-257.

Kilby, Christopher (2013). The Political Economy of Project Preparation: An Empirical Analysis of World Bank Projects. *Journal of Development Economics* 105: 211-225.

Kilby, Christopher (2015). Assessing the Impact of World Bank Preparation on Project Outcomes. *Journal of Development Economics* 115: 111-123.

Kilby, Christopher and Axel Dreher (2010). The impact of aid on growth revisited: Do donor motives matter? *Economics Letters* 107, 3: 338-340.

Kinsman, John, Jessica Nakiyingi-Miiro, A. Kamali and J Whitworth (2001). Evaluation of a comprehensive school-based AIDS education programme in rural Masaka, Uganda. *Health Education Research* 16: 85-100.

Kleemann, Linda, Peter Nunnenkamp and Rainer Thiele (2016). Gender Inequality, Female Leadership, and Aid Allocation: A Panel Analysis of Aid for Education. *Journal of International Development* 28, 3: 376-395.

Klöckner, Christian A. and Anke Blöbaum (2010). A comprehensive action determination model: Toward a broader understanding of ecological behaviour using the example of travel mode choice. *Journal of Environmental Psychology* 30, 4: 574-586.

Knack, Stephen, F. Halsey Rogers and Nicholas Eubank (2011). Aid Quality and Donor Rankings. *World Development* 39, 11: 1907-1917.

Koch, Michael T. and Sarah A. Fulton (2011). In the Defense of Women: Gender, Office Holding, and National Security Policy in Established Democracies. *Journal of Politics* 73, 1: 1-16.

Koremenos, Barbara (2013). What's Left Out and Why? Informal Provisions in Formal International Law. *Review of International Organizations* 8, 2: 137-162.

Kotsadam, Andreas, Gwudrun Østby, Siri A. Rustad, Andreas F. Tollefsen and Henrik Urdal (2018). Development aid and infant mortality. Micro-level evidence from Nigeria. *World Development* 105: 59-69.

Krupka, Erin L. and Roberto A. Weber (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association* 11, 3: 495–524.

Kuhnt, Jana and Sebastian Vollmer (2017). The Safe Childbirth Checklist: Can a Simple Tool Tackle the Maternal and Newborn Mortality and Morbidity Burden? Evidence from a Pair-Wise Matched Randomized Controlled Trial in Pakistan. *Mimeo.*

Kuziemko, Ilyana and Eric Werker (2006). How Much is a Seat on the Security Council Worth? Foreign Aid and Bribery at the United Nations. *Journal of Political Economy* 114, 5: 905-930.

Lai, Brian and Vanessa A. Lefler (2017). Examining the Role of Region and Elections on Representation in the UN Security Council. *Review of International Organizations* 12, 4: 585-611.

Landmann, Dirk, Carl J. Lagerkvist and Verena Otter (2017). *Determinants of Smallholder Farmers' Intention to Use Smartphones to Generate Agricultural Knowledge - Evidence from Rural India. Digitale Transformation - Wege in eine zukunftsfähige Landwirtschaft, Lecture Notes in Informatics (LNI)*, Bonn: Gesellschaft für Informatik.

Lang, Valentin F. and Andreas Presbitero (2017). The Political Economy of the World Bank–IMF Debt Sustainability Analysis. *Journal of Development Economics*. forthcoming.

Legovini, Arianna, Vincenzo Di Maro and Caio Piza (2015). Impact evaluation helps deliver development projects. *Policy Research Working Paper* 7157. World Bank Group.

Lichand, Guilherme and Anandi Mani (2016). Cognitive Droughts. *CAGE Online Working Paper Series*, Competitive Advantage in the Global Economy (CAGE).

Lott, John R. and Lawrence W. Kenny (1999). Did Women's Suffrage Change the Size and Scope of Government? *Journal of Political Economy* 107, 6: 1163-1198.

Lu, Kelan and Marijke Breuning (2014). Gender and Generosity: Does Women's Representation Affect Development Cooperation? *Politics, Groups, and Identities* 2, 3: 313-330.

Lugoe, Wycliffe and Jostein Rise (1999). Predicting Intended Condom Use among Tanzanian Students Using the Theory of Planned Behaviour. *Journal of Health Psychology* 4, 4: 497–506.

Lundsgaarde, Erik (2013). *The Domestic Politics of Foreign Aid*. London, UK: Routledge.

Lundsgaarde, Erik, Christian Breunig and Aseem Prakash (2007). Trade Versus Aid: Donor Generosity in an Era of Globalization. *Policy Sciences* 40, 2: 157-179.

Madrian, Brigitte C. (2014). Applying Insights from Behavioral Economics to Policy Design. *Annual Review of Economics* 6: 663-688.

Malan, Franck A. (2016). *Does Being an IMF Executive Board Member (Re-)Pay? An Examination of IMF Loans and Repayments*. Paper presented at the 65th Annual meeting of the French Economic Association.

Malik, Rabia and Randall W. Stone (2016). Corporate Influence in World Bank Lending. *Journal of Politics*, forthcoming.

Marchesi, Silvia and Tania Masi (2018). *What does influence the choice of the implementation level? Evidence from the World Bank.* Conference paper presented at the Political Economy of International Organizations conference, February 2018 in Madison, WI: USA.

Martens, Bertin, Uwe Mummert, Peter Murrell and Paul Seabright (2002). *The Institutional Economics of Foreign Aid.* Cambridge, UK: Cambridge University Press.

Mayer, Thierry and Soledad Zignago (2006). *Notes on CEPII's Distances Measures.* Available at: http://www.cepii.fr/anglaisgraph/bdd/distances.htm (accessed: May 2009).

McEachan, Rosemary Robin Charlotte, Mark Conner, Natalie Jayne Taylor and Rebecca Jane Lawton (2011). Prospective Prediction of Health-Related Behaviours with the Theory of Planned Behaviour: A Meta-Analysis. *Health Psychology Review* 5, 2: 97–144.

McGillivray, Fiona and Alastair Smith (2004). The Impact of Leadership Turnover on Trading Relations between States. *International Organization* 58, 3: 567-600.

McLean, Elena V. (2017). The Politics of Contract Allocation in the World Bank. *Review of International Organizations* 12, 2: 255-279.

Mikulaschek, Christoph (2017). Issue Linkage Across International Organizations: Does European Countries' Temporary Membership in the UN Security Council Increase Their Receipts from the EU Budget? *Review of International Organizations*. forthcoming.

Milner, Helen V., Daniel L. Nielson and Michael G. Findley (2016). Citizen Preferences and Public Goods: Comparing Preferences for Foreign Aid and Government Programs in Uganda. *Review of International Organizations* 11, 2: 219–45.

Minasyan, Anna and Peter Nunnenkamp (2016). Remittances and the effectiveness of foreign aid. *Review of Development Economics* 20, 3: 681-701.

Minasyan, Anna, Peter Nunnenkamp and Katharina Richert (2017). Does Aid Effectiveness Depend on the Quality of Donors? *World Development* 100: 16-30.

Minoiu, Camelia and Sanjay G. Reddy (2010). Development aid and economic growth: A positive long-run relation. Quarterly Review of Economics and Finance 50, 1: 27-39.

Moessinger, Marc-Daniel (2014). Do the Personal Characteristics of Finance Ministers Affect Changes in Public Debt? *Public Choice* 161, 1-2: 183-207.

Morrison, Kevin M. (2013). Membership No Longer has its Privileges: The Declining Informal Influence of Board Members on IDA lending. *Review of International Organizations* 8, 2: 291-312.

Moser, Christoph (2007). The Impact of Political Risk on Sovereign Bond Spreads - Evidence from Latin America. *Proceedings of the German Development Economics Conference 2007* 24, Verein für Socialpolitik, Research Committee Development Economics.

Moser, Christoph and Axel Dreher (2010). Do Markets Care about Central Bank Governor Changes? Evidence from Emerging Markets. *Journal of Money, Credit and Banking* 42, 8: 1589-1612.

Mosley, Paul (1985). The Political Economy of Foreign Aid: A Model of the Market for a Public Good. *Economic Development and Cultural Change* 33, 2: 373-393.

Nabunya, Proscovia, Fred M. Ssewamala and Vilma Ilicc (2014). Family economic strengthening and parenting stress among caregivers of AIDS-orphaned children: Results from a cluster randomized clinical trial in Uganda. *Children and youth services review* 44: 417-421.

Nayiga, Susan, Deborah DiLiberto, Lilian Taaka, Christine Nabirye, Ane Haaland, Sarah G. Staedke, Clare I.R. Chandler (2014). Strengthening patient-centred communication in rural Ugandan health centres: A theory-driven evaluation within a cluster randomized trial. *Evaluation (Lond)* 20(4): 471-491.

Neuenkirch, Matthias and Florian Neumeier (2015). Party Affiliation Rather than Former Occupation: The Background of Central Bank Governors and its Effects on Monetary Policy. *Applied Economics Letters* 22, 17: 1424-1429.

Neumayer, Eric (2003). *The pattern of aid giving: The impact of good governance on development assistance*. Routledge, London and New York.

Nickell, Stephen (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* 49, 6: 1417-1426.

Nizalova, Olena and Irina Murtazashvili (2014). Exogenous Treatment and Endogenous Factors: Vanishing of Omitted Variable Bias on the Interaction Term. *Journal of Econometric Methods* 5, 1: 71–77.

Novosad, Paul and Eric Werker (2018). Who Runs the International System? Power and the Staffing of the United Nations Secretariat. *Review of International Organizations*. forthcoming.

Nunn, Nathan and Nancy Qian (2014). U.S. Food Aid and Civil Conflict. *American Economic Review 104*, 6: 1630-1666.

Nunnenkamp, Peter and Hannes Öhler (2011). Throwing foreign aid at HIV/AIDS in developing countries: Missing the target? *World Development* 39, 10: 1704-1723.

Oatley, Thomas and Jason Yackee (2004). American Interests and IMF Lending. *International Politics* 41, 3: 415-29.

OECD (2008). *Is it ODA? Factsheet – November 2008.* Paris, France: Organisation for Economic Co-operation and Development. Available at http://www.oecd.org/investment/stats/34086975.pdf (accessed: 28 August 2014).

OECD (2008). *The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action*. Organisation for Economic Co-operation and Development. Paris.

OECD (2014). *Total Flows by Donor (ODA+OOF+Private) [DAC1]. Paris, France: Organisation for Economic Co-operation and Development.* Available at http://stats.oecd.org/ (accessed: 23 October 2014).

OECD (2017). *Aid (ODA) Disbursements to Countries and Regions [DAC2a].* Paris, France: Organisation for Economic Co-operation and Development. Available at http://stats.oecd.org/ (accessed: 08 February 2017).

OECD (2018). *Flows based on individual projects. Creditor Reporting System (CRS)-Data. Sectors monadic and dyadic.* Paris, France: Organisation for Economic Co-operation and Development. Available at https://stats.oecd.org/Index.aspx?DataSetCode=CRS1 (accessed: 22 March 2018).

Ogden, Jane (2003). Some Problems with Social Cognition Models: A Pragmatic and Conceptual Analysis. *Health Psychology* 22: 424–28.

Olsen-Telles, Erin (2013). Foreign Aid Voting in Congress: Does Gender Matter? *Paper presented at the 2013 Annual Conference of the Midwest Political Science Association*, Chicago, IL (April 11-14).

Parízek, Michal (2017). Control, Soft Information, and the Politics of International Organizations Staffing. *Review of International Organizations* 12, 4: 559-583.

Pasta, David J. (2009). Statistics and Data Analysis Learning When to Be Discrete: Continuous vs. Categorical Predictors.

Payne, John, Namika Sagara, Suzanne B. Shu, Kirstin Appelt and Eric Johnson (2013). Life Expectancy as a Constructed Belief: Evidence of a Live-to or Die-by Framing Effect. *Journal of Risk and Uncertainty* 46, 1: 27–50.

Perugini, Marco and Richard P. Bagozzi (2001). The Role of Desires and Anticipated Emotions in Goal-Directed Behaviours: Broadening and Deepening the Theory of Planned Behaviour. 40: 79–98.

Protogerou, Cleo, Alan J. Flisher, Leif Edvard Aar and Catherine Mathews (2012). The Theory of Planned Behaviour as a Framework for Predicting Sexual Risk Behaviour in Sub-Saharan African Youth: A Critical Review. *Journal of Child and Adolescent Mental Health* 24, 1: 15–35.

Rajan, Raghuram G. and Arvind Subramanian (2008). Aid and growth: What does the cross-country evidence really show? *Review of Economics and Statistics* 90, 4: 643-665.

Ravallion, M. (2012). Fighting poverty one experiment at a time: A review of Abhijit Banerjee and Esther Duflo's poor economics: A radical rethinking of the way to fight global poverty. *Journal of Economic Literature*, 50, 1: 103-114.

Ricardo Fuentes, Tobias Pfuetze, Papa Seck (2006). Does access to water and sanitation affect child survival? A five country analysis. No. HDOCPA-2006-04. *Human Development Report Office (HDRO)*, United Nations Development Programme (UNDP).

Roodman, David (2012). An index of donor performance. *Working Paper* 67 (October 2012 edition). Center for Global Development. Washington, DC.

Roodman, David (2015). A Replication of "Counting chicken when they hatch" (Economic Journal 2012). *Public Finance Review* 43: 282-286.

Rose, Andrew K (2007). The foreign service and foreign trade: Embassies as export promotion. *The World Economy* 30, 1: 22-38.

Round, Jeffery I. and Matthew Odedokun (2004). Aid Effort and Its Determinants. *International Review of Economics & Finance* 13, 3: 293-309.

Schneider, Tamera R., Peter Salovey, Anne Marie Apanovitch, Judith Pizarro, Danielle McCarthy, Janet Zullo and Alexander J. Rothman (2001). The Effects of Message Framing and Ethnic Targeting on Mammography Use Among Low-Income Women. *Health Psychology* 20, 4: 256–66.

Shah, Neil B., Paul Wang, Andrew Faker and Daniel Gastfriend (2015). Evaluations with impact: Decision-focused impact evaluation as a practical policymaking tool. *3ie Working Paper* 25. New Delhi: International Initiative for Impact Evaluation.

Simon, Herbert A. and Julian Feldman (1959). Theories of Decision-Making in Economics and Behavioral Science. Source: *The American Economic Review* 49, 3: 253–83.

Sligo, Frank X. and Anna M. Jameson (2000). The Knowledge — Behavior Gap in Use of Health Information. *Journal of the American Society for Information Science* 51, 9: 858–869.

Solnick, Sara J. (2001). Gender Differences in the Ultimatum Game. *Economic Inquiry* 39, 2: 189-200.

Spector, Jonathan M., Priya Agrawal, Bhala Kodkany, Stuart Lipsitz, Angela Lashoher, Gerald Dziekan Rajiv Bah, Mario Merialdi, Matthews Mathai, Claire Lemer and Atul Gawande (2012). Improving Quality of Care for Maternal and Newborn Health: Prospective Pilot Study of the Who Safe Childbirth Checklist Program. *PLoS ONE* 7, 5.

Spilimbergo, Antonio (2009). Democracy and Foreign Education. *American Economic Review* 99, 1: 528-543.

Staiger, Douglas and James H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 3: 557-586.

Steinwand, Martin C. (2015). Compete or Coordinate? Aid Fragmentation and Lead Donorship. *International Organization* 69, 2: 443-472.

Stone, Randall W. (2011). *Controlling Institutions: International Organizations and the Global Economy.* Cambridge University Press.

Stone, Randall W. (2013). Informal Governance in International Organizations: Introduction to the Special Issue. *Review of International Organizations* 8, 2: 121.

Svensson, Jakob (1999). Aid, Growth and Democracy. *Economics & Politics* 11, 3: 275–297.

Szent-Iványi, Balázs and Simon Lightfoot (2015). *New Europe's New Development Aid.* Abingdon, UK and New York, NY: Routledge.

Te Velde, Dirk Willem and Michael Warner (2007). *The Use of Subsidies by Development Finance Institutions in the Infrastructure Sector.* Overseas Development Institute, London, UK.

Thacker, Strom C. (1999). The High Politics of IMF lending. *World Politics* 52, 1: 38-75.

The World Bank (2009). Making Smart Policy: Using impact evaluation for policy making. Case studies on evaluations that influenced policy. *Doing Impact Evaluation Series* No. 14. Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation.

Thérien, Jean-Philippe and Alain Noël (2000). Political Parties and Foreign Aid. *American Political Science Review* 94, 1: 151-162.

Thorvaldsdottir, Svanhildur (2015). *Behave or Else? Aid and Voting on the UN Security Council.* unpublished manuscript.

Tingley, Dustin (2010). Donors and Domestic Politics: Political Influences on Foreign Aid Effort. *Quarterly Review of Economics and Finance* 50, 2: 40-49.

Togeby, Lisa (1994). The Gender Gap in Foreign Policy Attitudes. *Journal of Peace Research* 31, 4: 375-392.

Trautmann, Stefan T. and Gijs van de Kuilen (2015). Belief Elicitation: A Horse Race among Truth Serums. *Economic Journal* 125, 589: 2116–35.

Tversky, Amos and Daniel Kahneman (1981). The Framing of Decisions and the Psychology of Choice. *Science* 211, 4481: 453–58.

UN (2018). *Child Mortality Estimates.* Available at http://www.childmortality.org/ (accessed: 05 August 2018)

UNDP (2005). *Investing in development: A practical plan to achieve the Millennium Development Goals. Overview.* New York: UN Millennium Project.

UNICEF (2014). *Office of Research - Methodological Briefs Theory of Change.* http://socialinnovation.ca/about/theory-of-change.

Vaubel, Roland, Axel Dreher and Uğurlu Soylu (2007). Staff Growth in International Organizations: A Principal-Agent Problem? An Empirical Analysis. *Public Choice* 133, 3-4: 275-295.

Viceisza, Angelino C. G. (2016). Creating a Lab in the Field: Economics Experiments for Policymaking. *Journal of Economic Surveys* 30, 5: 835–54.

Vivalt Coville (2017). How do policymakers update. *Work in progress:* http://evavivalt.com/wp-content/uploads/How-Do-Policymakers-update.pdf (draft presented at CSAE 2018).

Vivalt, Eva (2015). Heterogeneous treatment effects in impact evaluation. *American Economic Review* 105, 5: 467-470.

Vogel, Tobias and Michaela Wanke (2016). *Attitudes and Attitude Change.* Second Edi. London and New York: Routledge.

Volkens, Andrea, Pola Lehmann, Theres Matthieß, Nicolas Merz and Sven Regel (2016). *The Manifesto Data Collection.* Manifesto Project (MRG/CMP/MARPOR). Version 2016b. Berlin, Germany: Wissenschaftszentrum Berlin für Sozialforschung.

Vreeland, James Raymond (2011). Foreign Aid and Global Governance: Buying Bretton Woods – The Swiss-bloc case. *Review of International Organizations* 6, 3-4: 369-391.

Vreeland, James Raymond and Axel Dreher (2014). *The Political Economy of the United Nations Security Council. Money and Influence.* Cambridge University Press.

Walrave, Michel, Wannes Heirman and Lara Hallam (2014). Under Pressure to Sext? Applying the Theory of Planned Behaviour to Adolescent Sexting. *Behaviour & Information Technology* 33, 1: 85–97.

WHO (2015). *Statistical profile for Uganda.* Available at http://www.who.int/gho/countries/uga.pdf?ua=1 (accessed: 05 August 2018).

WHO (2017). *Newborns: Reducing Mortality.* Fact Sheet.

WHO (2018). *Health statistics and information systems. Global health estimates. Metrics: Disability-Adjusted Life Year (DALY).* http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/ (accessed: 30 July 2018).

Wildschut, Tim, Martin Bruder, Sara Robertson, Wijnand A.P. van Tilburg, Constantine Sedikides (2014). Collective Nostalgia: A Group-Level Emotion That Confers Unique Benefits on the Group. *Journal of Personality and Social Psychology* 107, 5: 844–63.

Wilson, Dawn K., Robert M. Kaplan and Lawrence J. Schneiderman (1987). Framing of Decisions and Selections of Alternatives in Health Care. *Social Behaviour* 2, 1: 51–59.

Winters, Matthew S., Simone Dietrich and Minhaj Mahmud (2017). Perceptions of Foreign Aid Project Quality in Bangladesh. *Research and Politics*: 1–6.

Woldendorp, Jaap, Hans Keman and Ian Budge (2000). *Party Government in 48 Democracies*. Composition – Duration – Personnel. Dordrecht, Netherlands etc.: Kluwer Academic Publishers.

Wood, Bernard, Julia Betts, Florence Etta, Julian Gayfer, Dorte Kabell, Naomi Ngwira, Francisco Sagasti and Mallika Samaranayake (2011). *The Evaluation of the Paris Declaration*, Final Report, Copenhagen.

Workman, Michael, Michael F. Lesser and Joonmin Kim (2007). An Exploratory Study of Cognitive Load in Diagnosing Patient Conditions. *International Journal for Quality in Health Care* 19, 3: 127–33.

World Bank (1998). *Assessing aid: What works, what doesn't, and why?* Oxford: Oxford University Press.

World Bank (2006). *Budget Support as More Effective Aid? Recent Experiences and Emerging Lessons.* Washington, DC: The World Bank.

World Bank (2014). *World Development Indicators.* Washington, DC: The World Bank. Available at http://www.data.worldback.org/indicator (accessed: April 2014).

World Bank (2015). *World Development Report 2015 - Mind, Society, and Behavior.* Washington, DC: World Bank. http://www.worldbank.org/en/publication/wdr2015.

World Bank (2016). *World Development Indicators.* Washington, DC: The World Bank. Available at http://www.data.worldback.org/indicator (accessed: 06 April 2016).

World Development Indicators (2015). World Bank. Washington, DC.

World Health Organization (2015). *WHO Safe Childbirth Checklist Implementation Guide.* http://apps.who.int/iris/bitstream/10665/199177/1/9789241549455_eng.pdf?ua=1&ua=1

Younas, Javed (2008). Motivation for Bilateral Aid Allocation: Altruism or Trade Benefits. *European Journal of Political Economy* 24, 3: 661-674.

Young, Alwyn (2016). *Channeling Fisher : Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results.* London School of Economics, Draft.