The World Bank Research Observer is intended for anyone who has a professional interest in development. Observer articles are written to be accessible to nonspecialist readers; contributors examine key issues in development economics, survey the literature and the latest World Bank research, and debate issues of development policy. Articles are reviewed by an editorial board drawn from across the Bank and the international community of economists. Inconsistency with Bank policy is not grounds for rejection.

The journal welcomes editorial comments and responses, which will be considered for publication to the extent that space permits. On occasion the Observer considers unsolicited contributions. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages to the Editor. Please direct all editorial correspondence to the Editor, The World Bank Research Observer, 1818 H Street, NW, Washington, DC 20433, USA.

The views and interpretations expressed in this journal are those of the authors and do not necessarily represent the views and policies of the World Bank or of its Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of data included in this publication and accepts no responsibility whatsoever for any consequences of their use. When maps are used, the boundaries, denominations, and other information do not imply on the part of the World Bank Group any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

For more information, please visit the Web sites of the Research Observer at academic.oup.com/wbro, the World Bank at www.worldbank.org, and Oxford University Press at academic.oup.com.
Subscriptions
A subscription to The World Bank Research Observer (ISSN 0257-3032) comprises 2 issues. Prices include postage; for subscribers outside the Americas, issues are sent air freight.

Annual Subscription Rate (Volume 32, 2 issues, 2017)

Academic libraries
Print edition and site-wide online access: US$250/£167/€250
Print edition only: US$230/£152/€230
Site-wide online access only: US$184/£123/€184

Corporate
Print edition and site-wide online access: US$379/£252/€379
Print edition only: US$347/£231/€347
Site-wide online access only: US$280/£184/€280

Personal
Print edition and individual online access: US$69/£47/€69

Please note: US$ rate applies to US & Canada, Euros applies to Europe, UK£ applies to UK and Rest of World. Readers with mailing addresses in non-OECD countries and in socialist economies in transition are eligible to receive complimentary subscriptions on request by writing to the UK address below.

There may be other subscription rates available; for a complete listing, please visit https://academic.oup.com/wbro/subscribe.

Full pre-payment in the correct currency is required for all orders. Payment should be in US dollars for orders being delivered to the USA or Canada; Euros for orders being delivered within Europe (excluding the UK); GBP sterling for orders being delivered elsewhere (i.e., not being delivered to USA, Canada, or Europe). All orders should be accompanied by full payment and sent to your nearest Oxford Journals office. Subscriptions are accepted for complete volumes only. Orders are regarded as firm, and payments are not refundable. Our prices include Standard Air as postage outside of the UK. Claims must be notified within four months of despatch/order date (whichever is later). Subscriptions in the EEC may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Subscribers in Canada, please add GST to the prices quoted. Personal rate subscriptions are only available if payment is made by personal cheque or credit card, delivery is to a private address, and is for personal use only.

Back issues: The current year and two previous years’ issues are available from Oxford University Press. Previous volumes can be obtained from the Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. E-mail: psc@periodicals.com. Tel: (518) 537-4700. Fax: (518) 537-5899.

Contact information: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnlscust.serv@oup.com. Tel: +44 (0)1865 351907. Fax: +44 (0)1865 353485.

In the Americas, please contact: Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. E-mail: jnlorders@oup.com. Tel: (800) 852-7323 (toll-free in USA/Canada) or (919) 677-0977. Fax: (919) 677-1714.

In Japan, please contact: Journals Customer Service Department, Oxford University Press, 4-5-10-8F Shibya, Minato-ku, Tokyo, 108-8186, Japan. E-mail: custserv.jp@oup.com. Tel: +81 3 5444 5858. Fax: +81 3 3454 2929.


Environmental and ethical policies: Oxford Journals, a division of Oxford University Press, is committed to working with the global community to bring the highest quality research to the widest possible audience. Oxford Journals will protect the environment by implementing environmentally friendly policies and practices wherever possible. Please see academic.oup.com/journals/pages/authors.ethics for further information on environmental and ethical policies.

Digital Object Identifiers: For information on dois and to resolve them, please visit www.doi.org.

Permissions: For information on how to request permissions to reproduce articles or information from this journal, please visit academic.oup.com/journals/pages/authors.

Advertising: Advertising, inserts, and artwork enquiries should be addressed to Advertising and Special Sales, Oxford Journals, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. Tel: +44 (0)1865 354767; Fax: +44 (0)1865 353774; E-mail: julsadvertising@oup.com.

Disclaimer: Statements of fact and opinion in the articles in The World Bank Research Observer are those of the respective authors and contributors and not of the International Bank for Reconstruction and Development/ THE WORLD BANK or Oxford University Press. Neither Oxford University Press nor the International Bank for Reconstruction and Development/ THE WORLD BANK make any representation, express or implied, in respect of the accuracy of the material in this journal and cannot accept any legal responsibility or liability for any errors or omissions that may be made. The reader should make her or his own evaluation as to the appropriateness or otherwise of any experimental technique described.


Indexing and abstracting: The World Bank Research Observer is indexed and/or abstracted by ABI/INFORM, CAR Abstracts, Current Contents/Social and Behavioral Sciences, Journal of Economic Literature/EconLit, PAIS International, RePEc (Research in Economic Papers), Social Services Citation Index, and Wilson Business Abstracts.

Copyright © 2017 The International Bank for Reconstruction and Development/THE WORLD BANK. All rights reserved: no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission of the publisher or a license permitting restricted copying issued by the UK by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1P 9HE, or in the USA by the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. Typeset by Cenvoo Publisher Services, Bangalore, India; Printed by SCI, Secaucus, NJ.
How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence
David McKenzie

Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs
Abhijit V. Banerjee, Rema Hanna, Gabriel E. Kreindler, and Benjamin A. Olken

The Impacts of Fiscal Openness
Paolo de Renzio and Joachim Wehner

The Opportunities and Challenges of Digitizing Government-to-Person Payments
Leora Klapper and Dorothe Singer
How Effective Are Active Labor Market Policies in Developing Countries?  
A Critical Review of Recent Evidence

David McKenzie

In a well-functioning labor market, firms that want workers and workers who want jobs are able to find one another reasonably easily, and the only unemployment is low, frictional, and temporary. In such a world, the main area for government policy is passive policy, in which the government undertakes investments in infrastructure and provides the regulatory framework needed for the economy as a whole to grow and raise incomes, but does not intervene directly to help particular workers find jobs, or particular firms to find workers.

However, in practice, governments have long engaged in a variety of active labor market policies (ALMPs) that directly intervene in the labor market with the
aim of generating more and better employment opportunities for workers. Examples of such policies include training programs that aim to increase the skills of the labor supply, wage subsidies that aim to increase firms’ demand for labor, and job search and matching assistance that aims to better enable firms and workers to find and contract with one another.

Four recent global shocks and trends have increased the importance of jobs as a policy concern, and renewed interested in the effectiveness of ALMPs. First was the global financial crisis of 2007–2008, which increased unemployment in many countries worldwide. Second, rising demographic pressures in some parts of the developing world have led to headlines of a “jobs time bomb” with claims including that India needs to create 12 million new jobs annually (Kumar and Busvine 2014), 10 to 12 million young people are entering the job market annually in Africa (Mohammed 2015), and the Middle East and North Africa (MENA) region needs to create 100 million new jobs by 2020 (World Bank 2004). Third, high rates of youth unemployment, particularly in the MENA region, have raised fears of social unrest and large emigration flows (Kelly 2016). Finally, enormous progress in automation may mean that manufacturing jobs, which were vital to the growth of East Asian countries, may no longer be available for poorer countries as they develop. This has led to headlines like “Robots could eat all of Ethiopia’s jobs; South Africa, Nigeria and Angola not safe either” (Mwiti 2016).

While policymakers have long employed ALMPs and interest in their effectiveness has increased, until recently most of the evidence regarding their effectiveness came from developed and transition countries, and very little from experimental evidence. For example, Dar and Tzannatos (1999) cover 72 evaluations but only find Hungary and Poland among non-developed countries, and have evaluations for their programs based on matching participants with non-participants. The heavily cited update of this work by Betcherman, Olivas, and Dar (2004) added 39 additional evaluations from developing and transition countries, of which only four drew on randomized experiments, and of which only one (Galasso, Ravallion, and Salvia 2004) was published in an academic journal. The typical evaluation during this period used propensity-score matching to compare participants in an ALMP to non-participants using a relatively small number of cross-sectional observed characteristics to compare the two groups. There is continued debate about the extent to which matching can provide reliable estimates of program impacts, but estimates are likely to be more reliable when the selection process into programs is known and multiple periods of pre-program data are available for both treatment and control (Smith and Todd 2005), conditions that few of these non-experimental evaluations satisfy.

The last decade has seen growth in the number of experimental evaluations of ALMPs in developing countries. These new studies provide more rigorous evidence for the impacts of these programs, but still suffer from some of the same problems
faced by many non-experimental studies such as survey attrition, the difficulty of accounting for general equilibrium effects, and concerns with the right timing over which to measure impacts. I critically survey this recent literature and draw out lessons for the effectiveness of ALMPs.\textsuperscript{1} The general message is that traditional ALMPs that focus on skill training, wage subsidies, and job search assistance have at best modest impacts in most circumstances. I compare this to expectations of program impacts from participants and policymakers, and show that both groups tend to have over-optimistic expectations of how beneficial these programs can be. However, revealed preference also shows that many of the formal sector manufacturing jobs that these programs are intended to foster are not that highly valued by workers. I then turn to emerging evidence on the effectiveness of less traditional active labor market policy actions. I note the promise of policies that attempt to deal with sectoral and spatial mismatches in which workers are stuck in occupations or locations that differ from where demand is. Finally, I attempt to draw out lessons for new impact evaluations in this area, as well as some concluding lessons for policymakers.

The Rationale and Evidence for Traditional ALMPs as a Response

Traditional ALMPs are divided into three main categories. The first set of programs operate on the labor supply side, and aim to increase the employability of workers through vocational training (see McKenzie and Woodruff (2014) for a separate review of programs that foster self-employment through business training). A second set of programs aims to increase the demand for labor, through subsidizing the cost of labor to firms with employment subsidies.\textsuperscript{2} Finally, search and matching assistance programs aim to lower frictions that prevent demand from meeting supply in the labor market. I discuss the key economic rationale for each type of program, and the recent empirical evidence for each.

This recent evidence largely comes from randomized experiments that randomly assign a group of job-seekers or firms to a treatment group, who get offered the intervention, and a control group of similar individuals who do not. This approach has advantages to older, non-experimental evaluation methods like propensity-score matching that compared those who took up a program to those who did not, since we might worry that program participants may differ in a host of unobservable ways such as desire for work and search effort, alternative employment options, and talent, from those who do not take part.

Comparing employment outcomes for the treatment and control groups in a randomized experiment then gives an unbiased estimate of the impact of being offered the program under the stable unit treatment value assumption (SUTVA),
which requires that the outcomes for one individual are not affected by another individual’s treatment assignment. This is problematic if the treatment and control group both consist of job-seekers competing for the same set of jobs, in which case the ability of individuals in the control group to find jobs might be affected by the training or other support given to the treatment group. Most of the studies covered here deal with this issue by working with samples that are small relative to the overall labor market, so that direct job competition between the treatment and control groups may be negligible. Then the randomized experiment will yield an unbiased estimate of the private returns to individuals of participating in such programs. However, any gains to the treatment group may still come from displacing other job-seekers (not in the study), so that the public returns to such an intervention may be lower. This issue of general equilibrium effects needs to be considered when thinking of scaling up these programs, and I discuss this further when drawing policy conclusions from these studies. However, as a starting point, if programs do not pass a cost-benefit test based on private returns, then they are unlikely to be candidates for scale-up.

**Vocational Training**

Vocational training programs were the most common ALMP used by governments following the global financial crisis of 2007-08 (McKenzie and Robalino 2010). Blattman and Ralston (2015) note that the World Bank and its client governments invested nearly U.S. $1 billion per year between 2002 and 2012 on skills training programs. The premise of such programs is that a lack of certain technical skills is the reason that particular individuals are unemployed, and that these skills can be taught and learned in a relatively short period of time.

In practice, these programs are typically used with two target groups of beneficiaries. The first offers the program to the general population of unemployed workers. Although this is a common policy option, the only evaluation of such a program in a developing country setting is Hirshleifer et al. (2016), who conduct a randomized experiment to evaluate Turkey’s program for the unemployed. Typical programs here last a duration of three months, and cover a wide range of occupations.

The second approach is to more narrowly focus on low-income, or “at-risk” youth, where “youth” can include those aged 15 to 24, or even to 29, depending on the country. Programs focusing on youth have been particularly common in Latin America, and act as a substitute to the formal schooling system for youth who have dropped out. The standard model in the entra21, Jovenes, and Juventud y Empleo programs in Latin America has been to combine three months of classroom
training with 2–3 months of on-the-job training in the form of an internship. Some programs additionally provide life skills training.

There has been rapid growth in the number of randomized experiments evaluating these programs. I focus on evaluations of traditional vocational training programs, but there have also been several recent evaluations of bundled packages for adolescent girls that incorporate vocational training with other services such as business skills training, empowerment activities, and support in setting up businesses or finding jobs. An example is Adoho et al. (2014), who report early findings from Liberia. These authors find positive impacts of that program on employment, but estimate that it would take 12 years for participants to recoup the costs of the job skills training provided in that program.

Table 1 summarizes the results from 12 evaluations from eight countries. The typical evaluation measures impacts 12 to 18 months after the conclusion of the training program by using surveys administered to the treatment and control groups. The use of surveys to measure key employment outcomes raises several concerns. The most major concern is that of attrition, with all but one study having attrition rates of 18 percent or higher, ranging up to 46 percent in Cho et al. (2013). This attrition is a problem because we might expect the employment outcomes of individuals who refuse to be surveyed or who cannot be found to differ from those who are interviewed. A typical approach has been to compare attrition rates in the treatment and control groups, and then do a bounding exercise if the attrition rates vary (often the control group is slightly less likely to respond). But it is easy to think of problems that can arise even when the attrition rates are the same for both groups: for example, the attritors in the treatment group may be people who went through the training and did not find it useful and have still not found jobs, while those in the control group could be those who are too busy to answer surveys because they are employed in good jobs. This type of differential response would bias the estimated treatment effect upwards, overstating the impact of training.

A second issue with the use of survey measures of employment is the possibility that those in the treatment groups over-report their employment outcomes to express their appreciation for being given the program, while those in the control group potentially under-report these outcomes because they maintain some hope of still being given the program. Good survey design and survey framing can mitigate these issues. An alternative approach is then to use administrative data on employment from national social security or labor databases. These databases capture formally-registered employment and enable the trajectory of formal employment outcomes to be measured over longer time windows—including up to ten years after treatment in the case of Attanasio, Medina, and Meghir (2017). These databases are not subject to attrition, but because they only capture formal
Table 1: Summary of Vocational Training Program Impacts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>Hirshleifer et al. (2016)</td>
<td>Unemployed</td>
<td>5,902</td>
<td>6%</td>
<td>1 year</td>
<td>2.0</td>
<td>2.0 [-0.5, 4.4]</td>
<td>5.8 [-2.3, 13.8]</td>
<td>8.6 [-0.5, 17.7]</td>
<td>US$11.5</td>
<td>US$1700</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unemployed</td>
<td>0%</td>
<td>2.5 years</td>
<td>n.r</td>
<td>-0.1</td>
<td>n.r [-3.3, 1.5]</td>
<td>n.r [-7.9, 6.3]</td>
<td>-0.8 [-3.4, 4.4]</td>
<td>US$n[0.5, 4.4]$</td>
<td>US$n[7.9, 6.3]$</td>
</tr>
<tr>
<td>Argentina</td>
<td>Alzúa et al. (2016)</td>
<td>Low-income</td>
<td>407</td>
<td>0%</td>
<td>18 months</td>
<td>n.r.</td>
<td>8.0 [0.7, 15.3]</td>
<td>n.r. [17.1, 112.7]</td>
<td>64.9 [2.3, 13.8]</td>
<td>US$83 [0.5, 4.4]</td>
<td>US$1722</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low-income</td>
<td>0%</td>
<td>33 months</td>
<td>n.r.</td>
<td>4.3</td>
<td>n.r. [2.3, 6.7]</td>
<td>n.r. [10.0, 6.3]</td>
<td>23.1 [1.0, 8.0]</td>
<td>US$17.7</td>
<td>US$45</td>
</tr>
<tr>
<td>Colombia</td>
<td>Attanasio et al. (2011)</td>
<td>Low-income</td>
<td>4,350</td>
<td>18.5%</td>
<td>14 months</td>
<td>4.5</td>
<td>6.4 [1.0, 8.0]</td>
<td>11.6 [3.2, 9.6]</td>
<td>[4.5, 18.7] [12.8, 41.3]</td>
<td>US$12.8</td>
<td>US$750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Attanasio et al. (2017)</td>
<td>Low-income</td>
<td>0%</td>
<td>up to 10 years</td>
<td>n.r.</td>
<td>4.2</td>
<td>n.r. [1.0, 8.0]</td>
<td>n.r. [3.2, 9.6]</td>
<td>13.6 [4.5, 18.7]</td>
<td>US$17.7</td>
<td>US$45</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>Card et al. (2011)</td>
<td>Low-income</td>
<td>1,556</td>
<td>38%</td>
<td>12 months</td>
<td>0.7</td>
<td>2.2 [-4.6, 6.0]</td>
<td>10.8 [-2.3, 6.7]</td>
<td>[-4.2, 25.7] [-15.3, 61.5]</td>
<td>US$10 [0.5, 4.4]</td>
<td>US$330</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ibarrárán et al. (2014)</td>
<td>Low-income</td>
<td>5,000</td>
<td>20%</td>
<td>18 to 24 months</td>
<td>-1.3</td>
<td>1.8 [-4.8, 2.2]</td>
<td>6.5 [-0.3, 3.9]</td>
<td>[-4.8, 17.9] [-15.3, 61.5]</td>
<td>US$8.5 [0.5, 4.4]</td>
<td>US$700</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ibarrárán et al. (2015)</td>
<td>Low-income</td>
<td>5,000</td>
<td>34%</td>
<td>6 years</td>
<td>-1.4</td>
<td>2.6 [-4.4, 1.6]</td>
<td>1.9 [-0.5, 5.5]</td>
<td>[-10.0, 6.3] [-15.3, 61.5]</td>
<td>US$n[0.5, 4.4]$</td>
<td>US$n[7.9, 6.3]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acevedo et al. (2017)</td>
<td>Low-income</td>
<td>2,779</td>
<td>17.6%</td>
<td>3 years</td>
<td>0.7</td>
<td>n.r. [-4.0, 5.3]</td>
<td>n.r. [5.6, 186.0]</td>
<td>n.r. [1.0, 8.0]</td>
<td>US$n[0.5, 4.4]$</td>
<td>US$n[7.9, 6.3]$</td>
</tr>
<tr>
<td>India</td>
<td>Maitra and Mani (2017)</td>
<td>Low income</td>
<td>658</td>
<td>25%</td>
<td>18 months</td>
<td>8.1</td>
<td>n.r. [2.2, 14.0]</td>
<td>95.7 [5.6, 186.0]</td>
<td>n.r. [1.0, 8.0]</td>
<td>US$7.2 [0.5, 4.4]</td>
<td>US$39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Honorati (2015)</td>
<td>Low-income</td>
<td>2,100</td>
<td>23%</td>
<td>14 months</td>
<td>5.6</td>
<td>n.r. [0.9, 10.3]</td>
<td>29.7 [2.2, 14.0]</td>
<td>n.r. [1.0, 8.0]</td>
<td>US$47.5</td>
<td>US$1150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Malawi</td>
<td>Cho et al. (2013)</td>
<td>Low-income Youth</td>
<td>1,900</td>
<td>46%</td>
<td>4 months</td>
<td>n.r.</td>
<td>n.r.</td>
<td>−19.6</td>
<td>n.r.</td>
<td>−US$5</td>
<td>n.r.</td>
</tr>
<tr>
<td>Peru</td>
<td>Diaz and Rosas (2016)</td>
<td>Low-income Youth</td>
<td>4,509</td>
<td>35%</td>
<td>36 months</td>
<td>1.6</td>
<td>3.8</td>
<td>13.4</td>
<td>n.r.</td>
<td>n.r.</td>
<td>US$420</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7,151</td>
<td>0%</td>
<td>36 months</td>
<td>n.r</td>
<td>4.5</td>
<td>n.r</td>
<td>n.r</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
- Timeframe refers to time since the end of the intervention before measuring follow-up outcomes.
- n.r. denotes not recorded. Estimates are the Intention-to-Treat estimates reported in different studies.
- 95 percent confidence intervals shown in parentheses.
- (a) no impact on unconditional earnings reported. A negative and statistically significant impact on earnings conditional on working is reported.
- Impacts on employment are in terms of percentage points, impacts on earnings in terms of percentage growth relative to control mean.
- When study reports results for subgroups only, a weighted average is used to present the overall effect.
- Source: Author’s calculations based on studies cited in table.
employment, they may overestimate the employment impacts of the program if individuals simply shift from informal to formal employment.

With these caveats in mind, table 1 then provides an overview of the evidence from these recent studies. I consider two key outcomes: paid employment and earnings. I report here the intention-to-treat (ITT) effects estimated in the different studies. These effects show the impact of offering vocational training to target participants. Even though most programs require individuals to express interest and sign up for the training, not all of those selected for training complete it. In most of the programs here, between 70–85 percent of those selected for training complete it. The local-average-treatment effect, which is the effect of taking up training when it is offered, can then be obtained by multiplying these ITT impacts by between 1.2–1.4 in most cases.

We see that only three out of nine studies find a significant impact on employment. The simple unweighted average across the studies is a 2.3 percentage point increase in employment. That is, for every 100 people offered vocational training, fewer than three will find a job they would not have otherwise found. The last column of table 1 shows that the cost of these programs typically ranges from $500–to $1,700 per person trained (the exception being the tailoring and stitching training in India studied by Maitra and Mani (2017), which costs a remarkably low $39 per person trained). The result is an approximately $17,000–$60,000 cost per additional person employed.

A number of the studies have also considered formal employment as an outcome. This is of interest in its own right because of a belief that formal employment may offer additional benefits and stability to workers, as well as being a measure that can be obtained from administrative data. Studies that have measured both employment and formal employment have tended to find slightly larger impacts on formal employment, indicating that training helps shift workers towards more formal jobs. The average impact across the studies is 3.6 percentage points.

I consider the impact on earnings in terms of two measures. The first is the percentage increase in earnings relative to the earnings levels of the control group. The second is the absolute level increase in earnings relative to the control group in terms of U.S. dollars. Note that these comparisons are based on earnings alone, and do not include the value of benefits such as paid sick leave or pensions that may come from formal work, nor typically the costs of contributions to those programs. Only two out of nine studies find a statistically significant impact on earnings. However, all but two show positive point estimates, with a mean of a 17 percent increase and a median of 11 percent. The absolute change in monthly income ranges from -$5 to $83 per month, with a mean of $19.

Taken together, these studies show the potential of vocational training to have some impact on employment, but also that these impacts are modest in many cases. In order to get a sense of how to view the size of these effects, I find two
perspectives useful. The first is to consider vocational training as a substitute for schooling in building human capital. Standard estimates of the return to an additional year of schooling around the world show an average return of 10 percent, with returns to tertiary schooling averaging 21 percent in sub-Saharan Africa (Montenegro and Patrinos 2014). From this perspective, we might expect a three-month course to result in 3–7 percent higher earnings, and six months to result in 5–10 percent higher earnings. The earnings impacts in table 1 are largely within this order of magnitude, and are consistent with there being a return to human capital, but that vocational training should not be expected to deliver much different returns from schooling itself. An exception is Maitra and Mani (2017), where the increase in income represents a 95.7 percent increase on the control group’s mean. This reflects a situation where the women in their sample are unlikely to be working and have very low earnings, so this large relative increase is a small absolute increase of only $7.20 per month.

The second, more standard, perspective is that of cost-benefit. Comparing the cost of providing these programs to the monthly income gains shows that the cost of these programs averages 50 times the monthly income gain. Even adjusting for incomplete take-up (which means not having to pay the full costs for people who drop out), it will typically take three or four years at least for participants to recoup the cost of the program in income gains. This calculation also excludes the opportunity cost of income lost by the participants during the period they are trained. The result is that cost-benefit calculations for these programs rely on making assumptions of the trajectories of impact lasting for periods beyond which impacts have typically been measured. Some studies that have measured impacts over multiple time periods beyond a year after training (Alzúa, Cruces, and Lopez 2016; Hirshleifer et al. 2016; Acevedo et al. 2017) have tended to find impacts fall over time, making the assumption that short-term gains will necessarily persist, although others have found sustained impacts on formal employment for certain subgroups (Ibarrarán et al. 2015; Attanasio, Medina, and Meghir 2017).4 Further adding the need to discount the future at some rate, it is easy to arrive at the conclusion of Blattman and Ralston (2015), who state that “it is hard to find a skills training program that passes a simple cost-benefit test”.

In search of a more positive role for vocational training, researchers have pursued two approaches. The first is to find training programs that can be provided much more cheaply, such as the NGO program of Maitra and Mani (2017). If skills training can be delivered much cheaper, it does not need to deliver as large an income gain to be cost effective.

The second approach has been to investigate whether the returns to training might be different for some subgroups of the population or training types, to argue that targeted training might work. Foremost among this has been a focus on gender, and there appears to be a stylized fact in the literature that vocational training
has higher returns for women (e.g., Blattman and Ralston 2015). This appears to stem largely from the work in Colombia by Attanasio, Kugler, and Meghir (2011) and Attanasio, Medina, and Meghir (2017), who find significant impacts on employment for women but not for men. However, these authors never formally test for a difference in impact by gender, and indeed in their long-term follow-up, note that the magnitudes are similar for both men and women, but only statistically significant for women. Moreover, as table 2 shows, all of the studies that have formally tested for equality by gender can either not reject that impacts have been similar for men and women, or have found significantly higher impacts for men. Therefore, it should not be assumed that training is generally more effective for women.

Hirshleifer et al. (2016) investigate treatment heterogeneity by key characteristics of the type of training provided. These authors find some evidence that training is more effective when given by private providers rather than government training institutes. This is consistent with an increasing emphasis in policy towards better aligning training programs with private sector demand. However, these authors still find that even the impacts of privately-provided training are modest and fall off over time.

Table 2. The Vocational Training Works Better for Women Myth

<table>
<thead>
<tr>
<th>Country</th>
<th>Study</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkey</td>
<td>Hirshleifer et al. (2016)</td>
<td>Cannot reject equality of impacts by gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impacts only significant for males aged 25 and older</td>
</tr>
<tr>
<td>Argentina</td>
<td>Alzúa et al. (2016)</td>
<td>Impacts for men statistically different from women</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impacts are only significant for men</td>
</tr>
<tr>
<td>Colombia</td>
<td>Attanasio et al. (2011)</td>
<td>Does not test for equality by gender</td>
</tr>
<tr>
<td></td>
<td>Attanasio et al. (2017)</td>
<td>Impacts only significant for women</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Does not test for equality by gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impacts more significant for women</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>Card et al. (2011)</td>
<td>Cannot reject equality of impacts by gender</td>
</tr>
<tr>
<td></td>
<td>Ibarrarán et al. (2014, 2015)</td>
<td>No significant impact for either gender</td>
</tr>
<tr>
<td></td>
<td>Acevedo et al. (2017)</td>
<td>Does not test for equality by gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Significant impact on formal employment for men</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Does not test for equality by gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No significant long-run impact for either gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Finds significant impacts for both men and women</td>
</tr>
<tr>
<td>Malawi</td>
<td>Cho et al. (2013)</td>
<td>Cannot reject equality of impacts by gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No significant impact for either gender</td>
</tr>
<tr>
<td>Peru</td>
<td>Diaz and Rosas (2016)</td>
<td>Does not test for equality by gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Some significant impacts on formal employment for both</td>
</tr>
</tbody>
</table>
Finally, an important point to note with all of these evaluations is that they measure the private returns to vocational training, assuming that the treatment group and control group are not competing for the same jobs. Even if this is the case, and the estimates remain internally valid, the public policy question of whether to support such programs also depends on whether trained individuals get new jobs, or crowd out non-program participants who would have otherwise taken them. None of the studies were designed to look at this question, although Hirshleifer et al. (2016) and Attanasio et al. (2017) discuss it; both sets of authors examine whether impacts differ by the tightness of the labor market, but find no significant differences. This offers some comfort against the displacement concern, but it still seems likely that at least some of the modest gains shown by vocational training programs come from changing who gets particular jobs, rather than from generating new employment in the economy as a whole.

**Wage Subsidies**

In a simple model of the labor market, workers are paid their marginal product, and so if young workers are not very productive to begin with, they would simply be paid low wages. Indeed, in some African contexts under the apprenticeship system, workers actually receive negative wages, paying firms for the privilege of learning on the job. However, in many labor markets, minimum wages and subsistence constraints set a lower bar on the amount that firms can pay for labor, and additionally the presence of hiring and firing frictions means that if there is uncertainty about the productivity of a worker, firms may prefer not to hire them. The result is that individuals who are willing to work may be unemployed, particularly youth who are inexperienced and untested, and less able to signal productivity.

Wage subsidies are intended to help overcome these causes of unemployment. A temporary wage subsidy given to a worker lowers the cost to a firm of hiring that worker (although as Levinsohn and Pugatch (2014) show, workers may increase their reservation wages in response so the cost of labor need not fall by the full amount of the subsidy). This should then lead to an increase in employment for the period the subsidy is in effect. Moreover, there are several possible ways for this short-term subsidy to have a lasting impact on employment: the experience gained may act as a stepping stone to longer-term employment, workers may learn on the job and increase productivity to a level above minimum wages, and firms may learn about the quality of workers and be able to keep individuals who are good matches. Note that to the extent that these subsidies enable firms to form better matches and overcome training cost frictions, overall hiring may increase rather than merely having subsidized job-seekers completely displace unsubsidized ones.
Three studies have evaluated the impact of wage subsidies given to workers using randomized experiments in Argentina (table 3). The earliest was Galasso et al. (2004), who offered welfare recipients a wage subsidy voucher that was valid for up to 18 months, paying the firm up to $150 per month. However, employers had to formally register any workers hired with this subsidy, and would face severance charges if they fired the worker after the program, so only three workers in the treatment group were hired using the voucher. A similar situation arose in Levinsohn et al. (2014) in South Africa, in which youth were given vouchers that would pay the firm a monthly subsidy for up to six months if the firm formally registered the worker. Only 22 firms used the voucher, hiring only 30 workers out of 1,500 given the voucher. Both studies show the reluctance of firms to face the labor regulations associated with hiring workers.

In contrast, Groh et al. (2016a) did not require firms in Jordan to formally register the worker, following the norms of the labor market in which most employment was informal. These authors’ subsidy was also valid for six months. Half of the individuals given the voucher in their study used it, and there was a 38 percentage point increase in employment during the period the subsidy was in effect. However, as detailed in figure 1, once the subsidy ended this treatment effect disappeared quickly as firms fired workers, other workers quit, and the control group caught up a little. The result was no long-term significant impact on employment. Subsidies did not provide the stepping stone to additional work that theory might suggest.

Despite the lack of use of the vouchers in the Argentina and South Africa experiments, both studies do report significant impacts on wage employment (although no overall impact on employment in the Argentina case). The authors of both studies speculate that having the voucher gave job-seekers the confidence to approach more employers and exert more job search effort, which resulted in more employment, though just of an informal sort. If true, this would make the policy very cost-effective since hardly anyone cashed in the voucher. However, note that the attrition rates are high in both studies (23 percent in Argentina, and 39 percent by the two-year follow-up in South Africa). The South African study has a higher point estimate at two years than one year, but then shows the treatment effect decreasing over time when restricted to the sample present in both follow-up years. It seems highly likely that the employment outcomes of the attritors are different from those who responded to the survey, so extreme caution should be used in interpreting these treatment effects.

Moreover, as with vocational training, a key concern is that any gains to those receiving the vouchers come at the expense of others in the economy who would have otherwise been hired. Groh et al. (2016a) find suggestive evidence of this in Jordan. When they examine impacts by region and look at longer-term time trends, these authors find a lasting impact of the subsidy on employment in the
### Table 3. Summary of Wage Subsidy Impacts

<table>
<thead>
<tr>
<th>Country</th>
<th>Study</th>
<th>Population</th>
<th>Sample Size</th>
<th>Attrition</th>
<th>Time Frame</th>
<th>In Effect</th>
<th>Subsidy</th>
<th>Proportion using</th>
<th>Impact on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>Galasso et al. (2004)</td>
<td>Welfare recipients</td>
<td>548</td>
<td>22.5</td>
<td>18 months</td>
<td>Yes</td>
<td>0.011</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>Jordan</td>
<td>Groh et al. (2016a)</td>
<td>Female community college graduates</td>
<td>1349</td>
<td>8</td>
<td>6 months</td>
<td>Yes</td>
<td>0.503</td>
<td>38.4</td>
<td>228.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[33.3, 43.5]</td>
<td>[197, 260]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[2.8]</td>
<td>[15.9]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[−3.4, 9.1]</td>
<td>[−14.46]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[1.4]</td>
<td>[14.0]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[−3.2, 9.8]</td>
<td>[−17.45]</td>
</tr>
<tr>
<td>South Africa</td>
<td>Levinsohn et al. (2014)</td>
<td>Youth</td>
<td>3064</td>
<td>23.0</td>
<td>12 months</td>
<td>No</td>
<td>0.02</td>
<td>7.4</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[2.9, 11.9]</td>
<td>[−9.37]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[9.5]</td>
<td>[−19]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[3.6, 15.4]</td>
<td>[−72.34]</td>
</tr>
</tbody>
</table>

**Note:**

a. Paper did not provide standard errors, but this impact was not statistically significant.

Time Frame refers to timing since subsidy began. In Effect denotes whether subsidy still being paid at time of survey.

n.r. denotes not reported.

South African sample started with 4,009, but waited one year and re-interviewed before starting intervention. Estimates here based on 2010 sample.

Impacts on Employment are in terms of percentage points, on Earnings are Percent increase on Control Mean.
less-populated labor markets outside of Amman. But the control group ends up with a lower employment rate than other cohorts of graduates had received in recent years, and direct survey evidence suggests that they were competing directly with the treatment group for some jobs. The result is that wage subsidies do not seem likely to have increased aggregate employment in this case.

An alternative to providing subsidies to workers has been to give the subsidies to firms, to encourage them to hire more workers. De Mel, McKenzie, and Woodruff (2016) test the impact of wage subsidies given to microenterprises to encourage them to hire workers. These authors find that 24 percent of firms use the subsidy to hire a worker, resulting in an increase in employment while the subsidy is in effect. But the dynamics then look reasonably similar to those in figure 1, with much of this impact disappearing as soon as the subsidy is removed, and no long-term impact being felt after two years.

A final use of subsidies is to use them to help prevent liquidity-constrained firms from shedding workers during a temporary shock. This type of policy was another common response to the global financial crisis. The idea was that firms suffering a temporary demand shock and/or liquidity shock may fire workers who they would later want to hire back. A subsidy may prevent them from firing these workers in

---

**Figure 1.** Trajectory of Impact from a Wage Subsidy Program in Jordan

![Graph showing the impact of a wage subsidy program over time.](https://example.com/graph.png)

*Source: Groh et al. (2016a).* Figure shows month by month impacts of a wage subsidy on employment, along with 95 percent confidence intervals. The two vertical lines show the start and end of the subsidy period.
the first place and hasten the recovery of these firms if hiring and firing is costly. 
Bruhn (2016) evaluates a wage subsidy program that Mexico used during the glo-
bal financial crisis, using difference-in-difference analysis to compare the employ-
ment trajectories of durable manufacturers in industries eligible for the program to 
those in industries ineligible for the program. She finds employment to be 6–13 percent higher in the affected industries during the program, and to grow faster 
after the crisis, suggesting the program helped firms to recover more quickly from 
the shock.

This accumulated evidence suggests that wage subsidies are unlikely to be very effective in generating additional employment under standard labor market condi-
tions, and may also even not be very effective in playing a distributional role in determining which individuals get to access jobs. However, it also suggests two potential use cases. The first is during conditions of large, temporary shocks. Even if 
ALMPs like wage subsidies have no lasting impacts, from a social protection view-
point if they help households smooth temporary shocks then this might be justifi-
cation enough for their use. The difficulty here, of course, is in knowing whether or not the shocks are temporary or structural in nature, since there is a danger of trying to maintain employment in industries that economic shocks make perman-
ently less competitive. Secondly, the evidence suggests that wage subsidies may be useful for temporary employment creation. This might be important particularly in fragile economies where large youth unemployment raises other concerns. In this 
vein, short-term evidence from Yemen (McKenzie, Assaf, and Cusolito 2016) showed positive impacts of a youth internship program that subsidized firms to take on interns, although the outbreak of war prevented analysis of any lasting impacts.

Search and Matching Assistance

Many governments provide employment services in the form of helping job-seekers with preparing resumes, hosting labor exchanges, and helping to match firms with workers seeking employment. The review of Betcherman, Olivas and Dar (2004) was relatively favorable of these types of programs, and noted that since the costs are often low for providing such services, the cost-benefit ratios can be favorable. However, this recommendation was largely based on developing country evidence, and the review also noted, based on non-experimental evaluations from Brazil and Uruguay, that such programs may be less effective in countries with large informal sectors if workers typically use other channels to find jobs, and if they work at all, might work best for more educated job-seekers.

A competing view to this concern is that search and matching frictions may be greater in developing countries, leaving more scope for improvements.
The educational systems in many countries may not be very good at signaling quality, and may teach content that is very different from the skills that employers are looking for. Information about vacancies may be more difficult to come by if workers and firms are not all online, and match quality may be worse if informal networks are relied upon to fill vacancies. Improving this process could then reduce unemployment directly (by filling existing vacancies), as well as indirectly (by lowering hiring costs so firms create more vacancies).

Table 4 summarizes the results of nine recent randomized experiments that have tested various interventions designed to reduce information and search frictions, and to better match workers and firms together. These experiments incorporate several types of specific interventions. The work that tests public intermediation services most directly is Dammert, Galdo, and Galdo (2015), who worked with the public service provider in Peru to test whether providing information about job vacancies to registered job seekers improves employment, and additionally whether sending these announcements by text message helps further. Another example of providing information about job opportunities and recruiting services is Jensen (2012), who connected rural villages in India to experienced recruiters at the start of the business process outsourcing boom in India, and provided information about this new sector.

Two studies (Abebe et al. 2016b; Beam 2016) test the impact of job fairs, which bring firms and workers together. The idea here is to give both firms and workers the opportunity to assess a large number of possible matches at the same time, and become better informed about the range of job opportunities and worker types. Two studies (Franklin 2015; Abebe et al. 2016a) test the impact of reducing the monetary costs of search for job seekers by offering transport subsidies to allow them to travel to a different part of town where job opportunities are more commonly displayed.

The final approach used in four studies is to attempt to reduce the information frictions faced by firms by providing more information about job-seekers. Abel, Burger, and Piraino (2016) approach this by developing a standardized reference letter format, and encouraging job-seekers to get this reference from former employers. Groh et al. (2015), Abebe et al. (2016a), and Bassi and Nansamba (2017) instead develop their own tests of a variety of soft and hard skills that might otherwise be difficult for firms to observe, but which firms say they find valuable. Examples include information about math ability, creativity, teamwork, attendance rates, and communication skills.

These types of programs tend to be much cheaper than vocational training and wage subsidies (if taken up) in terms of cost per person invited to participate. The last column shows that all but one of the studies that provide cost information have costs of $25 or lower per person assisted. That is, the costs are one-fiftieth to one-hundredth of the cost of vocational training programs. The exception is Groh...
## Table 4. Evidence on Search and Matching Assistance

<table>
<thead>
<tr>
<th>Country</th>
<th>Study</th>
<th>Type of Assistance</th>
<th>Population</th>
<th>Sample Size</th>
<th>Attrition (%)</th>
<th>Time Frame</th>
<th>Employment Impact</th>
<th>Earnings Impact</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>Abebe et al. (2016a)</td>
<td>Transport subsidy</td>
<td>Young</td>
<td>2097</td>
<td>6.5</td>
<td>8 months</td>
<td>4.0</td>
<td>0.1</td>
<td>$7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Certifying hard skills</td>
<td>Young</td>
<td>1778</td>
<td>6.5</td>
<td>8 months</td>
<td>[−1.9, 9.9]</td>
<td>[−7.7]</td>
<td>n.r.</td>
</tr>
<tr>
<td></td>
<td>+ interview assistance</td>
<td>Unemployed Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[−4.0, 8.0]</td>
<td>[−9, 20]</td>
<td>$9</td>
</tr>
<tr>
<td></td>
<td>Franklin (2015)</td>
<td>Transport subsidy</td>
<td>Unemployed</td>
<td>877</td>
<td>31</td>
<td>10 months</td>
<td>6.8</td>
<td>n.r.</td>
<td>$14</td>
</tr>
<tr>
<td></td>
<td>Abebe et al. (2016b)</td>
<td>Job fair</td>
<td>18-29 year old</td>
<td>4059</td>
<td>6.5</td>
<td>4 months</td>
<td>−1.2</td>
<td>6.6</td>
<td>$14</td>
</tr>
<tr>
<td>India</td>
<td>Jensen (2012)</td>
<td>Connecting to recruiters</td>
<td>Young</td>
<td>1534</td>
<td>6</td>
<td>3 years</td>
<td>2.4</td>
<td>n.r.</td>
<td>$12</td>
</tr>
<tr>
<td></td>
<td>women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.2, 4.6]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td>Groh et al. (2015)</td>
<td>Certifying soft and hard</td>
<td>Unemployed</td>
<td>1354</td>
<td>19</td>
<td>5 months</td>
<td>2.4</td>
<td>n.r. (a)</td>
<td>$203</td>
</tr>
<tr>
<td></td>
<td>Skills &amp; matching</td>
<td>tertiary graduates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[−4.7, +9.4]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>Dammert et al. (2015)</td>
<td>information about job</td>
<td>Job-seekers</td>
<td>1280</td>
<td>7</td>
<td>1 month</td>
<td>6.2</td>
<td>n.r. (a)</td>
<td>$25</td>
</tr>
<tr>
<td></td>
<td>vacancy</td>
<td></td>
<td></td>
<td>1280</td>
<td>7</td>
<td>3 months</td>
<td>−0.2</td>
<td>n.r.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[−6.7, 6.3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>Beam (2016)</td>
<td>Attending job fair</td>
<td>20-35 year old</td>
<td>865</td>
<td>3.5</td>
<td>10 months</td>
<td>2.7</td>
<td>n.r.</td>
<td>$3.5</td>
</tr>
<tr>
<td>South Africa</td>
<td>Abel et al. (2016)</td>
<td>Encouraged to use reference letter</td>
<td>Unemployed</td>
<td>1267</td>
<td>17</td>
<td>3 months</td>
<td>2.0</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Uganda</td>
<td>Bassi and Nansamba (2017)</td>
<td>Certifying soft skills</td>
<td>Young</td>
<td>515</td>
<td>15.2</td>
<td>12 months</td>
<td>5.8</td>
<td>−6.3</td>
<td>$19</td>
</tr>
</tbody>
</table>

**Note:** Peru estimates are for pooled treatment effect across three subtreatments which all provided job vacancy information.

a. Study does not report unconditional earnings estimates. No significant impact found on earnings conditional on work.

b. No significant impact on log earnings conditional on work at 4 month follow-up.

ITT estimates reported except for Beam (2016) where several incentives used in encouragement design, and LATE impact of attending job fair after receiving attendance voucher is reported.

Impacts on employment are in terms of percentage points, impacts on earnings in terms of percentage growth relative to control mean.
et al. (2015), who had a cost of $203 per person, since their enrollment and testing procedure was rather expensive.

These lower costs certainly lower the bar in terms of treatment impacts needed in order for these programs to pass cost-benefit tests. However, as seen in table 9, out of the 10 different interventions, only one (Jensen 2012) finds a significant impact on employment, increasing employment by 2.4 percentage points over three years. Dammert et al. (2015) find that their intermediation services tend to speed up the process of finding a job, with a significant employment impact after one month, but by three months the control group has caught up. Many of the other studies have small but positive point estimates, with an average impact across the studies of 2.7 percentage points. However, it is also worth noting that, apart from Jensen (2012), none of the studies measures impacts beyond a year, so they cannot measure whether there is any sustained employment impact.

A number of studies deemphasize employment as an output, claiming that their intervention helps in improving the quality of jobs. These studies examine quality in different ways, sometimes defining quality jobs as “permanent” or “formal”, or simply as “wage employment” rather than self-employment. For example, Beam (2016) finds attending a job fair results in a 10 percentage point increase in formal employment, which is matched by a reduction in self-employment; Franklin (2015) finds more positive impacts on permanent employment and being employed in an office than on total employment; and Abebe et al. (2016a) find their job application workshop, which certified skills and provided interview preparation, led to a 6.9 percentage point increase in permanent employment.

However, there are two problems with justifying these programs on the basis of improved “job quality”. The first is that there is a large body of literature that debates the extent to which informality and self-employment are choices made by individuals, which have benefits associated with them such as flexible labor hours and less taxation, rather than reflecting exclusion from formal wage jobs (e.g., Maloney 2004). Indeed, Abebe et al. (2016a) find no significant change in job satisfaction from their treatment, despite the change to permanent employment. Secondly, as seen in table 9, none of these interventions show a significant impact in labor earnings. While the confidence intervals are wide in many cases, and therefore allow the possibility of these interventions passing cost-benefit tests, the short-time horizons and lack of significant impact on earnings means that there is currently no evidence that they do.

A further point to note is how few direct hires occur through many of these interventions, and how an important share of job offers are turned down by job-seekers. Groh et al. (2015) made more than 1,000 matches between firms and workers. Youths rejected the opportunity of a job interview 28 percent of the time, and when a job offer was received, they rejected the job offer or quit quickly 83 percent of time, resulting in only nine hires that lasted one month. Bassi and
Nansamba (2017) report that only 2–4 percent of their job matches resulted in a worker being hired, and few workers hired were still employed at the firm at their follow-up. Abebe et al. (2017b) invited 1,007 people to their job fairs (606 attended) but only 76 job offers were made and 14 people were hired. Beam (2016) reports only two respondents from her job fair were working for one of the employers that attended the fair at endline. As such, while the cost per person invited to treatment can be low, the cost per individual actually placed in a job can be substantially higher—Groh et al. (2015) estimate a cost of $22,000 in their case.

Several studies note that programs that allow workers to better certify their skill levels may have differential effects for those with low and high skills. Being able to signal your skills can be good if you have high skill levels, but disadvantageous if your skill levels are below those of other job-seekers. The result might be better-quality workers for firms, but simply a reallocation of who gets work from less- to more-skilled workers.

What Do Policymakers Expect of Such Programs and What Does Revealed Preference Show?

The above discussions show that traditional active labor market programs have had at most modest impacts on employment in most cases, with a typical intervention leading to a 2 percentage point increase in employment that is usually not statistically significantly different from zero. Cost-benefit calculations usually rely heavily on extrapolating statistically insignificant total earnings gains over periods well beyond the timeframe of the study.

These impacts are much lower than expected by policymakers and program participants in many cases. Hirshleifer et al. (2016) show this formally in the context of their vocational training experiment in Turkey. There was strong demand for this training from participants, with courses oversubscribed by a factor of two or more. Subjective expectations of the employment impact of the program elicited from participants show that they expected a 32 percentage point increase in the likelihood of employment, while staff in the government employment office expected the training to increase the likelihood of employment by 24 percentage points. These expectations far exceed the actual impact of 2 percentage points seen in table 1. Groh et al. (2016b) likewise show that policymakers in Jordan expected the wage subsidy program to have lasting impacts on youth employment, in contrast to the realized impacts.

Economists are also not immune to this tendency to think active labor market programs will be more effective than they typically are. A first testament to this comes from a number of the studies covered in this review being interventions designed by the researchers themselves, in addition to those evaluating programs
that governments were already going to implement. Secondly, Groh et al. (2016b) carried out an expectations elicitation exercise when presenting the results of their Jordan wage subsidy research. These authors find that development economists on average expected a 10 percentage point increase in employment after the subsidy had ended, compared to the 2.8 percentage point increase seen in table 3.

However, while revealed preference shows that there are participants who think these programs will be effective and therefore choose to participate in them, revealed preference also suggests that the types of formal jobs and manufacturing jobs that many of these programs think of as “successful” outcomes are not that valued by job-seekers. For example, Blattman and Dercon (2016) randomize job-seekers into industrial jobs in large formal firms in Ethiopia, and find that almost one-third of people offered a job quit in the first month, and 77 percent quit within the first year; further, workers experienced health problems from staying in this work. Similarly, Adhvaryu, Kala, and Nyshadham (2016) find that female garment workers in India have very high quit rates, losing almost 80 percent of the workers in their study over two years. These high rates of turnover are not consistent with formal jobs being so valuable and desired that workers never want to leave once they attain such positions.

The implicit assumption behind search and matching interventions in particular is often that search frictions make it costly and difficult for firms to find workers. Simple queries of firms often find firms saying that they find it hard to find the right workers. But one also sees firms being reluctant to raise wages or spend more money to get better matches. Groh et al. (2015) conducted a survey in Jordan where they tracked firms as they opened up job vacancies, and found that only 6 percent of the positions required more than 4 weeks to find a new employee, and most firms could, in fact, fill jobs quite quickly. Similarly, De Mel et al. (2016) find that firms in Sri Lanka say it would take seven days on average to fill positions. If it were particularly costly for firms to find and recruit workers, we might expect a range of market solutions to emerge to help them lower these costs. Indeed, there are a range of human resources consultants and executive talent firms that help firms fill skilled and unusual positions. But the lack of an existing market alternative to many of the interventions being tested may suggest that firms do not face large search costs for other entry-level positions.

What Types of Alternative Policies Show Promise?

Given the continued pressure for governments to be seen to be doing something to help people find jobs, this lack of empirical evidence for the effectiveness of many traditional programs is unlikely to be enough to cause them to be abandoned unless better alternatives can be found instead. What might these alternatives be?
One set of alternative policies is to move away from interventions on the labor supply side and focus more on policies to help firms overcome the obstacles they face in innovating, growing, and creating more jobs. Such private sector development programs also have a mixed record of success, but there are examples (e.g., McKenzie 2016 and the references therein) of programs that have generated new jobs. A related approach is to help firms overcome onerous regulations and labor laws that limit firms hiring and growing. Freeman (2010) provides a recent review of how labor market institutions and regulations, including the role of mandated benefits, minimum wages, and employment protection regulations affect employment. He notes that while the direct negative impacts on employment can be modest in many cases, in part due to incomplete enforcement, they do reduce labor mobility. In specific countries, particularly onerous regulations can have larger negative employment effects. Bertrand and Crépon (2016) find that teaching South African firms about labor laws and providing legal support to help them deal with these laws spurred new employment generation. Martin et al. (2017) show that removing regulations in India that limited the production of certain products to only small-scale firms resulted in a 6 percent increase in employment in affected districts.

On the labor supply side, the most promising interventions appear to be ones that help workers access different labor markets, as well as overcome sectoral and spatial mismatches. Sectorial mismatches arise when people are trapped in the wrong occupations as trade and technology change the demand for labor, or because of gender segregation in society. Campos et al. (2016) show that in Uganda women who cross over into male-dominated industries make three times as much as women who remain in female-dominated industries. Hendra et al. (2016) report that a demand-driven training program in the United States that aimed to train the unemployed in sectors which were in demand resulted in a 14 percent income gain after two years. However, these authors also note that these programs can be complex to run and require experienced providers.

The largest market failures in labor markets occur across space, with very different employment opportunities for the same skills depending on where individuals are located. We have seen that some of the more successful screening and matching interventions were ones that provided assistance with learning about job opportunities in a different location (Jensen 2012), or subsidizing job searches in different parts of the city (Franklin 2015; Abebe et al. 2016a). More striking evidence comes from Bryan, Chowdhury, and Mobarak (2014), who show that a small subsidy equal to the cost of a bus ticket spurred new seasonal migration in Bangladesh, increasing household consumption by 30–35 percent during the hungry season (they do not measure household income). Even larger gains can be had from facilitating international migration. Gibson and McKenzie (2014) show that sending seasonal workers to New Zealand increased per-capita incomes in Tonga and Vanuatu by more than 30 percent. Luthria and Malaulau (2013) discuss the
process of facilitation used by governments and the World Bank to allow this movement to happen. However, such facilitation is not always successful, especially if it focuses only on barriers on the worker side. For example, Beam, McKenzie, and Yang (2016) conducted several interventions in the rural Philippines to facilitate more international migration, and were unsuccessful in generating additional international employment.

Concluding Lessons for Impact Evaluations

The modal study surveyed in this review is from 2016, reflecting rapid recent growth in the body of evidence around active labor market interventions in developing countries. This body of work has generated substantial new knowledge, but also suffers from several limitations that future work can attempt to learn from:

1) Given the likely effect size of active labor market interventions, sample sizes may need to be a lot larger. Based on the current body of research, it seems many interventions may have only a modest impact on employment, such as a 2 percentage point increase. In some cases, for example expensive training programs, such an effect is too small to be economically meaningful. But cheaper programs such as search-and-matching assistance could still deliver gains that exceed the costs with these modest impacts. Taking as an example the 13 percent employment rate in the control group of Abel, Burger, and Piraino (2016), a study needs to have 6,424 individuals in the treatment group and 6,424 in the control group to detect a 2 percentage point employment impact with 80 percent statistical power. This is much larger than existing studies.

2) Measuring impacts over longer-time frames. Returns to these programs will differ substantially if they merely speed up the process of gaining employment versus having lasting impacts. Yet most studies measure impacts over at most 1–2 years, leaving them to speculate about cost-benefit on the basis of assumptions about how impacts vary over time. Tracking impacts over longer time periods is therefore needed. Studies that link participants to administrative records (such as Attanasio, Medina, and Meghir 2017) offer one promising way to do this.

3) Limiting attrition. When the likely impact on employment is only 2 to 3 percentage points, and attrition rates are 10, 20, or even more than 30 percent, any treatment effects are dwarfed by attrition, and bounds that incorporate this attrition will be completely uninformative. Limiting attrition is particularly difficult given that so many ALMPs focus on youth, who tend to be more mobile and difficult to track over time. Serious investment in limiting attrition, combined with the use of administrative data is needed.

4) Continued and improved careful measurement of costs. I was pleasantly surprised by the number of studies that did report the costs of the intervention, although a number
still lack this key information. More work is needed to make clear average versus marginal costs in understanding the cost structure as pilot programs expand.

5) **Pre-specifying outcomes and heterogeneity.** A number of studies fail to find a significant impact on either employment or earnings, but then emphasize impacts on a particular subgroup (such as one gender, or one skill level), or for one outcome (such as formal employment). Pre-specification of the primary outcomes and key heterogeneity of interest lessens concerns about multiple hypothesis testing.

6) **Testing placebo effects.** Several studies find impacts despite almost no direct hires through the program they study (e.g., Galasso, Ravallion, and Salvia 2004; Beam 2016; Levinsohn et al. 2014). These studies raise the possibility that simply doing anything to support job-seekers may encourage them to keep exerting effort and searching, so that what matters is their sense that someone wants them to succeed, not the particular policy pursued. Testing more formally this sort of placebo effect would be interesting in further work.

7) **Understanding general equilibrium better.** A key concern with many of these policies directed at particular job-seekers is that they merely change who gets the jobs that firms are advertising, without increasing the total number of jobs available. The ideal would be approaches like the Crepón et al. (2013) experiment in France, which randomized at the local labor market level. Abebe et al. (2016a) attempt this within clusters in Ethiopia. A second approach is to randomize also at the firm level, as in Groh et al. (2015) and Abebe et al. (2016b), to attempt to measure if firms increase hiring. Further methodological work to develop additional ways to examine these spillovers is needed.

**Concluding Lessons for Policy**

Given the importance of jobs for poverty reduction, productivity growth, and social cohesion (World Bank 2012), it is no surprise that policymakers have actively pursued policies to try to help job-seekers find jobs. But as this review has shown, an emerging body of evidence shows these policies to generally be far less effective than policymakers, program participants, and economists typically expect. It should be noted that this is not unique to ALMPs in developing countries: in their review of largely developed country evidence, Crépon and van den Berg (2016) conclude that “the general outlook for ALMPs is rather grim”.

One reason for this lack of effectiveness is a positive one: labor markets (at least in urban areas) in developing countries actually appear to work a lot better than is sometimes thought. It is easy to imagine all types of constraints that might inhibit the functioning of labor markets, but in practice firms appear to be able to fill many vacancies quite quickly, and workers turn down many job opportunities and quit jobs frequently in pursuit of better opportunities. These facts do not suggest that
workers and firms have great difficulties meeting one another, or that job-matches are so rare and scarce that workers cling to every job opportunity they receive. This is not to say that everyone who wants a well-paying wage job in the formal sector can get one, with many developing countries having high rates of youth unemployment, especially for the skilled. However, it appears to be other constraints that limit the number of jobs created, such as high minimum wages and inflexible labor laws, or lack of access to financing and infrastructure that prevent firm growth, with the solution to these issues lying outside of active labor market policies.

Nevertheless, while this suggests less of a role for traditional active labor market policies, there still appears to be significant scope for improvements in dealing with structural and spatial mismatches in labor. As the evidence here has shown, not everything that policymakers try works, and so these new policy innovations should be piloted against competing alternatives and accompanied by rigorous impact evaluations in order to test different approaches.

Notes

David McKenzie is a Lead Economist in the Development Research Group, World Bank; Email dmckenzie@worldbank.org. I thank Asli Demirguc-Kunt for encouraging me to give a policy talk on this topic, the editor and three anonymous referees for useful comments, and the authors of the different papers summarized for helpful clarifications and comments.

1. A complementary approach is meta-analysis, with Card et al. (2015) pooling together ALMP estimates from both developed and developing countries, including both randomized and quasi-experimental evaluations.

2. Another category of ALMPs that aims to increase labor demand are public works programs. There have been fewer recent experimental evaluations of these programs, although evaluations are in progress in Côte d’Ivoire and Sierra Leone. Blattman and Ralston (2015) and World Bank (2012) survey the existing evidence.

3. An exception could be in cases where unemployment has large negative social returns, such that governments would prefer to have someone in a job that costs more to provide than the individual earns.

4. In addition to the 33 month impact reported in table 1, Alzáia et al. (2016) also report an impact on formal employment (but not earnings) at 48 months. This is smaller still, at 1.4 percent, and not statistically significant. Acevedo et al. (2017) also report 12 month impacts, which are positive and significant on employment for women, and negative and significant on employment for men.

5. Acevedo et al. (2017) find no impact for either gender at 36 months, but do find stronger impacts for women at 12 months.

6. Franklin (2015) tests whether merely surveying people about job search leads to changes in behavior, and finds it does not.

References


Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs

Abhijit V. Banerjee, Rema Hanna, Gabriel E. Kreindler, and Benjamin A. Olken

Targeted transfer programs for poor citizens have become increasingly common in the developing world. Yet, a common concern among policy-makers and citizens is that such programs tend to discourage work. We re-analyze the data from seven randomized controlled trials of government-run cash transfer programs in six developing countries throughout the world, and find no systematic evidence that cash transfer programs discourage work. JEL codes: J22, I38, H53, C93

Governments in the developing world are increasingly providing social assistance programs for their poor and disadvantaged citizens. For example, in a recent review of programs worldwide, Gentilini, Honorati, and Yemtsov (2014) find that 119 developing countries have implemented at least one type of unconditional cash assistance program, and 52 countries have conditional cash transfer programs for poor households. Thus, on net, they find that 1 billion people in developing countries participate in at least one social safety net.1

These programs serve to transfer funds to low-income individuals and have been shown to reduce poverty (Fiszbein and Schady 2009) and to improve educational outcomes (Schultz 2004; Glewwe and Olinto 2004; Maluccio and Flores 2005) and access to health services (Gertler 2000, 2004; Attanasio et al. 2005). However, despite these proven gains, policy-makers and even the public at large often express concerns about whether transfer programs discourage work. In fact, these types of beliefs tend to be associated with less extensive and less generous social assistance programs: figure 1 shows a negative relationship between spending on cash transfers as a fraction of GDP and the share of the population in a country...
who believe that poverty is due to laziness (as opposed to an unfair society). But are these beliefs justified? Is this what the theory would predict? What does the evidence say?

On the one hand, transfer programs could reduce work incentives: individuals may not work—or exit visible forms of work—to ensure that they keep the benefits, or they may stop work simply through the income effect. On the other hand, these programs could have positive effects on work if they help relieve the credit constraints of the poor to allow them to invest in small enterprises or if they have spillover effects. Given that the theory has some ambiguity, it is then imperative to turn to the evidence. In developed country policy contexts, some transfer programs have indeed been shown to have small, but statistically significant, effects on work. However, there is little rigorous evidence showing that transfer countries in emerging and low-income countries actually lead to less work.

In this paper, we re-analyze the results of seven randomized controlled trials of government-run non-contributory cash transfer programs from six countries worldwide to examine their impacts on labor supply. Re-analyzing the data allows
us to make comparisons that are as similar as possible, using harmonized data definitions and empirical strategies. It also allows us to use a cutting-edge, statistical technique to pool the effects across studies to analyze in a systematic way the effects in different countries to obtain tighter statistical bounds than would be possible from any single study, while still allowing for the possibility that the different programs worldwide could have different treatment effects given the differing contexts.

We bring together data on this issue from all randomized control trials (RCT) that we identified that met three criteria: (1) it was an evaluation of a (conditional or unconditional) government-run cash transfer program in a low-income country that compared the program to a pure control group; (2) we could obtain micro data for both adult males and females from the evaluation; and (3) the randomization had at least 40 clusters. This yielded data for transfer programs from six countries: Honduras, Indonesia, Morocco, Mexico (two different programs), Nicaragua, and the Philippines.4 All of these programs are non-contributory transfers programs, rather than social insurance programs.

Across the seven programs, we find no systematic evidence of the cash transfer programs on either the propensity to work or the overall number of hours worked, for either men or women. This is a particularly stark finding, given the differences in context and program design across the differing settings. Importantly, pooling across the seven studies to maximize our statistical power to detect effects if they exist, we find no observable impacts on either work outcome. We can reject with high confidence any moderate negative effects for the elasticity of work outcomes with respect to income for men. If anything, the point estimates are positive. For women, more uncertainty persists even after aggregating: the point estimates are negative and small, with wide credible intervals that cover both negative and positive values. The overall low effects on work behavior may be, in part, due to the fact that the eligibility to receive (or stay on) one of the programs does not appear to be closely tied to current income levels.

Theoretically, the transfers could have different effects on work “outside the household” versus self-employment or work “within the household.” For example, one could imagine that the effect for the outside-work sector may be larger, as individuals fear—rationally or otherwise—that visible employment outside the household could disqualify them from receiving future transfers. Looking at the pooled sample, we find no aggregate effect on either outcome, although the analysis points to large dispersion in impacts across programs. Indeed, for most individual programs we do not find any significant effect for either outcome, and for one program we find a small shift towards work inside the household, while for another program we find a small shift towards work outside the household.

In short, despite the rhetoric that cash transfer programs lead to a massive exodus from the labor market, we do not find evidence to support these claims.
Coupled with the benefits of transfer programs that are well-documented in the literature, this further suggests that cash transfer programs can play an effective role in providing safety nets in developing and emerging countries.

Theoretical Frameworks and Existing Literature

While much of the discourse around transfer programs is centered on people working less, the theory is more ambiguous. On one hand, cash transfers may reduce work for two key reasons. First, these programs provide unearned income, and recipients may “spend” some of this extra income on leisure. That is, the pure income effect may lead recipients to work less if leisure is a normal good. Second, cash transfers may decrease labor supply if they act as a “tax” on labor earnings. Specifically, if people believe that higher earnings will disqualify them from receiving benefits, they will have a disincentive to work.

On the other hand, cash transfer could increase work through a number of mechanisms. First, cash transfers could help households escape the classic poverty trap problem elucidated by Dasgupta and Ray (1986) by allowing them to have a basic enough living standard to be productive workers. Second, an infusion of cash could reduce credit constraints to starting or growing a business. Indeed, Gertler, Martinez, and Rubio-Codina (2012) provide some evidence that Mexico’s Oportunidades program led poor households to be able to invest in productive assets. Third, cash transfers can also finance risky but profitable endeavors such as migration, which may lead to increases in the adult labor supply. For example, Ardington, Case, and Hosegood (2009) shows that the cash infusion from South African old-age social pension led to prime-aged adults having higher employment, mainly through migration. Finally, additional cash could have spillover effects within poor regions by providing additional cash that can spark increases in sales in local businesses.

The theoretical effect of transfers on work is thus ambiguous, suggesting that both the sign and magnitude of the treatment effects may be driven by the details of the program design (e.g., the targeting methods, the size of the transfers), as well as the underlying economic conditions (e.g., how cash constrained households are, how risk averse they are). Therefore, it is important to turn to the empirical evidence and to look at the evidence across a variety of contexts.

We now turn to evidence from previous studies on the impact of cash transfers on adult labor supply. Table 1 summarizes results from 21 studies, covering 17 conditional or unconditional cash transfer programs that do not have explicit work requirements. The last column summarizes the evidence on overall labor supply indicators, and on shifts in the allocation of labor supply. While not necessarily exhaustive, we included all published studies we could find with a rigorous
<table>
<thead>
<tr>
<th>Paper</th>
<th>Country and Program</th>
<th>Program Type</th>
<th>Research Design</th>
<th>Summary of Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garganta and Gasparini (2015)</td>
<td>Argentina, AUH</td>
<td>CCT</td>
<td>Difference-in-difference</td>
<td>The program reduces the proportion of informal households that acquire formal jobs, for families with children, relative to families without children.</td>
</tr>
<tr>
<td>Foguel and Barros (2010)</td>
<td>Brazil, Bolsa Familia</td>
<td>CCT</td>
<td>Difference-in-difference</td>
<td>Small increase in working probability of less than 1 percentage point for women and between 2 and 3 percentage points for men. Decrease of 0.6-2.6 hours of work per week for women, and an 0.6-1.6 hours increase for men.</td>
</tr>
<tr>
<td>Ribas and Soares (2011)</td>
<td>Brazil, Bolsa Familia</td>
<td>CCT</td>
<td>Propensity Score</td>
<td>No detectable effect on work probability or hours of work. Reduction in formal sector participation.</td>
</tr>
<tr>
<td>de Brauw et al. (2015)</td>
<td>Brazil, Bolsa Familia</td>
<td>CCT</td>
<td>Propensity Weighting</td>
<td>No detectable effect on work probability or hours of work. Shift of 8 hours per week of work away from the formal sector and into the informal sector.</td>
</tr>
<tr>
<td>Ferreira, Filmer and Schady (2009)</td>
<td>Cambodia, CESSP</td>
<td>CCT</td>
<td>Regression Discontinuity</td>
<td>No detectable effect on work probability or hours of work, both for pay and not for pay. Results on total income consistent with no behavioral response. Authors note that selection issues are not accurately resolved by matching.</td>
</tr>
<tr>
<td>Galiani and McEwan (2013)</td>
<td>Honduras, PRAF II</td>
<td>CCT</td>
<td>RCT and Regression Discontinuity in Municipality Poverty index</td>
<td>No detectable effect on work outside the household. Small increase in work inside the household for men, no detectable effect for women.</td>
</tr>
<tr>
<td>Alzua, Cruces, and Ripani (2013)</td>
<td>Honduras, PRAF II</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work or hours of work.</td>
</tr>
<tr>
<td>Paper</td>
<td>Country and Program</td>
<td>Program Type</td>
<td>Research Design</td>
<td>Summary of Findings</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------------------------------</td>
<td>--------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Asfaw et al. (2014)</td>
<td>Kenya, Cash Transfer for Orphans and Vulnerable Children</td>
<td>UCT</td>
<td>RCT and Propensity Score Matching</td>
<td>Reduction in wage work for men and women. Non-farm activities increase for women but decrease for men. No detectable effect on own farm work for either men or women. Effect on total work not reported.</td>
</tr>
<tr>
<td>Haushofer and Shapiro (2013)</td>
<td>Kenya, GiveDirectly</td>
<td>UCT</td>
<td>RCT</td>
<td>No detectable effects on whether primary income source is wage labor, own farm labor, or non-agricultural business.</td>
</tr>
<tr>
<td>Covarrubias et al. (2012)</td>
<td>Malawi, Social Cash Transfer (SCT)</td>
<td>UCT</td>
<td>RCT and Propensity Score Matching</td>
<td>Reduction in wage work between 3 and 5 days from a base of 7 days per month. No data on total work.</td>
</tr>
<tr>
<td>Skoufias et al. (2008)</td>
<td>Mexico, PAL</td>
<td>UCT and In-Kind</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work. Some evidence of substitution from agricultural to non-agricultural work.</td>
</tr>
<tr>
<td>Parker and Skoufias (2000)</td>
<td>Mexico, Progresa</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work.</td>
</tr>
<tr>
<td>Skoufias and Vincenzo Di Maro (2008)</td>
<td>Mexico, Progresa</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work or participation in wage work.</td>
</tr>
<tr>
<td>Alzua, Cruces, and Ripani (2013)</td>
<td>Mexico, Progresa</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work, or on agricultural employment. Small increase of hours of work for eligible women of 0.4 hours on a base of 42 hours per week</td>
</tr>
<tr>
<td>Maluccio and Flores (2005)</td>
<td>Nicaragua, RPS</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work. No significant effect on hours of work for men. Reduction of 5 hours of work per week for men.</td>
</tr>
<tr>
<td>Maluccio (2008)</td>
<td>Nicaragua, RPS</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work. Reduction of 4 hours of work per week for women. Reduction of 8 hours of work per week for men.</td>
</tr>
<tr>
<td>Paper</td>
<td>Country and Program</td>
<td>Program Type</td>
<td>Research Design</td>
<td>Summary of Findings</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Alzua, Cruces, and Ripani (2013)</td>
<td>Nicaragua, RPS</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on overall probability of work or hours of work, for men or women. For hours of work, large but statistically insignificant point estimates between $-1.5$ and $-2.7$ hours for men and between $-4$ and $-5.7$ hours for women. No detectable effect on agricultural employment.</td>
</tr>
<tr>
<td>Hasan (2010)</td>
<td>Pakistan, Punjab</td>
<td>CCT</td>
<td>Difference-in-difference</td>
<td>Decrease in time spent on paid work by 24-32 minutes, from a base of 47 minutes per day. Significant increase in the amount of housework by 100-120 minutes, from a base of 600 minutes per day. Effect on total work not reported.</td>
</tr>
<tr>
<td>Chaudhury et al. (2013)</td>
<td>Philippines, Pantawid Pamilya Program (PPP)</td>
<td>CCT</td>
<td>RCT</td>
<td>No detectable effect on work probability or hours of work.</td>
</tr>
<tr>
<td>Amarante et al. (2011)</td>
<td>Uruguay, PANES</td>
<td>UCT</td>
<td>Regression Discontinuity</td>
<td>The program reduces formal earnings. Data on informal work not available.</td>
</tr>
<tr>
<td>American Institutes for Research (2013)</td>
<td>Zambia, Child Grant Program</td>
<td>UCT</td>
<td>RCT</td>
<td>Significant decrease in wage labor, compensated by increase in participation in non-farm enterprises and labor on household farms.</td>
</tr>
</tbody>
</table>
experimental or natural-experiment-based research design. In terms of geographic cover, thirteen studies are from Latin America, four are from Africa, one is from South Asia, one is from China, and two are from South-East Asia.

Overall, these studies suggest little to no effects on overall labor supply. From among the fourteen studies with data on overall working probability or hours of work, nine do not find any significant effect, two find a combination of positive and null results, two find only negative results, and one finds a combination of positive and negative effects. For eight studies, we do not have explicit results on overall work probability or hours of work.

Those studies that do find an effect tend to find effects on the type of work done, rather than the total amount of work. For example, several studies have documented a shift from formal to informal labor for programs that explicitly exclude formal workers. Levy (2006), among others, argued that transfers targeted at informal workers discourage formalization. Evidence from Bolsa Família in Brazil, the Plan de Atención Nacional a la Emergencia Social (PANES) program in Uruguay, and the Universal Child Allowance in Argentina supports this hypothesis (Foguel and Paes de Barros 2008; Ribas and Soares 2011; de Brauw et al. 2015; Amarante et al. 2011; Garganta and Gasparini 2015). These studies find a reduction in formal work; when data is available, they also find no overall effect on work.

Several studies also document shifts away from work outside the household toward work within the household. Galiani and McEwan (2013) find a small switch to within-household work for men due to the PRAF program. Skoufias, Unar, and González-Cossío (2008) identify a switch from agricultural to non-agricultural work for the PAL program in Mexico.

Two studies in African countries find similar patterns of reductions in wage labor, together with increases in self-employed activities (Covarrubias et al. 2012; American Institutes for Research 2013). Hasan (2010) finds that a conditional cash transfer program in Pakistan decreased the time spent by mothers on paid work, while significantly increasing the amount of housework. Asfaw et al. (2014) also finds a large decrease in wage work, especially for men; nevertheless, there is little evidence of a compensatory increase in within-household work, especially for men.

Data, Empirical Strategy, and Sample Statistics

We now turn to systematically re-analyzing the labor supply effects of government-run transfer programs that have previously been experimentally evaluated. In this section, we first describe the data and then detail our empirical strategy. In the last sub-section, we provide sample statistics to provide a descriptive picture of each program area.
We began by identifying randomized evaluations of cash transfer programs in low-income and emerging nations. For a study to be included, it needed to have both a pure control group and at least one treatment arm of a conditional or unconditional cash transfer program.\footnote{10}

In total, we identified 18 randomized control trials that met the above criteria.\footnote{11} Of these, three were excluded because they did not include variables on both male and female adult labor supply in the public datasets,\footnote{12} three were excluded because the evaluated programs were not run by the government,\footnote{13} two were excluded due to baseline imbalance caused by a small number of clusters or different sampling in the control and treatment groups,\footnote{14} and we have been unable to obtain data for another three studies.\footnote{15} Online Appendix table 1 lists these excluded studies.

Therefore, we included seven RCTs in this analysis: Honduras’ Programa de Asignación Familiar II (PRAF II), Morocco’s Tayssir, Mexico’s Progresa and Programa de Apoyo Alimentario (PAL), Philippines’ Pantawid Pamilyang Filipino...
Program (PPPP), Indonesia’s Program Keluarga Harapan (PKH), and Nicaragua’s Red de Protección Social (RPS). A notable characteristic of all seven programs is that they are implemented by national governments (as opposed to NGOs) either as pilot or expansion programs, and thus are representative of “real-world” cash transfers. Figure 2 provides some details about the programs and evaluation data and provides references to key academic papers for each program (Online Appendix 2 provides additional information on the data).

In terms of program type, most of the programs that we include are conditional cash transfer (CCTs), where benefits are “conditional” on desirable social behaviors, such as ensuring that the recipient’s children attend school and get vaccinated. The two exceptions were: (1) Mexico’s PAL program, where benefits were not conditioned on behaviors, and (2) Morocco’s Tayssir program, which had two treatment arms consisting of a CCT and a “labeled” cash transfer in which the conditions were recommended but were explicitly not enforced. In general, it is important to note that there is considerable variation in how stringent conditions are enforced across countries, so that even in programs that are conditional “on the books”, beneficiaries may still receive the full stipend amount regardless of whether they meet them.

A first challenge in these types of programs is finding the poor (“targeting”). Unlike developed countries, where program eligibility can be verified from tax returns or employment records, developing country labor markets often lack formal records on income and employment, and thus alternative targeting methods must be used (see Alatas et al. 2012, for a description). For all of the programs in our study, regions were first geographically targeted based on some form of aggregate poverty data. After that, in five out of the seven programs, eligibility was determined by a demographic criterion (e.g., a woman in the household was pregnant or there were children below an age cutoff) and/or an asset-based means test (e.g., not owning land over a certain size).

Once a household becomes eligible for any of the programs that we study, the amount of benefit that one receives is the same regardless of actual income level and lasts at least a period between two and nine years, depending on the program. This differs from many U.S. transfer programs (e.g., Earned Income Tax Credit (EITC), Supplemental Nutrition Assistance Program (SNAP)), where the stipend depends (either positively or negatively) on family income, and is updated frequently. This discrepancy likely stems from the greater difficulty in ascertaining precise income levels in data-poor environments. However, similar to the U.S programs, the level of the transfer received was determined, at least in part, by the number of children in the family and their ages. On net, the programs were fairly generous, ranging from 4 percent of household consumption (Honduras’ PRAF II) to about 20 percent (Mexico’s Progresa), though all were intended to supplement other sources of income, rather than providing sufficient income that a household could subsist on the transfer alone.
For each evaluation, we obtained the raw evaluation micro-datasets from either online downloads or personal correspondence with the authors. Two features of the evaluation design affect the analysis. First, all of the studies that we consider are clustered-randomized designs, that is, the program was randomized over locations rather than individuals. Thus, in the analysis below, we cluster our standard errors by the randomization unit. Second, we obtained both baseline and endline data for five of the studies. Baseline data were not collected for the Philippines’ PPPP. Moreover, the baseline data for the treatment group of the Honduras’ PRAF II study was collected in a different agricultural season than for the control group (Glewwe and Olinto 2004). Alzua, Cruces, and Ripani (2013) point out that this leads to a small but statistically significant imbalance in labor supply between the two groups and, therefore, we decided not to use the baseline for this program. Therefore, as we discuss below, we use a different empirical strategy for the programs with baseline data and those without.

While some of the studies explored impacts on some of the work variables, the sample composition and work variable definitions varied across the studies. We therefore harmonized the datasets in several ways. First, we aimed to restrict our datasets to include all adult males and females aged 16 to 65 from eligible households. We have two exceptions to this, where we included adults in all surveyed households (regardless of eligibility status): First, Nicaragua’s RPS contains a random sample of households. About 6 percent of households were excluded from the cash transfer program based on a proxy means test, but we cannot identify them in the data. Second, Honduras’ PRAF II has a random sample from households in the geographically-targeted areas; we attempted to code the eligibility rules within the evaluation dataset, but did not feel fully confident in our ability to identify eligible households and thus include all individuals.

Next, for these samples we coded consistent variables for employment status and hours worked per week for each included individual. Importantly, our sample includes all individuals, regardless of whether or not they are in the labor force. Thus, if the cash transfer programs induce individuals to exit the labor force, this will be captured by our employment variable. Similarly, individuals who do not work are counted as “zero” hours of work in our analysis; thus, this variable is capturing both the decision to work (extensive margin) and the number of hours worked (intensive margin). Note that we lack information on hours of work for Indonesia’s PKH program, so it is only included in the analysis on employment status.

In the poor areas where the programs that we analyze are located, a significant share of people work in agriculture (in rural areas) or in self-employment. We include both these activities in the employment status, and we later analyze two outcome variables that differentiate between household work (any self-employed activity) and work outside the household (casual or permanent employment).
Empirical Strategy

We begin our analysis by first estimating the effect of being randomized to receive a transfer program on labor market outcomes, estimating the following regression:

\[ y_{ic} = \beta \text{Treat}_c + \mu_{s(c)} + \gamma \cdot X_{ic} + \epsilon_{ic} \]  

(1)

where \( i \) is an individual in cluster (randomization unit) \( c \), \( y_{ic} \) is individual \( i \)'s labor market outcome, either an indicator variable that takes the value of one if the individual is employed or a continuous variable on the hours an individual worked per week. Further, \( \text{Treat}_c \) is an individual variable that equals one if the individual was randomly assigned to the treatment group, and zero otherwise; \( \beta \) is the parameter of interest, providing the difference in work outcomes between the treatment and the control group. Given the randomization, the treatment and control groups should be similar along observable and unobservable baseline characteristics. Thus, \( \beta \) provides the causal estimate of the program on work outcomes.

Note two features of the specification. First, while the randomization should ensure that \( \beta \) captures the causal program impacts, we can include additional control variables to improve our statistical precision. Specifically, we include strata fixed effects (\( \mu_{s(c)} \)) and a number of individual-level control variables, (\( \gamma \cdot X_{ic} \)), including age, age squared, household size, years of education, and marital status dummies. For each control variable, we code missing values at the variable mean and include a dummy variable that indicates the observations with missing values. Standard errors are clustered at the randomization unit level.

We run this basic specification for the two programs for which we do not have reliable baseline data (Philippines’ PPPP and Honduras’ PRAF II). For the other five programs, we can take advantage of the fact that baseline data were also collected. Specifically, we stack the individual baseline and endline data and estimate the following difference-in-difference specification:

\[ y_{ict} = \mu_c + \text{Post}_t + \beta (\text{Treat}_c \times \text{Post}_t) + \gamma \cdot X_{ict} + \epsilon_{ict} \]  

(2)

where \( i \) is an individual in cluster \( c \) at time \( t \). While the randomization implies that equation (1) would provide a causal estimate of the program effect, the difference-in-difference specification allows us to better control for any baseline imbalances between the treatment and control group and thus provides us with even greater statistical precision. We now include the randomization unit fixed effects, \( \mu_c \), and all of the same control variables as before, and continue to cluster our standard errors at the randomization unit.\(^{18}\) The parameter of interest is again \( \beta \), which provides the difference in work outcomes across the treatment and control relative to their baseline values and conditional on our control variables.
A benefit of harmonizing and re-analyzing the various micro-datasets is that we can pool the data across studies and estimate an underlying treatment effect. This allows us to potentially generate tighter statistical bounds than would be possible from any one study, which is important if we want to try to identify a real zero—or very small effect—from just noise in the data. If cash transfers have the same impact across programs, then ordinary least squares analysis on the pooled data weighs the data optimally to estimate the underlying (universal) treatment effect.

However, it is unlikely that programs across different countries and contexts have the same effect, so our pooling approach needs to model this possibility explicitly. Therefore, we use a Bayesian hierarchical model to aggregate the results from the seven studies (Rubin 1981; Meager 2016). In this model, the treatment effect $\tau_p$ in program $p$ is allowed to vary across programs. Treatment effects corresponding to different programs are nevertheless related by a “parent distribution;” specifically, each $\tau_p$ is drawn iid from a normal distribution with mean $\tau$ and standard deviation $\sigma_\tau$, $\tau_p \sim N(\tau, \sigma_\tau)$. We aim to estimate the unknown parameters $\tau$ and $\sigma_\tau$ that describe the parent distribution; $\tau$ captures the mean treatment effect across the programs, and

### Table 2: Descriptive Statistics for Non-Program Areas

<table>
<thead>
<tr>
<th></th>
<th>Honduras PRAF (1)</th>
<th>Morocco Tayssir (2)</th>
<th>Philippines PPPP (3)</th>
<th>Mexico PAL (4)</th>
<th>Indonesia PKH (5)</th>
<th>Nicaragua RPS (6)</th>
<th>Mexico Progresa (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Work Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worked last week</td>
<td>0.59</td>
<td>0.63</td>
<td>0.56</td>
<td>0.52</td>
<td>0.61</td>
<td>0.55</td>
<td>0.48</td>
</tr>
<tr>
<td>Worked for Self/Family</td>
<td>0.42</td>
<td>0.51</td>
<td>0.26</td>
<td>0.17</td>
<td>0.26</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>Worked Out of HH</td>
<td>0.26</td>
<td>0.16</td>
<td>0.29</td>
<td>0.27</td>
<td>0.29</td>
<td>0.29</td>
<td>0.38</td>
</tr>
<tr>
<td>Hours/Week</td>
<td>19.80</td>
<td>20.86</td>
<td>22.73</td>
<td>21.63</td>
<td>23.63</td>
<td>23.63</td>
<td>17.87</td>
</tr>
<tr>
<td>Observations</td>
<td>4,174</td>
<td>2,757</td>
<td>2,293</td>
<td>3,567</td>
<td>20,246</td>
<td>4,183</td>
<td>53,226</td>
</tr>
<tr>
<td><strong>Panel B: Work Outcomes for Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worked last week</td>
<td>0.90</td>
<td>0.85</td>
<td>0.72</td>
<td>0.80</td>
<td>0.82</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td>Worked for Self/Family</td>
<td>0.67</td>
<td>0.63</td>
<td>0.31</td>
<td>0.30</td>
<td>0.46</td>
<td>0.46</td>
<td>0.10</td>
</tr>
<tr>
<td>Worked Out of HH</td>
<td>0.38</td>
<td>0.32</td>
<td>0.39</td>
<td>0.46</td>
<td>0.47</td>
<td>0.47</td>
<td>0.70</td>
</tr>
<tr>
<td>Hours/Week</td>
<td>31.70</td>
<td>34.29</td>
<td>29.51</td>
<td>35.80</td>
<td>39.51</td>
<td>39.51</td>
<td>34.56</td>
</tr>
<tr>
<td>Observations</td>
<td>2,132</td>
<td>1,272</td>
<td>1,215</td>
<td>1,647</td>
<td>10,198</td>
<td>2,131</td>
<td>25,850</td>
</tr>
<tr>
<td><strong>Panel C: Work Outcomes for Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worked last week</td>
<td>0.27</td>
<td>0.44</td>
<td>0.38</td>
<td>0.27</td>
<td>0.39</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>Worked for Self/Family</td>
<td>0.16</td>
<td>0.42</td>
<td>0.19</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>Worked Out of HH</td>
<td>0.13</td>
<td>0.02</td>
<td>0.18</td>
<td>0.10</td>
<td>0.11</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>Hours/Week</td>
<td>7.37</td>
<td>9.49</td>
<td>15.09</td>
<td>9.72</td>
<td>6.96</td>
<td>6.96</td>
<td>3.66</td>
</tr>
<tr>
<td>Observations</td>
<td>2,042</td>
<td>1,483</td>
<td>1,078</td>
<td>1,920</td>
<td>10,048</td>
<td>2,052</td>
<td>27,305</td>
</tr>
</tbody>
</table>

**Note:** This table reports descriptive statistics from the control group at endline. Panels A, B, and C restrict the sample respectively, to all adults, men, and women, between 16 and 65 years old. The binary work indicator is equal to 1 if the respondent reported working during the last week (last 30 days for Morocco Tayssir); the other work variables are reported for the same time frame.
\( \sigma \) captures the dispersion in the treatment effects. Intuitively, the hierarchical model allows the data to speak about the degree of similarly of the impacts across programs, while also reaping the benefits of improved precision from pooling the data. Details of this procedure can be found in Online Appendix 1.

Descriptive Picture

Table 2 provides descriptive statistics for the standardized work variables across the studies, using data from the control group at endline to show work outcomes in the absence of the program. Many of the program recipients would have worked in the absence of the program. Pre-program employment ranged from 48 percent in Mexico Progresa to 63 percent in Morocco, with a weighted mean of 56 percent across all programs. This figures includes all adults aged 16 to 65, including those not in the labor force due to being in school, having a disability, or being retired, and thus includes people who would likely not change their status, regardless of the presence of cash transfers. Across everyone regardless of employment status, we observe about 20 hours of work per week, implying about a 40-hour work week for those who are employed.

However, these means mask considerable heterogeneity in work patterns. First, male employment rates are high, with a weighted average of about 84 percent. In contrast, female employment rates tend to be much lower, ranging from 12 percent in Mexico Progresa to 44 percent in Morocco. Second, work outcomes tend to be split between self-employment/family work and outside work, with some exceptions: men in Honduras and both men and women in Morocco tend to be more engaged in work inside the house, while men in Mexico’s Progresa program tend to be more engaged in outside work.

Do Cash Transfers Reduce Work?

Overall Findings

Figure 3 provides a graphical summary of our main findings. In panel A, we graph the employment rate for all eligible adults in both the control and treatment arms for each evaluation. The evaluations are listed in order from the least generous in terms of benefits relative to consumption levels (Honduras’ PRAF) to the most generous (Nicaragua’s RPS and Mexico’s Progresa). Panel B replicates panel A, but for hours of work. The graphs suggest that the overall numbers for both employment rate and hours of work are similar across the treatment and control groups across all of the programs.
Table 3 provides the corresponding regression analysis underlying figure 3. Panel A presents the analysis for the binary employment outcome for each individual program, while panel B does so for hours of work per week. Remember that the hours of work variable captures both intensive and extensive work decisions, thereby providing the treatment effect on total work activity.

Consistent with figure 3, we do not observe a significant effect of belonging to a transfer program on employment in six of the programs (panel A). We only find an impact in one program: in Honduras—the least generous program—we find a 3 percentage point decrease in probability of work that is significant at the 10 percent level; note that when analyzing multiple coefficients, this is roughly what we may expect by pure chance. Panel B also shows no effect on hours worked per week: none of the individual coefficients are significant, even in the Honduras data where we observed a decrease in employment status.

Note: The “Control” (light, left) bars report the mean of the outcome variable (probability of work and hours worked in Panels A and B, respectively) in the control group, at endline. The “Treatment” (dark, right) bars report the control mean plus the treatment effect from in Table 3. The segments represent 95% confidence intervals.
Even if overall labor force participation did not change, the type of work that households participate in could change as a result of the transfers. In particular, households may choose not to work outside the household due to fears that this form of employment could disqualify them from receiving benefits, regardless of whether this fear is rational or irrational according to program rules. Therefore, in Table 4, we disaggregate work type by whether the work is self-employed/within the household (panel A) or outside of the household (panel B). We do this for all programs, except Indonesia’s PKH, where the disaggregated data do not exist.

Table 3. Experimental Estimates of the Impact of Cash Transfer Programs on Work Outcomes

<table>
<thead>
<tr>
<th>Program</th>
<th>Treatment Effect</th>
<th>Control Group Mean</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honduras PRAF</td>
<td>0.0295*</td>
<td>0.59</td>
<td>endline</td>
</tr>
<tr>
<td>Morocco Tayssir</td>
<td>0.0097</td>
<td>0.63</td>
<td>DD</td>
</tr>
<tr>
<td>Philippines PPPP</td>
<td>0.0096</td>
<td>0.56</td>
<td>endline</td>
</tr>
<tr>
<td>Mexico PAL</td>
<td>0.0135</td>
<td>0.52</td>
<td>DD</td>
</tr>
<tr>
<td>Indonesia PKH</td>
<td>0.0043</td>
<td>0.61</td>
<td>DD</td>
</tr>
<tr>
<td>Nicaragua RPS</td>
<td>0.0202</td>
<td>0.61</td>
<td>DD</td>
</tr>
<tr>
<td>Mexico Progresa</td>
<td>0.0089</td>
<td>0.55</td>
<td>DD</td>
</tr>
</tbody>
</table>

Panels A and B report the coefficients on Treatment x Follow-up for difference-in-difference (DD), and the coefficient on Treatment otherwise. Controls are age, age squared, years of education, marital status dummies (single, married or with partner, divorced or separated, and widow), household size, and survey wave fixed effects, as well as dummies for missing values for each control variable. Columns 1 and 3 include randomization strata fixed effects, columns 2, 4, 5, 6, and 7 include randomization unit (village) fixed effects. The sample is all adults between 16 and 65 years old, excluding domestic workers. The Control Group Mean reports the mean of the panel variable in the control group, at endline. Standard errors clustered at the randomization unit level are reported in round parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Even if overall labor force participation did not change, the type of work that households participate in could change as a result of the transfers. In particular, households may choose not to work outside the household due to fears that this form of employment could disqualify them from receiving benefits, regardless of whether this fear is rational or irrational according to program rules. Therefore, in Table 4, we disaggregate work type by whether the work is self-employed/within the household (panel A) or outside of the household (panel B). We do this for all programs, except Indonesia’s PKH, where the disaggregated data do not exist.

No clear systematic patterns emerge. In the four programs that had the least generous benefits (columns 1–4), we find no statistically observable impacts on either type of work. We find an increase in outside work and an associated decrease in within household work in Mexico’s Progresa program, but the opposite pattern holds for Nicaragua’s RPS program (which has a similar transfer size).

Finally, we consider men and woman separately, given the differences in baseline labor force participation. It is not clear ex ante whether we would expect
larger effects for men or women. For example, the additional income may allow a woman who previously had to work the ability to choose to stay home with the children if she prefers, or the additional income may make it possible for her to afford additional child care and actually work more. Moreover, the literature often paints a picture of the lazy male, who uses transfer stipends to shirk and instead waste money on cigarettes and alcohol, and thus it is important to understand if these stereotypes are borne out in the data.

Table 5 replicates table 3, but disaggregates by gender. Panels A and B report results on employment for men and women, and panels C and D report results on hours for the two groups. The impact of the cash transfer programs on men’s labor supply is only significantly different from zero in one program (Philippines), where it is positive. However, overall hours worked do not significantly change. For women, the impact is only significantly different from zero in one program (Honduras PRAF), where it is negative. However, none of the programs significantly affected hours worked. We also disaggregate the gender results by whether work is conducted within or outside the household (Online Appendix table 4). For men, we find a shift from working outside to inside the household in Nicaragua, but we find the exact opposite for Progresa. For women, we find slightly lower rates of working within the household in two of the six programs (Philippines PPPP and Mexico Progresa), and similarly lower rates of working outside the household in two programs (Honduras PRAF and Morocco Tayssir).

Table 4. Experimental Estimates of the Impact of Cash Transfer Programs on Household and Private Market Work Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Honduras PRAF</th>
<th>Morocco Tayssir</th>
<th>Philippines PPPP</th>
<th>Mexico PAL</th>
<th>Indonesia PKH</th>
<th>Nicaragua RPS</th>
<th>Mexico Progresa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
</tr>
<tr>
<td>Panel A. Worked in household</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Effect</td>
<td>0.0203</td>
<td>-0.0073</td>
<td>-0.0197</td>
<td>0.0054</td>
<td>–</td>
<td>0.0263**</td>
<td>-0.0235**</td>
</tr>
<tr>
<td>(0.0190)</td>
<td>(0.0252)</td>
<td>(0.0191)</td>
<td>(0.0168)</td>
<td>–</td>
<td>(0.0130)</td>
<td>(0.0096)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>8.483</td>
<td>29.832</td>
<td>4.527</td>
<td>15.598</td>
<td>–</td>
<td>12.979</td>
<td>182.533</td>
</tr>
<tr>
<td>Control Group Mean</td>
<td>0.42</td>
<td>0.51</td>
<td>0.26</td>
<td>0.17</td>
<td>–</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>Panel B. Worked outside the household</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Effect</td>
<td>-0.0335</td>
<td>-0.0035</td>
<td>0.0299</td>
<td>-0.0131</td>
<td>–</td>
<td>-0.0465*</td>
<td>0.0191*</td>
</tr>
<tr>
<td>(0.0254)</td>
<td>(0.0143)</td>
<td>(0.0197)</td>
<td>(0.0178)</td>
<td>–</td>
<td>(0.0235)</td>
<td>(0.0103)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>8.486</td>
<td>29.832</td>
<td>4.527</td>
<td>15.598</td>
<td>–</td>
<td>12.979</td>
<td>182.533</td>
</tr>
<tr>
<td>Control Group Mean</td>
<td>0.26</td>
<td>0.16</td>
<td>0.29</td>
<td>0.27</td>
<td>–</td>
<td>0.29</td>
<td>0.38</td>
</tr>
<tr>
<td>Method</td>
<td>endline</td>
<td>DD</td>
<td>endline</td>
<td>DD</td>
<td>DD</td>
<td>DD</td>
<td></td>
</tr>
</tbody>
</table>

Note: This table reports regression results of the impact of cash transfers on a dummy for working for self/family (panel A) and on a dummy for working outside the household (panel B). See table 3 notes for specification details. ***p < 0.01, **p < 0.05, *p < 0.1.
Pooling the Results

Table 5 reports the results for work outcomes from pooling the results for the seven programs using the Bayesian hierarchical model described above. In pooling the programs, to make them comparable we scale the estimated treatment effect for each program by the size of the transfer. The presented coefficients correspond to the impact of a hypothetical new cash transfer program worth 13.6 percent of household consumption, which is the average transfer size across the programs. Columns (2)–(4) provide effects on the work outcomes in levels. Columns (5)–(7) report the implied elasticities from the estimates in columns (2)–(4).

The pooled estimates further confirm little program impact on work. First, the estimated impact on the extensive margin decision to work in panel A is a decrease of 0.4 percentage points from a base of 56 percent. In fact, with 95
percent probability, a new program has an impact no lower than a 2.3 percentage points reduction in work status. Conversely, with 5 percent probability a new program will tend to increase work status by at least 1.4 percentage points. Similarly, for hours of work, the point estimate corresponds to a decrease of five minutes of work per week, from a base of 21 hours. With 95 percent probability, a new program will not reduce hours of work by more than 1 hour and 42 minutes per week.

In terms of elasticities, the estimates in columns (5)–(7) indicate that on average, a new program worth 10 percent more of household consumption will tend to reduce work status by 0.6 percent, and with 95 percent probability this effect will not be lower than a 3 percent decrease in work. For hours of work, on average such a program will tend to reduce work by 0.3 percent, and with 95 percent probability this effect is no lower than a 6 percent reduction in hours. These effects are broadly symmetric around zero, offering very little evidence of a negative impact of cash transfers on work outcomes.23

Looking at results separately by gender, for men the average effects are positive and more precise. Indeed, the results show a 0.1 percentage point increase in work status, and a positive elasticity of +0.01, while with 95 percent probability the

---

<table>
<thead>
<tr>
<th>Statistic of the posterior distribution:</th>
<th>(1) Weighted control mean</th>
<th>(2) Effect size ($\tau_p$)</th>
<th>(3) Mean 5th percentile</th>
<th>(4) 95th percentile</th>
<th>(5) Mean 5th percentile</th>
<th>(6) 95th percentile</th>
<th>(7) Mean 5th percentile</th>
<th>(8) 95th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Worked last week</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full sample:</td>
<td>0.56</td>
<td>−0.004</td>
<td>−0.023</td>
<td>0.014</td>
<td>−0.06</td>
<td>−0.30</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>For Men:</td>
<td>0.84</td>
<td>0.001</td>
<td>−0.020</td>
<td>0.026</td>
<td>0.01</td>
<td>−0.17</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>For Women:</td>
<td>0.29</td>
<td>−0.008</td>
<td>−0.039</td>
<td>0.024</td>
<td>−0.21</td>
<td>−0.99</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Panel B. Hours worked per week</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full sample:</td>
<td>21.1</td>
<td>−0.077</td>
<td>−1.734</td>
<td>1.356</td>
<td>−0.03</td>
<td>−0.60</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>For Men:</td>
<td>34.2</td>
<td>0.470</td>
<td>−1.702</td>
<td>2.965</td>
<td>0.10</td>
<td>−0.37</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>For Women:</td>
<td>8.7</td>
<td>−0.430</td>
<td>−2.588</td>
<td>1.611</td>
<td>−0.36</td>
<td>−2.18</td>
<td>1.36</td>
<td></td>
</tr>
</tbody>
</table>

Note: This table reports results from a Bayesian hierarchical model used to aggregate the results from the seven programs. The impact for each program from Table 3 or Table 5 is first scaled according to the size of the transfer, such that for each program the scaled coefficient corresponds to a transfer worth 13.6% of consumption. (The program transfer size is defined as the average transfer value relative to average consumption.) Column (1) reports the mean of the row variable in the control group at endline, averaged over the seven programs. Columns (2)-(4) present the mean, and the 5th and 95th percentiles of the posterior distribution of the site effect $\tau_p$, which measures the impact for a hypothetical new program. Columns (5)-(7) report the same statistics for the elasticity of the work outcome with respect to the size of the cash transfer. Bayesian posteriors are computed using the rstan package, 20,000 iterations on four chains, thinning the result by a factor of two.
impact in a new program will not reduce work by more than 2 percentage points, and the elasticity will not be lower than \(-0.17\). We further find a half hour increase per week due to cash transfers in panel B, and a positive elasticity of \(+0.10\). Once again, we can reject moderate negative effects with a high probability.

For women, the average effects are negative but small, corresponding to a 0.8 percentage point decrease in work status and half an hour of work less per week. Due to the low mean of these work outcome variables in the control group for women, the implied elasticities are moderately negative, between \(-0.2\) and \(-0.36\). However, the Bayesian meta-analysis points to significant uncertainty in the impact of a new cash transfer program for women, with estimates for work status in columns (3)–(4) and (6)–(7) between a 3.9 percentage point reduction and a 2.4 percentage point increase, and an elasticity between \(-1\) and \(+0.6\). Similar results for hours worked indicate that the existing data covers a high range of effects for women.

Understanding Mechanisms: Exploring the “Tax” Rate

As described above, transfer programs can have a negative effect on work for two reasons: (1) the income effect, and (2) individuals choosing to work less in fear of losing their benefits (“the tax rate” or “benefit withdrawal rate”). As we found little evidence of a systematic negative effect of the transfer programs across all of the countries that we examined, we now test to see whether this is rational given the expected “tax rates” of these programs.

To examine the tax rate, it is important to examine two aspects of the program. First, consider how individuals are added and subtracted from the list (“targeting”). In developed countries, programs are targeted based on income measured from administrative sources and recertified frequently. In contrast, obtaining frequent or real-time information about income is challenging in developing countries, and so targeting is often conducted infrequently through alternatives methods—proxy means tests, geographic targeting, etc. Bosch and Manacorda (2012), Grosh et al (2008), and Alderman and Yemtsov (2013), among others, have argued that the fact that targeting is less connected to current income suggests that taxes are low, and therefore, these programs are less likely to cause negative labor supply effects. Second, it is important to understand the size of the transfers. For example, Alderman and Yemtsov (2013) argue that the size of the transfer programs is often insufficient to live from, and thus, a small gain of income from the program is not enough to keep people out of the workforce.

Turning to the programs we consider, the way the targeting rules are designed suggests that the tax is, if anything, very small, since eligibly is rarely based directly on current observable income. In two out of the seven programs (Morocco Tayssir and Mexico PAL), targeting is purely geographic, meaning that everyone
within a chosen region received the program. This implies that any individual’s behavior is likely not to affect the probability of their receipt and thus the implied tax rate on labor income is effectively zero. Similarly, the Hondurans PRAF selects beneficiaries within geographically-targeted regions if the household includes a pregnant woman or children under age three, and so eligibility is not driven by work status. In the Nicaragua RPS, after the geographic targeting, a small fraction of households (6 percent) were excluded based on a simplified asset test, and thus most households are not going to lose eligibility status if they work more. In short, for about half of the programs, eligibility is not directly related to current employment or income, effectively implying no tax. Thus, one would expect close to no labor supply effects unless income effects were unusually large.

In the remaining three programs (Philippines PPPP, Indonesia PKH, and Mexico Progresa), beneficiaries are selected based on a full-fledged asset test (proxy means test or PMT). For two of these studies (Indonesia PKH and Mexico Progresa), we can examine the perceived implicit tax rate with respect to consumption by graphing the relationship between the expected total transfer for households at different consumption levels. The slope of the relationship represents the perceived, implicit tax rate with respect to consumption. Note that in so doing, we assume that households know exactly when they will be assessed for targeting purposes, so we assign as the ‘cost’ of working more the potential loss of the full net present value of the program for all the years they would then receive it.24

We document a weak relationship for both programs in figure 4, as households with higher consumption have only marginally lower total expected transfer size. Starting with Indonesia’s PKH, a household with Rp. 1,000 higher per capita annual consumption will receive in expectation Rp. 40 less in net present value transfers, calculated over a period of six years.25 This is not surprising, as only 4.5 percent of households receive the cash transfer, and among recipients the transfer is on average 10 percent of consumption.26 These factors attenuate the relationship between consumption and expected transfers.

In the case of Progresa, the fraction of eligible is higher (60 percent), the cash transfer is a larger fraction of household consumption (25 percent), and households receive benefits for nine years. Nevertheless, the implied tax rate is 15 percent.27 Thus, even for large transfers that cover about half the population, imprecise targeting attenuates the relationship between poverty and expected transfers.

In short, the targeting rules of these programs, coupled with the size of the transfers, provides one reason why we do not observe systematic negative effects of the transfer programs across the differing settings. Our findings on the implicit tax rates in Indonesia and Mexico echo Ravallion and Chen (2013), who measure the tax rate imposed by the Chinese Di Bao cash transfer program. The largest estimate that they find is 15 percent, much lower than the theoretical 100
percent tax rate implied by its goal of providing a means tested guaranteed minimum income.

Comparison with Asset Transfer Programs

Our analysis has focused on cash transfers programs that provide small amounts of money either monthly or quarterly to poor households. However, a policy alternative to cash transfers is an asset transfer program, which is typically a one-time intervention where the beneficiary receives a productive asset (or money to buy such an asset), with the idea that they will benefit from the asset’s future income stream. The labor supply effect of an asset program could be quite different from that of a cash transfer because it is a lump sum or a lumpy asset (e.g., livestock or tools for a business), an amount of which savings market failures might prevent households from accumulating from the transfer funds. If it is a productive asset that requires

Notes: Each panel displays a local linear regression of the expected transfer per capita as a function of consumption per capita (solid line), and a histogram of consumption per capita (bars). Panel (A) reports results from the PKH program in Indonesia using SUSENAS 2013 data in all districts where at least 1% of respondents report being PKH beneficiaries (N = 101,568); we coded the transfer size (in 2013) for each household, depending on whether they report receiving PKH, and on the number of children and their ages. The total transfer is the net present value of transfers over six years, assuming a discount factor δ = 0.9. Panel (B) reports results from the Progresa program in Mexico, using data from the October 1998 ENCEL survey (which is included in our main results). We use the eligibility variable together with the average per adult equivalent transfer size of 32.5 pesos per month reported in Angelucci and De Giorgi (2009). The total transfer is the net present value of transfers over nine years, assuming a discount factor δ = 0.9. We use per adult equivalent consumption from the same study. In both panels, we drop the top 5% of the consumption distribution.
complementary household labor to use, the presence of the asset would quite natu-
rally encourage additional work effort. Labor supply could also increase if the
household combines the lump sum with a loan to purchase a consumer durable
that complements the asset, but then needs to work harder to pay down the loan.

We can, thus, qualitatively compare the effects of cash transfers with these asset
programs. One version of the program is the so-called graduation model, developed
by BRAC in Bangladesh. Under this model, households, chosen for being the poor-
est members of poor communities, are given an asset of their choosing (from a set
of affordable assets) as well as some training and support, including a small income
stipend for a period of no more than six months. An RCT of this program by
Bandiera et al. (2013) reports, “After four years, eligible women work 170 fewer
hours per year in wage employment (a 26% reduction relative to baseline) and
388 more hours in self-employment (a 92% increase relative to baseline). Hence
total annual labor supply increases by an additional 218 hours which represents
an increase of 19% relative to baseline.” Another RCT by Banerjee et al. (2015) of
this program in six different countries (Ethiopia, Ghana, Honduras, India,
Pakistan, and Peru), reports that total labor supply across the six sites went up by
10 percent of the control group mean (or about 85 hours a year), two years after
the start of the program. Consistent with this, both the Bangladesh study and the
multi-country study also find increases in income and consumption of commensu-
rate magnitudes in these households.

There is also evidence from a small number of lump sum cash transfer pro-
grams. Blattman et al. (2016) carry out a randomized evaluation of a program
where women in Northern Uganda—most of whom had never run a business
before—were given a package comprised of $150 in cash, five days of business
training, and ongoing supervision. These authors find that hours worked per week
go up by a stunning 10 hours and, correspondingly, there is a doubling of new non-
farm enterprises and a significant rise in income. Blattman, Fiala, and Martinez
(2014) evaluate the Youth Opportunities Program (YOP), a government program in
northern Uganda designed to help unemployed adults become self-employed arti-
sans. The government invited young adults to form groups and prepare proposals
for how they would use a grant to train in and start independent trades. Funding
was randomly assigned among 535 screened, eligible applicant groups. Successful
proposals received a one-time unsupervised grants worth $7,500 on average—
about $382 per group member, roughly their average annual income. After four
years the treatment group had 57 percent greater capital stocks, 38 percent higher
earnings, and 17 percent more hours of work than did the control group.

Perhaps not surprisingly, these programs have a strong and clear positive effect
on labor supply, in contrast with the more or less zero effect we find from the in-
come support style cash transfer programs. However, it is very important to note
two aspects of these programs. First, all of these programs combined assets (or

Banerjee et al. 177
cash for assets) with training and support, and so the evidence is not yet available as to whether supervision is needed to achieve these increases in work or just the asset transfer would be enough. Moreover, it is likely that labor supply is a complementary input to the asset; for example, a cow or goat needs to be fed and taken care of. Future research is needed to disentangle the contributions of the various aspects of the programs. Second, in thinking through large-scale implementation across governments, physical assets (and in-kind transfers, in general) are often more expensive to distribute than cash. Moreover, we often observe leakages in the distribution of in-kind goods in many developing countries, with the goods never reaching program beneficiaries. New advances in technologies for distributing cash, such as mobile money, may make it easier to provide cash directly to beneficiaries with both potentially low leakage and low costs. Thus, research into understanding how large-scale physical asset distribution programs fare against these newer ways to distribute cash is also important for policy.

Conclusion

In recent years, there has been a large growth in transfer programs across the developing world. If anything, we might expect this trend to increase as countries grow: Chetty and Looney (2007) show that social insurance as a fraction of GDP rises as countries get richer, suggesting that safety nets may be increasingly important as countries grow and develop.

As transfer programs have increased, so has the debate about whether they simply discourage work, enabling a “lazy poor.” Aggregating evidence from randomized evaluations of seven government cash transfer programs, we find no systematic evidence of an impact of transfers on work behavior, either for men or women. Moreover, a 2014 review of transfer programs worldwide by Evans and Popova (2014) also shows no evidence—despite claims in the policy debate—that the transfers induce increases in spending on temptation goods, such as alcohol and tobacco. Thus, on net, the available evidence implies that cash transfer programs do not induce the “bad” behaviors that are often attributed to them in the policy space. Combined with the positive effects of transfer programs documented in the literature, this suggests that transfers can be an effective policy lever to help combat poverty and inequality.

Notes

Abhijit V. Banerjee, Gabriel E. Kreindler, and Benjamin A. Olken, Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142, USA; Rema Hanna, Harvard Kennedy School, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138, USA.
1. Note that this includes both in-kind and cash transfer programs.

2. See for example, the Ashenfelter and Plant (1990) analysis of the Seattle–Denver Maintenance Experiment, or Imbens, Rubin, Sacerdote (2001) estimates of the effect of unearned income on work from studying lottery winners.

3. This extends Alzua, Cruces, and Ripani (2013), which explores the program impacts on labor outcomes for three of the programs that we include. While we use slightly different specifications to harmonize across the full set of datasets that we include, our findings echo theirs.

4. Our sample covers countries from Latin America, Asia, and the Middle East. Unfortunately, randomized control trials for South Asia or for African countries do not exist, do not include labor supply information, or do not have publicly available data. This is an important area for future research to extend this type of analysis to these settings, which are on net lower income than the countries in our sample.

5. Note that households may not necessarily shift to “leisure,” but could shift to spending their time in productive ways. For example, a benefit of cash transfers could potentially be a reduction in child labor and a concurrent increase in the child’s education (see Behrman, Parker, and Todd 2011).

6. Evidence from developed countries tries to isolate this effect by looking at lottery winners. These studies generally find that the pure income effect on labor supply is modest (Imbens, Rubin, and Sacerdote 2001; Cesarini et al. 2015). In developing countries, Haushofer and Shapiro (2013) study a (non-governmental) large unconditional cash transfer in Kenya and do not detect any impact on total business profit or wage labor as primary income. Yang (2008) finds that in the Philippines there is no impact on aggregate household labor supply due to changes in remittances due to exchange rate shocks.

7. We have excluded studies of programs that contain explicit work requirements, such as India’s National Rural Employment Guarantee Act (NREGA) and Argentina’s Jefes y Jefas.


9. In India, cash transfers are extremely rare. Only 0.0035% of GDP goes to cash transfers (compared to 0.72% on social assistance in general), which ranks India below the 5th percentile in the ASPIRE dataset of 88 countries in terms of spending on cash transfers.

10. Some studies experimentally compare different ways of running a transfer program on recipients. While these provide valuable information on program design, they do not allow us to assess the full impact of introducing the program to begin with.

11. We apologize in advance if we have missed a particular study that meets our criterion. We tried to be as complete and systematic as possible.

12. Ecuador’s Bono de Desarrollo Humano (BDH) (Edmonds and Schady 2012; Schady and Caridad Araujo 2008), Nicaragua’s Atención a Crisis (Macours, Schady, and Vakis 2012), and Baird, McIntosh, and Ozler (2011) in Malawi.

13. Subsidios Condicionados a la Asistencia Escolar (SCAE) in Colombia (Barrera-Osorio et al. 2011), GiveDirectly in Kenya (Haushofer and Shapiro 2013), and a cash transfer for preschool in Uganda (Gilligan and Roy 2013).

14. Treatment status was randomized over eight communities in Malawi’s Social Cash Transfer Scheme (SCT) program (Covarrubias, Davis, and Winters 2012). Despite having a larger number of households, the small number of randomization units led to baseline imbalance on a number of indicators, and biases one towards not being able to measure a statistically significant effect unless the effect size is very large; therefore, we did not include this study. Sampling of beneficiaries and potential beneficiaries was done differently in treatment and control groups in Kenya’s CT-OVC (Asfaw et al. 2014), which led to large baseline imbalances.

15. The data for Tanzania’s Tanzania Social Action Fund (TASAF) (Evans, et al. 2014) is not yet available, and we were not able to obtain data for Burkina Faso’s Nahouri Cash Transfers Pilot...

16. Mexico’s PAL program also had an in-kind treatment, which we do not utilize for this analysis.

17. All programs except for Morocco’s ask about the number of hours worked during the last week. In Morocco, the reference period is the last 30 days, and we normalize the response by 7/30.

18. There are two additional differences across specifications. First, as Mexico’s Progresa includes three endline waves and Nicaragua’s RPS has two endline waves, we additionally include wave dummy variables in these specifications. Second, we weight observations in Morocco’s Tayssir to account for the sampling structure as in Benhassine et al. (2015).

19. We provide the control group statistics since we do not have baseline data for two of the programs and the definitions of work are not the same in the baseline and endline for one of the evaluations (Morocco’s Tayssir).

20. Appendix tables 2 and 3 report the baseline balance check by program, or in the case of the two programs without baseline, the balance on demographic characteristics at endline. With the exception of PAL and Progresa—for which the analysis in tables 3–6 uses the difference-in-difference specification—the joint significance tests do not reject balance.

21. Appendix figure 1 considers hours of work conditional on working status. The same pattern of results emerges.

22. These elasticities are $\frac{d \log \text{ProbabilityWork}}{d \log \text{Income}}$ and $\frac{d \log \text{HoursWorked}}{d \log \text{Income}}$ in panels A and B, respectively. To compute these elasticities, we take the estimated treatment effect in columns (2)–(4), divide by the mean of the outcome (probability work or hours worked) from column (1), and divide by the average increase in income due to the transfer (13.6 percent).

23. Online Appendix table 5 presents results for pooled results inside the household and work outside the household for men and women. The average impacts listed in column (2) are always close to zero, slightly positive for men and slightly negative for women. While there are a range of possible impacts on work (see columns 3 and 4), in all cases the zero effect is comfortably within the distribution of impacts. That is, there is no consistently negative effect of the transfer programs on work for any of the subgroups considered here.

24. An alternative assumption would be to assume that, ex-ante, households do not know in which year the targeting will take place. This assumption would yield effective tax rates that are 6 and 3-9 times smaller than the estimates reported here, for Indonesia and Mexico Progresa, respectively.

25. This is calculated over the steepest part of the graph in figure 3. We obtain essentially the same value when we include census area fixed effects.

26. This fraction is lower than the one reported in figure 1, because we use a different data source than for the main results, namely the SUSENAS national survey from 2013.

27. The fraction of eligible is larger than the one reported in figure 1, because we only used one of the follow up surveys (October 1998 ENCEL). Households are re-certified after three years, yet they continue to receive benefits for at least six more years. The implied tax rate with village fixed effects is 13 percent.

References


Banerjee et al.


The Impacts of Fiscal Openness

Paolo de Renzio and Joachim Wehner

Fiscal transparency and participation in government budgeting are widely promoted, yet claims about their benefits are rarely based on convincing evidence. We provide the first systematic review covering 38 empirical studies published between 1991 and early 2015. Increased budgetary disclosure and participation—which we call “fiscal openness”—are consistently associated with improvements in the quality of the budget, as well as governance and development outcomes. Only a handful of studies, however, convincingly identify causal effects, in the form of reduced corruption, enhanced electoral accountability, and improved allocation of resources. We highlight gaps and set out a research agenda that consists of: (a) disaggregating broad measures of budget transparency to uncover which specific disclosures are related to outcomes; (b) tracing causal mechanisms to connect fiscal openness interventions with ultimate impacts on human development; (c) investigating the relative effectiveness of alternative interventions; (d) examining the relationship between transparency and participation; and (e) clarifying the contextual conditions that support particular interventions. JEL codes: E62, H11, H61, H83

Transparency and participation are in vogue in international policy circles. These concepts form part of a new development consensus that has become “nearly universal” in the policy statements of major international organizations (Carothers and Brechenmacher 2014) and links them to a series of desirable outcomes. The fiscal policy arena is no exception. The website of the International Monetary Fund (IMF) claims that fiscal transparency “is critical for effective fiscal management and accountability”. The World Bank’s Budget Transparency Initiative (n.d.) asserts that budget transparency leads to less corruption, more efficient use of resources, more trust in government, and higher revenues. A number of organizations seem to promote transparency and participation in fiscal matters—which we
summarize with the term “fiscal openness”—based on an implicit theory of change that sees them leading to a variety of desirable impacts.

This positive view has engendered a growing set of international standards and norms. In its 2014 Fiscal Transparency Code, the IMF sets out benchmarks for fiscal reporting, forecasting, and budgeting, and the management of fiscal risks. For the first time since the inception of the transparency code in 1998, it also encourages governments to provide their citizens with “an opportunity to participate in budget deliberations” (IMF 2014). In 2015, the Council of the Organisation for Economic Co-operation and Development (OECD) approved a “Recommendation on Budgetary Governance”, which advocates for budget documents and data to be “open, transparent and accessible” and for budget debates to be “inclusive, participative and realistic” (OECD 2015). Finally, the multi-stakeholder Global Initiative for Fiscal Transparency (GIFT) has developed High-Level Principles that were endorsed by the United Nations General Assembly in December 2012. These principles enshrine the right of citizens to gain access to fiscal information and to have effective opportunities to participate in fiscal policymaking.

As with many norms and principles that gradually gain international acceptance, arguments in their favor can be distinguished as normative and instrumental. On the normative side, the intrinsic value of fiscal openness is increasingly recognized. Governments have started to translate related international norms and principles into domestic laws and practices, albeit unevenly. A review of budget laws in over 100 countries found that more than half explicitly mention transparency, at least as a key principle to guide fiscal policymaking. On the other hand, only seven budget laws included explicit provisions for citizen participation and engagement (de Renzio and Kroth 2011).

The instrumental side of the debate includes proponents of fiscal openness as well as skeptics. For instance, Heald (2003) points to a view that “over-exposure” to fiscal information may lead to “losses in effectiveness through high levels of transaction costs and excessive politicization”, while de Fine Licht et al. (2014) caution that transparency may engender frustration among citizens if not combined with “credible mechanisms for accountability”. Such concerns are legitimate. Ultimately, empirical evidence is required to understand whether, how, and when fiscal openness contributes to outcomes desired by governments, citizens, or market actors. The list of supposed benefits that proponents present is substantial, but these claims are rarely backed up by rigorous evidence. Too often, perceived positive impacts in a single case metamorphose into “best practice” examples that receive unquestioned support in policy reports, or statistical studies are cited without acknowledging potential threats to valid inference.

To provide a firmer empirical grounding for these debates, we carry out the first systematic review of published evidence on the impacts of transparency and participation in government budgeting. Others have reviewed the impact of “citizen
engagement” or “participatory governance” on improvements in governance and development (Gaventa and Barrett 2012; Speer 2012), leaving out questions linked to transparency, and without focusing specifically on government budgets. Other reviews (Fox 2014; Kosack and Fung 2014) are restricted to a small set of impact evaluation studies related to growing donor support for “social accountability” or “transparency and accountability” initiatives. Some reviews resemble our focus (Carlitz 2013; Ling and Roberts 2014) but consider a more limited range of evidence, and often group together a disparate set of interventions without sufficient conceptual underpinnings. Our review focuses squarely on fiscal openness, which is crucial at a time when international organizations are stepping up their efforts to promote fiscal disclosure and participation in budgeting as part of a package of reform initiatives.

This article proceeds as follows. We first describe our approach and analytic framework, followed by conceptual background and an overview of how “transparency” and “participation” in budgeting are operationalized in the literature. We then summarize the evidence across four broad categories of impacts, and assess its strength by focusing on whether studies can make a convincing claim to identify causal effects as well as their substantive contribution. Our conclusion develops potential directions for future research.

Our Approach and Framework

The scope of this review includes studies that (1) empirically evaluate a causal claim about the impact of an element of fiscal openness; (2) have achieved publication as a peer-reviewed academic article, or as a book with an academic press or well-known commercial publisher; and (3) are of sufficient length to qualify as a substantial piece of original research. We elaborate on some aspects of these criteria below, and describe the resulting set of studies for this review. We also set out a framework that guides our analysis in the following sections.

We examine empirical work that focuses on, or otherwise makes a significant contribution to, the evaluation of a causal argument about the impact of fiscal transparency or participation in budgeting. Hence, we do not cover purely theoretical work, although in some instances (e.g., Milesi-Ferretti 2003) this has laid the foundation for subsequent empirical studies we review. Scholars have also contributed conceptual analyses (e.g., Heald 2003), or described particular cases (e.g., de Sousa Santos 1998). We acknowledge that such work can make important contributions and spur critical reflection, but it falls outside the scope of our review.

In general, we excluded studies that are not precisely focused on an explanatory variable or intervention that falls under our definition of fiscal openness, which we discuss in detail in the following section. This may be the case when a measure of...
government disclosure of information is used that includes fiscal material, but is not limited to it. As a result, any reported effects cannot be precisely attributed to fiscal transparency. In instances where the link is plausibly strong, however, we have included the evidence in our review, while noting this limitation. We do not include the large body of literature on transparency and communication by central banks in the management of monetary—rather than fiscal—policy, and their impacts on financial markets (Blinder et al. 2008).

Our focus in this paper is on interventions that are due to government action, such as the publication of budget information or the provision of participation opportunities in the budget process. For this reason, we excluded a number of recent studies under the broad topic of “social accountability.” Few of these studies look at transparency and participation in budgetary matters, and many investigate interventions by non-governmental organizations or researchers rather than governments. In terms of publications covered, we considered limiting our analysis to books or peer-reviewed articles, but decided to make some reasonable exceptions. These include working papers from the IMF, the World Bank, and a few other institutions that have contributed significantly to the debate on fiscal openness and/or are frequently cited in other studies that we examine.

Following an initial sweep of the literature and applying the above criteria to filter the resulting list of over a hundred studies, we identified a core set of 38 papers published between 1991 and early 2015 as the basis for our review. Of these, 23 investigate the effects of variables related to fiscal transparency, and 14 relate to participation in budgetary decisions. Only a single study (Olken 2007) looks at both transparency and participation interventions to explore their relative impacts. About three-quarters of all studies use quantitative methods, most of them based on observational data. Only four studies are based on experimental designs, and three might be labeled quasi- or natural experiments. Twenty studies use evidence from a single country, and 18 are based on cross-national data. The appendix contains a listing with overview information.

Figure 1 provides a visual guide to our analysis. The boxes represent different sets of variables, and the arrows connections between them. Figure 1 should be read from the bottom up, indicating possible relationships between fiscal openness interventions, the budget process, and different types of impacts. Transparency and participation interventions potentially can relate to any of the four stages of the budget process: formulation within the executive branch (drafting stage), legislative review and approval, execution (when resources are raised and spent based on the approved budget), or ex post audit and evaluation. By altering the process of budgeting, a fiscal openness intervention may affect the quality of the budget in terms of its aggregates (for instance, how much is spent in total, or the size of the deficit), priorities, and operational efficiency (or service delivery: the conversion of inputs purchased with funds into tangible goods and services). This, in turn, may
affect governance outcomes (such as corruption, or the reelection prospects of politicians) and development outcomes. Proponents of fiscal openness also often posit that governance affects development. Hence, figure 1 shows a direct impact of the budget on development outcomes, as well as an indirect impact (via better governance). These relationships are embedded in particular contexts, suggesting possible scope conditions for the effectiveness of particular interventions and the specific nature of any impacts. The following sections flesh out the various elements in figure 1.

Our approach to assessing the contribution of these studies considers their substantive importance as well as the extent to which they identify causal impacts. Randomized evaluations have gained in importance in recent years precisely for their potential to deliver convincing evidence on the impact of policies and institutional reforms. Yet the substantive focus of our study means that experiments are not always possible, especially in relation to macroeconomic policy (Glennerster and Takavarasha 2013). For instance, governments would have to agree to the random assignment of fiscal transparency levels to evaluate its impact on, say, deficits or debt. This is highly unlikely, so observational data and the quality of their analysis will be crucial for gaining insights into these relationships. Other types of
outcomes in figure 1 are amenable to randomized evaluations, and we discuss relevant studies in detail, while acknowledging the contribution of other approaches.

Transparency and Participation as Independent Variables

We start by examining the key components of fiscal openness: transparency and participation. The IMF defines fiscal transparency as “the comprehensiveness, clarity, reliability, timeliness, and relevance of public reporting on the past, present, and future state of public finances”\(^{10}\). This definition captures much of what we consider important regarding the regular disclosure and dissemination of detailed and accessible information on all aspects of fiscal policy by the government (see also Kopits and Craig 1998; OECD 2001).

Existing definitions of public participation in budget processes are less well developed. In general, such definitions refer to a wide set of practices through which citizens, civil society organizations, and other non-state actors interact with public authorities to influence the design and execution of fiscal policies.\(^{11}\) This may occur at different stages of the budget cycle or in relation to specific service delivery or public investment issues.\(^{12}\)

In studies of fiscal transparency, the specificity of the independent variable of interest varies greatly. Least precise are measures of government transparency that include fiscal material, but are not limited to it (e.g., Bellver and Kaufmann 2005; Gelos and Wei 2005; Glennerster and Shin 2008; Lindstedt and Naurin 2010). Here, we cannot be certain whether an association is with fiscal transparency or some other aspect of government disclosure.

By far the most common operationalization of fiscal transparency in the literature is a broad approach that captures a wide range of disclosures with reference to the IMF Code and OECD Best Practices for Budget Transparency. Cross-national studies use data from IMF fiscal transparency assessments (e.g., Hameed 2005; Arbatli and Escolano 2012; Weber 2012), budget surveys by the OECD (e.g., Alt and Lassen 2006a and 2006b; Benito and Bastida 2009), the Open Budget Index (Hameed 2011; Blume and Voigt 2013; Alt, Lassen, and Wehner 2014), or an early budget transparency assessment for European Union (EU) countries (von Hagen and Harden 1994; Bernoth and Wolff 2008).\(^{13}\) This “broad” approach is rarely replicated at the subnational level, with few exceptions (Alt, Lassen, and Skilling 2002).

Work on specific components of fiscal transparency is rare. Three studies examine external auditing, one with cross-national data (Bernoth and Wolff 2008) and two in a single country (Olken 2007; Ferraz and Finan 2008). One well-known study looks at the publication of information on funds disbursed to local schools (Reinikka and Svensson 2005, 2011).\(^{14}\) Transparency interventions targeted at
the earlier drafting or approval stages of the budget process (see figure 1) are not examined in detail. Other important aspects are also overlooked. For example, we lack evidence on revenue transparency, including in relation to natural resources, and on transparency in public procurement. Overall, cross-national work on fiscal transparency tends to use broad and encompassing measures, while several single-country studies focus on selected disclosures.

Studies of participation in budget processes broadly belong to two groups. One looks at “participatory budgeting” as a specific mechanism first adopted in the southern Brazilian city of Porto Alegre, before spreading within Brazil and further afield. Participatory budgeting is a process of democratic deliberation consisting of organized assemblies through which citizen representatives are able to define and decide local public investment priorities. Some of the early literature (e.g., de Sousa Santos 1998) focused on its practice and potential for democratic development, and qualitative accounts on different countries (Ebdon and Franklin 2004; Wu and Wang 2011; Kasymova and Schachter 2014) discuss its consequences only in general terms. More recent quantitative studies assess the Brazilian experience by comparing municipalities that introduced participatory budgeting with those that did not (Boulding and Wampler 2010; Gonçalves 2014; Touchton and Wampler 2014).

A second group of papers looks at a variety of other participatory mechanisms adopted as part of decisions on city-level services, decentralization reforms (Heller, Harilal, and Chaudhuri 2007), public investment programs, or other similar initiatives that allow citizens to have a voice in determining resource allocation. These range from citizen surveys (Watson, Juster, and Johnson 1991; Simonsen and Robbins 2010) and citizens directly voting on budget priorities (Olken 2010; Beath, Christia, and Enikolopov 2012), to village forums shaped around voluntary or traditional practices (Diaz-Cayeros, Magaloni, and Ruiz-Euler 2010; Jaramillo and Wright 2015). These mechanisms are studied in sub-national contexts, with some papers comparing different mechanisms. We could not find any relevant studies that utilize cross-country data or look at national-level participation practices, except for the broad comparisons in Bräutigam (2004). Table 1 summarizes the main empirical measures of fiscal openness in the literature.

Summary of Findings

The 38 studies relate to different groups of outcomes represented in figure 1: (a) macro-fiscal, (b) allocation and service delivery, (c) governance, and (d) development outcomes. The first two categories relate to the quality of budgets, as assessed by public finance practitioners and scholars. The latter two categories look at the consequences of resource decisions and management. In category (a), 14 studies
look at fiscal performance, credit worthiness, and creative accounting (or “fiscal gimmickry”). Group (b) contains nine studies linking fiscal openness to the allocation of budget resources across different sectors or projects, and the delivery of public services. Category (c) has 11 studies that look at what we label governance outcomes, which range from corruption and political accountability to the mobilization of citizens. The final category (d) is small and considers development outcomes in areas such as health and education. Only a few studies look at several outcomes (e.g., Alt, Lassen, and Skilling 2002; Hameed 2005; Reinikka and Svensson 2011), and the appendix notes secondary impact categories. We examine each group in turn.

**Macro-fiscal Outcomes**

One of the most established areas of empirical research probes the relationship between budget transparency and fiscal outcomes, such as deficits or debt. An important early contribution by von Hagen and Harden (1995) develops several indices of the quality of budget institutions for 12 EU countries and documents an association with fiscal outcomes. However, their measure of the “informativeness” of the draft budget is only one component of their indices, which is not analyzed separately here or in later work (e.g., Hallerberg, Strauch, and von Hagen 2009).¹⁷

In pioneering work, Alt and Lassen (2006a, 2006b) examine the role of budget transparency in electoral budget cycles and its impact on public debt. In a panel of 19 OECD countries in the 1990s, these authors find large swings in the budget balance in low-transparency countries, where deficits are more than 1% of GDP lower in a post-election year than in an election year (Alt and Lassen 2006a). In a related study, these authors link this to higher levels of public debt (Alt and Lassen 2006b).

Benito and Bastida (2009) find a negative correlation of their budget transparency index with deficits, but not with debt levels, for a sample of up to 41 countries.

A second strand of research in this category links fiscal transparency to sovereign credit ratings and related variables. Hameed (2005) shows that transparency is associated with better credit ratings in a cross-section of 32 countries (see also Hameed 2011). Arbatli and Escolano (2012) confirm this association for a larger sample of up to 56 countries. These authors further present correlations suggesting that budget transparency works indirectly via its effect on fiscal outcomes for developed countries, whereas the effect on credit ratings is direct for developing countries.

Several papers consider the relationship between budget transparency and borrowing costs. Looking at EU countries and the United States, Bernoth and Wolff (2008) find that transparency mediates the association of detected creative accounting with risk premia in government bond markets. Wang, Shields, and Wang (2014) study 562 state bond issuances in the United States between 1986 and 2012, and find that both high and low transparency levels correlate with increased costs, but medium levels correlate with lower costs. Two related papers use broader measures of transparency, and find it appears to attract and retain equity fund investments in emerging markets (Gelos and Wei 2005) and to lower sovereign borrowing costs (Glennerster and Shin 2008). These patterns are broadly consistent with the work on transparency in budgeting, but we cannot isolate the contribution of fiscal disclosure.

A more recent approach is an empirical focus on the role of transparency in containing creative accounting or fiscal gimmicks, as proxied by “stock-flow adjustments”—the difference between the change in the stock of debt and annual deficits. Using IMF data for a sample of 87 countries, Weber (2012) finds that fiscal transparency correlates with decreased deviations. Alt, Lassen, and Wehner (2014) examine 14 EU countries from 1990 to 2007, showing that fiscal transparency dampens or eliminates the association with fiscal gimmicks of elections, fiscal rules (the Stability and Growth Pact), and economic downturns. These studies suggest that fiscal transparency may have both direct and indirect effects on creative accounting.

**Resource Allocation and Service Delivery Outcomes**

Evidence linking fiscal openness to shifts in resource allocation and improvements in the provision of public services is more recent. Relevant work focuses on
different participatory mechanisms rather than on fiscal disclosure. Unsurprisingly, much of this evidence is based on the pioneering Brazilian experience with participatory budgeting.

Goldfrank and Schneider (2006) document how after the introduction of participatory budgeting, the city of Porto Alegre increased the share of spending dedicated to the social sectors—by much more than in places where participatory institutions had not been established—and improved its performance in project completion. More recent papers exploit the widespread adoption of participatory budgeting practices across Brazil. Boulding and Wampler (2010) use a dataset covering 220 large Brazilian cities—64 of which introduced participatory budgeting between 1989 and 2000—to show that the adoption of participatory budgeting correlates with changes in resource allocation, especially increases in health and education programs (see also Touchton and Wampler 2014). Gonçalves (2014) estimates that health and sanitation spending increased by 20%–30% after municipalities introduced participatory budgeting.

The above papers all refer to a specific type of participatory institution and to a single country, albeit a very large one. We thus know little about the relevance of these results to other contexts. Bräutigam (2004) discusses the prospects for participatory budgeting around the world and looks at similar experiences across five countries. She suggests that its impact is often conditional on left-wing political parties winning power and using it to advance their progressive agenda, and on the existence of strong audit institutions as well as free, open, and well-informed public policy debates.

Another group of papers looks beyond participatory budgeting, drawing on experiences in a variety of countries and involving different practices. Simonsen and Robbins (2000) and Watson, Juster, and Johnson (1991) document how two U.S. cities have used citizen surveys to assess support for taxes for different services, or to help prioritize parts of the budget. Heller, Harilal, and Chaudhuri (2007) evaluate structures and processes introduced in 1996 in each panchayat (local government) of Kerala, India, to directly involve citizens in spending decisions. Survey respondents from 72 randomly selected panchayats indicated significant improvements, especially for roads, housing, and child services. Respondents also identified projects approved after the start of the campaign as much more appropriate and responsive to local needs. Looking at survey evidence across four states in South India, Besley, Pande, and Rao (2005) examine the role of village meetings, or gram sabhas, that discuss resource allocation. People from disadvantaged groups are more likely to attend than those from other groups, and these meetings appear to be associated with access to resources and services. However, women are less likely to participate than men, which affects representativeness.

Beath, Christia, and Enikolopov (2012) use a randomized field experiment in 250 villages across Afghanistan to assess the impact that different participatory mechanisms have on elite influence and resource allocation outcomes for local
development projects. All villages were part of a country-wide National Solidarity Program that provided grants for village projects. Project selection through a secret ballot was much less likely to be affected by elite preferences—and better reflected the needs of the majority of the population, as captured in villagers’ satisfaction levels—than projects in villages where decisions were taken in village council meetings, where elites could wield more influence. Another randomized evaluation (Olken 2010) finds a similar pattern in the Kecamatan Development Program in Indonesia, where a secret ballot to choose projects led to much higher satisfaction with the selected projects than representative village meetings. The conclusion from these two papers is that direct, rather than representative, participation promotes shifts in resource allocation.

Two additional papers look at the experience of Mexico and Peru in promoting participatory governance at the local level. Diaz-Cayeros, Magaloni, and Ruiz-Euler (2013) consider a reform in the state of Oaxaca in Mexico in 1995 that allowed indigenous communities to opt for a form of traditional governance called usos y costumbres, characterized by participatory practices in budgeting and implementation, rather than normal representation through political parties. The communities who adopted traditional governance structures increased energy provision and improved education and sewerage services much faster than municipalities governed by political parties, with decisions taken by politicians without citizen involvement. Jaramillo and Wright (2015) compare agricultural services under mandatory participatory budgeting introduced by the central government in Peru against those with voluntary participatory fora. Voluntary fora appear to facilitate flows of information and collective action and are associated with improvements in the quantity and quality of agricultural services.

**Governance Outcomes**

A large share of papers in this category deals with the question of whether transparency reduces corruption. A first group examines correlations in cross-country data. Bellver and Kaufmann (2005) develop a measure of transparency for 194 countries based on 20 independent sources looking at access to information, budget transparency, and press freedom, for example; their transparency index is negatively correlated with corruption. Hameed (2005) and Bastida and Benito (2007) detect similar correlations looking specifically at fiscal transparency and with smaller samples of countries. Lindstedt and Naurin (2010) reinvestigate Bellver and Kaufmann’s data and suggest that transparency is associated with lower corruption only when the information provided is accessible via a free press and can be utilized to hold governments to account in elections.

Two papers get closer to establishing a causal link between transparency and corruption. Reinikka and Svensson (2011) report on the following widely-cited
A 1996 survey by the World Bank and the Government of Uganda found that only a small percentage of funds released by the central government for supporting local schools with materials and equipment actually reached the schools. District officials diverted the rest through leakage and corruption. After the government started publishing details on these transfers in national newspapers and posting them on school notice boards, disbursements reaching schools shot up from an average of 25.4% in 1996 to 81.8% in 2001, most markedly in areas with better access to newspapers. This case popularized Public Expenditure Tracking Surveys (PETS) and is often cited as evidence that transparency can reduce corruption by enabling citizens to hold officials to account. Yet, replication attempts have been less successful (Sundet 2008). Hubbard (2007) points out that following the first survey, several education reforms also helped to reduce leakage. Hence, it is impossible to determine to what extent providing disbursement information to communities, rather than other changes, caused the improvements. While proximity to newspaper outlets is suggestive of such a mechanism, there are likely to be systematic differences between communities with and without easy access to media.

One unambiguously causal piece of evidence in this area is a field experiment carried out in Indonesia to test alternative approaches to lower corruption in village road projects (Olken 2007). Some randomly selected villages were told that their project would be audited by the central government audit agency, and audit findings were discussed at open village meetings. As a result, the amount of misused funds (measured as the difference between actual costs and an estimate provided by independent engineers) was eight percentage points lower compared to villages that did not receive the audit treatment. Elsewhere, villagers were invited to “accountability meetings” where they could query officials about project implementation, and provide anonymous comments. The participation treatment was associated with much smaller, and statistically insignificant, reductions in corruption. In addition to demonstrating that public audits mitigate corruption, this study calls into question the common assumption that community monitoring is a powerful deterrent to corrupt behavior. Although citizen engagement did have some effect on variations in project costs, its overall impact was far from decisive.

Some other studies link fiscal openness to electoral accountability. Alt, Lassen, and Skilling (2002) find that fiscal transparency correlates with higher gubernatorial popularity in U.S. states. Alt and Lowry (2010) extend this work and find no direct association of budget transparency with the retention of incumbent governors, but transparency dampens the negative correlation of tax increases with incumbent retention. A study by Ferraz and Finan (2008) examines how the public release of audit reports on federally transferred funds affects the reelection prospects of incumbent mayors. These authors exploit the fact that the timing of audits was randomized, with some municipalities audited prior to elections in 2004 and some afterwards, to identify causal effects. Where pre-election audits revealed violations indicating corruption,
their publication significantly reduced reelection probabilities, especially in municipalities with local radio stations that could publicize the audits.

Lastly, a few additional papers focus on other governance-related outcomes. Islam (2006) builds a transparency index that measures the timeliness of economic data (including government revenue and expenditure) published by the government, and shows that it is positively related to the quality of governance, as measured by the World Bank’s Worldwide Governance Indicators. Touchton and Wampler (2014) look at participatory budgeting in Brazilian municipalities, and report that it is associated with a statistically significant increase of 8% in the number of active civil society groups. These authors interpret this as evidence that participatory budgeting promotes collective action, citizen mobilization, and monitoring of state action (see also Goldfrank and Schneider 2006).

Development Outcomes

Finally, we turn to impacts on development. As far as fiscal transparency is concerned, evidence is very thin. Bellver and Kaufmann (2005) report a correlation between their transparency index and better socio-economic and human development indicators. Fukuda-Parr, Guyer, and Lawson-Remer (2011) also find that the Open Budget Index is positively associated with human development, but the correlation disappears once they include control variables in their regressions. Reinikka and Svensson (2011) find that access to budget information on school grants led to increases in school enrollment, and to some extent educational achievement, measured with exam scores. The authors argue that these results should be relevant for other countries with similar educational funding approaches, but they also note contextual factors that might limit external validity, including the role of parents in school management and the salience of primary education in public policy debates.

In addition, participatory budgeting has been linked to improved health indicators. Both Touchton and Wampler (2014) and Gonçalves (2014) report that Brazilian municipalities that introduced participatory budgeting saw their infant mortality rates drop significantly more than other municipalities. The main reasons, Gonçalves posits, is that citizen participation improves the targeting of policies and spending. Touchton and Wampler suggest this association strengthens over time.

Assessing the Strength of the Evidence

In this section, we assess the strength of the available empirical evidence against two criteria. One criterion is the degree to which a study minimizes threats to valid
inference due to the strength of its research design or methods. The second criterion is the substantive importance of a study, which has to do with the nature of the impact under investigation, as well as the degree to which we can draw broader implications from the results, including beyond the immediate empirical context.

Very few studies can plausibly claim to identify a causal effect of fiscal openness. Of the quantitative studies in our dataset, merely four are based on experimental designs (Beath, Christia, and Enikolopov 2012; Olken 2007 and 2010; Simonsen and Robbins 2000), while one exploits a natural experiment (Ferraz and Finan 2008) and another a quasi-experiment (Reinikka and Svensson 2011). In addition, the study by Glennerster and Shin (2008) is quasi-experimental, but it cannot claim to identify the causal effect of fiscal disclosure. While we would welcome more field experiments, this is not a feasible research design for evaluating some impacts of fiscal openness, in particular macro-fiscal outcomes, or when government bodies resist attempts to subject their initiatives to randomized evaluation. In the case of work on fiscal transparency and fiscal outcomes, while we lack randomized evaluations, a relatively large set of compatible results strongly suggests a general pattern.

The quality of non-experimental evidence varies greatly, and depends on the extent to which it addresses internal validity concerns. In particular, with cross-national data it is difficult to account for the full range of variables that may affect the relationship of interest, whereas subnational units within the same country are typically more comparable. Panel data, where units are observed repeatedly over time, can help to address such concerns by focusing on within-unit variation, as in several studies of participatory budget processes (e.g., Boulding and Wampler 2010; Diaz-Cayeros, Magaloni, and Ruiz-Euler 2010; Gonçalves 2014). A time dimension is also useful for investigating the direction of the relationship of interest. In the appendix, we summarize the empirical context of each study, and show that eight studies—all of them involving transparency—rely purely on cross-national data for a single year or time period. This is the least convincing approach for documenting causal impacts. In addition, measurement error is a source of bias, which is a special concern where budget transparency is assessed using survey responses supplied by governments directly and with limited quality control (see Wehner and de Renzio 2013). Some studies use instrumental variable strategies in attempts to address internal validity concerns (Alt and Lassen 2006b; Alt and Lowry 2010; Arbatli and Escolano 2012; Blume and Voigt 2013), but convincing instruments are exceedingly rare in this literature.

Case studies can help to clarify underlying mechanisms. However, in our context they often rely on broad arguments rather than detailed tracing of causal processes, usually pay little attention to constructing a counterfactual scenario,
and are sometimes commissioned by organizations that have an interest in claiming positive impacts. In some cases, variables are poorly conceptualized and operationalized. Moreover, case selection is a concern. As Carlitz (2013) notes, “successful initiatives have been examined in greater detail than unsuccessful ones”, which “can make it difficult to draw conclusions about the factors that lead to impact.” This is of course a wider problem, irrespective of methodological approach (Franco, Malhotra, and Simonovits 2014). Nonetheless, some studies (mostly employing case study approaches) do fruitfully explore the specific context and mechanisms in which fiscal openness initiatives have worked. These studies often produce insights and hypotheses that can then be tested elsewhere and help to advance knowledge by detailing how initiatives succeed or fail (see, e.g., the case studies in Khagram, Fung, and de Renzio 2013).

Yet, rigor in a study’s design and use of methods needs to be complemented by substantive importance, which can be limited for various reasons. One of these relates to how much we care about the documented impact. Take Simonsen and Robbins (2000), who examine how citizens react to different types of budget information. While the documentation of such public opinion impacts can be valuable—not least for governments that wish to manipulate public sentiments—arguably more important is whether people’s lives were materially affected. The dependent variables in Beath, Christia, and Enikolopov (2012) and Olken (2010) get closer, capturing impacts on project selection as well as villager satisfaction. Olken (2007) and Reinikka and Svensson (2011) document direct impacts on actual service provision, while Ferraz and Finan (2008) show effects on the ability of citizens to hold corrupt politicians to account via the ballot box. All of these studies come with the usual questions about generalizability, but a priori there are no strong grounds for assuming that the results are meaningless beyond the immediate empirical context in which they were obtained. Their wider relevance remains, above all, an empirical question.

The work by Olken (2007) stands out because not only is it empirically convincing due to its experimental design, but it is also substantively unique: this is the only study investigating the relative importance of fiscal transparency and participation in budgeting. The study suggests that “top-down” external auditing trumps “bottom-up” grassroots participation in project monitoring. Yet, this does not mean that any kind of participation has no effect. Moreover, the audit treatment included their delivery by the auditors to a special village meeting, which meant that “retribution from the village” was a potential sanction. Thus, a form of participation was built into the audit treatment, and cannot be disentangled from the mere disclosure of audit information. Further work is needed along these lines in order to rigorously assess different types of participation interventions (see also Olken 2010), and how they may or may not amplify any impacts of information disclosure.
A Guide for Future Research

This first systematic review of evidence on the impacts of fiscal openness assesses 38 studies that empirically investigate the effect of government disclosure of budgetary information or of mechanisms for public participation in the budget process on different outcomes, using qualitative or quantitative methods. Most studies fall into three main impact categories: macro-fiscal, resource allocation and service delivery, and governance. A fourth category—development outcomes—contains only a small number of studies.

The evidence on the link between fiscal openness and different macro-fiscal outcomes, such as indicators of fiscal discipline, is typically based on broad measures of transparency. This work links transparency to reduced deficits, debt, borrowing costs, and creative accounting. Only a handful of studies in this category investigate the role of specific components of budget transparency. Evidence on resource allocation and service delivery mainly considers participatory mechanisms. This work is no longer limited to the well-documented Brazilian experience with participatory budgeting, and provides insights on why certain mechanisms may work better than others.

Looking at governance, findings are more varied, also because studies use different definitions and measures. Cross-country studies document a negative correlation between (fiscal) transparency and corruption, but they cannot make strong causal claims. Evidence from Uganda, Indonesia, and Brazil provides much richer accounts of how specific disclosures of budget information can reduce corruption and promote accountability by incentivizing citizens to monitor governments, and public officials to refrain from corrupt behavior. It also appears that participatory mechanisms widen citizen involvement and mobilization. Findings on how fiscal openness affects development outcomes are scarce, but suggest impacts on education in Uganda and health in Brazil.

The quality of the evidence varies. The most convincing work—a small number of experiments and natural or quasi-experimental studies—documents impacts that many would consider beneficial: lower government borrowing costs due to macro-fiscal disclosure (Glennerster and Shin 2008), lower corruption due to audits (Olken 2007) and the publication of budget execution information (Reinikka and Svensson 2011), electoral consequences for politicians when audits suggest malfeasance (Ferraz and Finan 2008), and improved budget allocations due to citizen participation (Beath, Christia, and Enikolopov 2012; Olken 2010). While this is a surprisingly small collection of papers, the overall direction of this high-quality evidence is consistent.

How do these findings relate to the theory of change that we sketched earlier? Only some of the linkages in figure 1 are backed by sufficient evidence. The documented links of general forms of transparency are limited to intermediate steps, in
particular macro-fiscal outcomes, in the longer chain leading to development outcomes. Further down the chain, the evidence highlights how specific and locally-relevant disclosures, especially on budget execution and audits, can improve governance outcomes. Moreover, mechanisms enabling direct citizen participation in budgetary decisions provide feedback loops so that governments learn about citizens’ needs and priorities and can better respond to them. On the other hand, the existing literature sheds very little light on the link between transparency and participation, or between some of the earlier and later steps in the theory of change. For example, the evidence on participation offers little guidance on the types of fiscal disclosures that form the basis for engaging citizens, and the mechanisms leading to improved development outcomes are not clear. We identify several specific gaps and open questions for further research.

First, a largely untapped potential lies in disaggregating broad transparency measures to examine which specific disclosures are related to outcomes. For example, the Open Budget Index provides a detailed assessment of a package of eight budget documents. It is possible to distinguish fiscal disclosure across each document or stage of the budget process (drafting, approval, execution, and audit). Similarly, transparency on the revenue and expenditure sides could be distinguished (or specific disclosures within these, such as natural resource revenues). Without such disaggregation, the absolute and relative contribution of specific elements is unclear, and policy implications lack specificity. For instance, to contain public debt, should a government strengthen in-year execution updates or the quality of its budget proposal (and which part of the latter)? When reform has to be phased, it can be essential to target the greatest gains and quick results—not least to build the case for further changes. Similarly, what specific disclosures mobilize citizens and render participation more effective? Are citizens more interested in budget allocations, or in what was actually spent, and how do they react to and use such information in holding politicians and bureaucrats to account? Evidence on these questions would have immediate policy relevance.

A second and related point is that a greater focus on specific interventions may also enable scholars to better trace the mechanism through which they affect the quality of people’s lives. Perhaps the best example of this is the work by Reinikka and Svensson (2011), who investigate the chain from a transparency intervention in budget execution to impacts on school enrollment and learning outcomes. Most other research has focused on smaller segments of the causal chain represented in figure 1, and there is potential to follow the example of this work by examining how specific interventions affect development outcomes, including over longer periods of time.

Third, with very few exceptions (notably Olken 2007 and 2010), the evidence thus far tells us little about the relationships and trade-offs between different interventions, and their relative effectiveness. For example, how do different types of
participation (such as consultative or monitoring meetings) compare against particular disclosures (such as audits or in-year execution information) in terms of their impact on the leakage of funds and service delivery? Further, is the disclosure of execution information more effective in curbing corruption when it is complemented with a participatory monitoring opportunity? Finding answers to such questions requires studies of alternative interventions and their various combinations within the same research design. The role of such research in enhancing service delivery and development in poor countries is potentially large.

A fourth question is how fiscal transparency and participation in budgeting affect one another. Existing research into this complex area is limited, partly due to diverging views and emphases among major proponents of fiscal openness. For some, especially the IMF, macroeconomic stability is the ultimate goal. Here, the focus is on fiscal transparency and its effects on macro-fiscal outcomes. Others, such as development practitioners and civil society actors, are often more interested in promoting participation to affect allocations, service delivery, and governance. Partly as a result, research has tended to focus on one of the main components, less on how they affect one another, with only a few exceptions (Khagram, Fung, and de Renzio 2013). Fundamental questions remain in this area. For example, under what conditions does fiscal disclosure lead to greater participation in budgetary decisions? When and how does participation, in turn, lead to greater demand for fiscal transparency? Rigorous examination of this potentially reciprocal relationship could help to assess the case for enhancing participation in budgetary decisions.

Finally, future research should generate insights relating to the conditions under which fiscal openness interventions have a particular effect. Thus far, the cross-national research on macro-fiscal impacts of fiscal transparency, despite methodological limitations of this approach, yields arguably the strongest case that the results hold more generally. Yet, several studies suggest conditioning factors. Reinikka and Svensson (2011) and Ferraz and Finan (2008) show that the media can play a decisive mediating role. The impact of additional fiscal disclosures may also depend on the initial level of transparency (Heald 2003). With regard to participatory budgeting, Bräutigam (2004) notes that left-wing parties tend to favor pro-poor spending policies. Hence, they also have the strongest incentives to introduce participatory mechanisms that bring previously excluded groups into the decision-making process and legitimize reprioritization. It is less clear whether participatory budgeting is linked to similar outcomes in politically less receptive environments (see Kosack and Fung 2014). More explicitly comparative work could investigate crucial scope conditions.

Existing research documents a link between fiscal openness and a number of desirable outcomes. Yet, proponents of fiscal openness often struggle to convince
skeptical governments to take the risk of pursuing institutional changes in this area. Addressing the above gaps would go a long way in strengthening the evidence base.

Notes

Paolo de Renzio is a Senior Research Fellow at the International Budget Partnership. He can be contacted at pderenzio@internationalbudget.org. Joachim Wehner is Associate Professor in Public Policy at the London School of Economics and Political Science. He can be contacted at j.h.wehner@lse.ac.uk. The authors thank Hugh Batrouney for outstanding research assistance, as well as Jim Alt, Alta Fölscher, Jonathan Fox, Juan Pablo Guerrero, Tim Irwin, Sanjeev Khagram, Mareike Kleine, Steve Kosack, Victoria Louise Lemieux, Ian Lienert, Greg Michener, Murray Petrie, Nicola Smithers, Martin J. Williams, as well as three anonymous reviewers and the journal editor, Peter Lanjouw, for helpful comments and suggestions. Previous versions of this paper were presented at the 2015 Annual Meeting of the American Political Science Association, and at a workshop hosted by Greg Michener at the Fundação Getúlio Vargas in Rio de Janeiro. This work was supported by the Global Initiative for Fiscal Transparency (GIFT).

2. Recommendations are a formal instrument of the OECD, with which member countries should comply.
3. GIFT defines itself as “a multi-stakeholder action network working to advance and institutionalize global norms and significant, continuous improvements in fiscal transparency, participation, and accountability in countries around the world.” See http://fiscaltransparency.net/, accessed March 2, 2015.
4. For other skeptical views of transparency, see Bac (2001), Etzioni (2014), and comments by Francis Fukuyama that spurred a lively debate: http://www.the-american-interest.com/2015/01/04/the-limits-of-transparency/, accessed September 30, 2015.
5. The paper by Kosack and Fung was part of a research project funded by the Transparency and Accountability Initiative (TAI), a donor collaborative working to expand the impact and scale of such interventions. See http://www.transparency-initiative.org/about, accessed March 2, 2015. Fox’s work was commissioned by the Global Partnership for Social Accountability (GPSA), a World Bank initiative. See http://www.thegpsa.org/sa/, accessed March 2, 2015.
6. For reasons of time and capacity, we limit ourselves to research published in the English language. Moreover, all leading academic journals in the relevant disciplines are published in English.
7. Fox (2014), Joshi (2013), and Kosack and Fung (2014) review these studies.
8. The terms “quasi-experiment” and “natural experiment” are sometimes used interchangeably. However, as Dunning (2012) points out, only natural experiments are based on randomly assigned treatments, whereas quasi-experiments have nonrandom assignment. Hence, empirical evidence from the latter tends to be weaker.
9. Of course, economic or social development may also affect governance outcomes. In this instance, we highlight a core relationship posited, often implicitly, by proponents of fiscal openness.
11. The “Public Participation Spectrum” of the International Association for Public Participation (IAP2) orders different mechanisms based on their intensity of public involvement: http://www.iap2.org/au/resources/iap2s-public-participation-spectrum, accessed March 3, 2015. From weakest to strongest, participation can be used to inform, consult, involve, collaborate, and empower. The mechanisms in the surveyed literature span the whole range, with citizen surveys at the weaker end, and participatory budgeting as the strongest form that puts citizens in charge of specific decisions.
12. We thus exclude studies of legislative participation from the scope of this review, unless they examine public participation via a legislative body. Wehner (2014) reviews the literature on legislatures and public finance.

13. For example, the Open Budget Index assesses the availability and quality of the pre-budget statement, the executive budget proposal and supporting documents, the budget law as enacted by the legislature, in-year reports, a mid-year review, the year-end report, audit reports, and popular versions in the form of a “citizen budget” (International Budget Partnership 2012).

14. In this review, we count some related publications as a single study and cite the 2011 paper. Hameed (2005) develops sub-indices on data assurances, medium-term budgeting, budget execution reporting, and fiscal risk disclosure.

15. The World Bank’s (1998) Public Expenditure Management Handbook popularized three levels of budget outcomes: aggregate fiscal discipline; resource allocation and use based on strategic priorities; and efficiency and effectiveness of programs and service delivery. This is a reformulation of an older framework (Schick 1966). We summarize impacts in the first category separately, as they make up a significant share of the literature.

16. Elsewhere, Alesina et al. (1999) develop a ten-item index of budget institutions in Latin America. Their analysis includes a subindex they interpret as “an indirect measure of transparency”, which captures bailout practices and the borrowing autonomy of subnational governments and public enterprises. These are crucial aspects of fiscal management, but they are not closely related to the quality of budget information.

17. Alt et al. (2014) highlight transactions in shares and other equity, as well as the recording of “other accounts payable”. Consistent with the former result, Seiferling and Shamsuddin (2015) find that fully transparent governments generate between 6 and 8 percent higher returns on their equity portfolios than others (see also Seiferling 2013).

18. The nature of impacts and their substantive importance are linked to the nature of a fiscal openness intervention. Direct public involvement in budgetary decisions establishes a clear potential link to resource allocation and service delivery. However, it remains an empirical question to what extent “weaker” forms of participation, or fiscal transparency interventions, may yield similar impacts as well. Without further empirical research that identifies causal effects of a diverse set of fiscal openness interventions, it is impossible to judge their substantive importance in relative and absolute terms. We return to this point in the conclusion.

References


International Monetary Fund. 2014. Fiscal Transparency Code. Washington, DC, IMF.


Appendix: Overview of studies included in the review

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Independent variable category</th>
<th>Main impact category (secondary categories)</th>
<th>Quantitative</th>
<th>Empirical context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbatli &amp; Escolano</td>
<td>2012</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes\textsuperscript{a}</td>
<td>56 OECD and developing countries, 2010</td>
</tr>
<tr>
<td>Bastida &amp; Benito</td>
<td>2007</td>
<td>Transparency</td>
<td>Governance</td>
<td>Yes\textsuperscript{b}</td>
<td>41 OECD and non-OECD countries, 2003</td>
</tr>
<tr>
<td>Benito &amp; Bastida</td>
<td>2009</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes\textsuperscript{b}</td>
<td>41 OECD and non-OECD countries, 2003</td>
</tr>
<tr>
<td>Blume &amp; Voigt</td>
<td>2013</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes\textsuperscript{b}</td>
<td>47 OECD and non-OECD countries, 1990-2000 (cross-section)</td>
</tr>
<tr>
<td>Ferraz &amp; Finan</td>
<td>2008</td>
<td>Transparency</td>
<td>Governance</td>
<td>Yes\textsuperscript{**}</td>
<td>373 municipalities in Brazil, 2003-2005</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Independent variable category</th>
<th>Main impact category (secondary categories)</th>
<th>Quantitative</th>
<th>Empirical context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fukuda-Parr et al.</td>
<td>2011</td>
<td>Transparency</td>
<td>Development</td>
<td>Yes(^b)</td>
<td>84 countries, 2008</td>
</tr>
<tr>
<td>Glennster &amp; Shin</td>
<td>2008</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes(^***)</td>
<td>23 emerging market economies, 1999-2002 (quarterly data)</td>
</tr>
<tr>
<td>Hameed</td>
<td>2011</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes</td>
<td>68 countries, 2004-2009 (monthly data)</td>
</tr>
<tr>
<td>Hameed</td>
<td>2005</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes(^b)</td>
<td>57 countries, 1998-2002 (cross-section)</td>
</tr>
<tr>
<td>Hubbard</td>
<td>2007</td>
<td>Transparency</td>
<td>Governance</td>
<td>No</td>
<td>Uganda</td>
</tr>
<tr>
<td>Islam</td>
<td>2006</td>
<td>Transparency</td>
<td>Governance</td>
<td>Yes(^b)</td>
<td>170 countries, 2002</td>
</tr>
<tr>
<td>Lindstedt &amp; Naurin</td>
<td>2010</td>
<td>Transparency</td>
<td>Governance</td>
<td>Yes(^b)</td>
<td>110 countries, 2000s (cross-section)</td>
</tr>
<tr>
<td>Reinikka &amp; Svensson</td>
<td>2011</td>
<td>Transparency</td>
<td>Governance (development)</td>
<td>Yes(^***)</td>
<td>218 Ugandan primary schools, 1996 and 2002</td>
</tr>
<tr>
<td>von Hagen &amp; Harden</td>
<td>1995</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes(^b)</td>
<td>12 EU countries, 1981-1990 (cross-section)</td>
</tr>
<tr>
<td>Weber</td>
<td>2012</td>
<td>Transparency</td>
<td>Macro-fiscal outcomes</td>
<td>Yes</td>
<td>87 countries, 1980-2010</td>
</tr>
<tr>
<td>Besley et al.</td>
<td>2005</td>
<td>Participation</td>
<td>Allocation and delivery</td>
<td>Yes</td>
<td>522 villages in four states in South India, 2002</td>
</tr>
<tr>
<td>Bräutigam</td>
<td>2004</td>
<td>Participation</td>
<td>Allocation and delivery</td>
<td>No</td>
<td>Case studies of Brazil, Ireland, Chile, Mauritius, and Costa Rica</td>
</tr>
<tr>
<td>de Sousa Santos</td>
<td>1998</td>
<td>Participation</td>
<td>Allocation and delivery</td>
<td>No</td>
<td>Porto Alegre, Brazil</td>
</tr>
<tr>
<td>Diaz-Cayeros et al.</td>
<td>2010</td>
<td>Participation</td>
<td>Allocation and delivery</td>
<td>Yes</td>
<td>570 municipalities in Oaxaca, Mexico, 1990-2010</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Independent variable category</th>
<th>Main impact category (secondary categories)</th>
<th>Quantitative</th>
<th>Empirical context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldfrank &amp; Schneider Gonçalves</td>
<td>2006</td>
<td>Participation</td>
<td>Governance</td>
<td>No</td>
<td>Rio Grande do Sul, Brazil, 1999-2002</td>
</tr>
<tr>
<td>Heller et al.</td>
<td>2007</td>
<td>Participation</td>
<td>Development</td>
<td>Yes</td>
<td>72 local governments in Kerala, India, 2002</td>
</tr>
<tr>
<td>Jaramillo &amp; Wright</td>
<td>2014</td>
<td>Participation</td>
<td>Allocation and delivery</td>
<td>Yes</td>
<td>100 Peruvian municipalities, 2001 and 2007</td>
</tr>
<tr>
<td>Olken</td>
<td>2010</td>
<td>Participation</td>
<td>Allocation and delivery</td>
<td>Yes*</td>
<td>49 Indonesian villages, 2005-2006</td>
</tr>
<tr>
<td>Olken</td>
<td>2007</td>
<td>Transparency and participation</td>
<td>Governance</td>
<td>Yes*</td>
<td>608 Indonesian villages, 2003-2004</td>
</tr>
</tbody>
</table>

Note: Entries are grouped by independent variable category.
*Field experiment.
**Natural experiment (random assignment of treatment).
***Quasi-experiment (non-random assignment of treatment).
*Outcomes are measured for one year or time period, but the analysis accounts for initial conditions.
*bPurely cross-national analysis with no within-country variation over time.
The Opportunities and Challenges of Digitizing Government-to-Person Payments

Leora Klapper and Dorothe Singer

This paper reviews evidence on the benefits and challenges faced by governments migrating from cash to digital (electronic) government-to-person (G2P) payments. When supported by an appropriate consumer financial protection framework, digital payments enable confidential and convenient financial services, which can be especially important for women. By shifting government wages and social transfers into accounts, governments can lead by example. Digitizing G2P payments has the potential to dramatically reduce costs, increase efficiency and transparency, and help recipients build familiarity with digital payments. Digital wage and social transfer payments can also provide the on-ramp to financial inclusion and in many cases the first account that the recipient has in her own name and under her control. However, digitizing G2P payments is not without its challenges. Most importantly, digitization may require significant up-front investments in building an adequate physical payment infrastructure that is able to process such payments, as well as a financial identification system and a consumer protection and education framework to ensure that recipients have safe, reliable, and affordable access to the digital payment system. JEL codes: D14, G28, O16

More than 100 million poor people worldwide receive a government-to-person (G2P) payment (Demirgüç-Kunt et al. 2015). This includes, for example, government wages, government transfer payments (such as pensions, social benefits, and unemployment benefits), and tax refunds. While it is estimated that 90 percent of high-income countries make their G2P payments “mostly electronically”, over half of developing countries make their G2P payment in the form of cash or
paper-based payments such as checks (World Bank 2012a). Furthermore, 19 percent of adults report receiving a wage or social transfer payment from their government and 28 percent of recipients report receiving such payments in cash, including 39 percent of recipients in developing countries (Demirgüç-Kunt et al. 2015). As a result, governments, businesses, and individuals are bearing the often high cost of cash payments—costs associated with disbursing and receiving cash, the greater probability of “leakage” (fraud and corruption), and the higher incidence of associated crime.

Digital or electronic G2P payments can take different forms. Examples include direct deposits into bank accounts, transfers to pre-paid or stored-value cards that work as a virtual account, or mobile money transfers which may or may not be linked to a mobile money account. Depending on the type of digital payments, recipients can access the funds through an automated teller machine (ATM), at point-of-sale (POS) terminals, banking or mobile money agents, or other means.

One reason many governments still make their G2P payments mostly in cash or paper-based forms is that their digital payments infrastructure might be underdeveloped or have limited coverage in non-urban areas. However, innovations in technology and financial business models, such as mobile money service accounts and agent banking (which use mobile and/or Internet connections to provide real-time financial services), are expanding the reach of the electronic payments infrastructure in many countries. This in turn makes it increasingly feasible to digitize payments.

Governments play a pivotal role when it comes to digitizing payments in an economy. By shifting government wages and social transfers into accounts, governments can lead by example and play a catalytic role in building a digital payments infrastructure and ecosystem where all kinds of payments—including private-sector wages, payments for the sale of agricultural goods, utility bills, school fees, remittances, and everyday purchases—are done digitally. Governments also have an essential role to play in creating an enabling regulatory environment and promoting consumer protection and education to facilitate the shift to digital payments beyond government payments.

In this paper, we review the body of research that has emerged on digital G2P payments and suggest steps that governments can take to hasten the use of digital payments. The second section reviews the benefits for governments and recipients when government wages and social transfer payments are shifted from cash into accounts. Shifting to digital payments has many potential benefits, including lower costs, improved delivery speed, increased transparency, enhanced security, higher financial inclusion, and increased levels of women’s economic empowerment. The third section explores the challenges that countries around the world face as they look to shift their payments from cash into accounts—and the role governments can play in promoting a digital payments ecosystem. Putting in place a robust
system of digital payments requires significant physical infrastructure. The literature also shows that one cannot ignore the human element: new users of digital payments need to be educated on how to withdraw and send payments safely and cost-effectively, as well as on the benefits and risk of other financial services they might be offered (such as credit and insurance). Unless users can use the services comfortably and have confidence that financial service providers can be trusted, recipients of electronic payments will withdraw their payments and save the money and transact in cash—thereby losing the potential benefits of financial inclusion.2 The last section concludes.

Benefits of Digital G2P Payments

Shifting to digital G2P payments has many potential benefits, for both senders (governments) and receivers (persons): it can improve the efficiency of making payments by lowering the cost of disbursing and receiving them, and by increasing the speed of payments. Digitalization can increase the transparency of payments, and thus reduce the likelihood of leakage between the sender and receiver. Further, digitalization can enhance the security of payments and thus lower the incidence of associated crime. Shifting to digital payments can also provide an important first entry point into the formal financial system. Furthermore, by increasing the privacy of payments and increasing control over the funds received, shifting to digital payments can contribute to women’s economic empowerment.

Lower Costs for Governments

Shifting from cash to digital payments can lead to significant cost savings in the long term. The potential cost savings are especially striking when considering large-scale payments from government to people, such as government wage or social transfer payments. A rigorous evaluation of a social transfer program in Niger has shown that the variable cost of administering social transfer is 20 percent lower by mobile transfer than by manual cash distribution (Aker et al. 2013). In South Africa, the cost of disbursing social grants in 2011 by payment card was one-third that of manual cash disbursement (R13.50 compared to R35.92; Consultative Group to Assist the Poor 2011b). And in Mexico, a study estimates that the Mexican government’s shift to digital payments (which began in 1997) trimmed its spending on wages, pensions, and social welfare by 3.3 percent annually, or nearly $1.3 billion (Babatz 2013). The savings in Mexico are estimated to be primarily due to reduced losses due to unauthorized or incorrect payment. There are also some savings due to interest earned by not needing to deposit funds in advance of payments or to pay banks fees for distributing cash payments.
Lower Costs for Recipients

Shifting to digital payments can also result in direct cost savings for recipients. Recipients of cash payments for government wages or transfers (especially in rural areas) often have to travel a considerable distance to designated locations such as a government office or bank branch, either of which may only be available in a regional capital, in order to receive their payments. This can result in significant travel time and travel expenses, and is further costly in terms of lost wages or income forgone while traveling and waiting to collect a payment. In the past, recipients of digital payments may have needed to incur the cost and time to travel equally far distances to a bank branch to cash-out their electronic payment. However, technological innovations have made it possible to bring cash-out points ever closer to recipients in the form of ATMs, merchant point-of-sale (POS) terminals, and banking or mobile money agents. Increasingly, recipients can reduce the need to travel to a cash-out point altogether by making electronic payments directly from their accounts, such as using a card or phone to make retail and person-to-person payments, or paying bills or making purchases using a phone or over the Internet.

A randomized control trial in Niger, for example, found that administering social transfers by mobile transfer reduced the cost to recipients of collecting the transfer to one-quarter of the cost of collecting the manual cash transfer by reducing overall travel and wait time. Recipients of mobile transfers traveled shorter distances and waited for shorter periods of time to cash-out their electronic government transfer payments. For example, travel time to a cash-out point was reduced by 40 minutes, compared to manual cash distribution, plus an additional three hours wait time, on average, to receive a manual cash transfer. Based on average agricultural wages, the time savings attributable to the digital transfer channel for each payment translated into an amount large enough to feed a family of five for a day (Aker et al. 2013).

Improved Speed and Timely Delivery

In contrast to cash payments that travel at the speed of their carrier, digital G2P payments can be virtually instantaneous, regardless of whether the sender and receiver are in the same town or district. In digital form, government assistance payments in times of disaster can be made without delay when the need is greatest. For example, an analysis of several NGO programs in Haiti following the earthquake in 2010 found that mobile money payments were faster and safer than traditional physical cash delivery or voucher programs (Dalberg 2012). In addition, the Liberian government was able to quickly pay thousands of Ebola workers, often
working in rural and afflicted areas, by opening accounts for health workers and making payments digitally (Better Than Cash Alliance 2015).

Increased Transparency and Reduced Leakage

Digitizing G2P payments can also increase transparency and ensure that people receive wage or social transfer payments in full—and that only those eligible to receive payments do so. Given the fungibility and transactional anonymity of cash, cash payments are subject to “leakage” (payments that do not reach the recipient in full) and “ghost” (fake) recipients, particularly in the context of government transfers. By moving toward digital G2P payments, the traceability of the payment process is improved. First, recipients have digital records of the amount of the payments they are to receive, and the number of potential leakage points is lessened by reducing the number of people a payment needs to go through to reach the recipient. Second, digital payments generally require more stringent identification documentation of recipients to comply with documentation requirements for financial service providers, making it harder for ghost recipients to remain undetected.

Evidence from India shows that making social security pension (SSP) payments digitally via payment cards, compared to manual cash payout at the village level by a government official, results in a 1.8 percentage point lower incidence of bribe demands for obtaining the payment (compared to an incidence of 3.8 percentage points for manual cash payments, which results in a 47 percent reduction) and the incidence of ghost recipients fell by 1.1 percentage points (Muralidharan, Niehaus, and Sukhtankar 2014). Similarly, evidence from Argentina shows that depositing social transfer payments under the Plan Jefes program, a large national anti-poverty program, directly into an account instead of manual cash payout significantly reduced the demand for kickbacks from individuals or organizations that helped recipients into the program. When payments were made in cash, 4 percent of recipients reported paying kickbacks—after payments were digitized and made directly the percentage dropped to just 0.03 percent (Duryea and Schargrodsky, 2008).

Increased Security and Lower Crime

Recipients of G2P payments in cash often not only have to travel considerable distances to receive their payments, but are also particularly vulnerable to street crime once they carry the cash due to its transactional anonymity. While security is a concern when traveling with any large amount of cash, this concern is especially salient for regular cash payments, such as wage or social transfer payments that are received at publicly known points in time. Evidence from the United States
shows that when the federal government introduced the Electronic Benefit Transfer (EBT) in the mid-1990s, and thus switched from delivering social cash transfers by paper checks, which needed to be cashed, to electronic debit cards, the overall crime rate over the next 20 years was reduced by almost 10 percent as a direct result. This corresponded to 47 fewer crimes per 100,000 people per county per month as a direct result of switching welfare benefits from cash to credit (Wright et al. 2014).

When payments are received into accounts they can also be held more securely than manual cash payments. Recipients can store the payments in their accounts until needed and cash out smaller amounts at their convenience or make direct purchases using POS terminals in stores or directly transfer funds onwards to pay bills—assuming a digital payments infrastructure exists that makes this possible and convenient.

Increased Financial Inclusion

Another benefit of shifting G2P payments from cash into accounts is increased financial inclusion—which is broadly defined as both access to and use of appropriate, affordable, and accessible financial services. Empirical evidence at the micro and macro level shows that inclusive financial systems are an important component to economic and social progress on the development agenda (see Karlan and Morduch 2010; Cull et al. 2014; Beck 2015; Demirgüç-Kunt, Klapper, and Singer 2017 for overviews). Worldwide, 62 percent of adults have an account either at a financial institution or through a mobile money provider according to data from the 2014 Global Findex database (Demirgüç-Kunt et al. 2015). But while account ownership is nearly universal in high-income economies, with 94 percent of adults reporting having an account, only 54 percent of adults in developing economies report having an account at either a bank or other financial institution, or with a mobile money service provider. For women in developing countries, the situation is worse: Only 50 percent reported having an account, compared to 59 percent of men. And less than half (46 percent) of adults living in the poorest 40 percent of households within economies in developing countries have an account. Globally, 2 billion adults do not have an account.

Without access to the formal financial system, women, poor people, small businesses, and otherwise excluded people must rely on their own (limited) informal and semiformal savings and borrowing to finance educational, entrepreneurial, and other investments, thus making it harder to alleviate income inequality and spur broad-based economic growth. However, those who are excluded from the formal financial system are likely to be recipients of payments—including government wages and government-sponsored social transfers.
Digitizing G2P payments and shifting them into accounts presents an opportunity to expand account ownership among the unbanked. Shifting the payment of government wages from cash into accounts could increase the number of adults with an account by up to 35 million worldwide. Doing the same for government transfers could increase the number with an account by up to 130 million. Overall, digitizing G2P payments could increase the number of adults with an account by up to 160 million by bringing into the financial system the 8 percent of unbanked adults worldwide who receive either government wages or transfers only in cash. Indeed, digitizing G2P payments is a proven way to expand account ownership: in developing economies, about one-third of adults who received a government wage or transfer payment into an account in 2014 said they opened their first account specifically for that purpose. Overall, 5 percent of banked adults in developing countries opened their first account to receive public-sector wage payments or social benefits transfers (Demirgüç-Kunt et al. 2015).

Moreover, data for 123 countries show that digital payments that reduce the cost and increase the convenience of financial transactions may expand the pool of eligible account users and encourage existing account holders to use their accounts with greater frequency and for the purpose of saving (Allen et al. 2016). Studies in Mexico and Nepal show that following the provision of accounts to poor households, new account holders continued to deposit and maintain balances in their accounts, which led to a significant increase in household savings (Aportela 1999; and Prina 2012). From Mexico, there is also evidence that accounts opened through a social transfer program increased the frequency of remittances received through formal payment channels (Masino and Niño-Zarazúa 2014). Further, the randomized introduction of mobile money in rural Mozambique led to the substitution of mobile money for informal savings (Batista and Vicente 2013). However, payments into an account do not automatically translate into account use. As discussed in the next section, experience with social transfer programs in Brazil, Colombia, and Mexico has shown that recipients are unlikely to automatically use bank accounts for more than withdrawing benefits (Bold, Porteous, and Rotman 2012).

Increased Women’s Economic Empowerment

Evidence suggests that digitizing G2P payments might be especially valuable to women, who benefit from the greater confidentiality and control such payments offer, and can contribute to their economic empowerment within their households. In contrast to cash payments, the arrival of a digital payment is often private information that allows the recipient to conceal the payment, at least temporarily, from other household members or friends who may place demands on the use of the
money. Sociocultural norms and other factors might prevent women from controlling their own money and assets. But payments into an account make it harder for family and friends to access the funds and thus might give recipients greater control and agency with regard to how the money will be used. From the social cash transfer program in Niger, for instance, there is evidence that greater privacy and control of mobile transfers, compared to manual cash transfers, shifts intra-household decision-making in favor of women, that is, the recipients of the social cash transfer (Aker et al. 2013). A large body of empirical literature suggests that income in the hands of women, compared to men, is associated with improvements in children’s health and larger shares of household spending on nutritious food, healthcare, and housing (for an overview, see Duflo 2012).

The Challenges of Digitizing Payments—and What Governments Can Do

The benefits of digital payments go well beyond convenience: if provided efficiently and effectively, they can transform the financial lives of those who use them. But digitizing G2P payments is not without challenges. These challenges include making up-front investments in payments systems infrastructure, taking steps to guarantee a reliable and consistent digital payments experience, ensuring the regulatory framework conducive to building a digital payments ecosystem, and ensuring that recipients understand how digital payments work and can be accessed. It is also important to educate new digital payments recipients on the basic interactions involved in a digital payments system—using and remembering personal identification numbers (PINs), understanding how to access the payments, and knowing what to do when something goes wrong. The challenges also include fostering a digital payments ecosystem beyond G2P payments to reap the full benefits of digitizing G2P payments. And finally, digitizing G2P payments depends on the political will to do so.

An electronic payments system will not be effective, and could even have adverse effects, if it does not work well. Payment delays, network outages, or working with agent networks in which liquidity is a problem can undermine an entire electronic transfer program, as recipients fail to trust or understand the new system. A reliable payments system also needs to have effective safeguards in place to protect against fraud and data security breaches.

Governments, the private sector, and the international development community all have important roles in making digital payments systems more efficient and more accessible to low-income consumers. But governments play a pivotal role when it comes to digitizing payments in an economy. By shifting G2P payments into accounts, governments can play a catalytic role in building a cost-effective,
sustainable digital payments ecosystem. This is particularly relevant in countries where this has been a challenge due to low population density and low incomes. Governments can also help facilitate the shift to digital payments beyond government payments by creating an enabling legal and regulatory environment, improving identification, and promoting consumer education and protection.

The benefits of digitizing government payments must of course be weighed against the potential costs of the improvements to the payments infrastructure necessary to do so. But innovations in technology and business models in the financial sector in recent years continue to reduce the costs digitizing payments and provide private sector solutions for cash-out points.

The Digital Payments Ecosystem

More than 2 billion adults worldwide, in both developed and developing countries, are unbanked. Of this group, the majority is not excluded by choice; rather, cost, distance, documentation requirements, and other variables make it challenging or impossible to access banking services (Demirgüç-Kunt et al. 2015). Governments can encourage the expansion of account ownership and usage and the growth of digital payments by leading by example. The sheer volume of government payments, from salaries to pensions and social cash transfers, has the potential to add significant volumes of transactions to financial service providers. Electronic G2P payments could not only increase the number of adults with a new account, but also increase usage of existing accounts. About 270 million adults around the world receive government wage or transfer payments in cash according to data from the 2014 Global Findex database, including 110 million adults who already have an account (Demirgüç-Kunt et al. 2015). By shifting payments from cash into accounts, governments can make a critical contribution to the commercial viability of financial infrastructure in currently underserved areas such as rural locations, and can help reach especially low-income individuals. For instance, the financial and private sector might be incentivized to invest in POS networks that would allow government recipients to store money in their account and conveniently make payments directly using their debit card. This does not mean that governments will necessarily provide these digital payments directly by themselves. Rather, in partnering with private-sector payment service providers, governments can help jump start the expansion of the digital payments infrastructure.

Digital payments can take different forms. Examples include direct deposits into bank accounts, payment cards, and mobile payments. It is important for governments to carefully consider which type of digital payment channel is best suited for any particular case; this depends on a number of context- and country-specific factors including broad economic, demographic, and policy environment factors (Faz
and Moser 2013). For instance, in developed countries with advanced and broadly-used banking systems, digital payments by direct deposit into bank accounts are already common. In developing countries with more rudimentary financial systems that provide services to a limited segment of the population, and often primarily in urban areas, digital payment channels based on payment cards or mobile transfers may be more suitable.

The optimal digital payment channel may also vary within a country or within a specific payment type. For example, Brazil’s cash transfer program, Bolsa Família, which makes payments to more than 13 million families, allows recipients to choose whether to receive the cash transfer by payment cards, by direct deposit into a no-frills bank account, or, in rare circumstances, in the form of a manual cash payment (Consultative Group to Assist the Poor 2011a).

Shifting government payments into accounts is an important first step, but as long as digital payments are cashed out immediately upon receipt, their contribution toward financial inclusion, building a digital payments system, and reaping the full benefits of moving beyond a cash-based payments system will be limited. Only by building a digital payment ecosystem that encourages recipients to keep funds digitally by offering store-of-value or savings functionality, direct payment for purchases at POS terminals, and bill payment functionality will the full benefits of moving payments into accounts be realized. But payments into an account do not automatically translate into the use of accounts. Experience with social transfer programs in Brazil, Colombia, and Mexico has shown that recipients are unlikely to automatically use bank accounts for more than withdrawing benefits (Bold, Porteous, and Rotman 2012). This may be due to a lack of knowledge that the payment is not lost if not withdrawn in full, unfamiliarity with formal financial products and the benefits associated with them, uncertainty over whether there are costs associated with the use of the account, or a lack of trust in banks to keep funds safe. Realizing the full potential benefits of electronic payments via increased usage of payments and savings thus depends on products that allow for those uses, as well as on clear communication regarding these features.

Ensuring that government payment recipients can use the funds to easily and conveniently make digital payments to, and receive digital payments from, the many parties that they deal with financially—merchants, friends and family, employers, schools, utilities companies—will be key in building a digital payment ecosystem. Making digital payments cost effective and sustainable, especially for low-income and rural populations, will require leveraging a number of different technologies including mobile phones, ATMs, POS terminals, and online services. No one provider or sector can justify an investment in all of these elements or handle the contractual requirements of dealing with so many players. Rather, multiple players must be able to interconnect where necessary to provide individuals with a wide range of services, and must be able to do so on fair and equitable cost and access terms.
An additional benefit of building such a digital payment ecosystem is that the cash-out constraint will gradually lessen. This will be especially important in rural areas that are typically net-recipients of social transfer payments and where cash-management issues are a considerable challenge (Faz and Moser 2013; Bold, Porteous, and Rotman 2012).

The Physical Payments Infrastructure

National payments systems and the accompanying financial infrastructure are the backbone of digital payments. While digital G2P payments can be more cost effective in the long term, building an adequate physical infrastructure for reliable experiences with digital payments can require significant up-front investments (World Bank 2012b). Countries with advanced and broadly used payment and banking systems might already have a physical infrastructure in place to process digital payments. But in developing countries with more rudimentary payments systems and where such infrastructure is concentrated in urban areas, developing an adequate payment infrastructure, including a physical network, to deliver digital payments to all corners of the country is a significant challenge. These difficulties are often underestimated, as a study documenting the experience of digitizing G2P payments in four low-income countries (Haiti, Kenya, Uganda, and the Philippines) shows (Zimmerman, Bohling, and Rotman Parker 2014).

The high cost of traditional brick-and-mortar bank branches often concentrates financial access points in urban areas where higher population density and often more affluent customers make them profitable. However, in recent years, innovations such as mobile financial services and agent banking are increasingly providing ways to reach even rural and low-income individuals in a sustainable and cost-effective manner, making access to financial services through ATMs or POS terminals viable even in small communities. Furthermore, leveraging and modernizing existing infrastructure such as post offices can also provide new opportunities to reach financially underserved communities.

While the widespread use of mobile phones in developing countries seemingly suggests it would be easy to provide digital payments by mobile transfer even in countries with the most rudimentary banking systems, widespread mobile phone use is not sufficient. Reliable payments into mobile money accounts face significant infrastructure challenges. The lack of reliable electricity supply with which to power mobile phones, cell towers, and payments system IT infrastructure and limitations in mobile data network coverage are major obstacles to the expansion of digital financial services in rural areas. Even Kenya, which is well-regarded for its mobile money infrastructure, was unable to make a mobile money-based solution for a social transfer payment work from 2010 to 2012 due to network connectivity issues, and instead resorted to disbursing the transfer payments into accounts.
Digital Financial Identification

Central to financial transactions is the ability of financial service providers to accurately identify individuals, which is typically done via government-issued identification documents. In 2014, more than 2 billion people worldwide did not have any formal identification (Demirgüç-Kunt et al. 2015). Yet to open a bank or mobile money account or to make or receive most digital financial transactions—including government payments—customers generally need to submit relevant documents like government-issued identity cards or birth certificates.

Governments can create a biometric (digital) financial identification system to establish identity for customers who lack traditional paper documentation such as birth certificates. Biometric technology might also be leveraged to overcome the lower levels of technical adoption and literacy among government transfer recipients. For example, a thumbprint or retina scan might replace the need to remember long PIN numbers.

The Legal and Regulatory Environment

For the private sector to be able to provide digital payments solutions and contribute to building a digital ecosystem, it needs the space to develop innovative payment products. This means a regulatory environment that recognizes the contributions of all financial sector players, including non-banks such as payment

with debit cards issued by a financial institution (Zimmerman, Bohling, and Rotman Parker 2014).

Providing physical access to cash-in and cash-out points and ensuring sufficient liquidity at access points, including in rural areas, is one of the key challenges in moving toward digital payments. Even in a digital payments environment, cash-out points are a critical feature. While recipients of digital payments ideally keep their funds digitally and make purchases and pay bills electronically, the reality is that many countries, especially developing countries, still have a ways to go to become economies where digital payments are accepted for everyday purchases at local retail stores and markets. As a digital ecosystem evolves and allows recipients of digital payments to stay digital by making digital payments, cash-out constraints will lessen. However, in the meantime most people need to be able to cash-out at least part of the payments they receive; thus, people will look for a reliable cash-out experience, and financial systems will need to deliver one. Indeed, a reliable cash-out experience is key to the success of digital payments (Kendall and Voorhies 2014). Building an infrastructure that provides a reliable cash-out experience, however, remains a significant challenge, especially in rural areas that are typically net-recipients of government wage and transfer payments.
services providers and mobile network operators, which can play an important role in reaching traditionally financially underserved segments of the population such as the poor and those living in rural areas. Providing a clear and functional legal and regulatory framework for these new players is important to ensure both a level playing field between the different actors in the digital payment space and adequate protection of consumer funds. To that end, regulators have to define, among other things, who can provide financial services and act as an agent. Regulators also must find the appropriate balance between promoting interoperability and letting the market decide to foster a digital payments ecosystem. Regulators must also coordinate with each other, especially across complementary sectors such as financial services and telecommunications.

**Financial Consumer Protection and Education**

Consumer education is an important part of ensuring that recipients of digital G2P payments become familiar with digital payment systems and feel comfortable with the payment process and financial instrument. This includes understanding the program, payment process, payment conditionality (if applicable), and recourse mechanisms. If recipients do not understand how the program works or if payments are inconsistent, recipients will lose trust in the system.

Low-income recipients and those living in remote areas might not be familiar or comfortable with using a digital payment system. This is especially a challenge for social cash transfer programs that by definition often target the poorest segments of the population. Assuring basic financial literacy is necessary; for example, recipients should be educated about accessing and cashing-out their payment, using and remembering their PINs, understanding how much money they should receive at each payout period and how fees (if any) are incurred, and knowing what to do if something goes wrong (Zimmerman, Bohling, and Rotman Parker 2014).

Addressing these challenges is necessary for effective product adoption. A study of a government cash transfer program to low-income women in Pakistan illustrates some of the challenges that come with making digital payments to a population that is, for the most part, illiterate. Initially, many recipients did not understand the cash-out process at the banking agent, nor were they able to use an ATM on their own to withdraw payments due to insufficient communication and a product design that was not tailored to the needs of the recipients. Subsequent education efforts focusing on how to use the digital payment product, and adjustments in the design of the product, eventually led to an increase in the understanding and use of the product (West and Lehrer 2014). It needs to be stressed that the onus is also on providers of financial services, including the private sector, to design digital payment solutions that are tailored to the needs of individuals and are easy to understand.
Digitizing payments can bring people into the formal financial system for the first time and give them potential access to a range of financial products, which might be complex or bundled products, raising the associated risks for consumer segments with weaker financial capability. There are also significant issues concerning fraudulent, misleading, and unfair commercial practices, and consumers require the right to dispute any unauthorized transaction. Data privacy and security are important issues to be raised, and governments should safeguard personal information against loss or theft. Consumers should have access to appropriate—that is, independent, impartial, and free—redress mechanisms.

Political Economy Issues

Last but not least, political economy issues might present a significant challenge when it comes to shifting G2P payments from cash into accounts. A system that is hard to track, is less private, and entails the use of transitionally anonymous cash creates opportunities for individuals at every step of the money transfer to skim off some of the funds. Those benefitting from cash payments may thus work to obstruct the shift to digital G2P payments.

Conclusion

This paper reviews evidence on the benefits and challenges of governments migrating from cash to digital (electronic) payments. Digitizing G2P payments has the potential to dramatically reduce costs, increase efficiency and transparency, and help recipients build familiarity with digital payments. Doing so is not without challenges and may require significant up-front investments in building an adequate physical payment infrastructure and providing consumer education. But when supported by an appropriate financial consumer protection framework, digital payments enable confidential and convenient financial services, which can be especially important for women. Migrating to electronic G2P payments offers an opportunity to rapidly scale up access to financial services and provides an on-ramp to financial inclusion, and in many cases, leads to the first account that the recipient has in her own name and under her control.

Notes

Leora Klapper and Dorothe Singer are in the Finance & Private Sector Team, Development Research Group, World Bank. Corresponding author is Leora Klapper, email: lklapper@worldbank.org. The authors thank the UNCDF Better than Cash Alliance and the Bill & Melinda Gates Foundation for helpful comments.
1. A traditional bank account to hold the underlying balance is generally not necessary (World Bank 2012b).

2. For additional information, see the “Payment Aspects of Financial Inclusion (PAFI)” taskforce report on how payment systems and services promote financial inclusion efforts (Bank of International Settlements and World Bank Group 2016).

3. This comes at the potential risk that recipients might also withhold funds from which the entire household is entitled to benefit, such as in the case of certain social transfer payments.

References


How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence

Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs

The Impacts of Fiscal Openness

The Opportunities and Challenges of Digitizing Government-to-Person Payments