

# Essays on Financial Incentives to Healthcare Providers

Inaugural-Dissertation  
zur Erlangung des akademischen Grades eines Doktors  
der Wirtschafts- und Sozialwissenschaften  
der Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von  
Benjamin K. Chibuye, M.A.  
aus Lusaka, Zambia

Kiel, 2021



# Essays on Financial Incentives to Healthcare Providers

Inaugural-Dissertation  
zur Erlangung des akademischen Grades eines Doktors  
der Wirtschafts- und Sozialwissenschaften  
der Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von  
Benjamin K. Chibuye, M.A.  
aus Lusaka, Zambia

Kiel, 2021

Gedruckt mit Genehmigung der  
der Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

Dekan:	Prof. Dr. Kai Carstensen
Erstberichterstattender:	Prof. Dr. Menusch Khadjavi
Zweitberichterstattender:	Prof. Dr. Tobias Heidland

Tag der Abgabe der Arbeit:	6 July 2021
Tag der mündlichen Prüfung:	24 August 2021

# Dedication

*To my dear parents*

Colonel D.C. Chibuye (rtd) & Mrs Phyllis Chibuye

# Acknowledgements

This thesis is not only the culmination of years of my own dedicated effort, but also a reflection of the exceptional support, input and inspiration I have received from many over the years.

First and foremost, I wish to thank my supervisor Prof. Dr. Menusch Khadjavi for believing in me, for inspiring me and for supporting my doctoral journey. I have learnt so much from you and I could not have asked for a better advisor and mentor over the course of my doctoral journey. I also wish to thank Prof. Dr. Tobias Heidland for agreeing to serve as my second supervisor and thereby playing an equally critical role in my growth.

I wish to thank the German Academic Exchange Service (DAAD) for the generous doctoral research grant that has facilitated my doctoral studies. I wish to thank all DAAD staff particularly acknowledging Lorena Strikic, Claudia Nagler and Michael Hellenblink.

I have also benefited from tremendous support from numerous colleagues. I particularly wish to thank Patrick Nuess, Peter Hangoma, Torsten Figueiredo Walter, Twivwe Siwale, Jennifer Uju Okonkwo and Kacana Khadjavi. Among my many dear friends that have supported and encouraged me continuously, I would like to particularly thank Eliya Lungu, David Bvulani, Sydney Phiri, Peter Mulenga, Herryman Moono and Stanslous Saka.

My intellectual journey has been characterised by numerous phases. Particularly noteworthy is my time spent at the University of Zambia and the International Growth Centre. For the former phase of my journey, I particularly wish to thank Prof. Felix Masiye, Dr. Chrispin Mphuka (late) and Mr. Caesar Cheelo. For the latter phase, I particularly wish to thank Dr. Anand Rajaram and Dr. Dennis Chiwele.

Last but certainly not the least, I wish to particularly thank my family. Your love, encouragement and support are reasons why I am who I am today. I share this triumph with you all.

# Contents

<b>1</b>	<b>Introduction to the Dissertation</b>	<b>1</b>
<b>2</b>	<b>The Dark Side of Performance-Based Financial Incentives in Healthcare – Experimental Evidence from Zambia</b>	<b>15</b>
2.1	Introduction . . . . .	17
2.2	Experimental Design . . . . .	21
2.2.1	Tasks 1 and 2 . . . . .	21
2.2.2	Treatments . . . . .	24
2.2.3	Procedures . . . . .	28
2.3	Data . . . . .	30
2.4	Results . . . . .	30
2.5	Conclusion . . . . .	36
2.A	Appendix to Chapter Two . . . . .	47
2.A.1	Lubwa Receives Medicines . . . . .	47
2.A.2	Instructions . . . . .	48
<b>3</b>	<b>Comparing the Productivity of Performance-based and Unconditional Financial Incentives in Improving Infant Birthweight in Zambia</b>	<b>76</b>
3.1	Introduction . . . . .	78
3.2	Related Literature . . . . .	82
3.2.1	Literature on Incentivising Economic Agents . . . . .	82
3.2.2	Literature on Health and Social Programmes . . . . .	83
3.3	Context and Program Description . . . . .	85
3.4	Data and Descriptive Statistics . . . . .	86
3.5	Empirical Framework . . . . .	89
3.5.1	Identification Strategy . . . . .	89
3.5.2	Estimation Strategy . . . . .	91
3.6	Results . . . . .	95
3.6.1	Difference-in-Differences Estimates . . . . .	95
3.6.2	Regression Discontinuity Estimates . . . . .	100
3.7	Cost-Effectiveness . . . . .	101
3.8	Robustness Checks . . . . .	103
3.8.1	Original ZRBF Districts . . . . .	103
3.8.2	Triple Difference Estimates . . . . .	103
3.8.3	Parallel Trends . . . . .	104
3.9	Discussion and Conclusion . . . . .	106
3.A	Appendix to Chapter Three . . . . .	118
3.A.1	Missing Birthweight Observations . . . . .	119

3.A.2	Original ZRBF Districts . . . . .	120
3.A.3	Triple Difference and Low-Impact Sub-sample Estimates . . . . .	123
3.A.4	High-Impact Sub-sample ANC Estimates . . . . .	126
3.A.5	Difference-in-Difference Birthweight and Low Birthweight Estimates: District by Year Time-Trends . . . . .	129
3.A.6	Running Variable Density . . . . .	131
<b>4</b>	<b>Public Sector Compensation and Rural-Urban Gaps in Service Delivery: Evidence from Zambia</b>	<b>132</b>
4.1	Introduction . . . . .	134
4.2	Background . . . . .	136
4.2.1	The Rural-Urban Gap in Health Service Provision . . . . .	136
4.2.2	The Zambian Healthcare System . . . . .	137
4.2.3	The Rural Hardship Allowance Scheme . . . . .	138
4.3	Data . . . . .	139
4.4	Empirical Strategy . . . . .	144
4.4.1	Identification Strategy . . . . .	144
4.4.2	Estimation Strategy . . . . .	146
4.5	Results . . . . .	147
4.6	Robustness checks . . . . .	154
4.6.1	Bandwidth Choice . . . . .	154
4.7	Conclusion . . . . .	155
4.A	Appendices to Chapter Four . . . . .	166
4.A.1	Assistance at Delivery . . . . .	166
4.A.2	Placebo Outcomes and Bandwidth . . . . .	167

# List of Tables

2.1	Summary of Experimental Tasks and Treatments . . . . .	25
2.2	Summary Statistics of Covariates . . . . .	29
2.3	Regressions of Task 1 Score and Amoxil Allocations . . . . .	33
3.4.1	Baseline (2007) Means and Changes in Sample Characteristics . . . . .	87
3.6.1	Difference-in-Differences Estimates for Birthweight and low Birthweight. . .	96
3.6.2	Difference-in-Differences Estimates for the Number, Timing and Quality of Ante-natal Care Visits: Full Sample . . . . .	99
3.A.1	Missing Birthweight Observations by Treatment . . . . .	119
3.A.2	Baseline (2007) Means and Changes in Sample Characteristics for the 30 Original ZRBF Districts . . . . .	120
3.A.3	Difference-in-Differences Birthweight and Low Birthweight Estimates: 30 Original ZRBF Districts. . . . .	121
3.A.4	Difference-in-Differences Full Sample Estimates for Ante-natal Care Visits: 30 Original ZRBF Districts. . . . .	122
3.A.5	Triple Difference Birthweight and Low Birthweight Estimates. . . . .	124
3.A.6	Difference-in-Differences Birthweight and low Birthweight Heterogeneous Estimates: Low Impact Sub-samples. . . . .	125
3.A.7	Difference-in-Differences Estimates for Ante-natal Care Visits: Women with No or Only a Primary Education . . . . .	127
3.A.8	Difference-in-Differences Estimates for Ante-natal Care Visits: Women Aged 20 Years Old or Less or Older Than 35 years of Age. . . . .	128
3.A.9	Difference-in-Differences Birthweight and Low Birthweight Estimates: Dis- trict by Year Fixed Effects. . . . .	130
4.3.1	Health Facility Characteristics . . . . .	143
4.5.1	Receipt of Hardship Allowance: First Stage . . . . .	149
4.5.2	Reduced Form Estimates of the Impact of the Rural Hardship Allowance . .	151
4.5.3	Fuzzy RDD Estimates of the Impact of the Rural Hardship Allowance . . .	153
4.A.1	Impact of the Rural Hardship Allowance on Placebo Outcomes . . . . .	168
4.A.2	Impact of Eligibility on Health Facility Staffing: Robustness to Bandwidth Selection . . . . .	169

# List of Figures

2.1	Health-Framed Real Effort Task Tool . . . . .	22
2.2	Medicine Allocation Task Tool . . . . .	24
2.3	The Distribution of Task 1 Scores . . . . .	31
2.4	Health-Framed Real Effort Task Scores by Treatment . . . . .	31
2.5	Medicine Allocations by Treatment . . . . .	32
2.6	Participant Justifications for Medicine Allocations . . . . .	35
2.A.1	Lubwa Rural Clinic Receives of Donated Medicines. . . . .	47
3.3.1	Districts by Treatment . . . . .	85
3.5.1	Event Study Estimates . . . . .	94
3.6.1	Regression Discontinuity Plots . . . . .	102
3.8.1	Pre-trends Test using the Multiplegt Estimator . . . . .	105
3.A.1	Density of Running Variable at Treatment Threshold. . . . .	131
4.2.1	Rural-Urban Access to a Skilled Health Worker and Health Facility during Delivery . . . . .	137
4.2.2	Categorization of Districts . . . . .	139
4.2.3	Rural Hardship Allowance Eligibility and Receipt in Chipata district . . . .	140
4.4.1	Health Facility Density around the Eligibility Threshold . . . . .	145
4.5.1	Receipt of Hardship Allowance: First Stage . . . . .	150
4.5.2	Health Facility Staffing (Reduced Form) . . . . .	152
4.A.1	Assistance at Delivery: Friend, Relative or Traditional Birth Attendant. . .	166

# Chapter 1

## Introduction to the Dissertation

The critical role of the state in the development process has long since been established. What remains less clear, is how best to motivate and incentivise the front-line public sector employees tasked to deliver public services. In recent years, many low-income developing countries have sought to establish more productive, effective and equitable organisational structures. This has been motivated by the desire to foster growth by overcoming the inefficiency and low quality of service provision that have long since been associated with public service delivery. To this end, many such countries have sought to reform the rigid and formulaic civil service wage structures, traditionally based on education, seniority and rank with limited scope to reward performance (Finan et al., 2017). This reform has come in two common forms. The first, whose goal has been to enhance productivity, has involved augmentation of existing incentive structures through the use of performance-based financial incentives, which have been traditionally associated with the efficiency, productivity and profitability that exist in the private sector. The second, whose goal has been to enhance equity, has involved augmentation of existing incentive structures through the use of unconditional hardship-related pay increases, granted based on geographic location decisions to less attractive geographic regions.

While also having already inherent equity concerns, many low income developing countries have in recent years, also sought to enhance the productivity of one sector in particular - the health sector. Indeed, the relatively lower quality of healthcare in low income developing countries, has also been linked to low levels of healthcare provider morale, effort and a so called 'know-do gap', in which healthcare providers 'do' far less in the provision of healthcare that they 'know' through their training (Das and Hammer, 2014; Gertler and Vermeersch, 2012). The ultimate aim of these efforts has been to improve health outcomes in line with the global development agenda and the Sustainable Development Goals (SDGs). This has implied that the health sector in many low income developing countries has been among the sectors that have implemented both types of public sector incentive structure reform at a rapid pace. Given the pace at which these reforms have been undertaken, several critical questions have remained unanswered: Are performance-based financial incentives more productive relative to other types of financial incentives in the health sector? Could they be associated with adverse unintended effects? How do these high-powered financial incentives interact with the pro-social and other regarding nature of preferences in the health sector? What is the impact of unconditional pay increases granted to healthcare providers based on geographic location decisions?

This doctoral thesis is a cumulative thesis comprising three individual research articles. Taken both individually as well as together, they present evidence-based, rigorous answers to the above, as well as other related questions. This thesis thus contributes to several strands of the literature, with a focus on enhancing our collective understanding of the promises and pitfalls of incentivising healthcare providers.

In the discussion that follows, I provide a bird-eye view of each of the chapters of this thesis, particularly highlighting details on authorship, taking special care to highlight my own contribution. After this, I synthesize the contributions of the chapters to the literature.

Chapter one is titled, "The Dark Side of Financial Incentives in Healthcare - Experimental Evidence from Zambia." In the study underlying this chapter, we conduct a 'health-framed' field experiment with Zambian medical doctors and medical students, utilising two novel experimental tasks; a health-framed real-effort task and a medicine allocation task. The main contributions of this article to the literature come in the form of documenting the productivity and distortion effects of performance-based financial incentives and contrasting these effects with those of unconditional financial incentives all within a field experiment. Results-based Financing (RBF) health programmes, that have been implemented in close to 35 developing countries worldwide, have assumed that performance-based financial incentives yield a productivity effect, arising from greater healthcare provider motivation due to the performance-based financial incentives. The potential for a distortion effect between incentivised and non-incentivised aspects is assumed to be non-binding or inconsequential. The study, sought to establish whether a performance-based financial incentive could indeed lead to a productivity effect, without being accompanied by a distortion effect on healthcare provider decision making, all within the context of a low-income developing country such as Zambia that has implemented an RBF programme. We document three separate sets of results. Firstly, a strong distortion effect due to the performance-based financial incentive. This implies that the performance-based financial incentive is interpreted by healthcare providers as a payoff maximising incentive, distorting provider decision-making in favour of an incentivised medicine in the medicine allocation task. The second result is the absence of a productivity effect of the performance-based financial incentive. This arises because the performance-based financial incentive is not interpreted as a signal to compensate for the distortion effect with commensurately higher effort levels in the health framed real effort task. Owing to this fact, productivity levels in the health-framed real-effort task due to performance-based and unconditional financial incentives are virtually identical. Thirdly, three separate nudge treatments are unable to counteract the distortion effect of the performance-based financial incentive, nor do they support the productivity effect. These findings have important implications, both for the literature and for public policy dealing with the design of health programmes that use performance-based financial incentives to motivate healthcare providers.

Chapter one is jointly authored with Peter Hangoma and Menusch Khadjavi. I contributed to this article at all stages of its development, from conception of the research idea, design of the experimental tasks and data collection tools, securing of funding, implementation of the field experiment in Zambia, data analysis and preparation and revision of draft versions of the article. In addition to appearing in summarized form on the International Growth Centre (IGC) website, this article was presented at several international conferences. In 2019, these included; the 12th JDMX Meeting for Early Career Researchers; the TIBER Symposium on Psychology and Economics; the Economic Science Association (ESA) European meeting and the Health Economics Seminar, Erasmus University Rotterdam. In 2021, this article was presented at the Working Group on African Political Economy (WGAPE), Africa Evidence Summit.

Chapter two of the thesis is titled, "Comparing the Productivity of Performance-based and Unconditional Financial Incentives in Improving Infant Birthweight in Zambia." The aim of the study upon which chapter two is based is to complement the insights from chapter one by comparing the productivity effect of performance-based and unconditional financial incentives in improving health outcomes in the 'real world,' outside a field experimental setting. The study utilises a difference-in-differences research design, complemented with triple difference, event study and regression discontinuity designs and four waves of nationally representative Demographic and Health Survey (DHS) data from Zambia. It leverages an actual real world RBF programme, the Zambia RBF pilot, implemented between 2012 and 2014 and compares districts where healthcare providers were randomised to receive either performance-based or unconditional financial incentives and those that conducted business as usual under the Zambia RBF pilot.

The key contributions of chapter two are threefold. First, it documents that the weak productivity effect of performance-based financial incentives may also extend into the real world. This is evident from the fact that performance-based and unconditional financial incentives are found to have comparable effects in improving infant health outcomes. Secondly, the study shows that despite results revealing that performance-based financial incentives have marginally stronger effects, the fact that they are also by far more costly, implies that unconditional financial incentives should be given more attention by policymakers. This is evident from the fact that a cost-effectiveness exercise reveals that the unconditional financial incentives are considerably more cost-effective, relative to performance-based financial incentives. Thirdly, the study also documents evidence suggestive of a distortion effect in the performance-based financial incentive treatment condition. This is evident from the fact that, the number of Ante-natal Care (ANC) visits, a health service delivery indicator with a performance-based financial incentive of \$1.6 per visit attached to it, increases strongly under the performance-based financial incentive treatment condition. The quality of the ANC visits, composed of indicators not directly attached to the performance-based financial incentive, does not improve as strongly as the number of ANC visits do. This suggests that healthcare providers focus more on the visits and not to the same extent on the quality of the visits owing to the effects of the

performance-based financial incentive.

Chapter two is single authored. It has been presented at the 2020 WGAPE Berlin virtual conference 2020 and the Poverty Reduction Equity and Growth Network (PeGNet) virtual PhD conference 2021. I also acknowledge for helpful comments and feedback received from Willa Friedman, Jennifer Okonkwo and Patrick Nuess in refining the article.

Chapter three is titled, "Public Sector Compensation and Rural-Urban Gaps in Service Delivery: Evidence from Zambia." The study upon which this chapter is based evaluates the impact of the rural hardship allowance, an unconditional salary supplement of 20% paid by the Zambian government to public sector employees stationed in Zambian rural areas. The allowance aims to compensate public sector employees stationed in rural areas of Zambia for the lack of modern amenities in such areas. The study assesses the impact of the allowance on healthcare provider staffing levels in health facilities eligible to receive the allowance. It exploits spatial discontinuities in eligibility for the allowance to estimate fuzzy regression discontinuity models. Some studies in the literature exploit exogenous variation in wages created within the context of field experiments (Dal Bó et al., 2013; De Ree et al., 2018; Deserranno, 2019). However, this study contributes to the literature by leveraging administrative data from the universe of public healthcare providers in Zambia to study defacto health facility staffing levels. Additionally, the study contributes novel evidence of the effects of a hardship based unconditional financial incentive in the health sector within the context of a low income developing country such as Zambia. Most studies in the literature estimate the effects of similar hardship-based unconditional financial incentives on education outcomes (Neilson et al., 2021; Chelwa et al., 2019; Cabrera and Webbink, 2020; Pugatch and Schroeder, 2014).

The article documents evidence of substantial increases in health facility staffing on the eligible side of the allowance threshold for the allowance. The number of healthcare providers in health facilities eligible to receive the allowance increases by 57% once facilities meet the eligibility criteria to receive the allowance. Once the number of healthcare providers is adjusted for quality, the study finds that the quality-adjusted number of healthcare providers increases by 60% once facilities meet the eligibility criteria to receive the allowance.

Chapter three is jointly authored with Torsten Figueiredo Walter. I contributed to this article at all stages of its development, from conception of the research idea, securing of the administrative data from the Zambian government, analysis of the data and preparation and revision of draft versions of the article.

Taken in its entirety, this thesis contributes to several strands of the literature. I outline these contributions in the discussion below:

Firstly, this thesis contributes to a nascent literature on the personnel economics of the public sector in developing countries, a literature concerned with attracting, retaining

and motivating public sector employees. This literature has established, for instance, that public sector employees may be motivated by the mission of the organisation (Besley and Ghatak, 2005), by career incentives (Ashraf et al., 2020) or indeed by pro-social motivations such as reciprocity and intrinsic motivation (Ashraf and Bandiera, 2018; Besley and Ghatak, 2014). Dal Bó et al. (2013) and Deserranno (2019) have investigated the role of financial incentives on the applicant pool at the recruitment stage in Mexico and Uganda, respectively, finding strong effects of the incentives in attracting high quality applicants for public sector positions. While these papers focus on the extensive or recruitment margin, this thesis contributes to this literature at both the intensive and extensive margins of the effects of financial incentives in the public sector.

At the intensive margin, the contribution of this thesis comes in the form of utilising various rigorous methodological approaches to show that evidence for the productivity effect of performance-based financial incentives in the health sector may actually be much weaker than previously thought. Similarly, this thesis further shows that unconditional financial incentives to healthcare providers may actually be more productive than previously thought. This implies that greater consideration should be given to their usage, especially in low income developing country settings where resource constraints may be more binding, but where productivity increases in public service delivery may be even more critical. This contribution differs from studies in the education literature, which have revealed weak productivity effects of unconditional financial incentives in that sector (De Ree et al., 2018; Chelwa et al., 2019). This contribution, however parallels Ashraf et al. (2014) who find that non-financial incentives elicit higher productivity levels relative to performance-based financial incentives in Zambia. This implies that other incentive types such as unconditional financial incentives can be especially productive in the health sector, but not necessarily elsewhere.

At the extensive margin, this thesis provides novel evidence showing that an unconditional hardship-related financial incentive, can actually succeed in increasing the number of healthcare providers in under-served rural areas that are eligible to receive it. Such gains are, in-turn, critical in ensuring a more equitable distribution of healthcare providers between rural and urban areas that are characterised by inequalities in access to essential services such as high quality healthcare.

This thesis contributes to a strand of the economics literature on the signalling effects of incentives. In this literature, incentives may either crowd-in or crowd-out pro-social preferences by transmitting signals about the intentions of the principal providing the incentives, the ability of the agent performing the task, or indeed the ease or difficulty of the task itself (Bowles and Polania-Reyes, 2012; Benabou and Tirole, 2003; Deserranno, 2019). The contribution of this thesis to this literature is to show that performance-based financial incentives to healthcare providers may also contain a payoff maximising signal dominating the effort enhancing signal the incentives are usually thought to convey. However, unconditional financial incentives may not convey this same payoff maximising

signal and in this way, productivity levels owing to the two types of incentives may end up being quite comparable in a setting in which performance-based financial incentives would have been thought to be more productive.

This thesis also contributes to a strand of the literature in health economics which is concerned with RBF health programmes, implemented mostly in low income developing countries. Indeed, this literature is concerned with the implementation of these health programs, which have involved the provision of performance-based financial incentives to healthcare providers in an attempt to enhance their productivity and ultimately improve health outcomes. However, the sustainability and suitability of performance-based financial incentives to healthcare providers has been seriously questioned (James et al., 2020). A key contribution of this thesis to this literature is to show that more cost-effective incentive options other than performance-based financial incentives, should also be considered by policymakers in their quest to improve health outcomes. Similarly, given the predictions of agency theory on incentives in a multitasking framework (Holmstrom and Milgrom, 1991; Sherry, 2016) as regards the potential for multitasking effects, this thesis contributes evidence to this literature regarding the existence of this distortion effect among healthcare providers in a low income developing country setting such as Zambia. This distortion effect arises because healthcare providers are likely to interpret the performance-based financial incentive as a payoff maximising signal and not as an effort enhancing signal. An additional contribution of this thesis to the RBF literature is to generate evidence on the effects of a real world RBF programme on actual health outcomes. Indeed, most of this literature focuses on assessing the impact of RBF performance-based financial incentives on indicators of health service provision and not actual health outcomes (Basinga et al., 2011; Friedman et al., 2016; Paul et al., 2018; Bonfrer et al., 2014; Gage and Bauhoff, 2021). Generating evidence and shifting the policy discussion to also consider the effects of these programmes on health outcomes is critical, because the ultimate aim of these programmes is to actually improve population health outcomes.

Finally, this thesis also makes several contributions to the behavioural economics literature. Firstly, the incentivised experimental measures of healthcare provider pro-social, other-regarding preferences within the context of a low income developing country such as Zambia are, to the best of my knowledge, novel. Evidence existing in this literature has mainly come from subject pools in high income, developed country settings (Godager and Wiesen, 2013; Hennig-Schmidt et al., 2011). Secondly, the two experimental tasks at the heart of the field experiment in chapter one: the health-framed real-effort task and medicine allocation task, also represent a contribution of the thesis. These novel experimental tasks, are modified versions of effort tasks as well as the dictator game commonly used in the experimental economics literature. The novelty comes in the form of adapting these tasks so as to permit healthcare provider treatment decisions regarding real medicines that in-turn, affect the health of real patients in a rural Zambian clinic. Such an incentivised healthcare provider treatment decision making situation is, to the best of my knowledge, novel in the behavioural and experimental economics literature.

## Bibliography

- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics* 97(4), 543–569.
- Alkema, L., D. Chou, D. Hogan, S. Zhang, A.-B. Moller, A. Gemmill, D. M. Fat, T. Boerma, M. Temmerman, C. Mathers, et al. (2016). Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the un maternal mortality estimation inter-agency group. *The Lancet* 387(10017), 462–474.
- Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011). Inside the war on poverty: The impact of food stamps on birth outcomes. *The Review of Economics and Statistics* 93(2), 387–403.
- Amarante, V., M. Manacorda, E. Miguel, and A. Vigorito (2016). Do cash transfers improve birth outcomes? evidence from matched vital statistics, program, and social security data. *American Economic Journal: Economic Policy* 8(2), 1–43.
- Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics* 10, 439–463.
- Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review* 110(5), 1355–94.
- Ashraf, N., O. Bandiera, and B. K. Jack (2014). No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics* 120, 1–17.
- Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21(1), 1–42.
- Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014). Conditional, unconditional and everything in between: A systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness* 6(1), 1–43.
- Barber, S. L. and P. J. Gertler (2008). The impact of mexico’s conditional cash transfer programme, oportunitades, on birthweight. *Tropical Medicine and International Health* 13(11), 1405–1414.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. Soucat, J. Sturdy, and C. M. Vermeersch (2011). Effect on maternal and child health services in rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet* 377(9775), 1421–1428.
- Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies* 70(3), 489–520.

- Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *American Economic Review* 95(3), 616–636.
- Besley, T. J. and M. Ghatak (2014). Solving agency problems: Intrinsic motivation, incentives, and productivity. Technical report.
- Bonfrer, I., E. Van de Poel, and E. Van Doorslaer (2014). The effects of performance incentives on the utilization and quality of maternal and child care in burundi. *Social Science & Medicine* 123, 96–104.
- Bowles, S. and S. Polania-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature* 50(2), 368–425.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies, and D. Wiesen (2013). *How Effective are Pay-for-Performance Incentives for Physicians?—A Laboratory Experiment*. Number 413. Ruhr Economic Papers.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz, and D. Wiesen (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization* 131, 17–23.
- Cabrera, J. M. and D. Webbink (2020). Do higher salaries yield better teachers and better student outcomes? *Journal of Human Resources* 55(4), 1222–1257.
- Callaway, B. and P. H. Sant’Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.
- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal* 17(2), 372–404.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal* 14(4), 909–946.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6), 2295–2326.
- Cardona-Sosa, L. and C. Medina (2017). The effects of in utero programs on birth outcomes: The case of buen comienzo. *Journal of the Latin American and Caribbean Economic Association* 17(2), 93–134.
- Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal* 18(1), 234–261.
- Cattaneo, M. D., M. Jansson, and X. Ma (2019). lpdensity: Local polynomial density estimation and inference. *arXiv preprint arXiv:1906.06529*.
- Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531), 1449–1455.

- Chelwa, G., M. Pellicer, and M. Maboshe (2019). Teacher pay and educational outcomes: Evidence from the rural hardship allowance in zambia. *South African Journal of Economics* 87(3), 255–282.
- Chibuye, B., P. Hangoma, and M. Khadjavi (2021). The dark side of financial incentives in a health setting - experimental evidence from zambia. Working paper, Christian-Albrechts-University of Kiel.
- Chibwasha, C. J., A. Zanolini, M. Smid, B. Vwalika, M. P. Kasaro, M. Mwanahamuntu, J. S. Stringer, and E. M. Stringer (2016). Predictors and outcomes of low birth weight in lusaka, zambia. *International Journal of Gynecology & Obstetrics* 134(3), 309–314.
- Cross, S. S. and M. Baernholdt (2017). Identifying factors for worker motivation in zambia’s rural health facilities. *Journal of Health Care for the Poor and Underserved* 28(1), 63–70.
- Currie, J. and D. Almond (2011). Human capital development before age five. In *Handbook of Labor Economics*, Volume 4, pp. 1315–1486. Elsevier.
- Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics* 128(3), 1169–1218.
- Das, J. and J. Hammer (2014). Quality of primary care in low-income countries: Facts and economics. *Annual Review of Economics* 6(1), 525–553.
- de Bekker-Grob, E. W., M. Ryan, and K. Gerard (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21(2), 145–172.
- De Chaisemartin, C. and X. d’Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9), 2964–96.
- De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2018). Double for nothing? experimental evidence on an unconditional teacher salary increase in indonesia. *The Quarterly Journal of Economics* 133(2), 993–1039.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18(1), 105–128.
- Deci, E. L., R. Koestner, and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125(6), 627–693.
- Deserranno, E. (2019). Financial incentives as signals: Experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics* 11(1), 277–317.
- Díaz, J.-J. and V. Saldarriaga (2019). Encouraging use of prenatal care through conditional cash transfers: Evidence from juntos in peru. *Health Economics* 28(9), 1099–1113.

- Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics* 27(6), 1436–1450.
- Dussault, G. and M. C. Franceschini (2006). Not enough there, too many here: Understanding geographical imbalances in the distribution of the health workforce. *Human Resources for Health* 4(1), 1–16.
- Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics* 24(1), 211–223.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics* 108(2), 437–459.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review* 42(1), 1–34.
- Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy* 21(3), 673–705.
- Ferrinho, P., S. Siziya, F. Goma, and G. Dussault (2011). The human resource for health situation in zambia: Deficit and maldistribution. *Human Resources for Health* 9(1), 1–11.
- Figueiredo Walter, T. (2020). Misallocation in the public sector? cross-country evidence from two million primary schools. Technical report, STICERD Working Paper No. EOPP 70.
- Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. *Handbook of Economic Field Experiments* 2, 467–514.
- Friedman, J., J. N. Qamruddin, C. Chansa, and A. K. Das (2016). Impact evaluation of zambia's health results-based financing pilot project. Technical report, The World Bank.
- Gage, A. and S. Bauhoff (2021). The effects of performance-based financing on neonatal health outcomes in burundi, lesotho, senegal, zambia and zimbabwe. *Health Policy and Planning*.
- Gertler, P. and C. Vermeersch (2012). *Using Performance Incentives to Improve Health Outcomes*. The World Bank.
- Gneezy, U. and J. A. List (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5), 1365–1384.
- Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3), 791–810.
- Godager, G. and D. Wiesen (2013). Profit or patients' health benefit? exploring the heterogeneity in physician altruism. *Journal of Health Economics* 32(6), 1105–1116.

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.
- Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization* 106, 367–378.
- Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.
- Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature* 42(4), 1009–1055.
- Hennig-Schmidt, H., R. Selten, and D. Wiesen (2011). How payment systems affect physicians’ provision behaviour — an experimental investigation. *Journal of Health Economics* 30(4), 637–646.
- Herbst, C. H., M. Vledder, K. Campbell, M. Sjöblom, and A. Soucat (2011). The human resources for health crisis in zambia. *Washington (DC): World Bank*.
- Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.* 7, 24.
- Hoynes, H., D. Miller, and D. Simon (2015). Income, the earned income tax credit and infant health. *American Economic Journal: Economic Policy* 7(1), 172–211.
- Hoynes, H., M. Page, and A. H. Stevens (2011). Can targeted transfers improve birth outcomes?: Evidence from the introduction of the wic program. *Journal of Public Economics* 95(7-8), 813–827.
- Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635.
- James, N., K. Lawson, and Y. Acharya (2020). Evidence on result-based financing in maternal and child health in low-and middle-income countries: A systematic review. *Global Health Research and Policy* 5(1), 1–15.
- Kaarboe, O. and L. Siciliani (2011). Multi-tasking, quality and pay for performance. *Health Economics* 20(2), 225–238.
- Kesternich, I., H. Schumacher, and J. Winter (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? *Journal of Public Economics* 131, 1–11.
- Kramer, M. S. (1987). Determinants of low birth weight: Methodological assessment and meta-analysis. *Bulletin of the World Health Organization* 65(5), 663.
- Kramer, M. S. (2003). The epidemiology of adverse pregnancy outcomes: An overview. *The Journal of Nutrition* 133(5), 1592S–1596S.

- Lagarde, M. and D. Blaauw (2014). Pro-social preferences and self-selection into jobs: Evidence from south african nurses. *Journal of Economic Behavior & Organization* 107, 136–152.
- Lagarde, M. and D. Blaauw (2017). Physicians’ responses to financial and social incentives: A medically framed real effort experiment. *Social Science & Medicine* 179, 147–159.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review* 90(5), 1346–1361.
- Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281–355.
- Lemiere, C., C. Herbst, N. Jahanshahi, and E. Smith (2011). Reducing geographical imbalances of health workers in sub-saharan africa: A labor market perspective on what works, what does not, and why. world bank; 2011.
- Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2019). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly Journal of Economics* 134(3), 1627–1673.
- McCoy, D., S. Bennett, S. Witter, B. Pond, B. Baker, J. Gow, S. Chand, T. Ensor, and B. McPake (2008). Salaries and incomes of health workers in sub-saharan africa. *The Lancet* 371(9613), 675–681.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2), 698–714.
- Miller, G. and K. S. Babiarz (2013). Pay-for-performance incentives in low-and middle-income country health programs. Technical report, National Bureau of Economic Research.
- Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, and S. Rozelle (2012). Effectiveness of provider incentives for anaemia reduction in rural china: A cluster randomised trial. *Bmj* 345.
- Mullen, K. J., R. G. Frank, and M. B. Rosenthal (2010). Can you get what you pay for? pay-for-performance and the quality of healthcare providers. *The Rand Journal of Economics* 41(1), 64–91.
- Muralidharan, K. and N. Prakash (2017). Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics* 9(3), 321–50.
- Neilson, C., M. Bobba, T. Ederer, G. Leon-Ciliotta, and M. Nieddu (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in peru. Working paper, National Bureau of Economic Research.
- Ohlsson, A. and P. Shah (2008). *Determinants and Prevention of Low Birth Weight: A Synopsis of the Evidence*. Institute of Health Economics.

- Olken, B. A., J. Onishi, and S. Wong (2014). Should aid reward performance? evidence from a field experiment on health and education in indonesia. *American Economic Journal: Applied Economics* 6(4), 1–34.
- Paul, E., M. L. Dramé, J.-P. Kashala, A. E. Ndema, M. Kounnou, J. C. Aïssan, and K. Gyselincx (2018). Performance-based financing to strengthen the health system in benin: Challenging the mainstream approach. *International Journal of Health Policy and Management* 7(1), 35.
- Pega, F., S. Y. Liu, S. Walter, R. Pabayo, R. Saith, and S. K. Lhachimi (2017). Unconditional cash transfers for reducing poverty and vulnerabilities: Effect on use of health services and health outcomes in low-and middle-income countries. *Cochrane Database of Systematic Reviews* (11).
- Prust, M. L., A. Kamanga, L. Ngosa, C. McKay, C. M. Muzongwe, M. T. Mukubani, R. Chihinga, R. Misapa, J. W. van den Broek, and N. Wilmink (2019). Assessment of interventions to attract and retain health workers in rural zambia: A discrete choice experiment. *Human Resources for Health* 17(1), 26.
- Pugatch, T. and E. Schroeder (2014). Incentives for teacher relocation: Evidence from the gambian hardship allowance. *Economics of Education Review* 41, 120–136.
- Sappington, D. E. (1991). Incentives in principal-agent relationships. *Journal of Economic Perspectives* 5(2), 45–66.
- Shen, G. C., H. T. H. Nguyen, A. Das, N. Sachingongu, C. Chansa, J. Qamruddin, and J. Friedman (2017). Incentives to change: Effects of performance-based financing on health workers in zambia. *Human Resources for Health* 15(1), 20.
- Sherry, T. B. (2016). A note on the comparative statics of pay-for-performance in health care. *Health Economics* 25(5), 637–644.
- Sherry, T. B., S. Bauhoff, and M. Mohanan (2017). Multitasking and heterogeneous treatment effects in pay-for-performance in healthcare: Evidence from rwanda. *American Journal of Health Economics* 3(2), 192–226.
- Singh, P. (2015). Performance pay and information: Reducing child undernutrition in india. *Journal of Economic Behavior & Organization* 112, 141–163.
- Singh, P. and W. A. Masters (2020). Performance bonuses in the public sector: Winner-take-all prizes versus proportional payments to reduce child malnutrition in india. *Journal of Development Economics* 146, 102–295.
- United Nations, U. (2016). Transforming our world: The 2030 agenda for sustainable development.
- Zambia Ministry of Health, M. (2011). National health strategic plan (2011 - 2015). Technical report, Government of the Republic of Zambia.

Zambia Ministry of Health, M. (2017). National health sector strategic plan. Technical report, Government of the Republic of Zambia., Lusaka.

Zambia Ministry of Health, Z. S. A. and M. IFC (2019). Zambia demographic and health survey 2018. Technical report, Zambia Statistics Agency, Ministry of Health, and ICF, Lusaka, Zambia, and Rockville, Maryland, USA.

Zambia Ministry of National Development Planning, M. (2017). Seventh national development plan (sndp). Technical report, Zambia Ministry of National Development Planning.

Chapter 2

The Dark Side of  
Performance-Based Financial  
Incentives in Healthcare –  
Experimental Evidence from  
Zambia

## **Abstract**

The productivity of healthcare providers in a context with low quality of healthcare and weak healthcare outcomes in low-income developing countries remains a great concern. In response, health programmes try to improve productivity by using performance-based financial incentives. The payment of such incentives is based on performance on particular healthcare services. We employ a health-framed field experiment with medical students and doctors in Zambia to investigate the productivity and distortion effects of such incentives. We do not find evidence for the productivity effect, however, a significant distortion effect emerges. Three nudge treatments we employ in an effort to counteract the distortion effect, including allocation observability and emphasising medical norms, all fail to counteract it. Our findings highlight that performance-based incentives may signal payoff maximising behaviour to healthcare providers, without signalling commensurate productivity enhancement and should thus be used with caution. Policymakers are thus advised to explore the role of unconditional financial incentives, that may not signal distorting payoff maximising behaviour but lead to comparable levels of productivity.

*Keywords:* Performance-based financial incentives; Framed-field experiment; Decision making, Medical doctors; Zambia

*JEL Codes:* C93, I12, I15

## 2.1 Introduction

An effective healthcare system is crucial for the prosperity of a society and the functioning of its economy. This is evident in the focus of United Nations' Sustainable Development Goal 3 on ensuring 'good health and well-being'. In the midst of the current COVID-19 pandemic, an effective healthcare system is even more crucial. Numerous factors determine the effectiveness of healthcare systems. Beyond the level of funding and governance, the availability of infrastructure, medical supplies and equipment, the effectiveness of healthcare systems also crucially depends on the performance of public sector employees and medical staff. It is thus not surprising that the relatively lower quality that characterises healthcare in low income developing countries, has also been linked to low levels of healthcare provider effort and a so called 'know-do gap', in which healthcare providers 'do' far less in the provision of healthcare that they 'know' through their training (Das and Hammer, 2014; Gertler and Vermeersch, 2012). This has resulted in the popularity of Results Based Financing (RBF) health programmes in which healthcare providers are paid performance-based financial incentives with the ultimate goal of increasing their productivity. However, the appropriateness and sustainability of performance-based financial incentives for low income developing countries have been questioned (James et al., 2020). Similarly, performance-based financial incentives have also been linked to the potential for detrimental effects such as multitasking effects in the literature (Holmstrom and Milgrom, 1991; Sherry, 2016).

In this study we investigate the potential for a performance-based financial incentive to have two separate effects on healthcare provider decisions. A productivity effect, which may occur when a healthcare provider increases effort due to a performance-based financial incentive and a distortion effect, which may occur if a healthcare provider does not increase effort, but redirects decision-making solely towards the service that is incentivized. To this end, we implement a health-framed field experiment among medical students and current medical doctors in Zambia. Our experimental design incorporates two experimental tasks and six between-subjects treatments. Three aspects of our results are important to highlight. First, we do not find evidence of the productivity effect on healthcare provider decisions. Second, a substantial distortion effect on healthcare provider decisions emerges. Third, the distortion effect is unresponsive to three counteracting measures implemented. Our results imply that in certain instances, a performance-based financial incentive may be interpreted by healthcare providers as a pure payoff maximising incentive as opposed to an effort enhancing incentive. Our findings have important implications for real world health programmes aiming to use financial incentives to improve healthcare provider productivity. In such programmes, performance-based financial incentives should be used with caution, owing to the payoff maximising signal they send. Similarly, the use of unconditional financial incentives that do not contain this pure payoff maximising signal, should be considered.

Low income developing countries have among the weakest health outcomes worldwide. For example, approximately 99% of the 303,000 maternal deaths reported in 2015 occurred in these countries (Alkema et al., 2016). These challenges are even more evident in low income developing countries in Africa and are exemplified by the fact that rates of under age five mortality are eight times higher than those of OECD countries. Indeed, like many low-income developing countries in Africa, Zambia has been grappling with the challenges of improving the performance of its healthcare system and ultimately the health and productivity of its citizens (Zambia Ministry of Health, 2011). Stakeholders have acknowledged the problem of low productivity of the health sector and have implemented RBF programmes aimed at enhancing both health personnel effort and productivity (Friedman et al., 2016). One key distinguishing aspect of such productivity enhancement programmes is the earmarking of performance-based financial incentives to reward health facilities and individual health providers in these facilities for their performance. The performance-based financial incentives are tied to pre-agreed performance targets which are attached to a subset of specific health indicators (Basinga et al., 2011; Bonfrer et al., 2014; Friedman et al., 2016).

In this study we investigate whether such performance-based incentives can result in a productivity effect and whether they lead to a distortion effect on the allocation decisions of 359 healthcare providers, i.e. medical doctors and medical students at Zambia’s largest medical school. We employ a health-framed field experiment in which the health providers can exert real-effort and make real medical resource allocation decisions. We systematically vary how effort and allocation decisions are incentivized and test possible ways to counteract distortionary allocations.

Our experimental design employs two intertwined tasks: a health-framed real-effort task and a medicine allocation task. In the first step, the effort task, participants generate an effort-dependent medicine allocation budget. In the second step, the medicine allocation task, participants use their effort-dependent budgets to allocate essential medicines to a real rural Zambian health clinic. Participants use real health resource requirement data from the same rural clinic as a basis for their own medicine allocation decisions.

Our framed field experiment consists of six between-subjects treatments. In the *Baseline* treatment participants receive a fixed payment independent from their effort in the effort task and their decisions in the medicine allocation task. The *Baseline* treatment serves as a comparison group absent performance-based incentives. In two *Incentive* treatments participants receive a performance-based financial incentive: the greater the effort exerted in the effort task and the greater the monetary share allocated to purchase units of one incentivised medicine, relative to four non-incentivised medicines, which also results in a greater personal monetary payment. The two *Incentive* treatments each incentivise a different medicine to assess if allocation decisions depended in any way on which medicine was incentivised. Comparison of the two *Incentive* treatments with the *Baseline* thus helps us identify the presence of the productivity effect: whether participants’ effort could be

enhanced owing to the performance-based financial incentive. This comparison also helps us identify the presence of the distortion effect: whether allocation decisions would be distorted by the performance-based financial incentive, in favour of the incentivised medicine, at the expense of the non-incentivised medicines. Participating healthcare providers in the two *Incentive* treatments are thus presented with a decision situation in which they can either be consistent with what the rural Zambian health clinic needs or deviate from this and maximise their own monetary payoff.

Three separate nudging treatments investigate how the distortion effect of the performance-based financial incentive on medicine allocation decisions, can be counteracted. These treatments do so by combining the incentive component with the following elements respectively; (1) potential observability of medicine allocation decisions, i.e. there is a low chance that others learn your decision, (2) priming emphasising norms of professionalism and (3) information making the importance of a non-incentivised medicine very salient.

In our results, we do not find evidence for a productivity effect owing to the performance-based financial incentive. This implies that we do not see an increase in effort in the two *Incentive* treatments relative to the *Baseline* treatment, when effort is incentivized in the real-effort task. We, however, find evidence of a substantial distortion effect. This implies that healthcare provider medicine allocation decisions in the two *Incentive* treatments relative to the *Baseline*, are distorted in favour of the incentivised medicine at the expense of the non-incentivised medicines. Participating healthcare providers in the two *Incentive* treatments thus deviate from the requirements of the rural Zambian health clinic in such a way that their own monetary payoff is maximised. We also find that this distortion effect is remarkably unresponsive to the counteracting measures we implemented in three nudging treatments. The distortionary effect is also rationalised medically by many of our participants, suggesting a self-serving bias. Interestingly, the medical justifications in the two *Incentive* treatments also change accordingly, when a different medicine is incentivised.

Our findings may imply that performance-based financial incentives signal payoff maximising behaviour as opposed to productivity enhancement to healthcare providers (Bowles and Polania-Reyes, 2012; Deserranno, 2019). This is evident from the fact that the distortion effect overshadows the productivity effect in our field experiment. This evidence implies that performance-based financial incentives in healthcare settings should be used with caution, owing to this payoff maximising signal they send to healthcare providers. Similarly, policymakers implementing health programmes may need to explore the role of other incentive types that do not signal payoff maximisation to healthcare providers. Indeed, our field experiment reveals that healthcare providers in the *Baseline* treatment, receiving a financial incentive independent of their effort are able to achieve a level of productivity similar to the productivity level arising in the two *Incentive* treatments.

While healthcare providers in the real world may not necessarily distinguish patients who arrive in their facilities to receive care based on whether their disease condition is

financially incentivised or not, we believe that our findings are most relevant and applicable to provider decisions related to demand creation and outreach activities. Indeed, for the latter set of situations, healthcare providers must initiate the provision of healthcare. Thus, if some health conditions are attached to performance-based financial incentives for the providers while others are not, then the decision making of healthcare providers may be distorted to focus unequally on those health outcomes to which the financial incentives are attached, to the detriment of the non-incentivised health outcomes. An increase in provision productivity and overall healthcare cannot be expected – which questions the use of precious resources for such incentives, especially in low-income countries.

We provide evidence on the decision making of actual medical doctors and medical students in a health-framed field experiment with real health outcomes all in the context of a low income developing country, that has implemented an RBF programme. Our experiment further allows us to investigate the magnitudes of the productivity effect and the distortion effects. In this way, our paper makes contributions to several strands of the literature. First, our paper is connected to a strand of the literature on behavioural experiments in health economics.<sup>1</sup> The closest paper to ours in this literature is (Brosig-Koch et al., 2013), who investigate behavioural responses to performance-based payment in a laboratory setting. However, they do not investigate the interplay between a productivity effect and a distortion effect as we do. Our paper is also connected to another strand of the health economics literature investigating RBF programmes, implemented mostly in low income developing countries (Friedman et al., 2016; Basinga et al., 2011; Bonfrer et al., 2014).

Our paper is also connected to a strand of the literature investigating so-called multitasking effects (Holmstrom and Milgrom, 1991; Sherry, 2016; Eggleston, 2005).<sup>2</sup> This strand of the literature is concerned with job-design within principal-agent contexts and specifically, with the interplay between incentivised and non-incentivised aspects of agents job profiles. The distortion effect we investigate thus shares some similarities with this literature. We contribute to this literature using a field experiment to investigate a productivity and distortion effect of a performance-based financial incentive on healthcare provider decision making, taking a broader Judgement Decision-making (JDM) perspective.

---

<sup>1</sup>In this strand of the literature, laboratory and laboratory-in-the field experiments with medical and non-medical students as well as in artefactual field experiments with doctors, all performed in developed country settings, have provided numerous insights on physician behaviour. These studies find that health care workers provide more services when paid via fee-for-service relative to other payment mechanisms such as capitation. Medical doctors are equally behaviourally responsive based on payment mechanism (Brosig-Koch et al., 2016, 2013; Hennig-Schmidt et al., 2011; Lagarde and Blaauw, 2017; Green, 2014). Our findings are thus in line with this strand of the literature.

<sup>2</sup>The health sector presents an interesting case study of multitasking effects, given that multitasking has been noted to be especially prevalent in this setting (Eggleston, 2005). In fact, the exact channels through which it can occur have been modelled theoretically (Sherry, 2016; Eggleston, 2005; Kaarboe and Siciliani, 2011). The empirical literature is however mixed as to whether multitasking effects do indeed occur in health settings in practice. Some studies align with theoretical expectations (Feng Lu, 2012; Lagarde and Blaauw, 2017; Dumont et al., 2008), while others do not (Mullen et al., 2010; Sherry et al., 2017).

Finally, our paper is also connected to a strand of the literature investigating the signalling effect of financial incentives (Bowles and Polania-Reyes, 2012; Deserranno, 2019). In this literature, in a setting with limited information, financial incentives can signal certain attributes of the task being performed, the intentions of the principals or indeed, the agents perception of the principals intentions (Deci, 1971; Deci et al., 1999; Bowles and Polania-Reyes, 2012). Our contribution is to show that performance-based financial incentives, used within a health setting, can signal payoff maximising behaviour in a manner that leads to an additional effect on provider decision making - a distortion effect.

The rest of the paper is organised as follows, section 2.2 presents details of the experimental design, including the experimental tasks, treatments and procedures employed. Section 2.3 presents descriptive statistics of the data utilised in the study. Section 2.4 presents the descriptive and regression results of the study. Section 2.5 provides a discussion and concludes.

## 2.2 Experimental Design

In this section we present the experimental design, including the two tasks, six treatments and procedures of the data collection at the University of Zambia.

### 2.2.1 Tasks 1 and 2

The study was implemented as a health-framed field experiment with a between-subjects design (Harrison and List, 2004). All participants were requested to complete two tasks; a health-framed real effort task, labelled to participants as Task 1 and a medicine allocation task, Task 2.

We used Task 1 to measure participants' effort with and without a financial incentive. All participants were given exactly eight minutes within which to complete as many summations as possible in a sample Health Information Aggregation 1 (HIA.1) form. This is a commonly used form in the Zambian health system. In completing the HIA.1 form, participants were asked to sum data on the number of Out-Patient Department (OPD) First Attendances, In-Patient Department (IPD) Discharges and Deaths. Figure 2.1 depicts the HIA.1 form used in the health-framed real effort task.

Before beginning Task 1, instructions informed participants that the number of correct entries determined each participant's variable allocation budget for Task 2, labelled to participants as 'Budget 2'. Each correctly summed entry in the HIA.1 form added 50 Zambian Kwacha (ZMW), approx. 3.7 Euros to each participant's Budget 2.<sup>3</sup> Given that there were 60 possible entries in the HIA1 form, the size of each participant's Budget 2 could thus potentially fall between ZMW 0 and ZMW 3,000, approx. 220 Euros, with the exact amount depending on each participant's score in Task 1, multiplied by ZMW 50. Before the beginning of Task 1, instructions informed the participants that (1) the higher

---

<sup>3</sup>An exchange rate of 1 Euro = ZMW 13.6 applied at the time of the experiment.

**Disease Aggregation Form: HIA.1**

DIAGNOSES	OPD First Attendance				IPD Discharge				Deaths			
	<1 year	1-5 years	6+ Years	Total	<1 year	1-5 years	6+ Years	Total	<1 year	1-5 years	6+ Years	Total
<b>NOTIFIABLE DISEASES</b>												
Acute flaccid paralysis	12	29	87		61	24	69		89	72	64	
Cholera	47	20	46		50	2	47		4	58	38	
Measles	92	25	32		84	52	23		18	59	83	
Meningitis	5	45	72		44	71	11		100	34	43	
Neonatal tetanus	50	45	8		47	50	27		22	64	83	
Plague	14	12	46		99	63	2		29	67	61	
Rabies	78	87	91		96	93	63		79	26	54	
Dysentery	44	75	3		12	35	38		9	54	85	
Typhoid fever	90	33	0		85	97	22		99	54	2	
Yellow fever	33	77	1		46	98	83		1	73	71	
Any other unusual occurrence of disease or outbreak	34	72	27		37	40	76		87	73	90	
<b>SELECTED DISEASES</b>												
Ear Diseases	25	11	49		74	2	19		75	39	67	
Nose Diseases	33	31	69		89	56	0		62	71	87	
Throat Diseases	95	2	12		63	76	74		62	2	44	
Chronic Diseases	62	95	31		15	31	51		83	33	80	
Asthma	17	63	75		89	98	1		46	73	53	
Cardio-vascular diseases	68	44	93		6	59	28		82	5	47	
Diabetes	34	99	50		63	46	70		96	79	11	
Hypertension	23	12	46		29	93	17		5	11	13	
Eye diseases	36	15	35		35	35	35		31	72	6	

Figure 2.1: Health-Framed Real Effort Task Tool

*Notes:* This figure shows the data collection tool for Task 1, the health-framed real effort task. The design of the tool is based on the Health Information Aggregation 1 (HIA.1) form, a form commonly used to collect healthcare provision data in the Zambian health system. In the health-framed real effort task, participants filled in the grey-shaded regions of the form by summing the entries in the corresponding rows within exactly eight minutes. The number of correct entries out of the 60 possible entries was multiplied by ZMW 50 to generate each participants variable allocation budget, Budget 2. Budget 2, in-turn, was used as a variable allocation budget in the medicine allocation task, Task 2.

a participant's score the greater Budget 2 and, thus, the more money they could allocate for medicines to a real Zambian Rural Health Centre (RHC) and (2) in all treatments but *Baseline* a larger Budget 2, ceteris paribus, implied a larger payment for the participant

herself (below we will describe exactly how). While the chance to generate money to buy medicine for the RHC itself can be an incentive to exert effort in Task 1, we expected that the additional incentive to generate money for oneself at the same time, and not at the expense of the RHC, will result into even greater scores in all other treatments compared to *Baseline*. We regard it as crucial that both intrinsic and extrinsic motivation can play a role in Task 1, as they can both be important for effort of medical staff when working with patients as well. This set-up thus, allowed us to test for a productivity effect, in the form of higher participant scores in all other treatments relative to *Baseline* in Task 1.

In Task 2 participants were presented with a list of the consumed essential medicines in a real Zambian Rural Health Centre, namely Lubwa RHC, as well as an estimate of the percentage of the cost that each medicine represented (see the decision sheet in Figure 2.2). This medicine consumption data, which also presented a picture of the underlying disease burden in the catchment area of Lubwa RHC, represented a realistic picture of health resource needs in the RHC and was thus labelled as an ‘optimal allocation of health resources’ to this specific RHC.

We decided on Lubwa RHC as the recipient of the allocated medical supplies for several reasons: it is located in a very rural and remote area, a 12-hour drive away from Zambia’s capital Lusaka and there is no city and major infrastructure close to it, it is very unlikely that the medical students and doctors participating in our study have personal affiliations with the RHC and, given its remote location, it is common knowledge for our participants that there is typically a lack of the most basic medical supplies in a remote RHC like Lubwa RHC. These circumstances made the medicine allocation task to Lubwa RHC a high-stakes task regarding its economic and ethical impact.

In the process of completing Task 2, we provided participants with information from Lubwa RHC for them to base their own individual level allocation decisions on i.e. to choose their own percentage allocations. These allocation decisions were made as regards how to allocate their two budgets: Budget 2, determined in Task 1 of the experiment, and Budget 1 which was fixed at ZWM 1,000, approx. 74 Euros, to each of the medications in Figure 2.2.

To ensure incentive compatibility, all participants across all treatments were informed that the allocation decisions of one randomly drawn participant would actually be bought and donated to Lubwa RHC by the research team, exactly as per their own allocation decisions.<sup>4</sup> In this way, participants’ decisions in the experiment could affect the health of real patients in the real world – in a vastly under-financed health care system – and this ensured that there were actual costs to a distortion of allocation decisions. To this

---

<sup>4</sup>The *Baseline*, *Incentive-Amoxil* and *Observability* treatments were implemented in October 2018. The *Incentive-Flagyl*, *Prof-Norms* and *Need-ORS* treatments were implemented in June 2019. Budget constraints prevented the implementation of all participants’ resource allocation decisions. Given the under-financing and immediate need of the RHC to receive medications, we still regard the stakes as relevant.

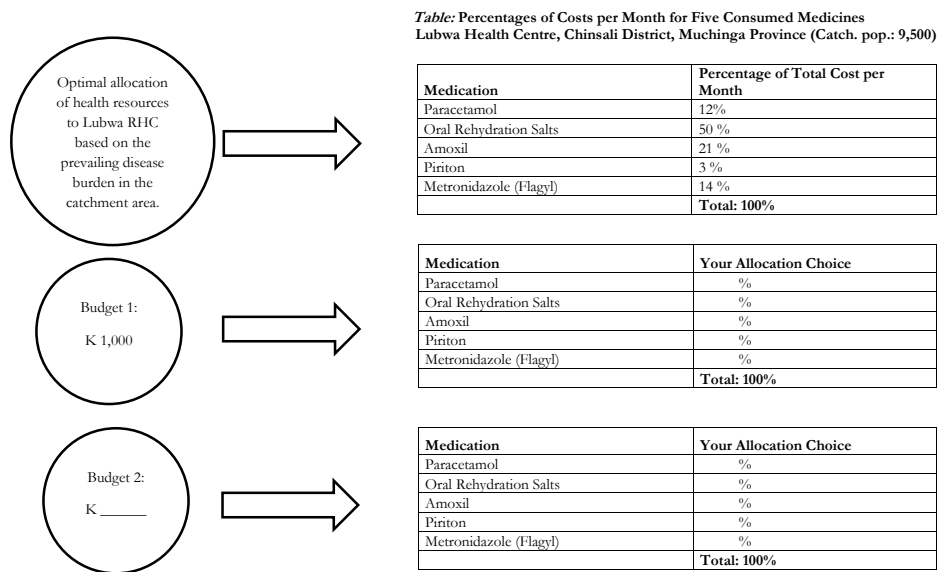


Figure 2.2: Medicine Allocation Task Tool

*Notes:* This figure shows the decision sheet used in the medicine allocation task, Task 2. In Task 2, participants made individual decisions as to how to allocate their Budget 1 and Budget 2, which was determined in the health-framed real effort task, Task 1, respectively. Participants decided on how to allocate Budgets 1 and 2 respectively, to each the five medications, using the data from Lubwa RHC on requirements for the same five medications, as a point of reference. Decisions were made when each participant filled-in their own percentages in the bottom two tables, allocating their Budgets 1 and 2, respectively across the five medicines.

end, we worked closely with the Zambian Ministry of Health, to purchase the medicines on the list of most consumed medicines and donate them to the RHC (see Figure 2.A.1 in appendix 4.A.2 for a picture of Lubwa RHC receiving the medicines).

### 2.2.2 Treatments

As noted above, participants completed Task 2 with experimental variation made salient, depending on which treatment they had been randomly allocated into. Table 2.1 presents a summary of the six treatments and how they correspond to the two experimental tasks discussed above. Instructions for all the treatments are shown in appendix 3.A.4.

As described above, participants in *Baseline* only received a fixed participation fee of ZMW 50 for completing both Task 1 and 2.

Table 2.1: Summary of Experimental Tasks and Treatments

Treatment	Task 1	Task 2				N
		Financial Incentive		Nudge	Nudge	
		Amoxil	Flagyl	Obs. decisions	Prof. norms	ORS Info.
<i>Baseline</i>	✓					48
<i>Incentive-Amoxil</i>	✓	✓				50
<i>Incentive-Flagyl</i>	✓		✓			42
<i>Observability</i>	✓	✓		✓		56
<i>Prof-Norms</i>	✓	✓			✓	50
<i>Need ORS</i>	✓	✓				✓ 49

*Notes:* This table contains a summary of the experimental tasks and treatments used in the field experiment. The symbol ✓ represents 'Yes' while empty cell represents 'No'. N shows the number of participants per treatment. Task 1 refers to the health-framed real effort task, while Task 2 refers to the medicine allocation task. 'Financial Incentive' refers to a performance-based financial incentive participants could receive, based on the percentage of Amoxil or Flagyl allocated to Lubwa RHC. 'Obs. decisions' refers to a nudge making the observability of allocation decisions salient. 'Prof. norms' refers to a nudge making norms of the medical profession salient. 'ORS info.' refers to a nudge making information on the importance of the non-incentivised medicine Oral Rehydration Salts (ORS) salient.

The participation fee was designed to mimic a fixed salary incentive scheme that Zambian health workers would ordinarily receive in the absence of a performance-based financial incentive. Participants' Task 1 scores in *Baseline* nevertheless still generated Budget 2 in Task 2 which was then used to allocate medicines.

Participants in all of the other five treatments also received a fixed participation fee of ZMW 50 and an additional performance-based payment. This performance-based payment depended on the score in Task 1 and on their allocation decisions to one of the five medications in Task 2. For the treatments in which Amoxil was incentivized, the following paragraph illustrates how we communicated the earnings functions to participants in the instructions:<sup>5</sup> The exact formula for your additional earnings is...

$$Earnings = \frac{Budget1}{4000} \times \%Amoxil_{Budget1} + \frac{Budget2}{4000} \times \%Amoxil_{Budget2} \quad (2.1)$$

Equation 2.2 below, thus represents the payoff function participants in all treatments in the field experiment faced:

$$\pi_i = \alpha + \frac{\mu}{\epsilon} \times \sigma_i + \frac{\gamma_i}{\epsilon} \times \theta_i \quad (2.2)$$

where  $\pi_i$  represents the monetary earnings of participant  $i$ ,  $\alpha$  represents the fixed participation fee of ZMW 50,  $\mu$  represents Budget 1, the fixed allocation budget,  $\sigma_i$  represents the percentage of Budget 1 participant  $i$  allocated to either Amoxil or Flagyl, the incentivised

<sup>5</sup>As per the instructions shown in appendix 3.A.4, sample calculations and control questions were also provided to the participants to help them understand the earnings formula. The formula was not shown to participants in the *Baseline* treatment

medicines.  $\gamma_i$  represents the size of the score-dependent Budget 2 while  $\theta_i$  represents the percentage of Budget 2 participant  $i$  allocated to the incentivised medicine.  $\epsilon$  is a scaling constant and equal to 4000 as shown in equation 2.1 above. In the *Baseline* treatment,  $\sigma_i$  and  $\theta_i$  are both equal to zero as participants only earned the fixed participation fee  $\alpha$ .

Instructions for both Task 1 and Task 2 were provided and read out aloud to all participants in each session. Hence, participants in all treatments except *Baseline*, knew that a greater score in Task 1 would directly translate into a larger Budget 2 for allocations to the RHC and more earnings for themselves at the end of the experiment. In the four treatments *Incentive-Amoxil*, *Observability*, *Prof-Norms* and *Need-ORS* this payment-relevant medication was Amoxil, while Flagyl was the payment-relevant medication in the *Incentive-Flagyl* treatment.<sup>6</sup>

This implies that participants could personally earn additional money in all treatments but *Baseline* when they scored higher in Task 1 and the more Amoxil (Flagyl in *Incentive-Flagyl*) they allocated to the RHC in Task 2. All treatments but *Baseline* therefore mimicked a salary plus performance-based incentive scheme. Our experimental design hence permits an investigation of whether the performance-based component in these treatments motivates higher effort and scores in Task 1 (i.e. more productivity) and distorts allocation decisions towards Amoxil (or Flagyl) in Task 2 – all compared to the *Baseline* treatment. We therefore formulate;

**Hypothesis 1, Productivity Effect: The average score in Task 1 is higher in all the performance-based incentive treatments compared to Baseline.**

Similarly, the treatments *Incentive-Amoxil* and *Incentive-Flagyl* allow us to test for a distortion effect in Task 2 compared to *Baseline*. We therefore also formulate;

**Hypothesis 2, Distortion Effect: The average share of Amoxil allocated to Lubwa RHC in Task 2 is greater in Incentive-Amoxil compared to Baseline (Flagyl in the Incentive-Flagyl treatment).**

In the light of our expectation of a distortion effect, we added three nudge treatments to our design to investigate what measures can be taken in order to curb a potential distortion effect cost-effectively.

First, the *Observability* treatment had all of the features of *Incentive-Amoxil* with one critical difference. Participants were also informed that there was a chance that their allocation decisions to the RHC would be made public to other participants in the experiment, i.e. persons with medical training, able to identify a distorted resource allocation. The chosen allocation decisions would be published on the central announcement board of the Department of Public Health, which students check for information frequently. This could

---

<sup>6</sup>This treatment served as a robustness check on whether medical (mis)allocation decisions were sensitive to which medicine was incentivised.

occur if their allocation decision was randomly selected for distribution to the RHC. In this way, their allocation decisions, *but not their identity*, were potentially observable to other participants. This additional observability component allowed the investigation of whether the distortion effect of the financial incentive could be counteracted through a potential public observability nudge, while keeping anonymity.

Second, the *Prof-Norms* treatment had all of the features of *Incentive-Amoxil* with another important difference. We tried to nudge participants with norms emphasising professionalism and patient regard before they made their allocation decisions in Task 2. Specifically, participants in *Prof-Norms* read a one page document that contained statements adapted from a publicly accessible Health Professionals Council of Zambia document<sup>7</sup> emphasising norms of professional medical ethics and patient regard. The one-pager also requested each participant to personally composed one sentence containing their personal reflections on the statements. The one page document contained statements such as:

*"A health professional should allocate available resources to the promotion of good medical practice and the maintenance of such health practitioner's professional competence,"*

*"A health professional should undertake to discharge his/her professional duties diligently and in a manner that best considers the welfare of the patient/client,"*

*"A health practitioner should not only choose but shall also be seen to be choosing the drug, appliance or goods which, in his/her independent professional judgement and having due regard to economy, will best serve the medical/health interest of his/her patient/client."*

Third, the *Need-ORS* treatment also contained all the features of *Incentive-Amoxil*, with one difference. We aimed to nudge participants with information that emphasised the salience of one of the non-incentivised medicines – oral rehydration salts (ORS).<sup>8</sup> Participants were presented with information on ‘knowledge of ORS’ as well as ‘usage of ORS,’ generally for Zambia, as well as for Muchinga Province, the province in which Lubwa RHC is located. This information was obtained from the Zambia Demographic and Health Survey (Zambia Ministry of Health and IFC, 2019). The information rationalised the relatively larger allocation to ORS evident from Figure 2.2 and tried to nudge participants on the clinic’s need for ORS, based on the relatively lower provision of ORS to under age five children in Muchinga Province, relative to the national average.

We implemented the three nudge treatments in an attempt to counteract a potential distortion effect in *Incentive-Amoxil* compared to *Baseline*. We therefore also formulate;

---

<sup>7</sup>Health Professions Council of Zambia (HPCZ), Professional Code of Ethics and Discipline: Fitness to Practice (2014). The HPCZ is legally mandated to register health practitioners in Zambia and regulate their professional conduct.

<sup>8</sup>ORS was chosen for this purpose because it was identified as the non-incentivised medicine participants distorted their allocations away from based on the data collected in October 2018.

**Hypothesis 3, Nudge Effect: The average share of Amoxil allocated to Lubwa RHC in Task 2 is lower in Observability, Prof-Norms and Need-ORS compared to Incentive-Amoxil.**

### 2.2.3 Procedures

We conducted our experiment in Lusaka, Zambia, at the University of Zambia Ridgeway Campus, within the vicinity of the University Teaching Hospital (UTH).<sup>9</sup> UTH is Zambia's largest tertiary hospital and it also serves as a teaching hospital for the University of Zambia's medical students. Being both Zambia's premier teaching hospital and also the highest referral hospital, UTH handles a portfolio of disease that is representative of the disease burden in the country. Ridgeway campus therefore gives us a unique opportunity for conducting this study.

Participants were informed about the study through announcements distributed to their medical student association, through social media platforms and hard copy posters distributed at Ridgeway campus. Participation was voluntary and interested students indicated their willingness to participate through Whatsapp, email and text messages. Participants were then randomly allocated into a session and informed as to the day and time they had been scheduled.

In total, 295 medical students were randomly allocated into one of the six treatments shown in table 2.1 above, while 64 participating medical doctors were randomised into two of the six treatments.<sup>10</sup> In October 2018, 154 medical students participated in 20 experimental sessions conducted while a different set of 141 medical students and 64 medical doctors participated in 15 experimental sessions conducted in May 2019. Each session lasted approximately 90 minutes and all sessions were conducted in a purpose fitted experimental laboratory in the School of Public Health within UTH. All experimental tasks were completed using paper forms.

As described above, all participants were given strictly and exactly eight minutes to complete Task 1. After this, Task 1 sheets were collected and corrected to determine the number of correct entries and the size of each participant's Budget 2. Thereafter, participants received the decision sheets for Task 2, indicating the respective sizes of Budgets 1 and 2. Participants in the *Prof-Norms* treatment read the page containing the nudge text before they completed Task 2. Once Task 2 had been completed, participants completed a questionnaire collecting demographic and socio-economic characteristics.<sup>11</sup>

---

<sup>9</sup>Our experiment received ethical approval from ERES in Zambia, reference number 2018-Aug-021, before conducting it.

<sup>10</sup>In our sample of medical doctors, only the *Incentive-Amoxil* and *Need-ORS* treatments were implemented as a robustness check of our main sample of medical students. The results presented below, replicate among the medical doctors and are available upon request from the authors.

<sup>11</sup>Summary statistics of participant demographic and socio-economic characteristics are shown in table 2.2 below.

Table 2.2: Summary Statistics of Covariates

Variable	Treatment							p-value
	All	<i>Baseline</i>	<i>Incentive-Amoxil</i>	<i>Incentive-Flagyl</i>	<i>Observability</i>	<i>Prof-Norms</i>	<i>Need ORS</i>	
Age	25.1	25.5	25.8	25.1	25.4	24.4	24.4	< 0.00
Male	75.9	83.3	84.0	71.4	87.5	64.0	63.2	0.12
Dist. to Lubwa	713.2	677.1	775.1	687.7	692.5	726.7	719.2	0.74
Year of study	5.5	5.8	5.6	5.5	5.5	5.3	5.2	<b>0.03</b>
Math grade	3.9	4.4	3.9	4.0	3.8	3.7	3.6	<b>0.01</b>
Physics grade	3.9	4.1	3.8	4.1	3.8	3.9	3.8	0.48
Experience	1.09	1.10	1.04	1.12	1.10	1.18	1.04	0.99
Alcohol use	1.65	1.77	1.76	1.71	1.63	1.56	1.47	0.70
Perceived relative income	2.52	2.63	2.50	2.55	2.48	2.60	2.37	0.87
Pro-social motivation	3.58	3.58	3.48	3.55	3.63	3.74	3.49	0.58
Residence	2.14	2.08	1.97	2.10	2.35	2.05	2.20	0.59
Smoke	4.07	8.33	2.00	4.76	5.36	2.00	2.04	0.99
Observations	295	48	50	42	56	50	49	

*Notes:* This table shows participants' mean demographic and socio-economic characteristics for the entire sample and by treatment. The p-values in the last column are based on Kruskal-Wallis tests including all six treatments, with p-values below 0.1 in bold. The variable distance to Lubwa reports the driving distance between Lubwa RHC and each participant's district of origin in Kilometers. The variable age is measured in years. The variable male reports the percentage of the sample that is male. The variable year of study categorises years of study, from second to seventh year. The variables Grade math and grade physics report each participants university mathematics and physics grades and are measured on a 6 point scale from low to high. The variable experience categorises the number of years, if any, that the participants have worked in the health sector, outside their studies, on a four point scale from low to high. The variable alcohol is represents each participants self-reported alcohol consumption and is measured on a five point scale from low to high. The variable 'perceived relative income' categorises each participants relative perception of their income level and is measured on a five point scale from low to high. The variable 'pro-social job motivation' categorises participants' self-reported beliefs about whether their choice of profession in the medical field will make a difference in people's lives, on a four point scale from low to high. The variable residence categorises socio-economic status of each participants area of residence on a four-point scale from low to high income. The variable smoke reports the percentage of participants that smoke.

Each participant then received her or his payment in private in Zambian Kwacha (ZMW). On average, participating medical students earned ZMW 83 (approx. 6 EUR) across all treatments while medical doctors on average earned ZMW 249 (approx. 18 EUR). These stakes were attractive for the respective groups.

## 2.3 Data

Table 2.2 provides summary statistics of participant demographic and socio-economic characteristics across all six treatments. It shows that most characteristics are balanced across the treatments, with the exception of age, years of study and math grades. While the magnitudes of differences for all three variables do not appear economically large, we will control for all the characteristics by using them as covariates in the regression analysis in section 2.4 below to ensure the robustness of findings. The table also reveals that the average participant was male, 25 years old, has studied medical studies for 5.5 years, perceived his income to be low, drank some alcohol, did not smoke and reported being intrinsically motivated to become a medical professional.

## 2.4 Results

In testing for hypothesis 1 on the productivity effect, we first begin by examining the distribution of Task 1 scores across all six treatments graphically in figure 2.3. Recall that the maximum achievable score in Task 1 was 60.<sup>12</sup> The figure thus reveals that Task 1 scores are well distributed between zero and the maximum score of 60, with most participant scores falling between approx. 20 and 30 and very few scores falling at the extremes. This implies that Task 1 successfully mimicked real world effort and was neither too easy nor too difficult for participants to complete and as with most real world tasks, the bulk of participants achieve a score close to the mean.

To further test the productivity effect, we compare average Task 1 scores among the six treatments. As noted in section 2.2 above, the greater the score in Task 1, the greater the personal earnings participants could receive in all treatments except *Baseline*, ceteris paribus. Of course, Task 1 scores still yielded a greater Budget 2 to allocate medicine to Lubwa RHC in *Baseline* as in all other treatments. However, this effect can only be explained by an intrinsic motivation to exert effort to generate more money to purchase more medicine for Lubwa RHC. Figure 2.4 depicts the mean Task 1 scores in all six treatments. We employ a series of pairwise two-sided Mann-Whitney tests to test whether we can find evidence for a productivity effect of the performance-based financial incentive in this task across the treatments. The pairwise Mann-Whitney tests yield only one out of five significant difference ( $p < 0.05$ ): greater Task 1 scores in *Incentive-Flagyl* than in *Baseline* (an average score of 28.64 vs. 25.43, respectively). Hence, there is only limited evidence for a productivity effect and in the majority of cases we cannot reject the null

---

<sup>12</sup>Before conducting the experiment, we pretested Task 1 with a number of volunteering colleagues. As with most of our colleagues, the eight-minute time-frame appeared successful in yielding a somewhat normal distribution and avoiding scores of 0 and 60.

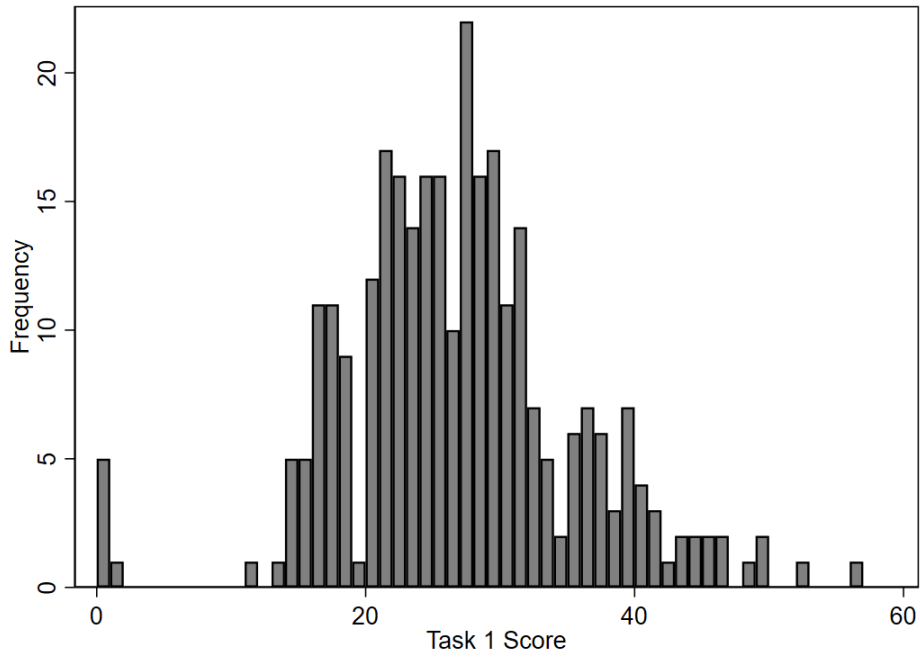


Figure 2.3: The Distribution of Task 1 Scores

*Notes:* This figure shows the distribution of participant scores in the health-framed real effort task, Task 1.

hypothesis. It may be that the intrinsic motivation of the participants is sufficient to motivate greater effort in Task 1. Thus, addition of the performance-based incentive in all other treatments, appears largely ineffective in increasing effort, relative to *Baseline*.

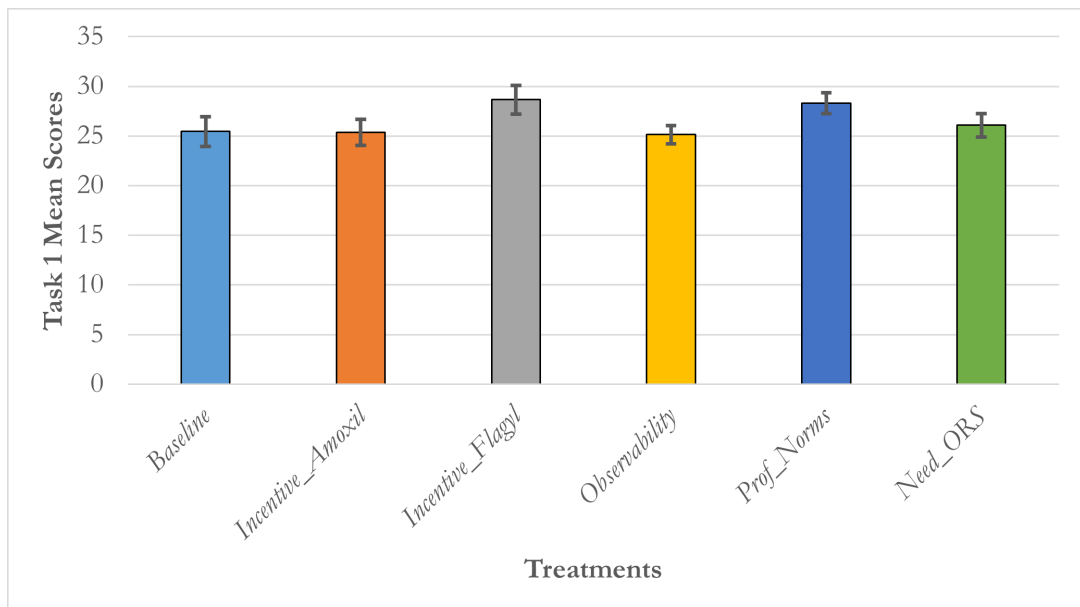


Figure 2.4: Health-Framed Real Effort Task Scores by Treatment

*Notes:* This figure shows mean participant scores by treatment in the health-framed real effort task, Task 1.

Next, we test for evidence for hypothesis 2, i.e. whether there is a distortion effect for the allocation of medicines towards the incentivized Amoxil and Flagyl (in *Incentive-Amoxil* and *Incentive-Flagyl* respectively) compared to Baseline.<sup>13</sup> Two Mann-Whitney tests clearly reject the null hypotheses in favor of greater Amoxil allocations for both Budgets 1 and 2 in *Incentive-Amoxil* compared to Baseline (approx. 55% vs. 32%,  $p < 0.000$  for both tests). Likewise, two further Mann-Whitney tests show that there are greater allocations of Budgets 1 and 2 to Flagyl in *Incentive-Flagyl* compared to Baseline (66-70% vs. 15%,  $p < 0.000$  for both tests). Hence, we find clear evidence in favor of hypothesis 2, i.e. that there is indeed a statistically and economically large distortion effect by the pay-for-performance incentives. Figure 2.5 graphically depicts the mean allocations of both Budget 1 and 2 to the medicines, across all six treatments in percentages.

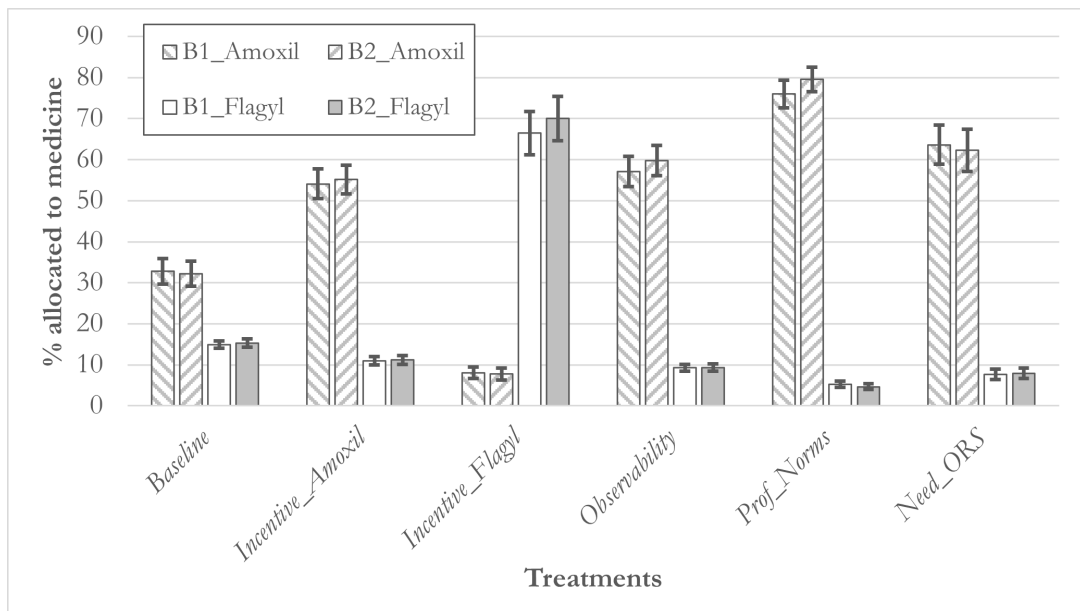


Figure 2.5: Medicine Allocations by Treatment

*Notes:* This figure shows the mean allocations across the incentivised and non-incentivised medicines by treatment in the medicine allocation task, Task 2.

Next, we assess whether any of the three nudge treatments are able to counter the distortion caused by the performance-based incentive as hypothesis 3 postulates. As Figure 2.5 suggests, this is not the case. When testing the average share of allocated Amoxil in *Observability*, *Prof-Norms* and *Need-ORS* pairwise against the average share in *Baseline*, the three respective two-sided Mann-Whitney tests reject hypothesis 3 at  $p < 0.000$ . That is, the three nudges are not able to counter the distortion effect of the performance-based incentive. Comparing the three nudge treatments, again pairwise, against *Incentive-Amoxil*, the mean shares of Amoxil in *Observability* and *Need-ORS* are not different from *Incentive-Amoxil* with  $p > 0.1$  for all tests and even higher in *Prof-Norms* with  $p < 0.01$  for both Budget 1 and 2.

<sup>13</sup>For this analysis we excluded 37 participants whose allocation percentages for Budget 1 and 2 did not add to 100%.

Table 2.3: Regressions of Task 1 Score and Amoxil Allocations

	(1)	(2)	(3)
	Task 1 score	% Amoxil Budget 1	% Amoxil Budget 2
<i>Baseline</i>	-1.146 (1.730)	-16.559 (5.016)***	-20.461 (4.951)***
<i>Incentive-Flagyl</i>	2.924 (1.818)	-43.595 (5.366)***	-46.635 (5.282)***
<i>Observability</i>	-0.214 (1.651)	4.982 (4.803)	4.595 (4.704)
<i>Prof-Norms</i>	1.757 (1.763)	25.259 (5.091)***	24.896 (5.041)***
<i>Need ORS</i>	-0.495 (1.776)	14.866 (5.212)***	9.784 (5.315)*
Task 1 score		0.201 (0.178)	0.293 (0.181)
Age	-0.807 (0.290)***	1.761 (0.854)**	2.062 (0.851)**
Male	-3.608 (1.312)***	0.295 (3.937)	5.201 (3.823)
Distance to Lubwa	-0.001 (0.001)	< 0.000 (0.004)	< 0.000 (0.004)
Year of study	1.305 (0.532)**	-3.128 (1.605)*	-3.603 (1.543)**
Math grade	1.836 (0.577)***	-2.420 (1.719)	-2.052 (1.666)
Physics grade	0.129 (0.644)	3.720 (1.848)**	2.141 (1.823)
Experience	-2.586 (1.064)**	0.915 (3.234)	0.382 (2.981)
Alcohol use	-0.174 (0.530)	-1.384 (1.535)	-1.014 (1.484)
Perceived relative income	0.426 (0.556)	-2.602 (1.595)	-2.838 (1.605)*
Pro-social job motivation	0.629 (0.816)	-1.218 (2.348)	-1.467 (2.319)
Smoke	-2.100 (2.806)	-4.401 (8.124)	-3.977 (8.052)
Constant	34.273 (7.617)***	25.050 (22.743)	23.674 (23.257)
Observations	276	256	257

*Notes:* This table shows Ordinary Least Squares (OLS) regressions of the impact of each treatment on participant score in the health-framed real effort task, Task 1, in column 1 and participant allocations to Amoxil, in percentages, from Budgets 1 and 2, in columns 2 and 3, respectively. *Incentive-Amoxil* is the omitted reference treatment. Fewer than 295 observations, are shown in some regressions due to missing questionnaire answers and participant allocation percentages not summing to 100%. Standard errors are reported in the parentheses.  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Hence, the three nudge treatments are ineffective to counter the distortion effect of the pay-for-performance incentive and the professional norms nudge even backfires and causes a greater distortion effect.

Next, we perform regression analyses based on equation 2.3 below. Recall that the Kruskal-Wallis tests for the balance of covariates revealed differences for age, year of study and math grades between treatments, as reported in Table 2.2 above. We thus control for all participant characteristics in the regressions, a move that also permits us to investigate correlations with the covariates.

$$y_i = \beta_0 + \beta_{1v}Treat_v + \beta_{2i}X_i + \gamma_t + \theta_s + \tau_i \quad (2.3)$$

In this equation,  $y_i$  represents our outcomes of interest, the score in Task 1 and the shares of Amoxil in Task 2, from Budgets 1 and 2 respectively, for participant  $i$ .  $Treat_v$  is a treatment indicator, with *Incentive-Amoxil* being the omitted reference treatment.  $X_i$

are the participant characteristics shown in table 2.2 above.<sup>14</sup>  $\gamma_t$  and  $\theta_s$  denote a time fixed effect for the second data collection period and session fixed effects, respectively, while  $\tau_i$  is the error term.

Table 2.3 below reports the results. In column 1, Task 1 score is the dependant variable, while in columns 2 and 3, the allocations to Amoxil from Budgets 1 and 2 respectively, in percentages, are the outcome variables. Column 1 of the table reveals no treatment effects of the various treatments on the Task 1 score. This is in line with the observations based on figure 2.4. However, rather interestingly, column 1 also reveals that older and male participants appear to have lower Task 1 scores, while more years of study and higher math grades predict higher scores in Task 1. Both positive correlations appears intuitive to us. Similarly, columns 2 and 3 show the the allocations of Amoxil of Budgets 1 and 2 and largely confirm the treatment effects discussed above based on the Mann-Whitney tests and figure 2.5 above.

One notable difference is that while the Mann-Whitney tests did not detect any differences in Amoxil allocations between *Incentive-Amoxil* and *Need-ORS*, regression models 2 and 3 indicate treatment effects in the direction of greater distortion in *Need-ORS* – in the same direction as in *Prof-Norms*. Hence, we find further evidence that two out of three nudges backfired. Covariates show positive correlations of a participant’s age with the percentage of Amoxil and negative correlations for years of study. It is likely that these two effects cancel out mechanically for many students as they progress, so we refrain from speculating about the underlying mechanisms.

In order to explore the nuances in participants’ allocation decisions based on the quantitative results above, we turn to an examination of the qualitative results. Recall from section 2.2 that at the end of each session, each participant completed an ex-post questionnaire collecting their demographic and socio-economic characteristics. In addition to this information, participants were also asked to explain the rationale behind their resource allocation decisions, in the event they decided to make resource allocation decisions that differed from the optimal allocation of resources data from Lubwa RHC as depicted in figure 2.2. This was an open-ended question to which participants were free to explain their decision making rationale and provide as many justifications as they wished.

Participant responses to this question were then transcribed and categorised into broad categories of justifications;<sup>15</sup>

1. Same allocations being chosen as in the optimal allocation data obtained from Lubwa RHC (‘Same as optimal budget’),

---

<sup>14</sup>We exclude the variable ‘residence’ given that many observations would drop from the analysis.

<sup>15</sup>These coded responses were then converted to dummy variables for each category, equalling 1 when a participant mentioned a specific category and 0 otherwise. The written responses are available from the authors upon request. Our coding also included the reasons ‘ORS can be made at home’, ‘Budget sizes’ and ‘Other medicines for disease burden’. We see no differences between treatments regarding the prevalence of these reason categories.

2. The financial incentive,
3. Amoxil being important given the disease burden in the catchment population of Lubwa RHC ('Amoxil for disease burden') and,
4. Flagyl being important given the disease burden in the catchment population of Lubwa RHC ('Flagyl for disease burden').

Figure 2.6 below presents these results. It shows that between 13% and 16% of participants explicitly mentioned the financial incentive as a justification for their allocation decisions in all treatments apart from *Baseline*. This justification seems to clearly reflect the performance-based framework built into the five treatments. The figure also reveals another noteworthy observation: medical justifications for allocation to the incentivised medications, Amoxil and Flagyl, respectively. Indeed more participants justify their allocations to Amoxil (Flagyl in *Incentive-Flagyl*) by mentioning that the disease burden at the clinic justifies their own increased allocation to Amoxil (Flagyl). The shares of participants providing this justification is significantly greater in all other treatments compared to *Baseline*, with  $p < 0.05$  for all cases. The *Incentive-Flagyl* treatment allows us to see how this self-serving justification clearly changes from one medicine to another, simply because the performance-based financial incentive is attached to that medicine. Hence, we detect self-serving justifications in all treatments with performance-based incentives.

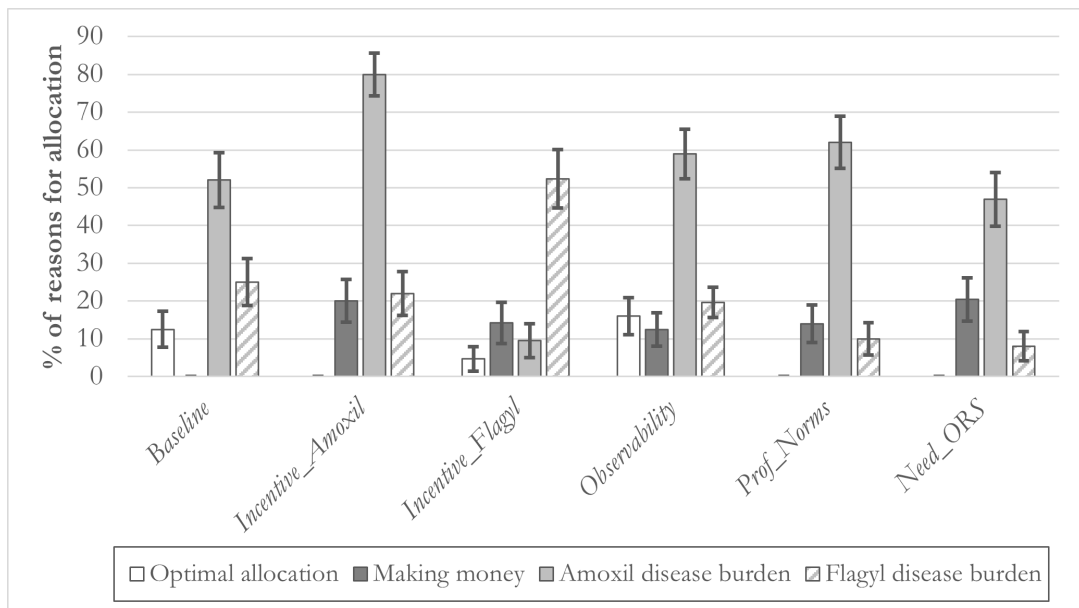


Figure 2.6: Participant Justifications for Medicine Allocations

*Notes:* This figure shows the proportion of participants providing each category of justification for their allocations across the incentivised and non-incentivised medicines by treatment in the medicine allocation task, Task 2.

We therefore find that increased allocations and medical justifications simply follow the medicine that is incentivised and it is not necessarily the characteristics of the medicine that determine increased allocations. The qualitative findings thus reveal a mechanism

steering the distortion effect: medical decision makers trying to convince others and potentially also themselves that their distorted decisions are medically justified.

## 2.5 Conclusion

The effectiveness of healthcare systems depends to a large extent on the performance of its healthcare providers. Health programmes in which healthcare providers are paid performance-based financial incentives have thus become common in low income developing countries. The goal of such programmes is increasing productivity, improving health outcomes and ultimately ensuring prosperity in such countries. However, evidence on the effects of performance-based financial incentives on healthcare providers decision making in low income developing country settings has remained scarce.

In this study, we investigate whether such performance-based incentives can result in two effects: a productivity effect, due to an increase in effort and a distortion effect, in which effort does not increase but decision making is solely redirected towards an incentivised service. We utilise a health-framed field experiment, with six between-subjects treatments, two experimental tasks and a sample of 359 Zambian medical students and medical doctors. In our field experiment, we create an incentivised decision-making situation involving the allocation of real and much-needed medicines to an actual rural health clinic. Healthcare providers in some treatment conditions thus face a decision situation in which they can either be consistent with the needs of a real rural Zambian health clinic or deviate from these needs so as to maximise their own payoff with the distortion effect being the result.

We find no evidence for a productivity effect, but a statistically and economically significant distortion effect. The distortion effect is remarkably unresponsive to counteracting measures in the form of three nudge treatments using observability of the allocation decision, emphasising norms of professionalism and flagging the importance of a non-incentivised health resource. The latter findings in particular, differ with evidence in the literature revealing that making the professional norms in the Hippocratic Oath salient is associated with enhanced altruism of participating medical student in dictator games (Kesternich et al., 2015).

Our findings therefore imply that the use of performance-based financial incentives in the healthcare sector should be cautious. This is owing to the fact that they may signal payoff maximising behaviours to healthcare providers, without signalling commensurate productivity enhancement. This payoff maximising behaviour may lead to a pure distortion effect towards the aspects attached to the performance-based financial incentives and away from those aspects to which no such incentive is attached. This is particularly important as the distortion effect in our experiment was not accompanied by greater productivity than attained by an unconditional fixed payment. This also implies that policymakers wishing to use financial incentives in their health programmes, are advised to explore the role of unconditional financial incentives, that do not signal distorting payoff

maximising behaviour, but also lead to comparable levels of productivity as performance-based financial incentives. The observation that the distortion effects of the financial incentive are so pervasive that the allocation decisions are unresponsive to counteracting nudging measures is noteworthy. It is also worrying that participants in our experiment give self-serving medical justifications for their distorted allocation decisions.

In our field experiment, participants made decisions concerning the allocation of essential medicines. Through the donation to Lubwa RHC their decisions relate to real health resources that have an effect on real patients outside the experimental setting. We do not believe that our findings imply that medical doctors necessarily distinguish arriving patients receiving urgent care in their facilities based on whether their disease condition is associated with a performance-based financial incentivised for themselves or not. We do, however, believe that our findings are relevant for medical staff's actual behavior. Specifically, our findings may be applicable to decision situations related to demand creation and outreach activities. For such situations, health care providers must take steps to initiate the provision of healthcare in their catchment areas. The presence of performance-based financial incentives in such situations may imply that healthcare providers may focus on payoff maximisation at the expense of catering to the most pressing health problems in their catchment areas by creating demand for health conditions attached to the incentives.

## Bibliography

- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics* 97(4), 543–569.
- Alkema, L., D. Chou, D. Hogan, S. Zhang, A.-B. Moller, A. Gemmill, D. M. Fat, T. Boerma, M. Temmerman, C. Mathers, et al. (2016). Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the un maternal mortality estimation inter-agency group. *The Lancet* 387(10017), 462–474.
- Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011). Inside the war on poverty: The impact of food stamps on birth outcomes. *The Review of Economics and Statistics* 93(2), 387–403.
- Amarante, V., M. Manacorda, E. Miguel, and A. Vigorito (2016). Do cash transfers improve birth outcomes? evidence from matched vital statistics, program, and social security data. *American Economic Journal: Economic Policy* 8(2), 1–43.
- Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics* 10, 439–463.
- Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review* 110(5), 1355–94.
- Ashraf, N., O. Bandiera, and B. K. Jack (2014). No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics* 120, 1–17.
- Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21(1), 1–42.
- Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014). Conditional, unconditional and everything in between: A systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness* 6(1), 1–43.
- Barber, S. L. and P. J. Gertler (2008). The impact of mexico’s conditional cash transfer programme, oportunitades, on birthweight. *Tropical Medicine and International Health* 13(11), 1405–1414.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. Soucat, J. Sturdy, and C. M. Vermeersch (2011). Effect on maternal and child health services in rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet* 377(9775), 1421–1428.
- Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies* 70(3), 489–520.

- Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *American Economic Review* 95(3), 616–636.
- Besley, T. J. and M. Ghatak (2014). Solving agency problems: Intrinsic motivation, incentives, and productivity. Technical report.
- Bonfrer, I., E. Van de Poel, and E. Van Doorslaer (2014). The effects of performance incentives on the utilization and quality of maternal and child care in burundi. *Social Science & Medicine* 123, 96–104.
- Bowles, S. and S. Polania-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature* 50(2), 368–425.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies, and D. Wiesen (2013). *How Effective are Pay-for-Performance Incentives for Physicians?—A Laboratory Experiment*. Number 413. Ruhr Economic Papers.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz, and D. Wiesen (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization* 131, 17–23.
- Cabrera, J. M. and D. Webbink (2020). Do higher salaries yield better teachers and better student outcomes? *Journal of Human Resources* 55(4), 1222–1257.
- Callaway, B. and P. H. Sant’Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.
- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal* 17(2), 372–404.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal* 14(4), 909–946.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6), 2295–2326.
- Cardona-Sosa, L. and C. Medina (2017). The effects of in utero programs on birth outcomes: The case of buen comienzo. *Journal of the Latin American and Caribbean Economic Association* 17(2), 93–134.
- Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal* 18(1), 234–261.
- Cattaneo, M. D., M. Jansson, and X. Ma (2019). lpdensity: Local polynomial density estimation and inference. *arXiv preprint arXiv:1906.06529*.
- Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531), 1449–1455.

- Chelwa, G., M. Pellicer, and M. Maboshe (2019). Teacher pay and educational outcomes: Evidence from the rural hardship allowance in zambia. *South African Journal of Economics* 87(3), 255–282.
- Chibuye, B., P. Hangoma, and M. Khadjavi (2021). The dark side of financial incentives in a health setting - experimental evidence from zambia. Working paper, Christian-Albrechts-University of Kiel.
- Chibwasha, C. J., A. Zanolini, M. Smid, B. Vwalika, M. P. Kasaro, M. Mwanahamuntu, J. S. Stringer, and E. M. Stringer (2016). Predictors and outcomes of low birth weight in lusaka, zambia. *International Journal of Gynecology & Obstetrics* 134(3), 309–314.
- Cross, S. S. and M. Baernholdt (2017). Identifying factors for worker motivation in zambia’s rural health facilities. *Journal of Health Care for the Poor and Underserved* 28(1), 63–70.
- Currie, J. and D. Almond (2011). Human capital development before age five. In *Handbook of Labor Economics*, Volume 4, pp. 1315–1486. Elsevier.
- Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics* 128(3), 1169–1218.
- Das, J. and J. Hammer (2014). Quality of primary care in low-income countries: Facts and economics. *Annual Review of Economics* 6(1), 525–553.
- de Bekker-Grob, E. W., M. Ryan, and K. Gerard (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21(2), 145–172.
- De Chaisemartin, C. and X. d’Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9), 2964–96.
- De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2018). Double for nothing? experimental evidence on an unconditional teacher salary increase in indonesia. *The Quarterly Journal of Economics* 133(2), 993–1039.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18(1), 105–128.
- Deci, E. L., R. Koestner, and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125(6), 627–693.
- Deserranno, E. (2019). Financial incentives as signals: Experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics* 11(1), 277–317.
- Díaz, J.-J. and V. Saldarriaga (2019). Encouraging use of prenatal care through conditional cash transfers: Evidence from juntos in peru. *Health Economics* 28(9), 1099–1113.

- Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics* 27(6), 1436–1450.
- Dussault, G. and M. C. Franceschini (2006). Not enough there, too many here: Understanding geographical imbalances in the distribution of the health workforce. *Human Resources for Health* 4(1), 1–16.
- Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics* 24(1), 211–223.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics* 108(2), 437–459.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review* 42(1), 1–34.
- Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy* 21(3), 673–705.
- Ferrinho, P., S. Siziya, F. Goma, and G. Dussault (2011). The human resource for health situation in zambia: Deficit and maldistribution. *Human Resources for Health* 9(1), 1–11.
- Figueiredo Walter, T. (2020). Misallocation in the public sector? cross-country evidence from two million primary schools. Technical report, STICERD Working Paper No. EOPP 70.
- Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. *Handbook of Economic Field Experiments* 2, 467–514.
- Friedman, J., J. N. Qamruddin, C. Chansa, and A. K. Das (2016). Impact evaluation of zambia's health results-based financing pilot project. Technical report, The World Bank.
- Gage, A. and S. Bauhoff (2021). The effects of performance-based financing on neonatal health outcomes in burundi, lesotho, senegal, zambia and zimbabwe. *Health Policy and Planning*.
- Gertler, P. and C. Vermeersch (2012). *Using Performance Incentives to Improve Health Outcomes*. The World Bank.
- Gneezy, U. and J. A. List (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5), 1365–1384.
- Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3), 791–810.
- Godager, G. and D. Wiesen (2013). Profit or patients' health benefit? exploring the heterogeneity in physician altruism. *Journal of Health Economics* 32(6), 1105–1116.

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.
- Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization* 106, 367–378.
- Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.
- Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature* 42(4), 1009–1055.
- Hennig-Schmidt, H., R. Selten, and D. Wiesen (2011). How payment systems affect physicians’ provision behaviour — an experimental investigation. *Journal of Health Economics* 30(4), 637–646.
- Herbst, C. H., M. Vledder, K. Campbell, M. Sjöblom, and A. Soucat (2011). The human resources for health crisis in zambia. *Washington (DC): World Bank*.
- Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.* 7, 24.
- Hoynes, H., D. Miller, and D. Simon (2015). Income, the earned income tax credit and infant health. *American Economic Journal: Economic Policy* 7(1), 172–211.
- Hoynes, H., M. Page, and A. H. Stevens (2011). Can targeted transfers improve birth outcomes?: Evidence from the introduction of the wic program. *Journal of Public Economics* 95(7-8), 813–827.
- Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635.
- James, N., K. Lawson, and Y. Acharya (2020). Evidence on result-based financing in maternal and child health in low-and middle-income countries: A systematic review. *Global Health Research and Policy* 5(1), 1–15.
- Kaarboe, O. and L. Siciliani (2011). Multi-tasking, quality and pay for performance. *Health Economics* 20(2), 225–238.
- Kesternich, I., H. Schumacher, and J. Winter (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? *Journal of Public Economics* 131, 1–11.
- Kramer, M. S. (1987). Determinants of low birth weight: Methodological assessment and meta-analysis. *Bulletin of the World Health Organization* 65(5), 663.
- Kramer, M. S. (2003). The epidemiology of adverse pregnancy outcomes: An overview. *The Journal of Nutrition* 133(5), 1592S–1596S.

- Lagarde, M. and D. Blaauw (2014). Pro-social preferences and self-selection into jobs: Evidence from south african nurses. *Journal of Economic Behavior & Organization* 107, 136–152.
- Lagarde, M. and D. Blaauw (2017). Physicians’ responses to financial and social incentives: A medically framed real effort experiment. *Social Science & Medicine* 179, 147–159.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review* 90(5), 1346–1361.
- Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281–355.
- Lemiere, C., C. Herbst, N. Jahanshahi, and E. Smith (2011). Reducing geographical imbalances of health workers in sub-saharan africa: A labor market perspective on what works, what does not, and why. world bank; 2011.
- Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2019). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly Journal of Economics* 134(3), 1627–1673.
- McCoy, D., S. Bennett, S. Witter, B. Pond, B. Baker, J. Gow, S. Chand, T. Ensor, and B. McPake (2008). Salaries and incomes of health workers in sub-saharan africa. *The Lancet* 371(9613), 675–681.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2), 698–714.
- Miller, G. and K. S. Babiarz (2013). Pay-for-performance incentives in low-and middle-income country health programs. Technical report, National Bureau of Economic Research.
- Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, and S. Rozelle (2012). Effectiveness of provider incentives for anaemia reduction in rural china: A cluster randomised trial. *Bmj* 345.
- Mullen, K. J., R. G. Frank, and M. B. Rosenthal (2010). Can you get what you pay for? pay-for-performance and the quality of healthcare providers. *The Rand Journal of Economics* 41(1), 64–91.
- Muralidharan, K. and N. Prakash (2017). Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics* 9(3), 321–50.
- Neilson, C., M. Bobba, T. Ederer, G. Leon-Ciliotta, and M. Nieddu (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in peru. Working paper, National Bureau of Economic Research.
- Ohlsson, A. and P. Shah (2008). *Determinants and Prevention of Low Birth Weight: A Synopsis of the Evidence*. Institute of Health Economics.

- Olken, B. A., J. Onishi, and S. Wong (2014). Should aid reward performance? evidence from a field experiment on health and education in indonesia. *American Economic Journal: Applied Economics* 6(4), 1–34.
- Paul, E., M. L. Dramé, J.-P. Kashala, A. E. Ndema, M. Kounnou, J. C. Aïssan, and K. Gyselincx (2018). Performance-based financing to strengthen the health system in benin: Challenging the mainstream approach. *International Journal of Health Policy and Management* 7(1), 35.
- Pega, F., S. Y. Liu, S. Walter, R. Pabayo, R. Saith, and S. K. Lhachimi (2017). Unconditional cash transfers for reducing poverty and vulnerabilities: Effect on use of health services and health outcomes in low-and middle-income countries. *Cochrane Database of Systematic Reviews* (11).
- Prust, M. L., A. Kamanga, L. Ngosa, C. McKay, C. M. Muzongwe, M. T. Mukubani, R. Chihinga, R. Misapa, J. W. van den Broek, and N. Wilmink (2019). Assessment of interventions to attract and retain health workers in rural zambia: A discrete choice experiment. *Human Resources for Health* 17(1), 26.
- Pugatch, T. and E. Schroeder (2014). Incentives for teacher relocation: Evidence from the gambian hardship allowance. *Economics of Education Review* 41, 120–136.
- Sappington, D. E. (1991). Incentives in principal-agent relationships. *Journal of Economic Perspectives* 5(2), 45–66.
- Shen, G. C., H. T. H. Nguyen, A. Das, N. Sachingongu, C. Chansa, J. Qamruddin, and J. Friedman (2017). Incentives to change: Effects of performance-based financing on health workers in zambia. *Human Resources for Health* 15(1), 20.
- Sherry, T. B. (2016). A note on the comparative statics of pay-for-performance in health care. *Health Economics* 25(5), 637–644.
- Sherry, T. B., S. Bauhoff, and M. Mohanan (2017). Multitasking and heterogeneous treatment effects in pay-for-performance in healthcare: Evidence from rwanda. *American Journal of Health Economics* 3(2), 192–226.
- Singh, P. (2015). Performance pay and information: Reducing child undernutrition in india. *Journal of Economic Behavior & Organization* 112, 141–163.
- Singh, P. and W. A. Masters (2020). Performance bonuses in the public sector: Winner-take-all prizes versus proportional payments to reduce child malnutrition in india. *Journal of Development Economics* 146, 102–295.
- United Nations, U. (2016). Transforming our world: The 2030 agenda for sustainable development.
- Zambia Ministry of Health, M. (2011). National health strategic plan (2011 - 2015). Technical report, Government of the Republic of Zambia.

Zambia Ministry of Health, M. (2017). National health sector strategic plan. Technical report, Government of the Republic of Zambia., Lusaka.

Zambia Ministry of Health, Z. S. A. and M. IFC (2019). Zambia demographic and health survey 2018. Technical report, Zambia Statistics Agency, Ministry of Health, and ICF, Lusaka, Zambia, and Rockville, Maryland, USA.

Zambia Ministry of National Development Planning, M. (2017). Seventh national development plan (sndp). Technical report, Zambia Ministry of National Development Planning.

# Appendices

## 2.A Appendix to Chapter Two

### 2.A.1 Lubwa Receives Medicines



Figure 2.A.1: Lubwa Rural Clinic Receives of Donated Medicines.

## 2.A.2 Instructions

*Instructions: Baseline*

ID # \_\_\_\_\_

General Instructions for the Participants

You are now taking part in an economic and health-related study. For this, it is important that you read the following instructions carefully.

The instructions you were given are exclusively for your private information and have to be given back after the study. During the study it is strictly forbidden to talk. If you have any questions, please raise your hand, and then one of the researchers will come to you and answer your questions in private.

Please be aware that your mobile phone and all other electronic devices have to be switched off. Any violation of these rules will lead to the exclusion from the study and from any payments.

All your decisions in the study are made anonymously. No other participant will be able to match your decisions with your identity.

For your participation in the study you will get a guaranteed show-up fee of 50 Kwacha.

There will be two decision-making tasks in this study, Task 1 and Task 2. These two tasks are interrelated and a budget that you generate in Task 1 will be carried over for the decision-making in Task 2. There is a chance that your decision in Task 2 will result in an actual, real allocation of medicine to a Rural Health Centre, so we ask you to pay close attention to the following instructions.

After all participants have made their decisions, we will ask you some questions in a questionnaire.

On the following pages we will describe the exact course of the study.

## Task 1

There are two budgets in this study, **Budget 1 and Budget 2**. In Task 2 you will use these two budgets to allocate actual medicine to a real rural health centre: Lubwa Rural Health Centre in Chinsali District, Muchinga Province.

**Budget 1 is always fixed at 1,000 Kwacha.**

However, **you** are able to determine the **size of Budget 2** in this Task 1.

Task 1 involves a sample Health Information Aggregation 1 (HIA.1) form. In completing the HIA.1 form, you will be asked to **sum the entries** in **each row** of the form and **enter the total** into the respective **grey-shaded boxes**.

**Each correctly summed row of entries** in the HIA.1 form **will add 50 Kwacha** to your **Budget 2**. That is, **the number** of all correctly summed entries will determine your **Budget 2**.

Example 1: If you fill in 15 grey-shaded boxes correctly, your Budget 2 will be  $15 \times 50$  Kwacha = 750 Kwacha.

Example 2: If you fill in 41 grey-shaded boxes correctly, your Budget 2 will be  $41 \times 50$  Kwacha = 2050 Kwacha.

In the example in Table 1, the correctly summed row total is  $2+6+3=11$ . This one correctly summed row would add 50 Kwacha to your Budget 2. **The more rows you sum correctly in the given time, the greater will be your Budget 2.**

**Table 1. Short Sample HIA.1 Form.**

	OPD First Attendance			
DIAGNOSES	Under 1 year	1-5 years	6+ Years	Total
<b>NOTIFIABLE DISEASES</b>				
Acute flaccid paralysis	2	6	3	<b>11</b>
Cholera	23	10	22	
...	...	...	...	

Each participant will be given **exactly 8 minutes** within which to complete as many summations as she/he likes and can manage.

Exercise:

**Question 1:** How large will **Budget 1** be if you sum **19 rows** correctly? Answer: K \_\_\_\_\_

**Question 2:** How large will **Budget 2** be if you sum **33 rows** correctly? Answer: K \_\_\_\_\_

## Task 2

In Task 2, you will be presented with a **list of consumed medicines** in a Zambian Rural Health Centre (RHC), as well as an estimate of the proportions (in %) of the costs of these medicines. Below we present you **actual medicine consumption data** from Lubwa RHC (catchment population: 9,500), Chinsali District, Muchinga Province, Zambia, in one month (Table 2).

**Please be aware that the distribution below represents an optimal allocation of health resources to Lubwa RHC based on the prevailing disease burden in the catchment area.**

**Table2: Percentages of Costs per Month for Five Consumed Medicines, Lubwa RHC.**

<b>Medication</b>	<b>Percentage of Total Cost per Month</b>
Paracetamol	12%
Oral Rehydration Salts	50 %
Amoxil	21 %
Piriton	3 %
Metronidazole (Flagyl)	14 %
	<b>Total: 100%</b>

In Task 2 we are asking you to **make your own allocation decisions** (choose your own percentages) to Lubwa RHC for the medicine items in Table 2, which could be the same as the optimal allocation in Table 2 or whatever you deem better. **Note that your allocation may be randomly selected, actually bought and sent to Lubwa RHC in October 2018.**

Recall from Task 1 that there are two budgets for which we ask you to make your allocation decisions: Budget 1 and Budget 2. You can **freely allocate between 0% and 100%** of the respective budget to **any of the five medicines on the list**, with the restriction that all percentages for a given budget **need to sum to exactly 100%**.

**Important:** At the end of all the sessions of this study, we will **randomly draw one** of the participants' ID numbers. We will then **actually buy and send** the corresponding allocated **medicine of both Budget 1 and Budget 2** to Lubwa RHC. The RHC is located in a rural area whose catchment population has very limited access to medical services. Your allocation choice will therefore have a **real impact on the available medicine in Lubwa RHC.**

All participants will be invited via e-mail or SMS to watch this random draw.

To this end, the research team will work closely with the Ministry of Health, to **purchase the medicines** on the list and donate them to the Lubwa RHC **exactly as per the randomly chosen participant's allocation decisions**. An official letter from the Ministry of Health will be available for all participants to see at the notice board of the School of Public Health (UTH).

This donation will be financed from research funds provided by German Academic Exchange Service (DAAD), Germany.

*Instructions: Incentive-Amoxil*

ID # \_\_\_\_\_

### General Instructions for the Participants

You are now taking part in an economic and health-related study. Depending on your decision-making and chance you can earn a substantial amount of money. For this, it is important that you read the following instructions carefully.

The instructions you were given are exclusively for your private information and have to be given back after the study. During the study it is strictly forbidden to talk. If you have any questions, please raise your hand, and then one of the researchers will come to you and answer your questions in private.

Please be aware that your mobile phone and all other electronic devices have to be switched off. Any violation of these rules will lead to the exclusion from the study and from any payments.

All your decisions in the study are made anonymously. No other participant will be able to match your decisions with your identity.

For your participation in the study you will get a guaranteed show-up fee of 50 Kwacha. Any additional payments depend on chance and your own decision-making.

There will be two decision-making tasks in this study, Task 1 and Task 2. These two tasks are interrelated and a budget that you generate in Task 1 will be carried over for the decision-making in Task 2. There is a chance that your decision in Task 2 will result in an actual, real allocation of medicine to a Rural Health Centre, so we ask you to pay close attention to the following instructions.

After all participants have made their decisions, we will ask you some questions in a questionnaire.

On the following pages we will describe the exact course of the study.

## Task 1

There are two budgets in this study, **Budget 1 and Budget 2**. In Task 2 you will use these two budgets to allocate actual medicine to a real rural health centre: Lubwa Rural Health Centre in Chinsali District, Muchinga Province.

**Budget 1 is always fixed at 1,000 Kwacha.**

However, **you** are able to determine the **size of Budget 2** in this Task 1.

Task 1 involves a sample Health Information Aggregation 1 (HIA.1) form. In completing the HIA.1 form, you will be asked to **sum the entries** in **each row** of the form and **enter the total** into the respective **grey-shaded boxes**.

**Each correctly summed row of entries** in the HIA.1 form **will add 50 Kwacha** to your **Budget 2**. That is, **the number** of all correctly summed entries will determine your **Budget 2**.

Example 1: If you fill in 15 grey-shaded boxes correctly, your Budget 2 will be  $15 \times 50$  Kwacha = 750 Kwacha.

Example 2: If you fill in 41 grey-shaded boxes correctly, your Budget 2 will be  $41 \times 50$  Kwacha = 2050 Kwacha.

In the example in Table 1, the correctly summed row total is  $2+6+3=11$ . This one correctly summed row would add 50 Kwacha to your Budget 2. **The more rows you sum correctly in the given time, the greater will be your Budget 2.**

**Table 1. Short Sample HIA.1 Form.**

	OPD First Attendance			
DIAGNOSES	Under 1 year	1-5 years	6+ Years	Total
<b>NOTIFIABLE DISEASES</b>				
Acute flaccid paralysis	2	6	3	11
Cholera	23	10	22	
...	...	...	...	

Each participant will be given **exactly 8 minutes** within which to complete as many summations as she/he likes and can manage.

Exercise:

**Question 1:** How large will **Budget 1** be if you sum **19 rows** correctly? Answer: K \_\_\_\_\_

**Question 2:** How large will **Budget 2** be if you sum **33 rows** correctly? Answer: K \_\_\_\_\_

## Task 2

In Task 2, you will be presented with a **list of consumed medicines** in a Zambian Rural Health Centre (RHC), as well as an estimate of the proportions (in %) of the costs of these medicines. Below we present you **actual medicine consumption data** from Lubwa RHC (catchment population: 9,500), Chinsali District, Muchinga Province, Zambia, in one month (Table 2).

**Please be aware that the distribution below represents an optimal allocation of health resources to Lubwa RHC based on the prevailing disease burden in the catchment area.**

**Table2: Percentages of Costs per Month for Five Consumed Medicines, Lubwa RHC.**

<b>Medication</b>	<b>Percentage of Total Cost per Month</b>
Paracetamol	12%
Oral Rehydration Salts	50 %
Amoxil	21 %
Piriton	3 %
Metronidazole (Flagyl)	14 %
	<b>Total: 100%</b>

In Task 2 we are asking you to **make your own allocation decisions** (choose your own percentages) to Lubwa RHC for the medicine items in Table 2, which could be the same as the optimal allocation in Table 2 or whatever you deem better. **Note that your allocation may be randomly selected, actually bought and sent to Lubwa RHC in October 2018.**

Recall from Task 1 that there are two budgets for which we ask you to make your allocation decisions: Budget 1 and Budget 2. You can **freely allocate between 0% and 100%** of the respective budget to **any of the five medicines on the list**, with the restriction that all percentages for a given budget **need to sum to exactly 100%**.

Let us explain how you can **earn extra money** in Task 2. **In addition to the 50 Kwacha participation fee**, you will have the **opportunity to earn extra money**, depending on your allocation decisions of medicines. There are **two factors which influence your earnings**:

- (1) First, your earnings depend on the **size of Budget 2** – that is, **the larger Budget 2** which you generated in Task 1, **the more money you can potentially earn**.
- (2) **The more Amoxil you allocate** to the RHC, **the more money you will earn**.

The exact formula for your additional earnings is

$$\frac{(\text{Budget 1})}{4000} * \% \text{ Amoxily}_{\text{Budget 1}} + \frac{(\text{Budget 2})}{4000} * \% \text{ Amoxily}_{\text{Budget 2}} = \text{Your Earnings}$$

**Example 1:** Let's assume that your Budget 2 is **2,000 Kwacha**. If you allocate **10%** of **Budget 1** to Amoxil and **20%** of **Budget 2** to Amoxil, then you will earn

$$\frac{(1,000 \text{ Kwacha})}{4,000} * 10 + \frac{(2,000 \text{ Kwacha})}{4,000} * 20 = 2.5 + 10 = \mathbf{12.5 \text{ Kwacha}}$$

**Example 2:** Let's assume that your Budget 2 is **3,000 Kwacha**. If you allocate **70%** of **Budget 1** to Amoxil and **20%** of **Budget 2** to Amoxil, then you will earn

$$\frac{(1,000 \text{ Kwacha})}{4,000} * 70 + \frac{(3,000 \text{ Kwacha})}{4,000} * 20 = 17.5 + 15 = \mathbf{32.5 \text{ Kwacha}}$$

**Important:** At the end of all the sessions of this study, we will **randomly draw one** of the participants' ID numbers. We will then **actually buy and send** the corresponding allocated **medicine of both Budget 1 and Budget 2** to Lubwa RHC. The RHC is located in a rural area whose catchment population has very limited access to medical services. Your allocation choice will therefore have a **real impact on the available medicine in Lubwa RHC**.

All participants will be invited via e-mail or SMS to watch this random draw.

To this end, the research team will work closely with the Ministry of Health, to **purchase the medicines** on the list and donate them to the Lubwa RHC **exactly as per the randomly chosen participant's allocation decisions**. An official letter from the Ministry of Health will be available for all participants to see at the notice board of the School of Public Health (UTH).

This donation will be financed from research funds provided by German Academic Exchange Service (DAAD), Germany.

Exercise: Assume your **Budget 2** is **2,000 Kwacha**. In addition to the 50 Kwacha for participation, **how much extra money** will you earn if you allocate **40%** of **Budget 1** to Amoxil and **10%** of **Budget 2** to Amoxil?

Please fill in the blanks:

$$\frac{(\quad)}{4,000} * (\quad) + \frac{(\quad)}{4,000} * (\quad) = (\quad) + (\quad) = \quad (\text{earnings})$$

*Instructions: Incentive-Flagyl*

ID # \_\_\_\_\_

### General Instructions for the Participants

You are now taking part in an economic and health-related study. Depending on your decision-making and chance you can earn a substantial amount of money. For this, it is important that you read the following instructions carefully.

The instructions you were given are exclusively for your private information and have to be given back after the study. During the study it is strictly forbidden to talk. If you have any questions, please raise your hand, and then one of the researchers will come to you and answer your questions in private.

Please be aware that your mobile phone and all other electronic devices have to be switched off. Any violation of these rules will lead to your exclusion from the study and from any payments.

All your decisions in the study are made anonymously. No other participant will be able to match your decisions with your identity.

For your participation in the study you will get a guaranteed show-up fee of 50 Kwacha. Any additional payments depend on chance and your own decision-making.

There will be two decision-making tasks in this study, Task 1 and Task 2. These two tasks are interrelated and a budget that you generate in Task 1 will be carried over for the decision-making in Task 2. There is a chance that your decision in Task 2 will result in an actual, real allocation of medicine to a Rural Health Centre, so we ask you to pay close attention to the following instructions.

After all participants have made their decisions, we will ask you some questions in a questionnaire.

On the following pages we will describe the exact course of the study.

## Task 1

There are two budgets in this study, **Budget 1 and Budget 2**. In Task 2 you will use these two budgets to allocate actual medicine to a real rural health centre: Lubwa Rural Health Centre in Chinsali District, Muchinga Province.

**Budget 1 is always fixed at 1,000 Kwacha.**

However, **you** are able to determine the **size of Budget 2** in this Task 1.

Task 1 involves a sample Health Information Aggregation 1 (HIA.1) form. In completing the HIA.1 form, you will be asked to **sum the entries** in **each row** of the form and **enter the total** into the respective **grey-shaded boxes**.

**Each correctly summed row of entries** in the HIA.1 form **will add 50 Kwacha** to your **Budget 2**. That is, **the number** of all correctly summed entries will determine your **Budget 2**.

Example 1: If you fill in 15 grey-shaded boxes correctly, your Budget 2 will be  $15 \times 50$  Kwacha = 750 Kwacha.

Example 2: If you fill in 41 grey-shaded boxes correctly, your Budget 2 will be  $41 \times 50$  Kwacha = 2050 Kwacha.

In the example in Table 1, the correctly summed row total is  $2+6+3=11$ . This one correctly summed row would add 50 Kwacha to your Budget 2. **The more rows you sum correctly in the given time, the greater will be your Budget 2.**

**Table 1. Short Sample HIA.1 Form.**

DIAGNOSES	OPD First Attendance			Total
	Under 1 year	1-5 years	6+ Years	
<b>NOTIFIABLE DISEASES</b>				
Acute flaccid paralysis	2	6	3	11
Cholera	23	10	22	
...	...	...	...	

Each participant will be given **exactly 8 minutes** within which to complete as many summations as she/he likes and can manage.

Exercise:

**Question 1:** How large will **Budget 1** be if you sum **19 rows** correctly? Answer: K \_\_\_\_\_

**Question 2:** How large will **Budget 2** be if you sum **33 rows** correctly? Answer: K \_\_\_\_\_

## Task 2

In Task 2, you will be presented with a **list of consumed medicines** in a Zambian Rural Health Centre (RHC), as well as an estimate of the proportions (in %) of the costs of these medicines. Below we present you **actual medicine consumption data** from Lubwa RHC (catchment population: 9,500), Chinsali District, Muchinga Province, Zambia, in one month (Table 2).

**Please be aware that the distribution below represents an optimal allocation of health resources to Lubwa RHC based on the prevailing disease burden in the catchment area.**

**Table2: Percentages of Costs per Month for Five Consumed Medicines, Lubwa RHC.**

<b>Medication</b>	<b>Percentage of Total Cost per Month</b>
Paracetamol	12%
Oral Rehydration Salts	50 %
Amoxil	21 %
Piriton	3 %
Metronidazole (Flagyl)	14 %
	<b>Total: 100%</b>

In Task 2 we are asking you to **make your own allocation decisions** (choose your own percentages) to Lubwa RHC for the medicine items in Table 2, which could be the same as the optimal allocation in Table 2 or whatever you deem better. **Note that your allocation may be randomly selected, actually bought and sent to Lubwa RHC in August 2019.**

Recall from Task 1 that there are two budgets for which we ask you to make your allocation decisions: Budget 1 and Budget 2. You can **freely allocate between 0% and 100%** of the respective budget to **any of the five medicines on the list**, with the restriction that all percentages for a given budget **need to sum to exactly 100%**.

Let us explain how you can **earn extra money** in Task 2. **In addition to the 50 Kwacha participation fee**, you will have the **opportunity to earn extra money**, depending on your allocation decisions of medicines. There are **two factors which influence your earnings**:

- (1) First, your earnings depend on the **size of Budget 2** – that is, **the larger Budget 2** which you generated in Task 1, **the more money you can potentially earn**.
- (2) **The more Metronidazole (Flagyl) you allocate to the RHC, the more money you will earn.**

The exact formula for your additional earnings is

$$\frac{(\text{Budget 1})}{4000} * \% \text{Flagyl}_{\text{Budget 1}} + \frac{(\text{Budget 2})}{4000} * \% \text{Flagyl}_{\text{Budget 2}} = \text{Your Earnings}$$

**Example 1:** Let's assume that your Budget 2 is **2,000 Kwacha**. If you allocate **10%** of **Budget 1** to Metronidazole (Flagyl) and **20%** of **Budget 2** to Metronidazole (Flagyl), then you will earn

$$\frac{(1,000 \text{ Kwacha})}{4,000} * 10 + \frac{(2,000 \text{ Kwacha})}{4,000} * 20 = 2.5 + 10 = \mathbf{12.5 \text{ Kwacha}}$$

**Example 2:** Let's assume that your Budget 2 is **3,000 Kwacha**. If you allocate **70%** of **Budget 1** to Metronidazole (Flagyl) and **20%** of **Budget 2** to Metronidazole (Flagyl), then you will earn

$$\frac{(1,000 \text{ Kwacha})}{4,000} * 70 + \frac{(3,000 \text{ Kwacha})}{4,000} * 20 = 17.5 + 15 = \mathbf{32.5 \text{ Kwacha}}$$

**Important:** At the end of all the sessions of this study, we will **randomly draw one** of the participants' ID numbers. We will then **actually buy and send** the corresponding allocated **medicine of both Budget 1 and Budget 2** to Lubwa RHC. The RHC is located in a rural area whose catchment population has very limited access to medical services. Your allocation choice will therefore have a **real impact on the available medicine in Lubwa RHC**.

All participants will be invited via e-mail or SMS to watch this random draw.

To this end, the research team will work closely with the Ministry of Health, to **purchase the medicines** on the list and donate them to the Lubwa RHC **exactly as per the randomly chosen participant's allocation decisions**. An official letter from the Ministry of Health will be available for all participants to see at the notice board of the School of Public Health (UTH).

This donation will be financed from research funds provided by German Academic Exchange Service (DAAD), Germany and the International Growth Centre (IGC).

Exercise: Assume your **Budget 2** is **2,000 Kwacha**. In addition to the 50 Kwacha for participation, **how much extra money** will you earn if you allocate **40%** of **Budget 1** to Metronidazole (Flagyl) and **10%** of **Budget 2** to Metronidazole (Flagyl)?

Please fill in the blanks:

$$\frac{(\quad)}{4,000} * (\quad) + \frac{(\quad)}{4,000} * (\quad) = (\quad) + (\quad) = \quad (\text{earnings})$$

*Instructions: Observability*

ID # \_\_\_\_\_

### General Instructions for the Participants

You are now taking part in an economic and health-related study. Depending on your decision-making and chance you can earn a substantial amount of money. For this, it is important that you read the following instructions carefully.

The instructions you were given are exclusively for your private information and have to be given back after the study. During the study it is strictly forbidden to talk. If you have any questions, please raise your hand, and then one of the researchers will come to you and answer your questions in private.

Please be aware that your mobile phone and all other electronic devices have to be switched off. Any violation of these rules will lead to your exclusion from the study and from any payments.

All your decisions in the study are made anonymously. No other participant will be able to match your decisions with your identity.

For your participation in the study you will get a guaranteed show-up fee of 50 Kwacha. Any additional payments depend on chance and your own decision-making.

There will be two decision-making tasks in this study, Task 1 and Task 2. These two tasks are interrelated and a budget that you generate in Task 1 will be carried over for the decision-making in Task 2. There is a chance that your decision in Task 2 will result in an actual, real allocation of medicine to a Rural Health Centre, so we ask you to pay close attention to the following instructions.

After all participants have made their decisions, we will ask you some questions in a questionnaire.

On the following pages we will describe the exact course of the study.

## Task 1

There are two budgets in this study, **Budget 1 and Budget 2**. In Task 2 you will use these two budgets to allocate actual medicine to a real rural health centre: Lubwa Rural Health Centre in Chinsali District, Muchinga Province.

**Budget 1 is always fixed at 1,000 Kwacha.**

However, **you** are able to determine the **size of Budget 2** in this Task 1.

Task 1 involves a sample Health Information Aggregation 1 (HIA.1) form. In completing the HIA.1 form, you will be asked to **sum the entries** in **each row** of the form and **enter the total** into the respective **grey-shaded boxes**.

**Each correctly summed row of entries** in the HIA.1 form **will add 50 Kwacha** to your **Budget 2**. That is, **the number** of all correctly summed entries will determine your **Budget 2**.

Example 1: If you fill in 15 grey-shaded boxes correctly, your Budget 2 will be  $15 \times 50$  Kwacha = 750 Kwacha.

Example 2: If you fill in 41 grey-shaded boxes correctly, your Budget 2 will be  $41 \times 50$  Kwacha = 2050 Kwacha.

In the example in Table 1, the correctly summed row total is  $2+6+3=11$ . This one correctly summed row would add 50 Kwacha to your Budget 2. **The more rows you sum correctly in the given time, the greater will be your Budget 2.**

**Table 1. Short Sample HIA.1 Form.**

	OPD First Attendance			
DIAGNOSES	Under 1 year	1-5 years	6+ Years	Total
<b>NOTIFIABLE DISEASES</b>				
Acute flaccid paralysis	2	6	3	11
Cholera	23	10	22	
...	...	...	...	

Each participant will be given **exactly 8 minutes** within which to complete as many summations as she/he likes and can manage.

Exercise:

**Question 1:** How large will **Budget 1** be if you sum **19 rows** correctly? Answer: K \_\_\_\_\_

**Question 2:** How large will **Budget 2** be if you sum **33 rows** correctly? Answer: K \_\_\_\_\_

## Task 2

In Task 2, you will be presented with a **list of consumed medicines** in a Zambian Rural Health Centre (RHC), as well as an estimate of the proportions (in %) of the costs of these medicines. Below we present you **actual medicine consumption data** from Lubwa RHC (catchment population: 9,500), Chinsali District, Muchinga Province, Zambia, in one month (Table 2).

**Please be aware that the distribution below represents an optimal allocation of health resources to Lubwa RHC based on the prevailing disease burden in the catchment area.**

**Table2: Percentages of Costs per Month for Five Consumed Medicines, Lubwa RHC.**

<b>Medication</b>	<b>Percentage of Total Cost per Month</b>
Paracetamol	12%
Oral Rehydration Salts	50 %
Amoxil	21 %
Piriton	3 %
Metronidazole (Flagyl)	14 %
	<b>Total: 100%</b>

In Task 2 we are asking you to **make your own allocation decisions** (choose your own percentages) to Lubwa RHC for the medicine items in Table 2, which could be the same as the optimal allocation in Table 2 or whatever you deem better. **Note that your allocation may be randomly selected, actually bought and sent to Lubwa RHC in October 2018.**

Recall from Task 1 that there are two budgets for which we ask you to make your allocation decisions: Budget 1 and Budget 2. You can **freely allocate between 0% and 100%** of the respective budget to **any of the five medicines on the list**, with the restriction that all percentages for a given budget **need to sum to exactly 100%**.

Let us explain how you can **earn extra money** in Task 2. **In addition to the 50 Kwacha participation fee**, you will have the **opportunity to earn extra money**, depending on your allocation decisions of medicines. There are **two factors which influence your earnings**:

- (1) First, your earnings depend on the **size of Budget 2** – that is, **the larger Budget 2** which you generated in Task 1, **the more money you can potentially earn**.
- (2) **The more Amoxil you allocate** to the RHC, **the more money you will earn**.

The exact formula for your additional earnings is:

$$\frac{(\text{Budget 1})}{4000} * \% \text{ Amoxil}_{\text{Budget 1}} + \frac{(\text{Budget 2})}{4000} * \% \text{ Amoxil}_{\text{Budget 2}} = \text{Your Earnings}$$

**Example 1:** Let's assume that your Budget 2 is **2,000 Kwacha**. If you allocate **10%** of **Budget 1** to Amoxil and **20%** of **Budget 2** to Amoxil, then you will earn

$$\frac{(1,000 \text{ Kwacha})}{4,000} * 10 + \frac{(2,000 \text{ Kwacha})}{4,000} * 20 = 2.5 + 10 = \mathbf{12.5 \text{ Kwacha}}$$

**Example 2:** Let's assume that your Budget 2 is **3,000 Kwacha**. If you allocate **70%** of **Budget 1** to Amoxil and **20%** of **Budget 2** to Amoxil, then you will earn

$$\frac{(1,000 \text{ Kwacha})}{4,000} * 70 + \frac{(3,000 \text{ Kwacha})}{4,000} * 20 = 17.5 + 15 = \mathbf{32.5 \text{ Kwacha}}$$

**Important:** As mentioned earlier, at the end of all the sessions of this study, we will **randomly draw one** of the participants' ID numbers. We will then **actually buy and send** the corresponding allocated **medicine of both Budget 1 and Budget 2** to Lubwa RHC. The RHC is located in a rural area whose catchment population has very limited access to medical services. Your allocation choice will therefore have a **real impact on the available medicine** in **Lubwa RHC**.

All participants will be invited via e-mail or SMS to watch this random draw. Afterwards, **we will inform all other participants via e-mail or SMS of the exact allocation of medicines (in %) of the randomly drawn participant**, if an allocation from this session is drawn.

To this end, the research team will work closely with the Ministry of Health, to **purchase the medicines** on the list and donate them to the Lubwa RHC **exactly as per the randomly chosen participant's allocation decisions**. An official letter from the Ministry of Health will be available for all participants to see at the notice board of the School of Public Health (UTH).

This donation will be financed from research funds provided by the German Academic Exchange Service (DAAD), Germany.

Exercise: Assume your **Budget 2** is **2,000 Kwacha**. In addition to the 50 Kwacha for participation, **how much extra money** will you earn if you allocate **40%** of **Budget 1** to Amoxil and **10%** of **Budget 2** to Amoxil?

Please fill in the blanks:

$$\frac{(\quad)}{4,000} * (\quad) + \frac{(\quad)}{4,000} * (\quad) = (\quad) + (\quad) = \quad (\text{earnings})$$

*Instructions: Prof-Norms*

ID Number \_\_\_\_\_

The following statements concerning the importance of professionalism for health professionals in Zambia, are taken from the Health Professions Council of Zambia, Professional Code of Ethics and Discipline: Fitness to Practice (2014).

- A health professional should allocate available resources to the promotion of good medical practice and the maintenance of such health practitioner’s professional competence.
- A health professional should undertake to discharge his/her professional duties diligently and in a manner that best considers the welfare of the patient/client.
- A health professional should observe proper standards of personal behaviour, not only in professional activities, but at all times.
- A health practitioner is required to uphold honest behaviour at all times.
- Health practitioners are expected to behave in ways acceptable for the conduct of their profession.
- A health practitioner should not only choose but shall also be seen to be choosing the drug, appliance or goods which, in his/her independent professional judgement and having due regard to economy, will best serve the medical/health interest of his/her patient/client.
- The professional relationship between the health practitioner and his/ her patient/client depends on trust and the assumption that the health practitioner will act in the best interests of the patient/client.

In light of the above, you are requested to compose one sentence, in your own words, reflecting why it is important for health professionals to always act in the best interests of their patients.

---

---

---

---

---

---

---

*Instructions: Need-ORS*

ID # \_\_\_\_\_

### General Instructions for the Participants

You are now taking part in an economic and health-related study. Depending on your decision-making and chance you can earn a substantial amount of money. For this, it is important that you read the following instructions carefully.

The instructions you were given are exclusively for your private information and have to be given back after the study. During the study it is strictly forbidden to talk. If you have any questions, please raise your hand, and then one of the researchers will come to you and answer your questions in private.

Please be aware that your mobile phone and all other electronic devices have to be switched off. Any violation of these rules will lead to your exclusion from the study and from any payments.

All your decisions in the study are made anonymously. No other participant will be able to match your decisions with your identity.

For your participation in the study you will get a guaranteed show-up fee of 50 Kwacha. Any additional payments depend on chance and your own decision-making.

There will be two decision-making tasks in this study, Task 1 and Task 2. These two tasks are interrelated and a budget that you generate in Task 1 will be carried over for the decision-making in Task 2. There is a chance that your decision in Task 2 will result in an actual, real allocation of medicine to a Rural Health Centre, so we ask you to pay close attention to the following instructions.

After all participants have made their decisions, we will ask you some questions in a questionnaire.

On the following pages we will describe the exact course of the study.

## Task 1

There are two budgets in this study, **Budget 1 and Budget 2**. In Task 2 you will use these two budgets to allocate actual medicine to a real rural health centre: Lubwa Rural Health Centre in Chinsali District, Muchinga Province.

**Budget 1 is always fixed at 1,000 Kwacha.**

However, **you** are able to determine the **size of Budget 2** in this Task 1.

Task 1 involves a sample Health Information Aggregation 1 (HIA.1) form. In completing the HIA.1 form, you will be asked to **sum the entries** in **each row** of the form and **enter the total** into the respective **grey-shaded boxes**.

**Each correctly summed row of entries** in the HIA.1 form **will add 50 Kwacha** to your **Budget 2**. That is, **the number** of all correctly summed entries will determine your **Budget 2**.

Example 1: If you fill in 15 grey-shaded boxes correctly, your Budget 2 will be  $15 \times 50$  Kwacha = 750 Kwacha.

Example 2: If you fill in 41 grey-shaded boxes correctly, your Budget 2 will be  $41 \times 50$  Kwacha = 2050 Kwacha.

In the example in Table 1, the correctly summed row total is  $2+6+3=11$ . This one correctly summed row would add 50 Kwacha to your Budget 2. **The more rows you sum correctly in the given time, the greater will be your Budget 2.**

**Table 1. Short Sample HIA.1 Form.**

DIAGNOSES	OPD First Attendance			Total
	Under 1 year	1-5 years	6+ Years	
<b>NOTIFIABLE DISEASES</b>				
Acute flaccid paralysis	2	6	3	11
Cholera	23	10	22	
...	...	...	...	

Each participant will be given **exactly 8 minutes** within which to complete as many summations as she/he likes and can manage.

Exercise:

**Question 1:** How large will **Budget 1** be if you sum **19 rows** correctly? Answer: K \_\_\_\_\_

**Question 2:** How large will **Budget 2** be if you sum **33 rows** correctly? Answer: K \_\_\_\_\_

## Task 2

In Task 2, you will be presented with a **list of consumed medicines** in a Zambian Rural Health Centre (RHC), as well as an estimate of the proportions (in %) of the costs of these medicines. Below we present you **actual medicine consumption data** from Lubwa RHC (catchment population: 9,500), Chinsali District, Muchinga Province, Zambia, in one month (Table 2).

**Please be aware that the distribution below represents an optimal allocation of health resources to Lubwa RHC based on the prevailing disease burden in the catchment area.**

**Table 2: Percentages of Costs per Month for Five Consumed Medicines, Lubwa RHC.**

<b>Medication</b>	<b>Percentage of Total Cost per Month</b>
Paracetamol	12%
Oral Rehydration Salts	50 %
Amoxil	21 %
Piriton	3 %
Metronidazole (Flagyl)	14 %
	<b>Total: 100%</b>

In Task 2 we are asking you to **make your own allocation decisions** (choose your own percentages) to Lubwa RHC for the medicine items in Table 2, which could be the same as the optimal allocation in Table 2 or whatever you deem better. **Note that your allocation may be randomly selected, actually bought and sent to Lubwa RHC in August 2019.**

Recall from Task 1 that there are two budgets for which we ask you to make your allocation decisions: Budget 1 and Budget 2. You can **freely allocate between 0% and 100%** of the respective budget to **any of the five medicines on the list**, with the restriction that all percentages for a given budget **need to sum to exactly 100%**.

Table 3 below shows the percentage of women with knowledge of Oral Rehydration Salts (ORS) and the percentage of children with diarrhoea actually given ORS. This information is reported both generally for Zambia as well as for Muchinga province, the province where Lubwa RHC is located. The table reveals that while the percentage of women with knowledge of ORS seems to point to it being a good treatment option for diarrheal diseases, the percentage of children with diarrhoea actually given ORS is lower than the national average and seems to point to a need for more ORS to be provided.

**Table 3: Knowledge and Use of ORS: Zambia and Muchinga province**

	Percentage of women with knowledge of ORS packets	Percentage of children with diarrhoea given ORS
Muchinga province	95%	51%
Zambia	97%	64%

Source: Central Statistical Office

Now, let us explain how you can **earn extra money** in Task 2. **In addition to the 50 Kwacha participation fee**, you will have the **opportunity to earn extra money**, depending on your allocation decisions of medicines. There are **two factors which influence your earnings**:

- (1) First, your earnings depend on the **size of Budget 2** – that is, **the larger Budget 2** which you generated in Task 1, **the more money you can potentially earn**.
- (2) **The more Amoxil you allocate** to the RHC, **the more money you will earn**.

The exact formula for your additional earnings is

$$\frac{(Budget\ 1)}{4000} * \% Amoxil_{Budget\ 1} + \frac{(Budget\ 2)}{4000} * \% Amoxil_{Budget\ 2} = Your\ Earnings$$

**Example 1:** Let's assume that your Budget 2 is **2,000 Kwacha**. If you allocate **10%** of **Budget 1** to Amoxil and **20%** of **Budget 2** to Amoxil, then you will earn

$$\frac{(1,000\ Kwacha)}{4,000} * 10 + \frac{(2,000\ Kwacha)}{4,000} * 20 = 2.5 + 10 = \mathbf{12.5\ Kwacha}$$

**Example 2:** Let's assume that your Budget 2 is **3,000 Kwacha**. If you allocate **70%** of **Budget 1** to Amoxil and **20%** of **Budget 2** to Amoxil, then you will earn

$$\frac{(1,000\ Kwacha)}{4,000} * 70 + \frac{(3,000\ Kwacha)}{4,000} * 20 = 17.5 + 15 = \mathbf{32.5\ Kwacha}$$

**Important:** At the end of all the sessions of this study, we will **randomly draw one** of the participants' ID numbers. We will then **actually buy and send** the corresponding allocated **medicine of both Budget 1 and Budget 2** to Lubwa RHC. The RHC is located in a rural area whose catchment population has very limited access to medical services. Your allocation choice will therefore have a **real impact on the available medicine** in **Lubwa RHC**.

All participants will be invited via e-mail or SMS to watch this random draw.

To this end, the research team will work closely with the Ministry of Health, to **purchase the medicines** on the list and donate them to the Lubwa RHC **exactly as per the randomly chosen participant's allocation decisions**. An official letter from the Ministry of Health will be available for all participants to see at the notice board of the School of Public Health (UTH).

This donation will be financed from research funds provided by German Academic Exchange Service (DAAD), Germany and the International Growth Centre (IGC).

Exercise: Assume your **Budget 2** is **2,000 Kwacha**. In addition to the 50 Kwacha for participation, **how much extra money** will you earn if you allocate **40%** of **Budget 1** to Amoxil and **10% of Budget 2 to Amoxil**?

Please fill in the blanks:

$$\frac{(\quad)}{4,000} * (\quad) + \frac{(\quad)}{4,000} * (\quad) = (\quad) + (\quad) = \quad \text{(earnings)}$$

## Chapter 3

# Comparing the Productivity of Performance-based and Unconditional Financial Incentives in Improving Infant Birthweight in Zambia

## **Abstract**

Despite health programmes providing performance-based financial incentives to health-care providers being a popular policy approach, evidence documenting their effectiveness relative to alternative types of financial incentives, remains scarce. This study compares infant birth outcomes in districts of Zambia where healthcare providers were randomised to receive either performance-based financial incentives, unconditional financial transfers, or conducted business as usual during the Results Based Financing pilot. It uses a Difference-in-Differences design and four waves of nationally representative Demographic and Health Survey data. Results reveal that performance-based and unconditional financial incentives have remarkably similar effects on birth outcomes. Benefits are concentrated in babies of women at higher risk of bearing low birthweight children. Triple difference and Regression Discontinuity results are qualitatively similar. The findings imply that unconditional financial incentives, which are more cost-effective, can also be an effective tool in improving child health outcomes.

*Keywords:* Infant birth weight; Difference-in-differences; Unconditional and performance based financial incentives

*JEL Codes:* I15, I12, J13

### 3.1 Introduction

Poor maternal and child health outcomes, particularly in low income developing countries, have propelled efforts to improve maternal and child health to the very top of the global development agenda. This is exemplified by specific commitments made in the Sustainable Development Goals (SDGs), such as reducing neonatal and under five mortality to no more than 12 and 25 deaths per 100,000 live births by 2030 (United Nations, 2016). In a quest to achieve this ambitious developmental agenda, health programmes involving the provision of performance-based financial incentives to healthcare providers have emerged as a common policy tool. Informed by insights from principal-agent theory, an inherent assumption within this policy response has been that paying healthcare providers based on their performance is a superior work effort solicitation approach, relative to other types of financial incentives. However, despite the popularity of this approach, evidence documenting the effectiveness and sustainability of performance-based financial incentives, particularly relative to alternative types of financial incentives, remains scarce.

This study investigates whether paying healthcare providers performance-based financial incentives is more effective in improving infant birth outcomes relative to paying unconditional financial incentives in the context of a low income developing country, Zambia. It exploits the district randomised rollout of the Results Based Financing (RBF) programme in Zambia and utilises multiple rounds of Demographic and Health Survey (DHS) data to estimate Difference-in-Differences (DD), Regression Discontinuity (RD) and Triple-Difference (DDD) models. Results reveal that improvements in birthweight and the probability of a low birthweight birth owing to the two types of financial incentives, are comparable. Owing to this fact, unconditional financial incentives, which are less costly, are found to be more cost-effective. Comparable gains in the quality of Antenatal Care (ANC) owing to the two types of financial incentives, are identified as a potential channel driving the results.

Like other low income developing countries, Zambia has continued to grapple with poor maternal and child health outcomes. For example, approximately 27 infants per 1,000 live births die within the first 28 days after birth. Similarly, Zambia has among the highest rates of maternal mortality worldwide, with approximately 278 women dying while carrying and delivering their babies per 100,000 live births. Approximately 10% of infants are born with low birthweight (Zambia Ministry of Health and IFC, 2019). Among other factors, these maternal and child health outcomes have been linked to an unavailability of essential medicines and medical supplies as well as low levels of motivation of healthcare providers (Zambia Ministry of Health, 2017; Friedman et al., 2016). A leading policy response has come in the form of health programmes that have involved performance-based financial incentives being paid to healthcare providers. One such programme is the Results Based Financing (RBF) programme, with an estimated \$385.6 million being committed for its implementation during the SDG (and their predecessor Millennium Development

Goals, MDGs) implementation period.<sup>1</sup> In this vein, the Zambia RBF pilot project was implemented between 2012 and 2014. The ZRBF pilot aimed to enhance the quality of maternal and child healthcare provision and increase provider morale, effort and ultimately productivity primarily through the provision of performance-based financial incentives to healthcare providers. This implies that during the ZRBF pilot, healthcare providers in randomly selected Zambian districts received performance-based financial incentives based on the achievement of pre-agreed targets on nine maternal and child health indicators, including the provision of ANC visits to pregnant women. In-turn, healthcare providers in yet another set of randomly selected districts received financial transfers independent of their performance (unconditional financial incentives), while others received no financial incentives at all and simply conducted business as usual (Shen et al., 2017; Friedman et al., 2016).

Alternative insights from the theoretical and empirical literature, however, imply that the productivity effects of the various types of financial incentives provided under the ZRBF are a-priori unclear and are thus an open empirical issue. Standard principal-agent theory in economics predicts that attaching an agent's or worker's compensation directly to her effort ought to result in an increase in effort and thus productivity (Sappington, 1991). However, evidence has revealed that such incentives can potentially have unintended, distortionary effects, a phenomenon commonly known in the literature as multitasking (Holmstrom and Milgrom, 1991), with similar distortion effects arising among Zambian healthcare providers (Chibuye et al., 2021). Similarly, efficiency-wage theory in economics asserts that workers granted above market clearing wages, that are unconditionally higher and independent of their effort may respond with a higher level of effort (Akerlof, 1982). This implies that contrary to the policy rationale underlying RBF programmes, unconditional financial incentives may also induce work effort. This makes a comparison of the productivity of the two types of financial incentives in improving health outcomes an important undertaking with implications for the design of health programmes involving the provision of financial incentives to healthcare providers and ultimately, the achievement of the SDGs by 2030.

This study exploits the district randomised roll out of the ZRBF project to answer the following question: when measured by increased birthweight and reduced low birthweight, is it more productive to provide performance-based or unconditional financial incentives to healthcare providers, relative to providing no incentives at all? The study utilises a DD design, complemented with DDD and RD designs and four waves of nationally representative Zambia DHS data. Specifically, the study compares birthweight and the probability of a low birthweight birth in districts in which healthcare providers were randomised to receive performance-based or unconditional financial incentives to those districts in which healthcare providers conducted business as usual. The study also explores the role of the number, timing and quality of ANC visits, measured by the number of tetanus injections and whether or not a woman was weighed and had her blood pressure taken during

---

<sup>1</sup><https://www.rbfhealth.org/mission>.

pregnancy, as potential mechanisms driving the results.

To preview the results, the study finds that performance-based and unconditional financial incentives to healthcare providers have comparable effects in improving infant birth outcomes. Specifically, DD estimates suggest that birthweight increases by 32 grams in districts receiving performance-based financial incentives and by 45 grams in districts receiving unconditional financial incentives. Similarly, while estimates suggest a 1.3 percentage-point reduction in the probability of a low birthweight birth in districts receiving performance-based financial incentives, the same effect is 0.6 percentage-points in districts receiving unconditional financial incentives. The comparability of these effects implies that they are found to be statistically indistinguishable. However, none of the above estimates are statistically significant. Furthermore, the study finds that improvements in birthweight and low birthweight are concentrated in children born to women who are at higher risk of bearing low birthweight children. This is evident from the fact that estimates are larger and more statistically significant in these high impact sub-samples of children born to women with either no education or only a primary education and women younger than 20 years of age or older than 35 years of age. Indeed, estimates in these sub-samples of children reveal increases in birthweight in both TU and TP districts of between 29 and 207 grams and reductions in the probability of a low birthweight birth of between 4 and 10 percentage points, respectively.

To investigate a channel potentially driving the above results, the study also shows that improvements in the number and timing of ANC visits are mainly concentrated in districts receiving performance-based financial incentives, with 0.4 more ANC visits, conducted approximately 0.2 months earlier per pregnant woman, occurring in this group of districts. However, the only dimension of ANC quality showing an improvement in these districts is the probability of a woman having her blood pressure taken during pregnancy, which increases by 6 percentage points. Similar effects are evident in districts receiving unconditional financial incentives, with 0.3 more ANC visits conducted and 0.1 more tetanus injections being administered per pregnant woman and a 4 percentage point increase in the probability of a woman being weighed during pregnancy. Thus, while districts receiving performance-based financial incentives had a larger number of earlier initiated ANC visits, the treatment may not have translated into commensurate improvements in the quality of ANC visits.

One potential driver behind the pattern of results noted above could be multitasking (Holmstrom and Milgrom, 1991). While healthcare providers received an explicit financial incentive of \$1.6 per ANC visit in performance-based districts, the quality of ANC was only indirectly incentivised through a quality multiplier. Healthcare providers could have thus focused more on the number of ANC visits, the aspect of their tasks directly attached to the performance-based financial incentive. An improvement in the number and timing of ANC visits without a commensurate improvement in the quality of the ANC visits thus implies that birth outcomes in performance-based districts did not improve any more than

they did in unconditional districts.

A critical implication of the above findings is that the gap between the productivity of performance-based financial incentives to healthcare providers relative to unconditional financial incentives, may not be as large as previously thought. Indeed, back of the envelope cost-effectiveness calculations actually reveal that performance-based financial incentives would have to be at least 5 times more effective, i.e. increase birthweight by 758 grams on average, an implausibly large increase, to achieve a level of cost-effectiveness comparable to that achieved by the unconditional financial incentives. To make performance-based financial incentive programmes more effective, policymakers may have to incur additional costs of directly incentivising the quality of healthcare provision. Alternatively, consideration can be given to unconditional financial incentives to achieve comparable improvements in such health outcomes. The above insights have profound implications for the design of health programs involving provision of financial incentives to healthcare providers and by implication, the achievement of the SDGs by 2030.

The study focuses on birthweight given the importance of this metric as an overall indicator for infant health. Annually, an estimated 16% of all infants globally and an estimated 10% in Zambia, are born with low birthweight (Ohlsson and Shah, 2008; Zambia Ministry of Health and IFC, 2019). Infants born with low birthweight have been found to be at a higher risk of stillbirth, to be more likely to require neonatal intensive care, have worse education outcomes and developmental difficulties and to be at higher risk of death, later in their childhood (Zambia Ministry of Health and IFC, 2019; Chibwesha et al., 2016; Currie and Almond, 2011). Improvements in birthweight thus represent a single poignant mechanism in the quest to improve other important child health and developmental outcomes.

This study contributes to three strands of the literature. Firstly, the study contributes to a growing literature on healthcare provider financial incentives by investigating the impact of performance-based financial incentives on actual health outcomes - birthweight and low birthweight. Most studies in this literature investigate the impact of these programmes on indicators of health service provision and not health outcomes (Basinga et al., 2011; Friedman et al., 2016; Bonfrer et al., 2014; Miller et al., 2012). Secondly, by comparing the effectiveness of performance-based to unconditional financial incentives, this study also contributes to an important growing literature comparing the effectiveness of alternative mechanisms of delivering social programmes such as conditional relative to unconditional cash transfers on health outcomes (Pega et al., 2017) and on education outcomes (Baird et al., 2014). A related literature also compares the effectiveness of different incentive types such as performance-based relative to unconditional transfers to schools on education outcomes (Mbiti et al., 2019), winner-take-all contests relative to proportional rewards (Singh and Masters, 2020), career relative to pro-social incentives in improving health outcomes (Ashraf et al., 2020), performance-based relative to non-financial incentives (Ashraf et al., 2014) and block grants relative to performance-based incentives to communities (Olken

et al., 2014). Thirdly, by assessing the impact of an RBF programme on birthweight and low birthweight, this study also contributes to a growing literature assessing the impact of social programmes such as conditional cash transfers on birth outcomes (Amarante et al., 2016; Cardona-Sosa and Medina, 2017; Almond et al., 2011; Hoynes et al., 2011, 2015).

The rest of the paper is organised as follows: section 3.2 discusses literature related to the study, section 3.3 presents the context in which the ZRBF pilot project was undertaken. Section 3.4 presents the data and descriptive statistics while section 3.5 presents the identification strategy and empirical framework utilised in the analysis. Section 3.6 presents the main results of the study, while section 3.7 presents back of the envelope cost-effectiveness calculations. Section 3.8 presents various sets of robustness checks, while section 3.9 discusses the results and also concludes.

## 3.2 Related Literature

A key underlying hypothesis this study seeks to investigate, is whether the use of performance-based and unconditional financial incentives, relative to conducting business as usual, can result in healthcare providers having more morale and exerting greater work effort, with these effects, in turn, improving the quality of healthcare provision and ultimately resulting in improvements in infant birth outcomes. To this end, an examination of three aspects of the literature, may enlighten this investigation.

### 3.2.1 Literature on Incentivising Economic Agents

As noted above, principal-agent theory in economics predicts that attaching an agents compensation directly to their effort results in an increase in the agents effort and thus productivity (Sappington, 1991). This has been found to hold in private firms (Lazear, 2000). These insights have contributed to an expansion in the use of performance-based financial incentives in the public sector in general and in the public health sector in particular. However, challenges identified with these incentives have encompassed multitasking effects discussed above, but have also included observations of the financial incentives not being large enough to induce meaningful change (Friedman et al., 2016; Gneezy and Rustichini, 2000).

A growing literature has investigated the use of performance-based financial incentives to healthcare providers in RBF programmes (Basinga et al., 2011; Friedman et al., 2016; Paul et al., 2018; Bonfrer et al., 2014; Gage and Bauhoff, 2021). However, most studies use indicators of health service provision and quality as opposed to actual health outcomes as the main outcome variables of interest.<sup>2</sup> Indeed, even the original impact evaluation of the Zambia RBF, Friedman et al. (2016), falls within this category. This implies that there is relatively less evidence regarding the effects of performance-based financial incentives

---

<sup>2</sup>Utilising indicators of health service provision as opposed to health outcomes in provider performance contracts has been argued to have some advantages in the contract theory literature (see for example, the discussion in Miller and Babiarz (2013)).

on actual health outcomes on the supply-side. An exception is Gertler and Vermeersch (2012) who examine the effects of performance-based financial incentives to healthcare providers on child weight-for-age and height-for-age in Rwanda. Gertler and Vermeersch (2012) find a 0.53 of a standard deviation gain in weight-for-age and 0.25 of a standard deviation gain in height-for-age, respectively.<sup>3</sup> However, in addition to not investigating infant birthweight, the design of the Rwanda RBF programme leveraged by Gertler and Vermeersch (2012) did not permit a comparison of performance-based and unconditional financial incentive arms to a business as usual control, as conducted in this study.

Despite unconditional financial incentives being less emphasised in the literature investigating RBF programmes, economic theory suggests that they too can potentially have work effort inducing effects. For example, the gift exchange hypothesis, emanating from efficiency-wage theory mentioned previously, has been established empirically in laboratory settings (Fehr et al., 1993, 1998) as well as in field settings (Gneezy and List, 2006). Similarly, the behavioural economics literature has built on these insights and revealed that public sector workers in the health sector may be intrinsically motivated for instance by the mission of the organisation (Besley and Ghatak, 2005) or may have other regarding preferences such as altruism (Godager and Wiesen, 2013) and other pro-social preferences (Lagarde and Blaauw, 2014) and may thus not even require financial incentives in order to exert higher effort levels, enhance their productivity and improve health outcomes. The above thus points to the importance of a comparison of alternative modes of financial incentivisation of healthcare providers.

### **3.2.2 Literature on Health and Social Programmes**

To appreciate the link between improvements in the quality of healthcare provision and improvements in infant birth outcomes, an appraisal of a strand of the health economics literature may be instructive. Indeed, this literature puts forth two main factors that influence infant birthweight; the length of time spent in the mother's body (gestational length) and the rate at which the foetus grows while in the mothers body or a combination of both (Kramer, 2003, 1987). This implies that an infant may be born with low birthweight, which refers to a birthweight less than 2,500 grams, either owing to an insufficient amount of time being spent in the mothers body, termed a preterm birth or, to a slower than normal rate of growth while in the mothers body, a phenomenon known in the literature as Intra-Uterine Growth Restriction - IUGR. The primary cause of low birthweight in developing countries has been identified to be IUGR, with the deficiencies in IUGR being exacerbated by factors such as maternal age and education. Specifically children born to women younger than 20 years of age or older than 35 years of age as well as women with low levels of education may be especially at risk (Kramer, 2003). This insight has resulted in social and development programmes with a focus on alleviat-

---

<sup>3</sup>Additional studies investigate the productivity of other types of financial incentives in improving actual health outcomes in other settings. For instance, Singh and Masters (2020) and Singh (2015) assess the productivity of winner-take-all relative to proportional payments in improving child weight-for-age malnutrition in India. Similarly, Miller et al. (2012) assess the productivity of performance-based provider incentives to school principals to reduce child anaemia in China.

ing so-called modifiable factors causing IUGR, being emphasised in developing countries.<sup>4</sup> These programmes have included components emphasizing the provision of higher quality healthcare in general and ANC in particular and the improvement of maternal nutrition (Kramer (1987)). Indeed, by facilitating both early identification and timely treatment of potentially harmful conditions during pregnancy, high quality ANC has been argued to be a key factor in alleviating the effects of IUGR and thus improving both gestational length, the rate of foetal growth and by implication, birthweight (Kramer, 1987). Thus, healthcare providers exerting greater work effort in the provision of healthcare in response to different financial incentives, can result in higher productivity in the alleviation of modifiable factors causing IUGR and ultimately to improvements in infant health outcomes. An examination of the ANC channel is thus critical in understanding any improvements in infant birth outcomes.

Building on insights from the health economics literature, numerous social programmes have focused on the demand side, using financial and non-financial incentives such as conditional or unconditional cash and in-kind transfers to encourage the utilisation of higher quality healthcare in general and ANC in particular, improve maternal nutrition and by implication, all associated maternal and child health outcomes. Closer to the current study, several studies have investigated the impact of such demand-side health programmes on infant birthweight.<sup>5</sup> Almond et al. (2011) and Hoynes et al. (2011) assess the impacts of the Food Stamp Programme (FSP) and the Supplemental Programme for Women, Infants and Children (WIC), respectively, on birthweight in the United States. Both studies utilise DD designs and exploit the differences in the timing of receipt of the respective programmes across US counties overtime. Almond et al. (2011) find an increase in birthweight for infants exposed to FSP during the third trimester of pregnancy, of approximately 2 grams for white and 5 grams for African-American infants. Similarly, Hoynes et al. (2011) find that WIC participation increases infant birthweight by approximately 2 grams, with no impact on the infant low birthweight. An inherent weakness of such studies that use a DD design and exploit variation in the timing of treatment as a basis of identification, is that the variance weighted nature of the two-way fixed effects DD estimator implies that with many groups and time periods and heterogeneous treatment effects overtime, some units might receive negative weights, resulting in bias in the computation of the two-way fixed effects DD treatment effect estimator (De Chaisemartin and d’Haultfoeulle, 2020; Goodman-Bacon, 2021).<sup>6</sup> By relying on the randomised roll-out of the original ZRBF pilot instead of variation in the timing of treatment across districts, this study is not plagued by this potential bias.

---

<sup>4</sup>Some factors, while important determinants of IUGR, are, unfortunately, be less amenable to modification. These include maternal and paternal height and maternal psychological factors.

<sup>5</sup>Another set of studies, for instance, Hoynes et al. (2015) investigate the effects of non-health programs such as the Earned Income Tax Credit (EITC), a U.S programme that utilises the tax system to provide a refundable income transfer to low-income households with children, on birthweight, finding a reduction in low birthweight of between 0.17 to 0.31 percentage points and an increase in mean birthweight of around 6.4 grams.

<sup>6</sup>Callaway and Sant’Anna (2020) proposes an alternative estimator robust to bias resulting from heterogeneity of treatment effects in DD designs with variation in the timing of treatment. The source of exogenous variation this study relies on does not require the use of this estimator.

### 3.3 Context and Program Description

Despite recording encouraging progress in the recent past, Zambia, like other low-income developing countries, continues to face numerous health system challenges. These are reflected in the fact that indicators of health status are disconcertingly high and the pace of their improvement overtime has been lacklustre. For instance, as of 2018, the rate of neonatal deaths,<sup>7</sup> stood at 27 deaths per 1,000 live births, an increase from 24 deaths per 1,000 live births in 2014. Similarly, under-age 5 mortality stood at 61 deaths per 1,000 live births, a decline from 75 deaths per 1,000 live births in 2014. The maternal mortality ratio stood at 278 deaths per 100,000 live births, a decline from 398 deaths per 100,000 live births in 2014, while approximately 10% of infants were born with low birthweight (Zambia Ministry of Health and IFC, 2019). Factors that have been identified as underlying the lacklustre performance of the Zambian health system include; a shortage of appropriately qualified human resource, poor motivation of said human resource, often manifesting in the form of high rates of absenteeism, lack of punctuality and low morale as well as inconsistent availability of essential medications and medical supplies (Zambia Ministry of Health, 2017; Friedman et al., 2016).

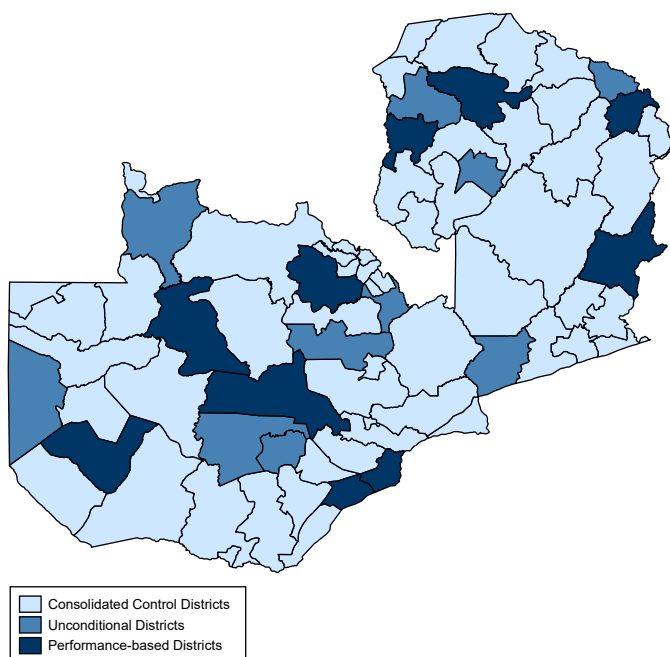


Figure 3.3.1: Districts by Treatment

In an attempt to improve the performance of the health sector, boost health outcomes and improve progress towards the attainment of MDG four and five, the ZRBF pilot project was implemented between April 2012 and October 2014. The RBF approach involves a principal entity providing financial or in-kind rewards to an agent, contingent on the agent achieving pre-specified targets (Friedman et al., 2016). The ZRBF pilot was accompanied by an impact evaluation in which 30 districts in Zambia were randomised

<sup>7</sup>An infant death within the first 28 days after live birth.

into three conditions; in ten districts, individual healthcare providers, their health facilities and their respective responsible district health administrations, were randomised to receive performance-based financial incentives (TP). In another ten districts, healthcare providers were randomised to receive unconditional financial incentives (TU) and in another ten districts, healthcare providers received no financial incentives of any sort at all.<sup>8</sup> The distribution of districts by treatment is shown graphically in figure 3.3.1. The ultimate aim of the ZRBF pilot was to improve health outcomes by among other things, increasing the productivity of existing healthcare providers. While this remained the stated ultimate goal, in reality, the programme and impact evaluation instead only focused on the indicators of healthcare utilisation selected for incentivisation. Among the nine incentivised indicators of healthcare utilisation incentivised in TP districts, were ANC visits.<sup>9</sup> This implies that in TP districts, healthcare providers received \$1.6 dollars per pregnant women per ANC visit conducted and thus, were specifically incentivised to increase utilisation of ANC, a key component in the improvement of infant birthweight. Providers in TU districts on the other hand, unconditionally received increased financial incentives and the remaining districts received no financial incentives to improve ANC utilisation. This explicit incentivisation of ANC in TP districts and the lack thereof in TU and control districts, implies that healthcare providers effort and morale, the quantity and quality of ANC and in turn, birth weight and the prevalence of low birthweight, should have improved much more in TP districts relative to both TU and control districts. It is this hypothesized response to different types of financial incentives that this study sets out to empirically investigate.

### 3.4 Data and Descriptive Statistics

The main data source used in this study is the Zambia Demographic and Health Survey (ZDHS). Specifically, four waves of nationally representative ZDHS datasets over the years 1996, 2001 and 2007, before the ZRBF and 2014, at the end of ZRBF implementation, are utilised. The DHS is a nationally representative household survey funded by the United States Agency for International Development (USAID) and implemented in collaboration with national governments in over 90 developing countries. The survey collects indicators on the health and nutrition of the population, with a specific focus on women aged 15-49 years old and their children.

---

<sup>8</sup>Healthcare providers in TP districts received both individual and health facility-level performance contracts. Health facilities could utilise a minimum of 40% of their performance-based earnings to improve healthcare provision within their health facilities. A maximum of 60% of their performance-based earnings could be given to the individual healthcare providers, depending on their individual performance. Healthcare providers in unconditional districts were guaranteed to receive approximately 50% of the financial incentives received by the TP districts, irrespective of their performance. Both TP and TU type districts also received maternal and child health medical equipment to improve healthcare provision (Friedman et al., 2016).

<sup>9</sup>The other incentivised indicators included; curative consultation, institutional deliveries by a skilled provider, post-natal visits, full immunization of children aged at most one year, pregnant women receiving three doses of malaria prophylaxis, the number of women using modern contraceptive methods, pregnant women tested for HIV and the number of HIV positive pregnant women given anti-retroviral therapy prophylaxis.

Table 3.4.1: Baseline (2007) Means and Changes in Sample Characteristics

Variable	TU	TP	Cons. Control	P-value: $\Delta$ TP = $\Delta$ TU
Low birth weight	0.08 (0.27)	0.09 (0.29)	0.09 (0.29)	0.79
Birth weight (Grams)	3390.43 (735.60)	3243.51 (636.09)	3219.85 (645.09)	0.48
Multiple-births	0.02 (0.16)	0.05 (0.22)	0.04 (0.19)	0.04
Child aged U-3	0.26 (0.44)	0.28 (0.45)	0.26 (0.44)	0.06
Child Female	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.69
Birth Order	3.40 (2.23)	3.40 (2.29)	3.28 (2.23)	0.92
Mothers age	34.93 (8.23)	34.49 (8.06)	34.51 (8.14)	0.25
Numb. ANC visits	4.08 (1.49)	4.08 (1.53)	3.97 (1.65)	0.80
Numb. HH members	6.19 (2.29)	6.59 (2.88)	6.66 (2.74)	0.03
Numb. Children aged U-5	1.50 (0.96)	1.60 (1.02)	1.49 (1.06)	0.04
Improved water	0.22 (0.41)	0.27 (0.44)	0.45 (0.50)	0.95
Female HH head	0.25 (0.44)	0.20 (0.40)	0.22 (0.42)	0.60
Age of HH head	41.69 (10.78)	41.80 (11.44)	41.80 (10.96)	0.58
Urban residence	0.16 (0.36)	0.13 (0.34)	0.41 (0.49)	0.24

*Notes:* This table shows means for TP, TU and consolidated control districts for sample characteristics at baseline (2007). The p-values in column 4 are generated from linear regressions with each respective variable in the row as the outcome variable regressed on a treatment indicator and year indicator interaction and district fixed effects (standard errors are clustered at the district level). The p-values are generated from t-tests of the difference between the TP and TU coefficients, where each coefficient measures the change in the outcome between each treatment and the control group. The categorical variables not shown in the table are; mothers highest level of education (primary across all treatments and control) and whether the woman has currently/formerly or has never been in a marital union (currently in a union across treatments and control). Birth order refers to the order in which a woman's children are born, beginning from her first birth.

The DHS survey collects information on a wide array of topics ranging from fertility, maternal and child health, HIV/AIDS and nutrition.<sup>10</sup> Specifically, the birth recode (BR) of the various Demographic and Health Surveys is used in this study. The BR contains information on the health status, including the birthweight, of each and every birth (child born) to surveyed women. A recall period of five years is used as a point of reference for each respondent in the survey. Combining the four ZDHS surveys, the final sample of observations utilised for the main outcomes comprises observations on 15,365 children born as singleton births, with non missing birthweight observations.<sup>11</sup>

Table 3.4.1 summarises baseline (2007) characteristics of the sample by treatment and compares changes in these characteristics between 2007 and 2014, after treatment for the full sample of observations.<sup>12</sup> Columns 1, 2 and 3 in table 3.4.1 reveals that TP, TU and the control districts were indeed quite similar at baseline. However, consolidated control districts are somewhat more urban.

The p-values in column 4 of table 3.4.1 are generated from linear regressions with each respective variable in the row as outcome variable, regressed on a treatment indicator and year interaction and district fixed effects, with 2007 as the reference year and standard errors clustered at the district level. The p-values are generated from t-tests of the difference between the TP and TU coefficients, with each coefficient measuring the change in the outcome between each treatment and the control group. Column 4 reveals that between 2007 and 2014, there are a few differences in sample characteristics between the changes in TP districts relative to the control districts and changes in TU relative to the control districts. Specifically, there are differences in the change in the proportions of children aged under age-3, the number of household members, the number of children aged under age-5 as well as in the proportions of multiple births between TP and TU, respectively. Owing to these differences, the main empirical analysis below controls for child (age, sex, birth order), mother (age, highest level of education and whether she is currently/formerly/has never been in a marital union) and household (number of household members, children aged under age 5, age of household head and indicators for availability of an improved water source, female headed household and urban residence) characteristics.

---

<sup>10</sup><https://www.dhsprogram.com/Who-We-Are/About-Us.cfm>.

<sup>11</sup>The entire sample size across all four surveys comprises observations on 119,036 children. However, 72% of children in the sample have missing birthweights. This fact, does not represent a weakness of the data source utilised but is a reflection of the underlying underdevelopment of the country as a whole, with infants not being born in health facilities and thus not being weighed at birth. Table 3.A.1 in annex 3.A.1 shows that birthweight observations are equally likely to be missing across TU and TP treatments, a fact which reduces the likelihood of bias in the analysis owing to missing birthweight observations.

<sup>12</sup>Consolidated control districts in table 3.4.1 combine both the 10 business as usual control districts from the ZRBF as well as the remaining 42 districts not included in the ZRBF. Table 3.A.2 in annex 3.A.2 presents similar descriptive statistics for the 10 ZRBF control as well as TU and TP districts, revealing that utilising the consolidated control districts does not bias the analysis. The sample used in the main analysis is further discussed in section 3.5 below.

## 3.5 Empirical Framework

### 3.5.1 Identification Strategy

This study utilises a Difference-in-Differences (DD) design and exploits both the timing of childrens' births and the district-level randomised roll out of the original ZRBF pilot (Friedman et al., 2016; Shen et al., 2017). Specifically, the study compares changes in birthweight and the probability of low birthweight for children born after the start of the ZRBF in April 2012 and those born before the start of treatment on this date. Treatment is constructed as the interaction of two variables: The first variable is constructed to represent an 'intensity' of treatment in-utero,<sup>13</sup> with the variable being equal to zero for all children born before the start of treatment in April 2012, equal to one for all children born nine months or more after the start of treatment and equal to the fraction of the nine month in-utero period, spent under treatment for children born within nine months of the start of treatment.<sup>14</sup> The second treatment variable captures infants born to mothers residing either in performance-based (TP) districts, unconditional (TU) district or consolidated control districts.

The change in birthweight and the probability of low birthweight among these groups of children is compared in the ten districts in which healthcare providers were randomised to receive TP, the ten districts in which healthcare providers were randomised to receive TU incentives and the remaining 52 districts, in which healthcare providers received no financial incentives at all. The consolidated control group utilised in the main analysis can thus be thought of as an 'expanded' control group as it consolidates not just the ten 'business as usual' control districts included in the original ZRBF pilot, but also includes the remaining 42 districts not included in the ZRBF pilot. The inclusion of the 42 additional districts permits an increase in statistical power of the current study and is valid as long as the parallel trends identifying assumption, critical for the DD strategy, discussed below, is fulfilled.<sup>15</sup> Heterogeneity in the treatment effects overtime and across districts, as well as the presence and size of negative weights, are assessed using the multiple groups and time-periods estimator (De Chaisemartin and d'Haultfoeuille, 2020).

The effects of healthcare utilisation on health outcomes can be potentially plagued by bias arising from non-random selection effects. By leveraging the randomised assignment to treatment inherent within the ZRBF pilot, this study is able to identify the causal

---

<sup>13</sup>Modelling the intensity of treatment in this manner is motivated by the observation in the literature that the timing of ANC in pregnancy is critical, with beneficial effects of ANC being largest, the earlier ANC is initiated in pregnancy (see for instance the discussion in Díaz and Saldarriaga (2019)).

<sup>14</sup>Each child's date of birth is normalised relative to the start of treatment to equal the number of months to/from April 2012. Thus, for an infant born one month after the start of treatment in May 2012, intensity of treatment in-utero would be 0.11 (i.e. 1/9). Similarly, for an infant born eight months after the start of treatment, intensity of treatment in-utero would be 0.89 (i.e. 8/9). Intensity of treatment in-utero for children born nine or more months after the start of treatment equals one.

<sup>15</sup>Tables 3.A.2, 3.A.3 and 3.A.4 in annexes 3.A.1 and 3.A.2 present the descriptive statistics, birthweight, low birthweight and ANC results respectively for the 30 ZRBF districts and show that while results are robust to the exclusion of the 42 additional districts, they are less precisely estimated, owing to the analysis only being based on 30 districts/clusters.

effects of the two treatments on infant birth outcomes. By taking eligibility for treatment, represented by infants being born to mothers residing either in TP or TU districts to be equivalent to treatment, all estimates in the study thus represent an Intention to Treat (ITT) estimate of the effects of TP and TU incentives on infant birth outcomes.

The key identifying assumption is that the difference in birthweight and the probability of low birthweight of infants in TP districts, followed the same trend as in TU and consolidated control districts in the years prior to treatment. In the absence of the two treatments, outcomes would have continued following the same trend in TP and TU as that in consolidated control districts between 2012 and 2014, were it not for the two treatments. Any deviations from the parallel/common trends between 2012 and 2014 are thus, interpreted as the causal effect of the two types of financial incentives. An approach to the assessment of the plausibility of the parallel trends assumption in the literature, which is by definition untestable, involves assessing pre-treatment treatment effects i.e. assessing the significance of treatment coefficient estimates for the years preceding treatment (Almond et al., 2011; Autor, 2003; Díaz and Saldarriaga, 2019; Hoynes et al., 2011, 2015; Muralidharan and Prakash, 2017). In addition to the above, the plausibility of the parallel trends assumptions can be assessed by using the multiple groups and time period (multiple) estimator of De Chaisemartin and d'Haultfoeuille (2020). Any unobservable differences between the three types of groups are assumed to be time invariant. Observable differences between the three types of groups are assumed to be captured by the inclusion of observable time varying covariates shown in table 3.4.1, in the analysis. The analysis also assumes that spillover effects among the districts randomised into the various treatments are absent. This latter assumption is also plausible, given that treatment was only assigned at the district level.

The primary DD analysis above is supplemented by two additional analyses. These analyses both alter the control groups used and in so doing, rely on slightly different sources of exogenous variation, relative to the DD analysis. The first analysis utilises a triple differences (DDD) design. Specifically, the DDD analysis compares changes in birthweight and the probability of low birthweight for those children aged at most three years old in 2014. This group of children, born during the ZRBF pilot projects implementation period were potential beneficiaries during their in-utero period. However, the older cohort of children aged over age-3 in 2014, were born before the ZRBF implementation period and were thus, too old to benefit in-utero from treatment. The over age-3 cohort thus represent a within-district control group leveraged in the DDD analysis. The outcomes for those children aged at most and over age-3 in the years 1996, 2001, 2007, before treatment are compared to those in 2014, after treatment. The differences in birthweight between these two groups overtime in the ten districts in which healthcare providers were randomised to receive TP and the ten districts in which healthcare providers were randomised to receive TU, are compared to the change in the remaining 52 consolidated control districts, in which healthcare providers conducted business as usual.

The second secondary analysis utilises an RD design. This analysis complements the DD and DDD analyses by relying on different identifying assumptions and a slightly different definition of control group. This sub-analysis utilises the 2014 DHS dataset only and exploits the five year recall period for all births captured in the DHS to compare birthweight and the probability of low birthweight for those infants born on either side of the qualifying date for the receipt of treatment: April 2012 (Friedman et al., 2016). This analysis utilises date of birth as a running variable and is conducted separately for TP and TU districts. The RD analysis also utilises a within-district control group: children born on the non-qualifying side of the treatment threshold within each district. This group of children thus experienced similar within-district conditions in-utero. A key identifying assumption in this RD framework is that of continuity of the running variable, which implies that each child’s date of birth is as good as randomly assigned within a close neighbourhood of the treatment threshold (Hahn et al., 2001). This assumption, requires that in a close neighbourhood of the treatment threshold, there is no possibility of children’s dates of birth being systematically and precisely manipulated such that children are more likely to be born after the qualifying threshold, relative to before this date or vice versa.<sup>16</sup>

### 3.5.2 Estimation Strategy

The main DD estimating equation of interest is equation 3.1 below,

$$y_{irvt} = \beta_0 + \beta_1 TU_r \times Int_i + \beta_2 TP_r \times Int_i + \gamma_r Dist_r + \lambda_v Byr_v + \delta X_{it} + \kappa_{rt} Dist_r \times year_t + \epsilon_{irvt} \quad (3.1)$$

$y_{irvt}$  represents each respective outcome of interest - birthweight or the probability of a low birthweight birth for child  $i$  in district  $r$ , birth cohort  $v$  and DHS survey year  $t$ .  $TU_r$  and  $TP_r$  are treatment status indicators equal to 1 if child  $i$  was born to a woman residing in either an unconditional or performance-based district, respectively, with consolidated control districts serving as the reference category.  $Int_i$  represents the intensity of treatment in-utero variable elaborated above. Some terms are omitted from equation 3.1, in order to conserve space. The preferred and most conservative specification including both district,  $Dist_r$  and birth cohort/year,  $Byr_v$  fixed effects, is estimated. Some specifications assess the robustness of the results to linear violations of parallel trends by including district by survey year time trends,  $Dist_r * year_t$ .  $X_{it}$  are time varying child (age, sex, birth order), mother (age, highest level of education and whether she is currently/formerly/has never been in a marital union) and household (number of household members, children aged under age 5, age of household head and indicators for availability of an improved water source, female headed household and urban residence) characteristics, as shown in table 3.4.1. The main coefficients of interest in equation 3.1, capturing the causal effects of TU and TP are  $\beta_1$  and  $\beta_2$ , respectively. Standard errors are clustered at the district level, the

---

<sup>16</sup>Figure 3.A.1 in annex 3.A.6, reveals the absence of heaping of observations immediately to the right of the qualifying threshold for treatment, providing evidence supporting the plausibility of continuity of the running variable around the threshold.

level of treatment.

In order to investigate the channels driving effects for the main outcome variables, a variant of Equation 3.1 is also estimated. In this case  $y_{irvt}$  is the outcome of interest for woman  $i$ , residing in district  $r$ , with a child in birth cohort  $v$  and in survey year  $t$ . The outcomes of interest are the number of ANC visits, the timing of the first ANC visit, in months and the quality of ANC, measured by the number of tetanus injections received during pregnancy and indicator variables of whether the woman was weighed and whether she had her blood pressure taken during pregnancy, respectively.

In addition to an analysis conducted on the full sample of observations using equation 3.1, a heterogeneity analysis of the effects of TP and TU on the main outcomes: birthweight and the probability of low birthweight is also conducted for high impact sub-samples of children. These high impact sub-samples of children are identified based on their mothers age and education level. Specifically, a heterogeneity analysis of children born to women with either no education or only a primary education and women younger than 20 years of age and older than 35 years of age is conducted. Children born to women with these characteristics have been identified as being at a higher risk of being born with low birthweight and thus most likely to benefit, in-utero, from treatment (Hoynes et al., 2015; Cardona-Sosa and Medina, 2017; Hoynes et al., 2011; Kramer, 1987; Amarante et al., 2016).

$$\begin{aligned}
y_{irvt} = & \alpha_0 + \alpha_1 U3_i \times TU_r \times year_{2014} + \alpha_2 U3_i \times TU_r \times year_{2001} \\
& + \alpha_3 U3_i \times TU_r \times year_{1996} + \alpha_4 U3_i \times TP_r \times year_{2014} \\
& + \alpha_5 U3_i \times TP_r \times year_{2001} + \alpha_6 U3_i \times TP_r \times year_{1996} \\
& + \gamma_r Dist_r + \lambda_v Byr_v + \delta X_{it} + \kappa_{rt} Dist_r \times year_t \\
& + \epsilon_{irvt} \tag{3.2}
\end{aligned}$$

The DD analysis in equation 3.1 is complemented with a DDD analysis based on equation 3.2 above. All variables in equation 3.2 are as defined in equation 3.1, with the exception of  $U3_i$  which is an indicator for child  $i$  being aged under-age 3 in each survey year,  $year_t$ , 1996, 2001, 2007 and 2014, respectively. Despite the ZRBF beginning in 2012, 2007 is taken as the baseline year in the DDD analysis, given that this is the earliest year before the start of the ZRBF for which the DHS was conducted. 2007 thus serves as the omitted reference category in this equation. Additional terms in this equation are also omitted so as to conserve space. The coefficients of interest in equation 3.2, capturing the causal effects of TU and TP are the respective 2014 coefficients  $\alpha_1$  and  $\alpha_4$ .

$$y_{irv} = \beta_0 + \sum_v \Psi_v Byr_v \times TU_r + \sum_v \Theta_v Byr_v \times TP_r + \gamma_r Dist_r + \delta X_i + \epsilon_{irv} \tag{3.3}$$

Additionally, pre-treatment treatment effects and the plausibility of the parallel trends assumption are assessed using the event-study specification 3.3 above. Variables in this

equation are as defined above, with the exception that the treatment status indicators,  $TU_r$  and  $TP_r$ , are interacted with the birth year indicators  $Byr_v$ . In this analysis, birth year 2011 is the reference year, year -1, relative to which estimates are made. Birth year 2012 is year 0, the 'event' year in which treatment begins. Thus, this analysis runs from event year -21, which is birth year 1991, before the start of treatment, to event year 3, which is birth year 2014, after the start of treatment. Event-study estimates from this equation are presented graphically in figure 3.5.1. The figure reveals that for the years preceding treatment, coefficients on the pre-treatment treatment effects are not statistically significant. This is true for mean birthweight in TU districts in panel A and TP districts in panel B, as well as the probability of low birthweight in TU districts in panel C and TP districts in panel D, respectively. This observation thus supports the plausibility of the parallel trends assumption in this setting.<sup>17</sup>

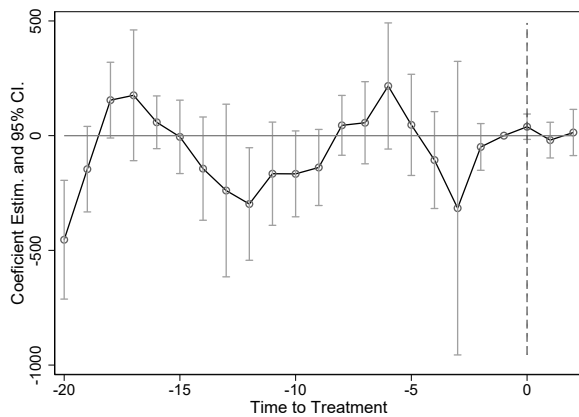
RD estimates are based on equation 3.4.<sup>18</sup> Equation 3.4 is similar to equation 3.1, except that the former includes the running variable  $Tlapse$ , measuring child  $i$ 's date of birth normalised to equal the number of months before or after the start of the treatment in April 2012. The function  $f(Tlapse)$  is a flexible function of the normalised date of birth, estimated by polynomials of order  $J$ . Equation 3.4 is estimated separately for  $j = TU, TP$ , respectively. The treatment indicators in this case equal one if child  $i$  is born after the start of the treatment. An interaction term of the normalised running variable and the treatment status indicator,  $f(Tlapse_i) \times Trt_j$ , is included in some specifications to allow for differing slopes on either side of the threshold.

$$y_{irv} = \tau_0 + \tau_1 f(Tlapse_i) + \tau_2 Trt_{jr} + \tau_3 f(Tlapse_i) \times Trt_{jr} + \gamma_r Dist_r + \lambda_v Byr_v + \delta X_i + \epsilon_{irv} \quad (3.4)$$

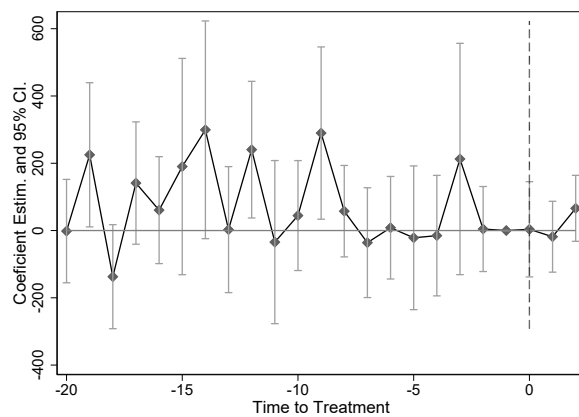
Assuming a linear model, equations 3.1, 3.2 and 3.3 can be estimated via linear regression, implemented via Ordinary Least Squares (OLS). Local polynomial regression is used to estimate equation 3.4, using robust bias-corrected estimation and data-driven bandwidth selection procedures for each outcome as proposed in Calonico et al. (2014, 2017). Higher order polynomials of the running variable are used in the analysis to account for potential non-linearities. A triangular kernel, emphasising observations near the treatment threshold, is also utilised. The continuity assumption is assessed using the discontinuity in density testing procedure proposed in Cattaneo et al. (2020) and estimated using the procedure proposed in Cattaneo et al. (2019).

<sup>17</sup>The robustness of the plausibility of the parallel trends assumption is further investigated using an alternative approach, the multiplept estimator in section 3.8.

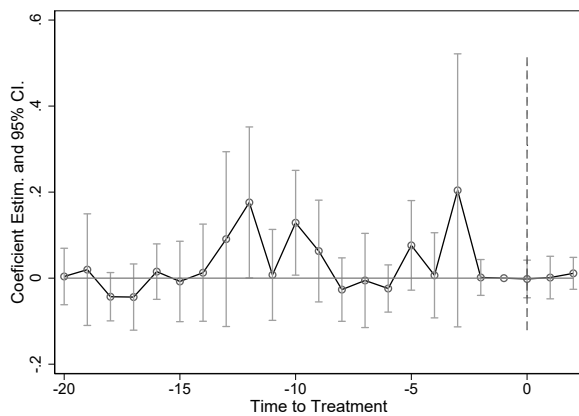
<sup>18</sup>Given the potential for non-random selection effects of healthcare utilisation and the ITT nature of the DD estimates noted above, the RD model in this case can be considered a fuzzy RD with one-sided perfect compliance. Taken in this way, equation 3.4 is thus a reduced form equation estimated as a sharp RD model which also yields ITT estimates of treatment effects (Lee and Lemieux, 2010).



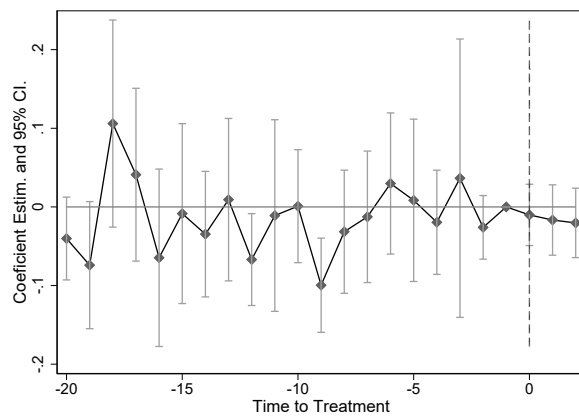
(a) Birthweight (Grams), Unconditional Districts



(b) Birthweight (Grams), Performance-based Districts



(c) Low Birthweight, Unconditional Districts



(d) Low Birthweight, Performance-based Districts

Figure 3.5.1: Event Study Estimates

*Notes:* The figure plots event-study coefficient estimates and associated 95% confidence intervals based on the event-study equation 3.3. Panels A and B show event-study estimates for birthweight in unconditional and performance-based districts, respectively. Panels C and D show estimates of the probability of a low birthweight birth in unconditional and performance-based districts, respectively. The estimates are calculated relative to event time -1, which represents birth year 2011 and run from event time -21, representing birth year 1991, to event time 3, representing birth year 2014.

## 3.6 Results

### 3.6.1 Difference-in-Differences Estimates

DD estimates from equation 3.1 are presented in table 3.6.1. Each column in the table contains results from a separate DD regression, with the dependent variable in each regression shown in the second row of the table. In columns 1-3, birthweight is the dependent variable, while in columns 4-6, the low birthweight indicator is the dependent variable. All regressions are based on the most conservative specification with both district and year of birth fixed effects and standard errors clustered at the district level.<sup>19</sup> The p-values in row 5 are based on a t-test of the significance of the difference between each pair of corresponding TP and TU coefficients in each respective column. Columns 1 and 4 show estimates for mean birthweight and low birthweight, respectively, for the full sample observations. Columns 2, 3 and 5 and 6, in turn, contain DD estimates of the heterogeneous effects of TP and TU on mean birthweight, in columns 2 and 3 and low birthweight in columns 5 and 6, for the high-risk sub-samples of children, respectively.

The DD coefficient estimates in column 1 of the table reveal that relative to the change in mean birthweight in control districts, the change in mean birthweight for the full sample of children born in TU districts was 45 grams higher, post treatment. Similarly, mean birthweight in TP districts was 32 grams higher, post treatment. Neither the TP nor the TU coefficient is, however, statistically significant. Column 2 of table 3.6.1 contains DD estimates of the heterogeneous effects of TP and TU on mean birthweight of children born to women with either no education or only a primary education. Results reveal that the change in mean birth weight of children in TU districts for this sub-sample of children was 103 grams higher, post treatment. In the case of TP districts, children born to this high impact sub-sample of women had birthweight 29 grams higher on average, post treatment. The TU coefficient is statistically significant at the 5% level, while the TP coefficient is not statistically significant.<sup>20</sup> Presentation of the heterogeneous effects of TU and TP on mean birthweight concludes in column 3 of table 3.6.1 which contains estimates for children born to the high impact sub-sample of women aged at most 20 years old and those aged 35 years and older, respectively. Results reveal that the change in mean birth weight in TU districts was 207 grams higher, post treatment for this sub-sample of children. Similarly, mean birthweight in TP districts was 163 grams higher, post treatment in this sub-sample. Both coefficient estimates are statistically significant at the 5% level.<sup>21</sup>

---

<sup>19</sup>DD estimates similar to those in table 3.6.1 but estimated with district by year fixed effects and not district and birth cohort fixed effects are qualitatively similar to those in table 3.6.1 and are shown in table 3.A.9 in annex 3.A.5.

<sup>20</sup>A similar regression in column 1 of table 3.A.6 in annex 4.A.2 for the sub-sample of women with a secondary education or higher reveals qualitatively similar but somewhat smaller coefficient estimates.

<sup>21</sup>A similar regression shown in column 2 of table 3.A.6 in annex 4.A.2 on a low impact sub-sample of women older than 20 years and 35 years or younger, reveals qualitatively similar but somewhat smaller TU and TP coefficient estimates, respectively.

Table 3.6.1: Difference-in-Differences Estimates for Birthweight and low Birthweight.

	(1)	(2)	(3)	(4)	(5)	(6)
	birth weight	birth weight	birth weight	Low birth weight	Low birth weight	Low birth weight
<b>Unconditional</b> $\times$ <b>Int</b>	45.37 (42.25)	102.7** (41.57)	207.0** (94.88)	-0.00617 (0.0180)	-0.0433* (0.0226)	-0.0998*** (0.0359)
<b>Performance-based</b> $\times$ <b>Int</b>	32.18 (29.34)	28.57 (39.26)	162.5** (69.47)	-0.0134 (0.0114)	-0.0351** (0.0143)	-0.0993*** (0.0264)
P-value Coef.(TP - TU)	0.785	0.156	0.696	0.722	0.736	0.990
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Sub(Educ)	Sub(Age)	Full	Sub(Educ)	Sub(Age)
Control Group Mean	3091.0*** (46.77)	3249.9*** (61.38)	3070.4*** (69.45)	0.139*** (0.0202)	0.0911*** (0.0252)	0.160*** (0.0352)
Observations	15365	9504	4409	15365	9504	4409
$R^2$	0.050	0.051	0.079	0.016	0.019	0.034

*Notes:* The table shows DD estimates of the main (columns 1 and 4) and heterogeneous (columns 2 and 3; 5 and 6) impacts of unconditional and performance-based financial incentives on mean birthweight and low birthweight for singleton births, for children in unconditional and performance-based districts, relative to consolidated control districts. Infant birthweight is measured in grams, while low birthweight is an indicator variable for infant birth weight less than 2,500 grams. Each column shows a separate regression with some coefficients omitted to conserve space. The intensity (int) variable equals zero for all children born before treatment (before April 2012), one for all children born nine months or more after the start of treatment and equals the fraction of the in-utero (nine month) period, potentially spent under treatment for children born within nine months of the start of treatment. Heterogeneous effects are estimated based on mothers education in column 2 and 5 (no education and only primary education) and mothers age in column 3 and 6 (mothers  $\leq 20$  and those  $>35$  years old). The preferred and most conservative specification, including both district and birth cohort fixed effects, is used in all regressions. Standard errors, clustered at the district level, are in parentheses;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Estimates of the impact of TU and TP on the probability of a low birthweight infant in the main sample and the high impact sub-samples are shown in columns 4, 5 and 6 of table 3.6.1, respectively. Column 4 of the table reveals that for the full sample of children, the change in the probability of an infant being born with low birthweight in TU districts reduced by 0.6 percentage points, while the same probability in TP districts reduced by 1.3 percentage points, post treatment. Estimates in column 4 for the probability low birthweight are, however, not statistically significant. Column 5 of table 3.6.1 shows that the probability of an infant being born with low birthweight in the no or at most primary education sub-sample reduced by 4 percentage points, post treatment, both in TU and TP districts. The TP coefficient is significant at the 5% level while the TU coefficient is only significant at the 10% level.<sup>22</sup> Similarly, column 6 of table 3.6.1 reveals that for children born to the high impact sub-sample of women aged either at most 20 years old or aged 35 years and older, the probability of an infant being born with low birthweight reduced by approximately 10 percentage points in both TU and TP districts respectively, post treatment. Both the TU and TP coefficients are statistically significant at the 1% level.<sup>23</sup>

Several aspects of the DD results above are worth noting at this juncture. Firstly, TU and TP coefficients are quite similar in size and thus, p-values of the significance of the difference between them in row 5 of table 3.6.1 reveal that they are statistically indistinguishable. Secondly, the results differ in important ways from some studies in the social programme/cash transfer literature assessing infant health outcomes. Most papers in this literature find more statistically significant effects of these social programmes on mean infant birthweight and low birthweight births (Almond et al., 2011; Amarante et al., 2016; Barber and Gertler, 2008; Hoynes et al., 2011, 2015). In the current study, similar effects are mostly restricted to children in high impact sub-samples most likely to benefit from treatment. The findings above may imply that when financial incentives are provided to healthcare providers on the supply-side, the providers are still able to ensure that the greatest productivity benefits in-terms of improved child health outcomes still accrue to high impact subgroups of the population most likely to benefit from treatment.

As noted previously, the receipt of Antenatal Care (ANC) in pregnancy that is initiated early and is of high quality is a key channel both in reducing the likelihood of low birthweight and increasing birthweight. The ZRBF pilot thus included an explicit incentivisation of ANC visits for healthcare providers in TP districts of \$1.6, per woman, per visit conducted, while similar performance-based financial incentives were absent in TU and control districts. Thus, in order to gain further insights into the potential channels driving the above results in table 3.6.1, an assessment of the impact of TP and TU on the

---

<sup>22</sup>A similar regression shown in column 3 of table 3.A.6 in annex 4.A.2 for the low impact sub-sample of women a secondary education or higher, reveals qualitatively similar but somewhat smaller TU and TP coefficient estimates, respectively.

<sup>23</sup>A similar regression shown in column 4 of table 3.A.6 in annex 4.A.2 on a low impact sub-sample of women older than 20 years and 35 years or younger, reveals qualitatively similar but smaller TU and TP coefficient estimates.

number, timing and quality of ANC visits, is warranted. To this end, table 3.6.2 below presents estimates based on equation 3.1 for the full sample of observations.<sup>24</sup>

Column 1 of the table contains DD estimates with the number of ANC visits as outcome variable while in column 2, the timing of the first ANC visit, in months, is the dependent variable. In columns 3-5, estimates with the quality of ANC measured as; the number of tetanus injections received during pregnancy, whether the woman was weighed and whether she had her blood pressure taken during pregnancy, respectively, as dependant variable, are shown. All regressions are based on the most conservative specification with both district and year of birth fixed effects and standard errors clustered at the district level. The p-value in row 5 is based on a t-test of the significance of the difference between each pair of TP and TU coefficients in each respective column.

Column 1 of table 3.6.2 shows that relative to control districts pre and post treatment, the change in the number of ANC visits in TU districts was 0.3 visits higher, post treatment. In the case of TP districts, 0.4 more ANC visits were conducted, post treatment. The latter effect in TP districts is significant at the 1% level, while the former effect in TU districts is only significant at the 10% level. Column 2 of table 3.6.2 shows that the timing of the first ANC visit was 0.04 months earlier in TU districts, post treatment. Similarly, in the case of TP districts, the timing of the first ANC visit was 0.2 months earlier. Only the latter effect in TP districts is significant at the 10% level.

Column 3 of table 3.6.2 reveals that 0.1 more tetanus injections were administered to women during pregnancy in TU districts, post treatment. Similarly, in TP districts, 0.1 more tetanus injections were administered during pregnancy, post treatment. The TU coefficient is significant at the 5% level while the TP coefficient is not statistically significant. In column 4, we see a 3.7 percentage point increase in the probability of a pregnant woman being weighed during pregnancy in TU and a 0.3 percentage point increase in this probability in TP districts, post treatment. The TU coefficient is significant at the 10% level while the TP coefficient estimate is not statistically significant.

Lastly, column 5 of table 3.6.2 reveals that the probability that a woman had her blood pressure checked during pregnancy increased by 3.2 percentage, post treatment, in TU districts. In the case of TP districts, this probability increased by 6 percentage points, post treatment. The former effect in TU districts is not statistically significant, while the latter effect in TP districts is statistically significant at the 5% level.

Several aspects of the results in table 3.6.2 are noteworthy. Firstly, the larger and statistically significant coefficient estimates for performance-based districts in columns 1 and 2, are consistent with healthcare providers being provided with an explicit

---

<sup>24</sup>Similar estimates for the high impact sub-samples are shown in table 3.A.7 for the 'no or at most primary education' sub-sample of women and in table 3.A.8 for women younger than 20 years old or older than 35 years old.

Table 3.6.2: Difference-in-Differences Estimates for the Number, Timing and Quality of Ante-natal Care Visits: Full Sample

	(1)	(2)	(3)	(4)	(5)
	num. ANC visits	timing 1st ANC	num. tetanus injections	weighed	blood pressure
<b>Unconditional</b> $\times$ <b>Int</b>	0.264*	-0.0411	0.116**	0.0372*	0.0322
	(0.147)	(0.124)	(0.0544)	(0.0215)	(0.0292)
<b>Performance-based</b> $\times$ <b>Int</b>	0.403***	-0.242*	0.131	0.00269	0.0602**
	(0.128)	(0.145)	(0.0868)	(0.0251)	(0.0239)
P-value Coef.(TU - TP)	0.419	0.271	0.875	0.279	0.412
District fixed effects	Yes	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes	Yes
Control Group Mean	3.353***	4.641***	1.738***	0.896***	0.715***
	(0.133)	(0.0689)	(0.0425)	(0.0100)	(0.0182)
Observations	24073	23637	24262	16876	16872
$R^2$	0.146	0.067	0.075	0.053	0.105

*Notes:* The table shows DD regressions of the number, timing and quality of Ante-natal Care (ANC) visits in unconditional and performance-based districts, relative to consolidated control districts, for the full sample of observations. Each column shows a separate DD regression, with some coefficients omitted to conserve space. The intensity (Int) variable is zero for all children born before treatment (before April 2012), one for all infants born nine months or more after the start of treatment and is set to equal the fraction of the in-utero (nine month) period, spent under treatment for children born within nine months of the start of treatment. The preferred and most conservative specification, including both district and birth cohort fixed effects, is used in all regressions. Consolidated control districts are the reference category. ANC quality is measured by the number of tetanus injections received by a women before giving birth, and dummy variables indicating; whether the woman was weighed and had her blood pressure taken, respectively, during pregnancy. The regressions in columns 4 and 5 contain fewer observations because in the ZDHS 1996, respondents were not asked about being weighed or having their blood pressure taken at ANC. Standard errors, clustered at the district level, are in parentheses.  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

performance-based financial incentive of \$1.6 per women per ANC visit offered in TP districts during the ZRBF. The effect on the number of ANC visits in table 3.6.2, is consistent with healthcare providers responding to this performance-based financial incentive with greater effort and greater morale in the provision of ANC, with this effect extending to providers encouraging women to not only attend more ANC visits, but also initiate these earlier in pregnancy as seen in column 2. Rather interestingly, similar effects are also evident in TU districts, despite healthcare providers in these districts only receiving unconditional financial incentives independent of both effort and the number of ANC visits conducted. The effects on the number and timing of ANC visits in TP districts do not, however, seem to translate into a corresponding increase in the quality of the ANC visits as observed in columns 3 and 4 (as noted above, column 5 shows a statistically significant effect of TP on the probability that a pregnant woman had her blood pressure checked during pregnancy). This observation could be a reflection of the fact that the quality of healthcare provision in TP districts was not directly incentivised but was only indirectly incentivised through a quality multiplier. This observation, could thus explain the weak improvements in both birthweight and low birthweight in table 3.6.1 above. Additionally, as in table 3.6.1, the TU and TP coefficients are quite similar and thus virtually statistically identical. This is evident from the p-values from t-tests of the significance of the difference between these coefficients in row 5 of table 3.6.2.

The observation of a statistically significant effect on the timing of the first ANC visit in TP districts is also in line with the findings of the original impact evaluation, which found a statistically significant reduction in the timing of the first ANC of 0.4 months relative to districts defined in the current study as TU districts (Friedman et al., 2016).

Estimates of the impact of TU and TP on the number, timing and quality of ANC in the high impact sub-samples of women with no education or only a primary education and those women aged 20 years old or younger or older than 35 years of age are presented respectively in table 3.A.7 and table 3.A.8 in annex 3.A.4. Overall, both sets of estimates are qualitatively similar to those in table 3.6.2 above and reveal gains in the number of and timing of ANC visits in TP districts larger than those in TU districts.

### **3.6.2 Regression Discontinuity Estimates**

As noted in section 3.5, this section complements the DD analysis with an RD analysis. The RD analysis utilises date of birth as the running variable and exploits the 5 year recall period for all births in the 2014 DHS to compare the birth outcomes of children born on either side of the eligibility threshold for TU and TP treatment - April 2012. The analysis is conducted separately for each treatment and utilises as the control group, children born in the same district but on the ineligible side of the treatment threshold. The validity of the RD design is based on the fulfilment of the continuity assumption (Hahn et al., 2001). Figure 3.A.1 in annex 3.A.6, shows a histogram of the density of births around the eligibility threshold with a superimposed local polynomial estimate of the density function (Cattaneo et al., 2020, 2019). Both reveal the absence of an unusual clustering of births

on either side of the treatment threshold, supporting the plausibility of continuity of the running variable around the threshold in this case.

Figure 3.6.1 graphically plots RD estimates based on equation 3.4 for birthweight in panel A and the probability of a low birthweight birth in panel C for TU districts. Similarly, panels B and D plot RD birthweight and low birthweight estimates for TP districts, respectively. Given that the RD analysis utilises only 1 of the 4 waves of ZDHS data sets utilised in the DD analysis and that the analysis is conducted separately per treatment, the number of observations on either side of the cut off in each instance is thus, quite limited. The RD results in figure 3.6.1 should thus, be interpreted with this consideration in mind.

The analysis reveals broadly similar conclusions to those reached based on the DD analysis in table 3.6.1. Specifically, panels A and B of figure 3.6.1 reveal an increase in birthweight in TU and TP districts, respectively, at the eligibility threshold. Similarly, panels C and D reveal a decrease in the probability of a low birthweight in TU and TP districts, respectively, at the eligibility threshold.

### **3.7 Cost-Effectiveness**

This section presents illustrative back of the envelope cost-effectiveness calculations for the case of birthweight. Based on the original ZRBF pilot, the per-capita cost of implementing TP was estimated at \$7.91 while that of implementing TU was estimated at \$2.16 (Friedman et al., 2016). These cost estimates are combined with the DD birthweight coefficient estimates from the sub-sample of children born to women aged less than 20 years old or older than 35 years old in table 3.6.1. This sub-sample of children are noted to experience increases in birthweight of 207 and 163 grams for TU and TP, respectively, that are also statistically significant and thus, also potentially cost-effective (Mbiti et al., 2019).

The results of the back of the envelope cost effectiveness exercise reveal that the cost of increasing birthweight by 1 gram in TU districts is \$0.010 while that in TP districts is \$0.049. This implies that unconditional financial incentives may be more cost effective per gram of birthweight gained. Alternatively, for performance-based financial incentives to be at least as cost effective as unconditional financial incentives, given their per-capita cost of \$7.91, they would need to increase birthweight by at least 758 grams on average, a 5 fold increase in birthweight. Such an increase seems implausible given that estimates in the current study and the literature only range between 2 grams and 207 grams (Almond et al., 2011; Amarante et al., 2016; Barber and Gertler, 2008; Cardona-Sosa and Medina, 2017; Hoynes et al., 2011, 2015).

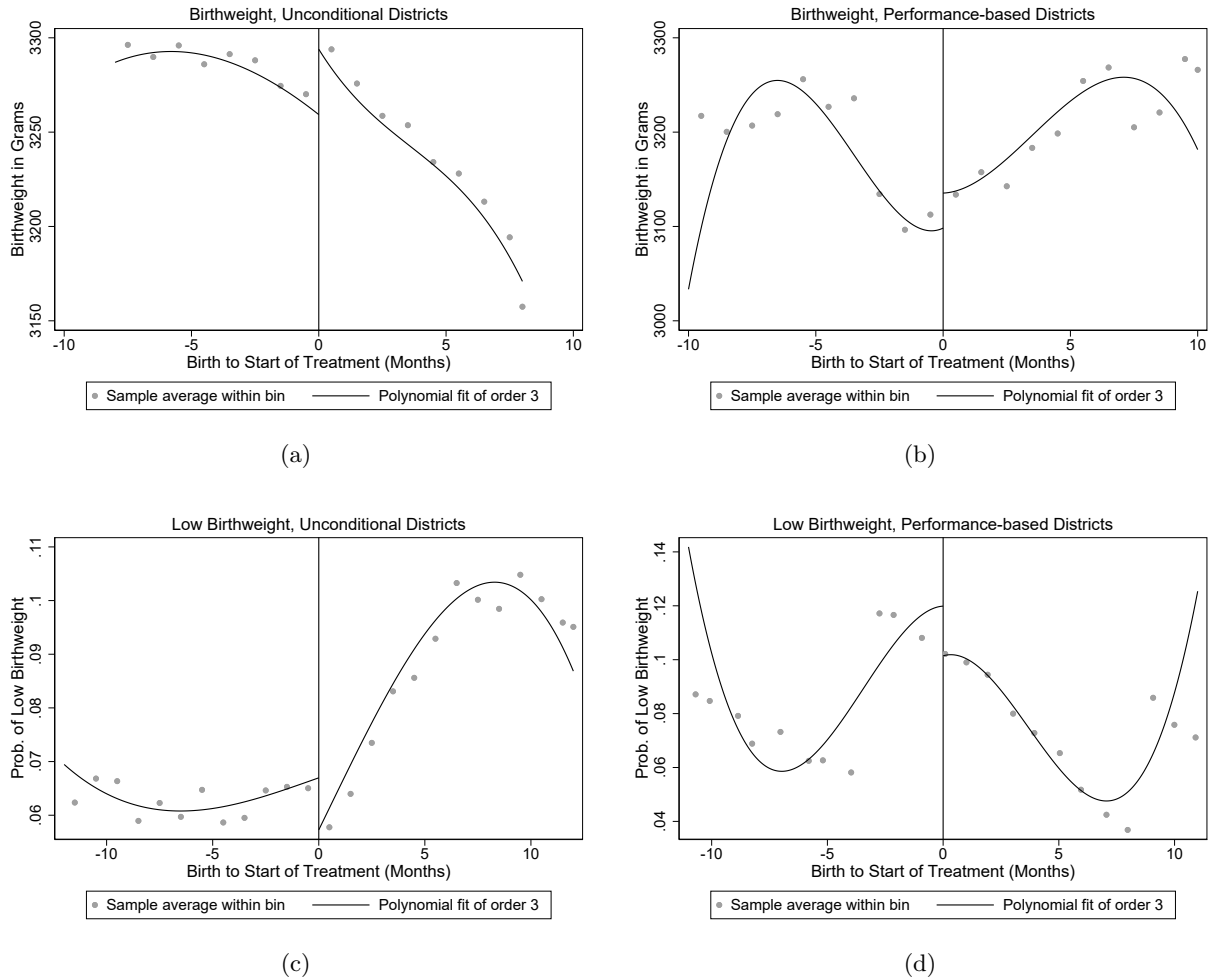


Figure 3.6.1: Regression Discontinuity Plots

*Notes:* The figure shows local polynomial RD estimates for birthweight and the likelihood of a low birthweight birth in TU districts in panels A and C and TP districts in panels B and D, respectively. The horizontal axis plots the running variable, date of birth, normalised such that the treatment threshold, April 2012, is at zero. Bias-robust estimation and data-driven bandwidth selection procedures for each outcome as proposed in Calonico et al. (2014, 2017), are used. Gray dots are quantile-spaced within-bin averages, mimicking the underlying variance of the data. Polynomials of orders 3, that account for potential bias due to non-linearities in the data, are shown. All specifications utilise a triangular weighting kernel.

## 3.8 Robustness Checks

### 3.8.1 Original ZRBF Districts

As noted above, the original ZRBF pilot randomised 30 Zambian districts to either receive TP, TU or conduct business as usual. As outlined above, the main analysis in this study exploits the nationally representative nature of the ZDHS to include the 42 districts that also continued to conduct business as usual, but were not included in the original ZRBF. The analysis in this study thus compares the 20 TU and TP districts included in the ZRBF to a consolidated set of 52 districts that conducted business as usual. While the inclusion of the additional 42 control districts to the original 10 ZRBF control districts potentially increases statistical power owing to treatment being assigned at the district level, this could potentially result in bias if the 42 business as usual districts excluded in the ZRBF are systematically different from those districts that were included. To assess the potential of any such bias, the analysis in annex 3.A.2 is restricted only to the original 30 ZRBF districts. Table 3.A.2 presents descriptive statistics for these districts and reveals quite similar characteristics to those presented in table 3.4.1 in section 3.4. However, while 41% of consolidated control districts are urban in table 3.4.1, table 3.A.2 reveals that only 21% of the ZRBF control districts are urban.

Tables 3.A.3 and 3.A.4 in annex 3.A.2 continue in this vein and present analyses for the main outcomes as well as the channels of impact for the 30 ZRBF districts. Table 3.A.3 shows that coefficient estimates for birthweight and the probability of a low birthweight birth, are qualitatively similar to those presented in table 3.6.1 in section 3.6. While coefficients in the full samples are suggestive of improvements in birthweight and the probability of a low birthweight birth, coefficients also become larger and more statistically significant in the same high impact sub-samples of children as in table 3.6.1. Similarly table 3.A.4 in annex 3.A.2 presents results for the channels of impact - the quantity and quality of ANC for the 30 ZRBF districts. While results also share qualitative similarities with those presented in table 3.6.2, the former are less precisely estimated owing to the fewer observations and clusters (only 30) on which they are based.

Importantly, the discussion above thus reveals that the main results are robust to the exclusion of the 42 business as usual districts not included in the ZRBF and their inclusion does not introduce bias in the analysis. In-fact, the inclusion of the additional districts or clusters and their corresponding observations only serves to increase the statistical power and precision of the results presented in the main analysis.

### 3.8.2 Triple Difference Estimates

As noted in section 3.5, the DD analysis is also complemented with a DDD analysis, estimated based on equation 3.2. Recall that the DDD specification in equation 3.2, compares birthweight and the probability of a low birthweight birth of children aged under age-3, with children aged over age-3 serving as an additional, within district control

group, in TP and TU districts and consolidated control districts in the years 1996, 2001 and 2007, prior to treatment and in 2014, after treatment.

Results, presented in annex 4.A.2, are qualitatively similar to the DD results presented in subsection 3.6.1 and the RD results in subsection 3.6.2, but are somewhat less precisely estimated. Specifically, similar to the DD results presented in table 3.6.1, columns 1 and 4 of table 3.A.5 reveal insignificant effects of TU and TP on infant birthweight and low birthweight, respectively, for the full sample of observations. Similarly, coefficient estimates are larger and gain some statistical significance in the same high impact sub-samples as those examined in table 3.6.1. Whereas all the mean birthweight coefficients in the DD results in table 3.6.1 are positive, columns 1 and 3 of table 3.A.5, however, reveals negative but insignificant coefficient for TU districts in the full sample of observations. All other DDD and DD estimates for the main sample are qualitatively similar.

The lower precision in the DDD estimates may be due to several reasons, two of which are discussed below. Firstly, with standard errors clustered at the district level, the level of treatment, statistical power is closely connected to the number of district clusters: 72 in this case. This implies that the estimation of additional coefficients necessitated by the DDD estimation with numerous sets of fixed effects, results in a loss of statistical power. Secondly, the definition of the population potentially exposed to treatment as comprising children aged at most 3 years old, as opposed to the intensity of treatment definition of exposure to treatment used in the DD analysis, could imply that the treatment variable is measured with a degree of measurement error in the DDD analysis, whereas no such error arises in the DD analysis.

### 3.8.3 Parallel Trends

As noted in section 3.5, the validity of the crucial parallel trends assumption, whose plausibility was established in figure 3.5.1, is further assessed using an additional approach from the more recent Differences-in-Differences literature: the multiple group and time (multiplegt) estimator.

As noted by De Chaisemartin and d’Haultfoeuille (2020), the multiplegt estimator (or  $DD_M$ ) is useful in DD type settings in which there are many groups and time-periods being assessed and it is suspected that the treatment effect of interest might not be constant, but may actually vary across the different groups and time periods. In settings with multiple groups and time periods, the conventional two-way fixed effects DD estimator provides a weighted average treatment effect across all of the groups and time-periods and the weights associated with some of these treatment effects may be negative. With varying treatment effects across groups and time-periods, the presence of negative weights may imply that when estimated, the conventional two-way fixed effects DD estimator may turn out negative when in fact, all of the underlying average treatment effects themselves are actually positive (De Chaisemartin and d’Haultfoeuille, 2020). The multiplegt estimator has been argued to be robust and valid even in the presence of such treatment effect

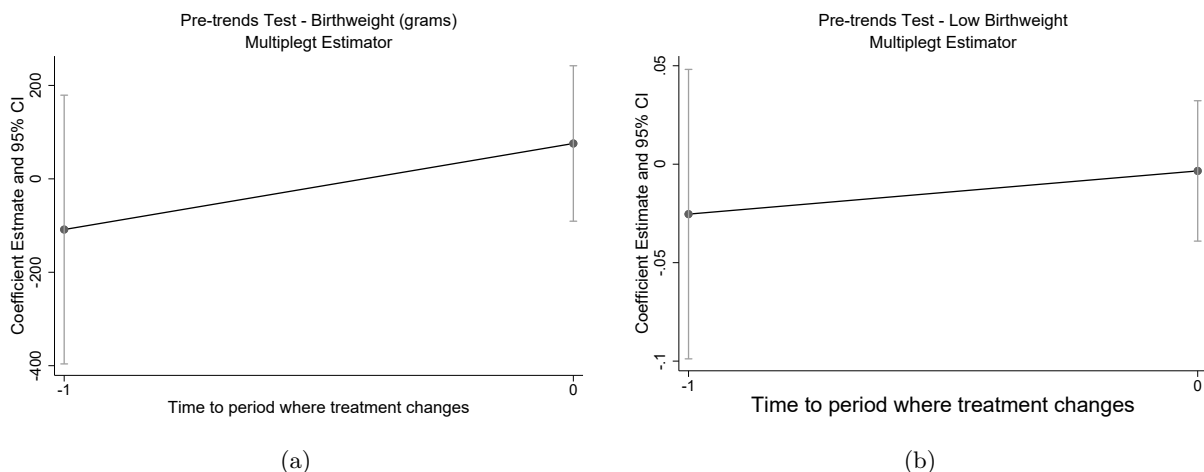


Figure 3.8.1: Pre-trends Test using the Multiplegt Estimator

*Notes:* The figure shows a test for parallel trends using the Multiplegt estimator proposed in De Chaisemartin and d'Haultfoeuille (2020). It compares pre-treatment treatment effects in the combined treated groups, TU and TP districts to consolidated control districts. Pre-treatment treatment effects are calculated for  $n-2$  time periods, where  $n$  equals the total number of time periods (four survey years in this case). Panel A shows estimates for mean birth weight while panel B shows estimates for low birthweight. Time 0 refers to the baseline year, 2007, while time -1 refers to the first pre-baseline survey year, 2001.

variability or heterogeneity over the groups and time-periods being examined.<sup>25</sup>

Given that the validity of the multiplegt estimator hinges on the plausibility of the parallel trends assumption, it therefore presents an alternative test for the plausibility of the parallel trends assumption, one that is exploited in the current analysis and presented in figure 3.8.1. A test of the parallel trends assumption using the multiplegt estimator amounts to comparing pre-treatment treatment effects in consolidated control districts relative to TU and TP districts combined, the groups that change treatment status. The estimator calculates the effects for  $n - 2$  time periods, with  $n$  being the total number of time periods. In the present case  $n$  equals the total number of survey years, four, implying that  $n - 2$  equals two (i.e. the survey years 2007 and 2001).

Panel A of figure 3.8.1 reveals insignificant pre-treatment treatment effects in the case of birthweight while panel B reveals similar insignificant pre-treatment treatment effects in the case of low birthweight.<sup>26</sup> Thus, the observations based on the multiplegt estimator seem to support the plausibility of the parallel trends assumption, a similar conclusion to that reached based on figure 3.5.1 in section 3.6.

<sup>25</sup>An associated analysis to determine the presence of negative weights reveals that there are no negative weights associated with the individual average treatment effects across the 3 groups and over the survey years in the current analysis. This implies that the conventional two-way fixed effects DD estimator is sufficient and the multiplegt estimator is therefore not used to calculate treatment effects during the post-treatment period.

<sup>26</sup>Taking the case of birthweight for illustrative purposes,  $DID_M^{pl,0}$  equals 75.73 (standard error of 75.62), while  $DID_M^{pl,-1}$  equals -108.53 (standard error of 114.31).

### 3.9 Discussion and Conclusion

Similar to other low income developing countries, efforts to improve maternal and child health outcomes in Zambia remain at the very top of the development agenda. A popular policy response has involved the implementation of health programmes providing performance-based financial incentives to healthcare providers. However, despite the popularity of this approach, evidence documenting the relative effectiveness of performance-based financial incentives to alternative types of financial incentives, remains scarce.

This study compares the relative effects on birthweight and the probability of a low birthweight birth of performance-based (TP) or unconditional (TU) financial incentives to healthcare providers. The study exploits the district randomised roll out of the original Zambia Results Based Financing (ZRBF) pilot project, implemented between 2012 and 2014 in Zambia and compares those districts in which healthcare providers were randomised to receive either TP or TU to control districts in which healthcare providers conducted 'business as usual'. It utilises four waves of nationally representative Demographic and Health Survey (ZDHS) data and a Difference-in-Differences (DD) research design, complemented with triple differences (DDD) and Regression Discontinuity (RD) designs.

DD estimates of the study suggest that birthweight in TP districts increases by 32 grams while the probability of a low birthweight birth reduces by 1.3 percentage-points. TU districts experience comparable improvements in infant birth outcomes with estimates suggesting a 45 gram increase in birthweight and a 0.6 percentage-point reduction in the probability of a low birthweight birth. However, these estimates are not statistically significant. The comparability of the gains in birth outcomes across the two sets of treatments also implies that all estimates are statistically indistinguishable. The study also examines birth outcomes in 'high impact' sub-samples of children born to women identified as being at higher risk of bearing low birthweight children. These are women with either no education or only a primary education and women younger than 20 years of age or older than 35 years of age. DD estimates in these sub-samples of children reveals both larger and more statistically significant increases in mean birthweight in both TU and TP districts of between 29 and 207 grams and reductions in the probability of a low birthweight birth of between 4 and 10 percentage points, respectively. Overall, magnitudes of these TU and TP coefficients are also remarkably similar and thus statistically indistinguishable. Back of the envelope cost effectiveness calculations reveal that each gram increase in birthweight is achieved at \$0.049 for TP and \$0.010 in the case of TU. Performance-based financial incentives would thus have to be an implausible 5 times more effective i.e. increase birthweight by 758 grams, to be as cost-effective as unconditional financial incentives.

The study also sheds light on a potential mechanism driving the above results - the number, timing and quality of antenatal care (ANC) visits, measured by the number of tetanus injections and whether a woman was weighed and had her blood pressure

taken during pregnancy. During the ZRBF pilot, the number of ANC visits conducted by healthcare providers was attached to an explicit performance-based financial incentive for healthcare providers of \$1.6 per woman per visit. DD estimates reveal that this explicit incentivisation resulted in 0.4 more ANC visits occurring approximately 0.2 months earlier, with the probability of a woman having her blood pressure taken during pregnancy also increasing by 6 percentage points in TP districts. While the above effects in TP districts are statistically significant, the corresponding effects in TU districts, are qualitatively similar. Interestingly, estimates for the quality of ANC measured by the number of tetanus injections and the probability a woman was weighed during pregnancy, are statistically significant in TU but not in TP districts. The above implies that while TP resulted in a larger number of earlier initiated ANC visits, it did not adequately translate into commensurate improvements in the quality of ANC visits. This could thus account for the comparable improvements in birthweight and low birthweight in TU and TP districts.

One potential driver behind the pattern of results noted above could be multitasking. While healthcare providers received the above noted explicit financial incentive of \$1.6 per ANC visit in TP districts, the quality of ANC in TP districts was only indirectly attached to a performance-based financial incentive, through a quality multiplier. Healthcare providers could have thus focused on the number of ANC visits, that were attached to the performance-based financial incentive, to the detriment of the quality of the ANC visits. This narrative is consistent with the literature on multitasking in general (Holmstrom and Milgrom, 1991) and with Zambian healthcare providers in particular (Chibuye et al., 2021). This could thus explain why, even though coefficient estimates are suggestive of improvements, the quality of ANC did not commensurately improve in TP districts. An improvement in the number and timing of ANC visits without a commensurate improvement in the quality of the ANC visits thus implies that health outcomes such as birthweight and low birthweight in TP districts did not improve any more than they did in TU districts.

Results of this study would thus imply that where improvements in child health outcomes such as birthweight and low birth weight are concerned, performance-based financial incentives may not be as productive as previously thought. Similarly, unconditional financial incentives would also seem to achieve a comparable level of productivity. Because the latter incentive type can be implemented at lower cost, they may actually be more cost-effective. This finding differs from other studies in the literature such as Mbiti et al. (2019), who find no impacts on student education outcomes of providing schools with unconditional transfers, positive impacts of providing performance-based financial incentives to teachers and both positive and significant effects of providing both types of incentives to schools.

One limitation of this study lies in the fact that the data source used, assembles multi-round ZDHS data on 119,036 children in order to have 15,365 birthweight observations for the full sample. Birthweight observations are, however, shown to be equally likely to be

missing in both TU and TP districts overall and also for each of the respective survey years included in the analysis, reducing the likelihood of bias in the analysis. The large number of missing observations are, however, a reflection of the underlying underdevelopment of the setting in which the study is conducted. However, it is precisely in such low income resource constrained developing country settings that such evidence is most critical.

In order to improve health outcomes such as birthweight and low birthweight, policy-makers thus have two choices: implementing performance-based financial incentive programmes and incurring additional costs of directly incentivising the quality of healthcare provision or using more cost-effective unconditional financial incentives to achieve comparable improvements in such health outcomes. The above insights, while emanating from a programme implemented during the Millennium Development Goals (MDGs) period with the aim of achieving the original maternal and child health MDGs are quite relevant to the post-MDG development agenda represented by the Sustainable Development Goals (SDGs).

## Bibliography

- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics* 97(4), 543–569.
- Alkema, L., D. Chou, D. Hogan, S. Zhang, A.-B. Moller, A. Gemmill, D. M. Fat, T. Boerma, M. Temmerman, C. Mathers, et al. (2016). Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the un maternal mortality estimation inter-agency group. *The Lancet* 387(10017), 462–474.
- Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011). Inside the war on poverty: The impact of food stamps on birth outcomes. *The Review of Economics and Statistics* 93(2), 387–403.
- Amarante, V., M. Manacorda, E. Miguel, and A. Vigorito (2016). Do cash transfers improve birth outcomes? evidence from matched vital statistics, program, and social security data. *American Economic Journal: Economic Policy* 8(2), 1–43.
- Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics* 10, 439–463.
- Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review* 110(5), 1355–94.
- Ashraf, N., O. Bandiera, and B. K. Jack (2014). No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics* 120, 1–17.
- Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21(1), 1–42.
- Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014). Conditional, unconditional and everything in between: A systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness* 6(1), 1–43.
- Barber, S. L. and P. J. Gertler (2008). The impact of mexico’s conditional cash transfer programme, oportunitades, on birthweight. *Tropical Medicine and International Health* 13(11), 1405–1414.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. Soucat, J. Sturdy, and C. M. Vermeersch (2011). Effect on maternal and child health services in rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet* 377(9775), 1421–1428.
- Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies* 70(3), 489–520.

- Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *American Economic Review* 95(3), 616–636.
- Besley, T. J. and M. Ghatak (2014). Solving agency problems: Intrinsic motivation, incentives, and productivity. Technical report.
- Bonfrer, I., E. Van de Poel, and E. Van Doorslaer (2014). The effects of performance incentives on the utilization and quality of maternal and child care in burundi. *Social Science & Medicine* 123, 96–104.
- Bowles, S. and S. Polania-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature* 50(2), 368–425.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies, and D. Wiesen (2013). *How Effective are Pay-for-Performance Incentives for Physicians?—A Laboratory Experiment*. Number 413. Ruhr Economic Papers.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz, and D. Wiesen (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization* 131, 17–23.
- Cabrera, J. M. and D. Webbink (2020). Do higher salaries yield better teachers and better student outcomes? *Journal of Human Resources* 55(4), 1222–1257.
- Callaway, B. and P. H. Sant’Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.
- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal* 17(2), 372–404.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal* 14(4), 909–946.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6), 2295–2326.
- Cardona-Sosa, L. and C. Medina (2017). The effects of in utero programs on birth outcomes: The case of buen comienzo. *Journal of the Latin American and Caribbean Economic Association* 17(2), 93–134.
- Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal* 18(1), 234–261.
- Cattaneo, M. D., M. Jansson, and X. Ma (2019). lpdensity: Local polynomial density estimation and inference. *arXiv preprint arXiv:1906.06529*.
- Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531), 1449–1455.

- Chelwa, G., M. Pellicer, and M. Maboshe (2019). Teacher pay and educational outcomes: Evidence from the rural hardship allowance in zambia. *South African Journal of Economics* 87(3), 255–282.
- Chibuye, B., P. Hangoma, and M. Khadjavi (2021). The dark side of financial incentives in a health setting - experimental evidence from zambia. Working paper, Christian-Albrechts-University of Kiel.
- Chibwasha, C. J., A. Zanolini, M. Smid, B. Vwalika, M. P. Kasaro, M. Mwanahamuntu, J. S. Stringer, and E. M. Stringer (2016). Predictors and outcomes of low birth weight in lusaka, zambia. *International Journal of Gynecology & Obstetrics* 134(3), 309–314.
- Cross, S. S. and M. Baernholdt (2017). Identifying factors for worker motivation in zambia’s rural health facilities. *Journal of Health Care for the Poor and Underserved* 28(1), 63–70.
- Currie, J. and D. Almond (2011). Human capital development before age five. In *Handbook of Labor Economics*, Volume 4, pp. 1315–1486. Elsevier.
- Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics* 128(3), 1169–1218.
- Das, J. and J. Hammer (2014). Quality of primary care in low-income countries: Facts and economics. *Annual Review of Economics* 6(1), 525–553.
- de Bekker-Grob, E. W., M. Ryan, and K. Gerard (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21(2), 145–172.
- De Chaisemartin, C. and X. d’Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9), 2964–96.
- De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2018). Double for nothing? experimental evidence on an unconditional teacher salary increase in indonesia. *The Quarterly Journal of Economics* 133(2), 993–1039.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18(1), 105–128.
- Deci, E. L., R. Koestner, and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125(6), 627–693.
- Deserranno, E. (2019). Financial incentives as signals: Experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics* 11(1), 277–317.
- Díaz, J.-J. and V. Saldarriaga (2019). Encouraging use of prenatal care through conditional cash transfers: Evidence from juntos in peru. *Health Economics* 28(9), 1099–1113.

- Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics* 27(6), 1436–1450.
- Dussault, G. and M. C. Franceschini (2006). Not enough there, too many here: Understanding geographical imbalances in the distribution of the health workforce. *Human Resources for Health* 4(1), 1–16.
- Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics* 24(1), 211–223.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics* 108(2), 437–459.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review* 42(1), 1–34.
- Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy* 21(3), 673–705.
- Ferrinho, P., S. Siziya, F. Goma, and G. Dussault (2011). The human resource for health situation in zambia: Deficit and maldistribution. *Human Resources for Health* 9(1), 1–11.
- Figueiredo Walter, T. (2020). Misallocation in the public sector? cross-country evidence from two million primary schools. Technical report, STICERD Working Paper No. EOPP 70.
- Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. *Handbook of Economic Field Experiments* 2, 467–514.
- Friedman, J., J. N. Qamruddin, C. Chansa, and A. K. Das (2016). Impact evaluation of zambia's health results-based financing pilot project. Technical report, The World Bank.
- Gage, A. and S. Bauhoff (2021). The effects of performance-based financing on neonatal health outcomes in burundi, lesotho, senegal, zambia and zimbabwe. *Health Policy and Planning*.
- Gertler, P. and C. Vermeersch (2012). *Using Performance Incentives to Improve Health Outcomes*. The World Bank.
- Gneezy, U. and J. A. List (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5), 1365–1384.
- Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3), 791–810.
- Godager, G. and D. Wiesen (2013). Profit or patients' health benefit? exploring the heterogeneity in physician altruism. *Journal of Health Economics* 32(6), 1105–1116.

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.
- Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization* 106, 367–378.
- Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.
- Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature* 42(4), 1009–1055.
- Hennig-Schmidt, H., R. Selten, and D. Wiesen (2011). How payment systems affect physicians’ provision behaviour — an experimental investigation. *Journal of Health Economics* 30(4), 637–646.
- Herbst, C. H., M. Vledder, K. Campbell, M. Sjöblom, and A. Soucat (2011). The human resources for health crisis in zambia. *Washington (DC): World Bank*.
- Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.* 7, 24.
- Hoynes, H., D. Miller, and D. Simon (2015). Income, the earned income tax credit and infant health. *American Economic Journal: Economic Policy* 7(1), 172–211.
- Hoynes, H., M. Page, and A. H. Stevens (2011). Can targeted transfers improve birth outcomes?: Evidence from the introduction of the wic program. *Journal of Public Economics* 95(7-8), 813–827.
- Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635.
- James, N., K. Lawson, and Y. Acharya (2020). Evidence on result-based financing in maternal and child health in low-and middle-income countries: A systematic review. *Global Health Research and Policy* 5(1), 1–15.
- Kaarboe, O. and L. Siciliani (2011). Multi-tasking, quality and pay for performance. *Health Economics* 20(2), 225–238.
- Kesternich, I., H. Schumacher, and J. Winter (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? *Journal of Public Economics* 131, 1–11.
- Kramer, M. S. (1987). Determinants of low birth weight: Methodological assessment and meta-analysis. *Bulletin of the World Health Organization* 65(5), 663.
- Kramer, M. S. (2003). The epidemiology of adverse pregnancy outcomes: An overview. *The Journal of Nutrition* 133(5), 1592S–1596S.

- Lagarde, M. and D. Blaauw (2014). Pro-social preferences and self-selection into jobs: Evidence from south african nurses. *Journal of Economic Behavior & Organization* 107, 136–152.
- Lagarde, M. and D. Blaauw (2017). Physicians’ responses to financial and social incentives: A medically framed real effort experiment. *Social Science & Medicine* 179, 147–159.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review* 90(5), 1346–1361.
- Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281–355.
- Lemiere, C., C. Herbst, N. Jahanshahi, and E. Smith (2011). Reducing geographical imbalances of health workers in sub-saharan africa: A labor market perspective on what works, what does not, and why. world bank; 2011.
- Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2019). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly Journal of Economics* 134(3), 1627–1673.
- McCoy, D., S. Bennett, S. Witter, B. Pond, B. Baker, J. Gow, S. Chand, T. Ensor, and B. McPake (2008). Salaries and incomes of health workers in sub-saharan africa. *The Lancet* 371(9613), 675–681.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2), 698–714.
- Miller, G. and K. S. Babiarz (2013). Pay-for-performance incentives in low-and middle-income country health programs. Technical report, National Bureau of Economic Research.
- Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, and S. Rozelle (2012). Effectiveness of provider incentives for anaemia reduction in rural china: A cluster randomised trial. *Bmj* 345.
- Mullen, K. J., R. G. Frank, and M. B. Rosenthal (2010). Can you get what you pay for? pay-for-performance and the quality of healthcare providers. *The Rand Journal of Economics* 41(1), 64–91.
- Muralidharan, K. and N. Prakash (2017). Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics* 9(3), 321–50.
- Neilson, C., M. Bobba, T. Ederer, G. Leon-Ciliotta, and M. Nieddu (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in peru. Working paper, National Bureau of Economic Research.
- Ohlsson, A. and P. Shah (2008). *Determinants and Prevention of Low Birth Weight: A Synopsis of the Evidence*. Institute of Health Economics.

- Olken, B. A., J. Onishi, and S. Wong (2014). Should aid reward performance? evidence from a field experiment on health and education in indonesia. *American Economic Journal: Applied Economics* 6(4), 1–34.
- Paul, E., M. L. Dramé, J.-P. Kashala, A. E. Ndema, M. Kounnou, J. C. Aïssan, and K. Gyselincx (2018). Performance-based financing to strengthen the health system in benin: Challenging the mainstream approach. *International Journal of Health Policy and Management* 7(1), 35.
- Pega, F., S. Y. Liu, S. Walter, R. Pabayo, R. Saith, and S. K. Lhachimi (2017). Unconditional cash transfers for reducing poverty and vulnerabilities: Effect on use of health services and health outcomes in low-and middle-income countries. *Cochrane Database of Systematic Reviews* (11).
- Prust, M. L., A. Kamanga, L. Ngosa, C. McKay, C. M. Muzongwe, M. T. Mukubani, R. Chihinga, R. Misapa, J. W. van den Broek, and N. Wilmink (2019). Assessment of interventions to attract and retain health workers in rural zambia: A discrete choice experiment. *Human Resources for Health* 17(1), 26.
- Pugatch, T. and E. Schroeder (2014). Incentives for teacher relocation: Evidence from the gambian hardship allowance. *Economics of Education Review* 41, 120–136.
- Sappington, D. E. (1991). Incentives in principal-agent relationships. *Journal of Economic Perspectives* 5(2), 45–66.
- Shen, G. C., H. T. H. Nguyen, A. Das, N. Sachingongu, C. Chansa, J. Qamruddin, and J. Friedman (2017). Incentives to change: Effects of performance-based financing on health workers in zambia. *Human Resources for Health* 15(1), 20.
- Sherry, T. B. (2016). A note on the comparative statics of pay-for-performance in health care. *Health Economics* 25(5), 637–644.
- Sherry, T. B., S. Bauhoff, and M. Mohanan (2017). Multitasking and heterogeneous treatment effects in pay-for-performance in healthcare: Evidence from rwanda. *American Journal of Health Economics* 3(2), 192–226.
- Singh, P. (2015). Performance pay and information: Reducing child undernutrition in india. *Journal of Economic Behavior & Organization* 112, 141–163.
- Singh, P. and W. A. Masters (2020). Performance bonuses in the public sector: Winner-take-all prizes versus proportional payments to reduce child malnutrition in india. *Journal of Development Economics* 146, 102–295.
- United Nations, U. (2016). Transforming our world: The 2030 agenda for sustainable development.
- Zambia Ministry of Health, M. (2011). National health strategic plan (2011 - 2015). Technical report, Government of the Republic of Zambia.

Zambia Ministry of Health, M. (2017). National health sector strategic plan. Technical report, Government of the Republic of Zambia., Lusaka.

Zambia Ministry of Health, Z. S. A. and M. IFC (2019). Zambia demographic and health survey 2018. Technical report, Zambia Statistics Agency, Ministry of Health, and ICF, Lusaka, Zambia, and Rockville, Maryland, USA.

Zambia Ministry of National Development Planning, M. (2017). Seventh national development plan (sndp). Technical report, Zambia Ministry of National Development Planning.

# Appendices

### 3.A Appendix to Chapter Three

### 3.A.1 Missing Birthweight Observations

Table 3.A.1: Missing Birthweight Observations by Treatment

Panel A	Treatment			(4)
	(1) Cons. Control	(2) TU	(3) TP	
Missing Birthweight	.66 (77,716)	.094 (11,626)	.096 (11,950)	
Panel B	Treatment			P-value (TU-TP)
Missing Birthweight (Year)	Cons. Control	TU	TP	
2014	.82 (37,083)	.83 (6,037)	.82 (5,923)	.773
2007	.84 (16,788)	.88 (2,336)	.87 (2,244)	.982
2001	.87 (17,823)	.94 (3,060)	.93 (2,923)	.982
1996	.87 (20,310)	.92 (1,885)	.93 (2,624)	.598

*Notes:* This table analyses the proportion of missing birthweight observations by consolidated control, unconditional (TU) and performance-based (TP) treatments, first for the entire sample in panel A and then by survey year in panel B (the number of missing observations are in parentheses). Panel A of the table reveals that birthweight observations are equally likely to be missing in TU and TP treatments (columns 2 and 3). Similarly, panel B reveals that birthweight observations are also equally likely to be missing in TU and TP for each of the 4 survey years presented. This is clearly reflected by the p-values in column 4 of panel B. The p-values are generated from a linear regression of a missing birthweight indicator on a treatment and survey year indicator interaction and district fixed effects (standard errors are clustered at the district level). The p-values are generated from a t-test of the significance of the difference between the TP and TU coefficients and reveal that the null hypothesis that the difference between the proportion of missing birthweight observations in TU relative to the consolidated control districts, is not significant to that between TP and consolidated control districts. This implies that the missing birthweight observations are unlikely to result in bias in the main analysis.

### 3.A.2 Original ZRBF Districts

Table 3.A.2: Baseline (2007) Means and Changes in Sample Characteristics for the 30 Original ZRBF Districts

Variable	TU	TP	Cons. Control	P-value: $\Delta$ TP = $\Delta$ TU
Low birth weight	0.09 (0.29)	0.09 (0.29)	0.10 (0.30)	0.956
Birth weight (Grams)	3368.45 (746.53)	3243.51 (636.09)	3199.37 (660.84)	0.380
Multiple-births	0.03 (0.16)	0.05 (0.22)	0.03 (0.18)	0.046
Child aged U-3	0.26 (0.44)	0.28 (0.45)	0.26 (0.44)	0.085
Child Female	0.51 (0.50)	0.51 (0.50)	0.52 (0.50)	0.673
Birth Order	3.39 (2.23)	3.40 (2.29)	3.37 (2.30)	0.96
Mothers age	34.89 (8.15)	34.49 (8.06)	34.40 (8.14)	0.245
Numb. ANC visits	4.08 (1.52)	4.07 (1.54)	3.78 (1.69)	0.437
Numb. HH members	6.26 (2.29)	6.59 (2.88)	6.73 (2.78)	0.000
Numb. Children aged U-5	1.51 (0.97)	1.60 (1.02)	1.55 (1.02)	0.044
Improved water	0.23 (0.42)	0.27 (0.44)	0.37 (0.48)	0.688
Female HH head	0.24 (0.43)	0.20 (0.40)	0.20 (0.40)	0.965
Age of HH head	41.79 (10.90)	41.81 (11.45)	41.31 (11.13)	0.950
Urban residence	0.19 (0.39)	0.13 (0.34)	0.21 (0.41)	0.272

*Notes:* This table shows means for TP, TU and control districts for sample characteristics at baseline (2007) for the 30 (10 per treatment) Original Zambia Results Based Financing (ZRBF) districts. The p-values in column 4 are generated from linear regressions with each respective variable in the row as the outcome variable regressed on a treatment indicator and year indicator interaction and district fixed effects (standard errors are clustered at the district level). The categorical variables not shown in the table are; mothers highest level of education (primary across all treatments and control) and whether the woman has currently/formerly or has never been in a marital union (currently in a union across treatments and control). Birth order refers to the order in which a woman's children are born, beginning from her first birth.

Table 3.A.3: Difference-in-Differences Birthweight and Low Birthweight Estimates: 30 Original ZRBF Districts.

	(1)	(2)	(3)	(4)	(5)	(6)
	birth weight	birth weight	birth weight	Low birth weight	Low birth weight	Low birth weight
<b>Unconditional</b> $\times$ <b>Int</b>	41.48 (71.33)	38.49 (81.11)	145.3 (106.5)	0.0206 (0.0235)	0.00172 (0.0310)	-0.0952** (0.0445)
<b>Performance-based</b> $\times$ <b>Int</b>	42.80 (64.46)	10.36 (76.98)	118.6 (76.46)	0.00837 (0.0184)	0.00493 (0.0235)	-0.100*** (0.0323)
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Sub(Educ)	Sub(Age)	Full	Sub(Educ)	Sub(Age)
Control Group Mean	3093.6*** (86.53)	3301.6*** (113.7)	3170.7*** (96.53)	0.115*** (0.0333)	0.102** (0.0387)	0.124** (0.0481)
Observations	4119	2765	2999	4119	2765	2999
$R^2$	0.048	0.055	0.076	0.022	0.030	0.043

*Notes:* The table shows DD regressions of the main and heterogeneous impacts of unconditional and performance-based financial incentives on infant birthweight and low birthweight for singleton births, for children in the in the 30 Original Zambia Results Based Financing (ZRBF) unconditional and performance-based, relative to control districts. Infant birthweight is measured in grams, while low birthweight is an indicator variable for infant birthweight less than 2,500 grams. Each column shows a separate regression with some coefficients omitted to conserve space. The intensity-post variable is zero for all children born before treatment (before April 2012), 1 for all children born 9 months or more after the start of treatment and is set to equal the fraction of the in-utero (9 month) period, (potentially) spent under treatment for children born within 9 months of the start of treatment. Heterogeneous effects are assessed based on mothers mothers education in column 3 (no education and only primary education) and mothers age in column 4 (mothers  $\leq 20$  and those  $> 35$  years old). The preferred and most conservative specification, including both district and birth cohort fixed effects, is used in all regressions 2007, the baseline year for the analysis, is also the reference year in the regressions. Standard errors, clustered at the district level, are in parentheses;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Table 3.A.4: Difference-in-Differences Full Sample Estimates for Ante-natal Care Visits: 30 Original ZRBF Districts.

	(1)	(2)	(3)	(4)	(5)
	num. ANC visits	timing 1st ANC	num. tetanus injections	weighed	blood pressure
<b>Unconditional</b> $\times$ <b>Int</b>	-0.188 (0.153)	0.0559 (0.144)	0.256*** (0.0600)	0.00977 (0.0242)	0.0428 (0.0317)
<b>Performance-based</b> $\times$ <b>Int</b>	0.0282 (0.133)	-0.255 (0.159)	0.228*** (0.0789)	-0.0116 (0.0292)	0.0691** (0.0303)
District fixed effects	Yes	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes	Yes
Control Group Mean	3.353*** (0.134)	4.526*** (0.101)	1.645*** (0.0657)	0.890*** (0.0210)	0.709*** (0.0350)
Observations	7216	7019	7261	5658	5657
$R^2$	0.098	0.066	0.075	0.048	0.088

*Notes:* The table shows DD regressions of the number, timing and quality of Ante-natal Care visits in the 30 Original Zambia Results Based Financing (ZRBF) unconditional unconditional and performance-based, relative to control districts. Each column shows a separate DD regression, with some coefficients omitted to conserve space. The intensity (Int) variable is zero for all children born before treatment (before April 2012), 1 for all infants born 9 months or more after the start of treatment and is set to equal the fraction of the in-utero (9 month) period, spent under treatment for children born within 9 months of the start of treatment. The preferred and most conservative specification, including both district and birth cohort fixed effects, is used in all regressions. Consolidated control districts are the reference category. ANC quality is measured by the number of tetanus injections received by a women before giving birth, and dummy variables indicating; whether the woman was weighed and had her blood pressure taken, respectively, during pregnancy. The regressions in columns 4 and 5 contain fewer observations because in the ZDHS 1996, respondents were not asked about being weighed or having their blood pressure taken at ANC. Standard errors, clustered at the district level, are in parentheses.  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### 3.A.3 Triple Difference and Low-Impact Sub-sample Estimates

Table 3.A.5: Triple Difference Birthweight and Low Birthweight Estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
	birth weight	birth weight	birth weight	Low birth weight	Low birth weight	Low birth weight
<b>TU × 2014 × U3</b>	-71.96 (169.3)	120.5 (132.0)	7.830 (264.6)	-0.0139 (0.0682)	-0.125*** (0.0411)	-0.206** (0.102)
<b>TP × 2014 × U3</b>	113.6 (106.6)	244.1** (94.22)	289.8 (250.5)	-0.0177 (0.0354)	-0.0634* (0.0377)	-0.0580 (0.0849)
TU × 2001 × U3	-139.4 (129.7)	49.76 (153.3)	205.0 (492.0)	0.0642 (0.0824)	0.0346 (0.0839)	-0.197 (0.287)
TP × 2001 × U3	86.02 (115.0)	250.7 (202.7)	-36.16 (326.6)	-0.0607 (0.0473)	-0.138*** (0.0500)	0.0905 (0.0973)
TU × 1996 × U3	93.09 (187.1)	159.3 (216.1)	-316.3 (196.4)	-0.0553 (0.103)	-0.140 (0.120)	-0.111 (0.0772)
TP × 1996 × U3	70.34 (152.1)	3.377 (221.6)	24.97 (230.1)	0.0558 (0.0449)	0.0973* (0.0490)	0.129 (0.111)
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Control Group Mean	3046.7*** (79.39)	3129.2*** (111.8)	2988.3*** (104.0)	0.0978*** (0.0356)	0.0798* (0.0457)	0.227*** (0.0587)
Observations	15365	9504	4409	15365	9504	4409

*Notes:* The table shows DDD regressions of the main and heterogeneous impacts of performance-based and unconditional financial incentives on infant birthweight (columns 1-3) and low birthweight (columns 4-6) for singleton births, for children aged under 3 years old in unconditional and performance-based districts, relative to children aged above 3 years old in consolidated control districts. Each column contains a separate DDD regression (some coefficients are omitted to conserve space). Heterogeneous effects are assessed based on mothers education in columns 2 and 5 (no education and only primary education) and mothers age in columns 3 and 6 (mothers  $\leq 20$  and those  $> 35$  years old). 2007, the baseline year for the analysis, is also the reference year in the regressions. Standard errors, clustered at the district level, are in parentheses;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Table 3.A.6: Difference-in-Differences Birthweight and low Birthweight Heterogeneous Estimates: Low Impact Sub-samples.

	(1)	(2)	(3)	(4)
	birth weight	birth weight	Low birthweight	Low birthweight
<b>Unconditional</b> $\times$ <b>Int</b>	58.96 (95.84)	73.15* (42.38)	-0.00837 (0.0281)	-0.0256 (0.0167)
<b>Performance-based</b> $\times$ <b>Int</b>	70.82** (33.03)	40.67 (30.99)	-0.0344 (0.0213)	-0.0292** (0.0119)
District fixed effects	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes
Sample	Sub(Educ)	Sub(Age)	Sub(Educ)	Sub(Age)
Control Group Mean	2967.8*** (85.29)	3110.1*** (46.66)	0.126*** (0.0373)	0.131*** (0.0205)
Observations	6160	15734	6160	15734
$R^2$	0.062	0.049	0.032	0.016

*Notes:* The table shows DD regressions of heterogeneous impacts of unconditional and performance-based financial incentives on infant birthweight and low birthweight for singleton births, for children in unconditional and performance-based districts, relative to consolidated control districts. Infant birthweight is measured in grams, while low birthweight is an indicator variable for infant birthweight less than 2,500 grams. Each column shows a separate regression with some coefficients omitted to conserve space. The intensity-post variable is zero for all children born before treatment (before April 2012), 1 for all children born 9 months or more after the start of treatment and is set to equal the fraction of the in-utero (9 month) period, (potentially) spent under treatment for children born within 9 months of the start of treatment. Heterogeneous effects are assessed based on mothers education in columns 1 and 3 (secondary education or higher) and mothers age in columns 2 and 4 (mothers > 20 and those < 35 years old). The preferred and most conservative specification, including both district and birth cohort fixed effects, is used in all regressions. Standard errors, clustered at the district level, are in parentheses.;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

### 3.A.4 High-Impact Sub-sample ANC Estimates

Table 3.A.7: Difference-in-Differences Estimates for Ante-natal Care Visits: Women with No or Only a Primary Education

	(1)	(2)	(3)	(4)	(5)
	num. ANC visits	timing 1st ANC	num. tetanus injections	weighed	blood pressure
<b>Unconditional</b> $\times$ <b>Int</b>	0.226 (0.169)	-0.0940 (0.130)	0.141** (0.0605)	0.0280 (0.0271)	0.0490 (0.0411)
<b>Performance-based</b> $\times$ <b>Int</b>	0.399*** (0.140)	-0.226 (0.167)	0.139 (0.0983)	0.00329 (0.0305)	0.0773** (0.0313)
District fixed effects	Yes	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes	Yes
Control Group Mean	3.763*** (0.0975)	4.398*** (0.0722)	1.744*** (0.0436)	0.889*** (0.0142)	0.686*** (0.0223)
Observations	17488	16998	17608	11622	11617
$R^2$	0.112	0.073	0.077	0.053	0.097

*Notes:* The table shows DD regressions of the number, timing and quality of Ante-natal Care visits in unconditional and performance-based districts, relative to consolidated control districts, for the high impact sub-sample of women either with no education or only a primary education. Each column shows a separate regression with some coefficients omitted to conserve space). The intensity-post variable is zero for all children born before treatment (before April 2012), 1 for all infants born 9 months or more after the start of treatment and is set to equal the fraction of the in-utero (9 month) period, spent under treatment for children born within 9 months of the start of treatment. ANC quality is measured by the number of tetanus injections received by a women before giving birth, and dummy variables indicating; whether the woman was weighed, had her blood pressure taken, was told about complications and received iron tablets/syrup, respectively, during pregnancy. The regressions in columns 4 and 5 contain fewer observations because in the ZDHS 1996 and 2001, respondents were not asked about being weighed or having their blood pressure taken at ANC. Standard errors, clustered at the district level, are in parentheses.  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 3.A.8: Difference-in-Differences Estimates for Ante-natal Care Visits: Women Aged 20 Years Old or Less or Older Than 35 years of Age.

	(1)	(2)	(3)	(4)	(5)
	num. ANC visits	timing 1st ANC	num. tetanus injections	weighed	blood pressure
<b>Unconditional</b> × <b>Int</b>	0.134 (0.171)	0.0588 (0.150)	0.0783 (0.0831)	0.0587** (0.0292)	-0.00326 (0.0555)
<b>Performance-based</b> × <b>Int</b>	0.410*** (0.147)	-0.201 (0.144)	0.203* (0.112)	-0.00485 (0.0361)	0.0377 (0.0410)
District fixed effects	Yes	Yes	Yes	Yes	Yes
Birth Year fixed effects	Yes	Yes	Yes	Yes	Yes
Control Group Mean	3.287*** (0.166)	4.707*** (0.0960)	1.750*** (0.0605)	0.861*** (0.0176)	0.726*** (0.0263)
Observations	7668	7454	7706	5563	5561
$R^2$	0.140	0.076	0.091	0.055	0.101

*Notes:* The table shows DD regressions of the number, timing and quality of Ante-natal Care visits in unconditional and performance-based districts, relative to consolidated control districts, for the high impact sub-sample of women aged 20 years old or less or older than 35 years of age. Each column shows a separate regression with some coefficients omitted to conserve space). The intensity-post variable is zero for all children born before treatment (before April 2012), 1 for all infants born 9 months or more after the start of treatment and is set to equal the fraction of the in-utero (9 month) period, spent under treatment for children born within 9 months of the start of treatment. The preferred and most conservative specification, including both district and birth cohort fixed effects, is used in all regressions. 2007, the baseline year, is also the reference year in all regressions. Consolidated control districts are the reference category. ANC quality is measured by the number of tetanus injections received by a women before giving birth, and dummy variables indicating; whether the woman was weighed, had her blood pressure taken, was told about complications and received iron tablets/syrup, respectively, during pregnancy. The regressions in columns 4 and 5 contain fewer observations because in the ZDHS 1996 and 2001, respondents were not asked about being weighed or having their blood pressure taken at ANC. Standard errors, clustered at the district level, are in parentheses.  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### 3.A.5 Difference-in-Difference Birthweight and Low Birthweight Estimates: District by Year Time-Trends

Table 3.A.9: Difference-in-Differences Birthweight and Low Birthweight Estimates: District by Year Fixed Effects.

	(1)	(2)	(3)	(4)	(5)	(6)
	birth weight	birth weight	birth weight	Low birth weight	Low birth weight	Low birth weight
<b>Unconditional</b> × <b>Int</b>	38.12 (39.08)	106.5*** (36.84)	231.4*** (77.90)	0.00194 (0.0185)	-0.0414** (0.0159)	-0.0908*** (0.0319)
<b>Performance-based</b> × <b>Int</b>	58.74* (34.30)	76.23** (37.86)	229.3*** (83.28)	-0.00978 (0.0158)	-0.0352** (0.0150)	-0.111*** (0.0349)
District by year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Sub(Educ)	Sub(Age)	Full	Sub(Educ)	Sub(Age)
Control Group Mean	3231.7*** (38.97)	3437.8*** (51.30)	2639.3*** (64.43)	0.0417** (0.0158)	0.0269 (0.0238)	0.152*** (0.0303)
Observations	15365	9504	4409	15365	9504	4409
$R^2$	0.055	0.056	0.091	0.019	0.024	0.048

*Notes:* The table shows DD regressions of the main and heterogeneous impacts of unconditional and performance-based financial incentives on infant birthweight and low birthweight for singleton births, for children in unconditional and performance-based districts, relative to consolidated control districts. Infant birthweight is measured in grams, while low birthweight is an indicator variable for infant birthweight less than 2,500 grams. Each column shows a separate regression with some coefficients omitted to conserve space. The intensity-post variable is zero for all children born before treatment (before April 2012), 1 for all children born 9 months or more after the start of treatment and is set to equal the fraction of the in-utero (9 month) period, (potentially) spent under treatment for children born within 9 months of the start of treatment. Heterogeneous effects are assessed based on mothers education in column 3 (no education and only primary education) and mothers age in column 4 (mothers  $\leq 20$  and those  $> 35$  years old). All regressions include district by year fixed effects. 2007, the baseline year for the analysis, is also the reference year in the regressions. Standard errors, clustered at the district level, are in parentheses.;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

### 3.A.6 Running Variable Density

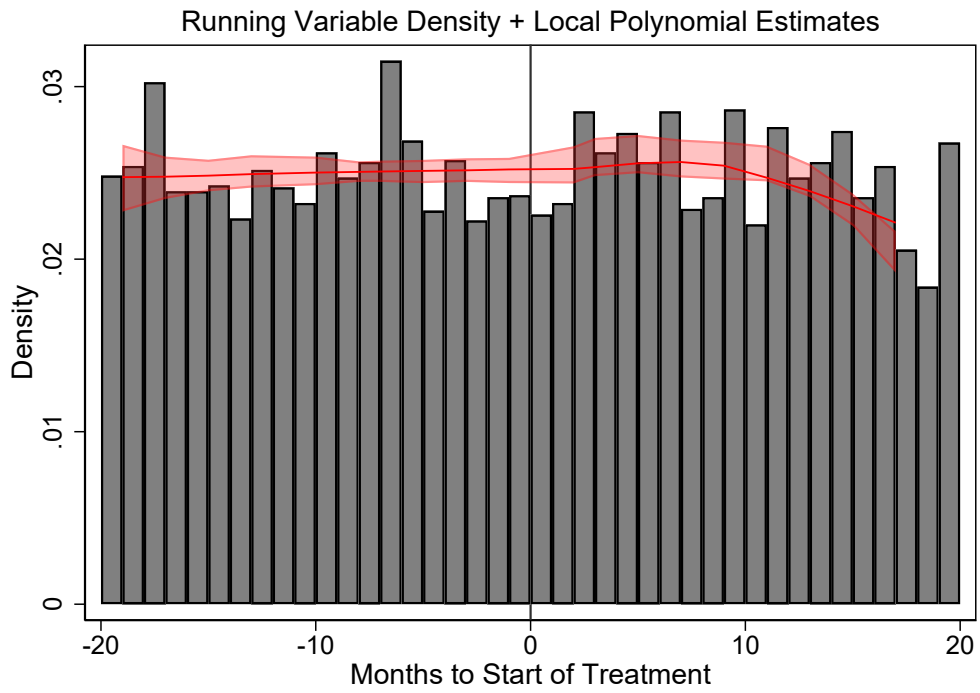


Figure 3.A.1: Density of Running Variable at Treatment Threshold.

## Chapter 4

# Public Sector Compensation and Rural-Urban Gaps in Service Delivery: Evidence from Zambia

## **Abstract**

This paper studies the effect of unconditional wage increases in rural health facilities on staffing levels. First, we document large differences in access to and utilization of health services between rural and urban areas in the context of our study, Zambia. Second, we leverage a spatial discontinuity in the eligibility for a rural hardship allowance to estimate the causal effect of the resulting wage increase on health facility staffing levels. We find that the unconditional wage increase of 20% induced by the allowance leads to a 51% increase in the number of health workers in health facilities eligible to receive it. Our findings thus illustrate the poignancy of unconditional hardship allowance schemes as a policy tool in addressing inequalities in access to high quality health workers between rural and urban areas in low income developing country settings.

*Keywords:* Inequality, Health, Financial Incentives, Development.

*JEL Codes:* I1, I14, D04.

## 4.1 Introduction

Rural areas of developing countries are often characterised by low levels of access to public services. A particular concern is the absence of qualified personnel for service delivery in these areas. Both the quantity and the quality of key service delivery agents, such as health workers and teachers, is typically much lower than in urban areas.<sup>1</sup> While the rigidity of the wage structure in the public sector limits the ability of governments to attract workers to locations with less desirable working conditions, hardship allowance schemes have emerged as a common policy tool to address staffing inequalities. These schemes offer unconditional wage supplements for working in rural areas. Despite their popularity, evidence on their effectiveness remains limited, especially in the health sector.<sup>2</sup>

This paper studies the effect of a rural hardship allowance scheme on health facility staffing in Zambia. First, we document large differences in access to and utilization of healthcare services between urban and rural areas. Second, we exploit a spatial discontinuity in the eligibility for the hardship allowance to estimate its causal effect on facility staffing levels. The allowance scheme implies a 20% higher pay for health workers stationed at facilities more than a predetermined distance from the respective district capital. Comparing facilities on either side of the eligibility threshold, we find that the allowance leads to a 51% increase in the number of health workers in eligible facilities.

Poor health outcomes have long been attributed to health worker shortages in Zambia (Ferrinho et al., 2011; Herbst et al., 2011; Zambia Ministry of Health, 2017). Large workforce deficits in rural areas, where some facilities have no qualified staff at all, are associated with particularly low levels of health (Zambia Ministry of Health, 2011). As a mitigation strategy, the Zambian government provides an unconditional salary supplement of 20% over the basic salary to health workers stationed in rural areas. This supplement, known as the rural hardship allowance, is paid to workers at health facilities located outside a pre-specified radius from their respective district capital, where the radius depends on the government’s classification of the remoteness of a district.<sup>3</sup> The aim of the allowance is to compensate workers for the inconveniences associated with a lack of modern amenities in far flung rural areas.<sup>4</sup>

This study utilises a spatial fuzzy regression discontinuity design within districts to determine whether the unconditional salary increase induced by the rural hardship allowance, results in an increase in health workers in rural health facilities. More specifically, we estimate the causal effect of the rural hardship allowance on the number of health workers

---

<sup>1</sup>See Dussault and Franceschini (2006) or Lemiere et al. (2011) for a documentation of this phenomenon in the health sector. See Figueiredo Walter (2020) for evidence of staffing inequalities across public primary schools in low-income countries.

<sup>2</sup>We discuss the related literature in detail below.

<sup>3</sup>The radius is larger in more urban districts. Details are discussed in section 4.2.

<sup>4</sup>Health workers have been noted to especially shun rural areas owing to factors such as a general lack of social and economic infrastructure and amenities and fewer opportunities for training and income supplementation in rural areas (McCoy et al., 2008).

stationed at a health facility, comparing staffing levels of facilities on either side of the eligibility threshold. Additionally, we evaluate the impact of the allowance on the quality-adjusted number of health workers, as measured by the total facility wage bill, excluding allowance payments.

To preview our results, we find that the probability of the payment of the hardship allowances increases by 71 percentage points at the eligibility threshold, from a baseline probability of 20% among ineligible facilities near the threshold. Assessing the impact of the allowance on staffing levels among facilities eligible to receive the allowance reveals that the number of health workers increases by 51%, while the quality-adjusted number of health workers increases by 50%. Instrumenting the payment of the allowance with eligibility for the allowance among facilities near the threshold, we estimate that the allowance increases the number of health workers by 80% and the quality-adjusted number of health workers by 86% among eligible facilities that actually receive the allowance.

This paper primarily speaks to an emerging literature on the role of financial incentives for recruitment and retention of public sector employees. Existing evidence is largely in line with our findings, showing that financial incentives can help the state bridge staffing gaps between more and less attractive locations. For example, Dal Bó et al. (2013) randomize wage offers to community development agents across Mexican municipalities and document that higher wages reduce the recruitment gap between more and less developed municipalities. Similarly, hardship allowances have been shown to increase the quality of teachers at less attractive schools in multiple contexts, notably in Peru (Neilson et al., 2021), the Gambia (Pugatch and Schroeder, 2014) and Uruguay (Cabrera and Webbink, 2020). In Zambia, however, Chelwa et al. (2019) do not find evidence of a significant effect of the rural hardship allowance scheme on school staffing levels and teacher characteristics in rural schools. In contrast to the existing literature, this paper focuses on the health sector and documents a substantial effect of higher wages on staffing levels in rural health facilities - even though the examined allowance scheme does not appear to succeed at closing staffing gaps between urban and rural schools in the same setting.

A second strand of literature this paper is related to is concerned with health workers' preferences over job attributes, typically leveraging hypothetical discrete choice experiments to derive wage differentials required to compensate for unattractive job attributes (de Bekker-Grob et al., 2012). Of particular relevance for this paper is the work by Prust et al. (2019) in the Zambian context. They elicit preferences over rural and urban positions while varying other job attributes, including salary (a base salary and a 20%, 25% and 30% rural allowance, respectively), among 474 current and future health workers. The results from their experiment suggest that hardship allowances can increase the acceptance of rural positions. The hypothetical and non-incentivised nature of participant responses as well as the underlying convenience sample of health workers and students, however, cast doubt on the external validity of their quantitative estimates for actual job take up decisions among the population of health workers. This paper circumvents these concerns

by leveraging administrative data from the universe of public health workers in Zambia to study de facto health facility staffing levels.

The remainder of this paper is organised as follows. Section 4.2 provides background information on the Zambian healthcare system and the rural hardship allowance scheme. Section 4.3 outlines the data sources used in this paper and provides descriptive statistics. Section 4.4 explains the estimation strategy, while section 4.5 presents the results. Robustness checks are documented in section 4.6 and section 4.7 concludes.

## 4.2 Background

### 4.2.1 The Rural-Urban Gap in Health Service Provision

Despite Zambia being listed as one of the most urbanised countries in Sub-Saharan Africa, the majority of the country's population, approximately 60%, still reside in rural areas. However, it is rural areas that continue to suffer the adverse effects of a rural 'development deficit' that encompasses a lack of access to basic economic infrastructure, such as roads and electricity, and also extends to basic social services (Zambia Ministry of National Development Planning, 2017). This is well exemplified by rural-urban inequalities in access to health services, which remain large. Figure 4.2.1, which utilises data from the 2018 Demographic and Health Survey (DHS), illustrates these inequalities (Zambia Ministry of Health and IFC, 2019). Panel A of the figure shows the proportion of urban and rural women respectively, who report having access to a skilled health worker at delivery, where a skilled health worker is defined as either a Medical Doctor, Nurse, Midwife, Clinical Officer or Community Health Assistant.<sup>5</sup> Similarly, panel B of figure 4.2.1 shows the proportion of urban and rural women respectively, that report delivering their babies from a health facility. Panel A reveals that while 93% of urban women report having been assisted by a skilled health worker at delivery, only 75% of rural women report having similar access to a skilled health worker at delivery. Similarly, panel B reveals that while 93% of urban women report delivering their babies in a health facility, only 78% of rural women report having similar access to a health facility at delivery. The relatively limited access to health services in rural areas implies that rural women are forced to develop coping strategies, which may include turning to informal channels. This is illustrated in figure 4.A.1 in annex 4.A.1, which reveals that 21% of rural women report having been assisted by a friend, relative or traditional birth attendant during delivery. In the case of urban women, however, only 6% report having received similar assistance during the process of delivering their babies.

The above exercise, while being only descriptive and illustrative, is certainly poignant. This is because it serves to highlight the nature of the rural-urban inequalities in access to

---

<sup>5</sup>Sustainable Development Goal (SDG) 3, aims to, "Ensure healthy lives and promote well-being for all at all ages." Increasing the proportion of births attended by skilled healthcare workers is among the key indicators for achieving this goal (United Nations, 2016).

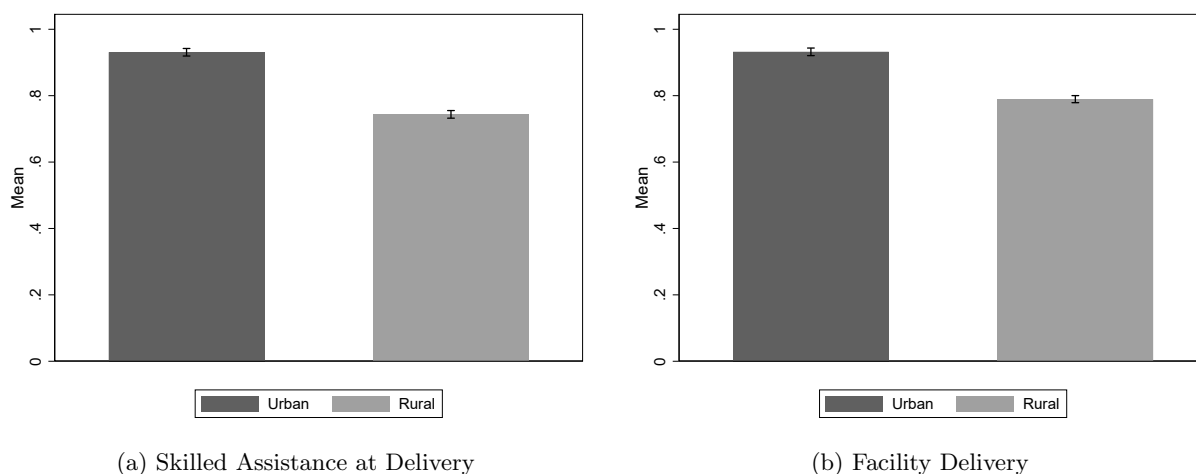


Figure 4.2.1: Rural-Urban Access to a Skilled Health Worker and Health Facility during Delivery

*Notes:* This figure plots the proportion of women with access to a skilled healthcare worker and to a health facility at delivery in rural and urban areas respectively. A skilled healthcare worker is defined as either a Medical Doctor, Nurse, Midwife, Clinical Officer or Community Health Assistant. The figures shown are the authors calculations based on data from the 2018 Demographic and Health Survey.

health services that the provision of the rural hardship allowance to health workers aims at addressing.

#### 4.2.2 The Zambian Healthcare System

The Zambian healthcare system comprises four tiers or levels of care. The highest tier comprises level-3 hospitals that are at the apex of the health system, providing the highest level of care and catering for the most complex of disease conditions. These are followed by level-2 and then by level-1 hospitals, respectively. Urban and rural health centres and health posts comprise the lowest and most basic level of care, providing mainly health promotion, disease prevention and basic curative services. The health system is based on a referral system, whereby more complex cases, that cannot be handled by health centres and health posts, are referred to level-1, then level-2 and finally, to level-3 hospitals. Hospitals are located in urban areas while health posts are typically located in rural areas. Health centres can be found throughout the country, in both urban and rural areas. As of 2017, Zambia had a total of 2,933 health facilities, comprising eight level-3 hospitals, 34 level-2 hospitals, 99 level-1 hospitals, 1,839 health centres and 953 health posts (Zambia Ministry of Health, 2017).

The government is the main provider of healthcare in Zambia and the largest employer of health workers, owning and operating approximately 90% of health facilities at all levels of the healthcare system under the Ministry of Health (MoH). Health workers are officially employed by the Zambian civil service but are managed by the MoH and are deployed to locations of critical need at recruitment. Upon confirmation of their employment status in the Zambian civil service, health workers, like other civil servants, are eligible to initiate a

transfer process, whereby they can potentially be reassigned to other health facilities (possibly in another location) upon approval of their transfer request. Rural-urban dynamics in this process are such that health workers are first deployed to health facilities in rural areas, independent of their preferences, owing to the greater need in such areas. However, evidence suggests that due to hardship concerns many workers quickly initiate the process of requesting for transfers to health facilities in urban areas (Cross and Baernholdt, 2017; McCoy et al., 2008).

The policy goal of the Zambian MoH is to ensure adequate numbers of well trained health workers, equitably distributed nationwide to meet the health needs of the population. The human resource situation reached crisis levels, with only 43% of the World Health Organisation (WHO) recommended staffing levels of 2.28 health workers per 1,000 of the population, as recently as 2010. In response to this, the National Training Operational Plan and the National Human Resources for Health Strategic plan were implemented between 2011 and 2016. Efforts were focused at numerous levels including opening of new training institutions, training of more health workers and their recruitment into the civil service and the establishment of a Community Health Assistant (CHA) cadre. Despite these efforts, the deficit of health workers still persists with only 68% of approved positions being filled. Rural areas are especially disadvantaged in this regard owing to challenges related to relatively lower health worker retention relative to urban areas (Zambia Ministry of Health, 2017).

### 4.2.3 The Rural Hardship Allowance Scheme

The rural hardship allowance in its current form was streamlined in 2010.<sup>6</sup> It is earmarked for health workers stationed in health facilities outside a pre-specified radius from the capital of their respective district. For this purpose, Zambian districts are classified into four different categories. For the most urban category-A districts, health workers in health facilities at least 30 kilometres (km) away from the district capital, qualify to receive the allowance in the form of an unconditional 20% supplement to their basic salaries. For more rural category-B and C districts, health workers stationed in facilities at least 20 and 25 km, respectively, from their district capital are eligible to receive the same 20% salary supplement. In the case of the most rural category-D districts, health workers in all health facilities are eligible to receive the allowance. Figure 4.2.2 shows the distribution of Zambian districts by category as well as each district's respective capital. Figure 4.2.3 illustrates the distance-to-district-capital eligibility criteria for the case of Chipata district, a category-B district.

The distance of each health facility from its respective district capital and thus eligibility to receive the allowance are determined firstly by the Office of the Surveyor General, a department within the Zambian Ministry of Lands responsible for collecting Global

---

<sup>6</sup>For a further discussion on the background and evolution of the rural hardship allowance, see the account in Chelwa et al. (2019).

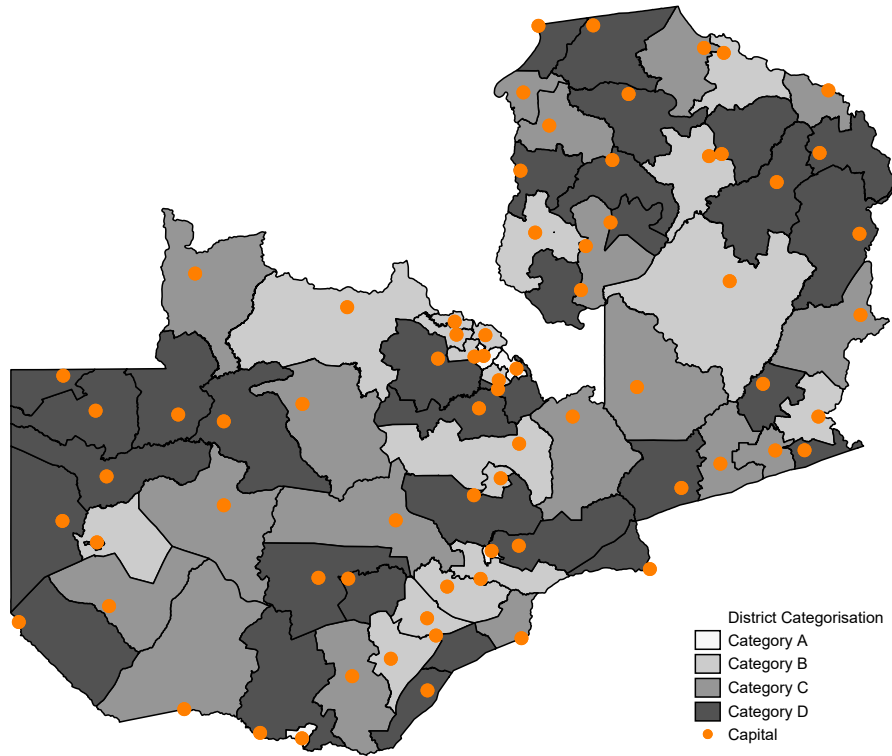


Figure 4.2.2: Categorization of Districts

*Notes:* This figure shows the categorization of Zambian districts for purposes of the rural hardship allowance payment. Different categories are displayed in different shades of grey. District capitals are marked by yellow circles.

Positioning System (GPS) coordinates of health facilities and district capitals. This information is then passed onto the Payroll Management Establishment Control (PMEC) unit at the heart of the Zambian civil service, which is responsible for the management of the payroll for all Zambian civil servants. Eligibility for receipt of the rural hardship allowance is only partly determined by the GPS coordinates and district categorisation, however, as health facilities judged as especially isolated owing to being separated from their respective district capitals by natural barriers such as mountains or rivers, can be re-classified as eligible to receive the allowance in the event they are deemed ineligible, based on GPS coordinates alone (Chelwa et al., 2019). These classification and re-classification dynamics are also evident in figure 4.2.3. In the figure, three health facilities are displayed as receiving the allowance despite falling within the 'urban zone' that is less than 25km from the capital of Chipata district. This illustrates the type of ex-post reclassification discussed above. However, as can be noted in the figure, these re-classifications are the exception as opposed to the rule with health facilities that are outside the urban zone actually being the majority of facilities that receive the rural hardship allowance.

### 4.3 Data

This study utilises several administrative data sets. The first data set is the Japan International Cooperation Agency (JICA) health facility survey conducted in Zambia in 2017.

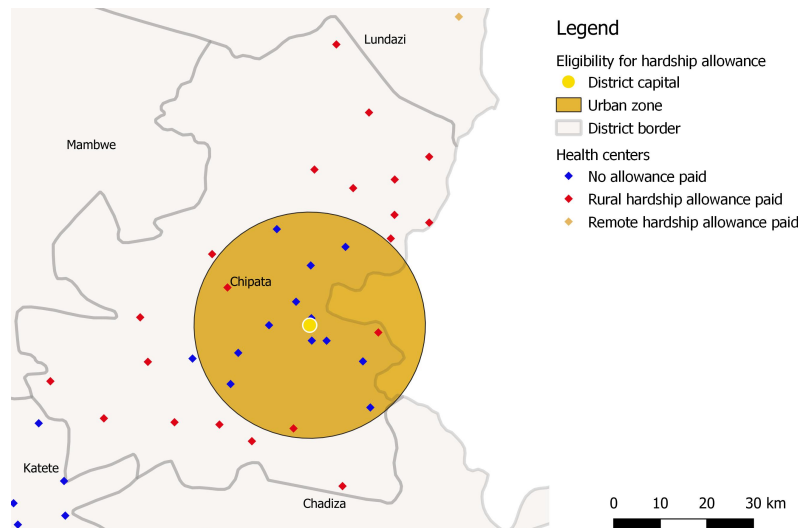


Figure 4.2.3: Rural Hardship Allowance Eligibility and Receipt in Chipata district

*Notes:* This figure illustrates rural hardship allowance eligibility based on the distance-to-district-capital eligibility criteria and actual receipt of the allowance for health facilities in Chipata district, a category-B district.

The JICA facility survey collected data on 2,395 government owned or supported health facilities at all levels of care in Zambia, with the aim of assessing health worker staffing levels in these facilities. As noted in section 4.2, Zambia had a total of 2,933 health facilities as of 2017, with approximately 90% being government owned or supported facilities. This implies that the JICA survey contains data on approximately 82% of government owned or supported health facilities at all levels of care in the country. The JICA survey collected information on such aspects as facility infrastructure, including the number of inpatient and maternal beds and a head count of the catchment population of each facility. Additionally, the number, type, specialisation and gender of health workers stationed at the facility was collected. The types of health workers listed are the following: Adherence Officer, Clinical Officer, Community Health Assistant, Dental Therapist, Elimination of Mother to Child Transmission (EMTCT) officer, Environmental Health Technician, Laboratory Technologist, Medical Doctor, Medical Licentiate, Nurse, Nutritionist, Pharmacist, Pharmacy Technician, Physiotherapist, Psycho-social Councillor, Public Health Inspector or Radiographer. The specialisation of health workers list their speciality. Taking nurses as a case in point, their specializations are listed as, for instance; Nurse-in-charge, Public Health Nurse, Registered Nurse, Registered Diagnostic Nurse, Registered Ophthalmic Nurse, Registered Paediatric Nurse, Registered Theatre Nurse, Registered Nurse Midwife, Zambia Enrolled Nurse, Zambia Enrolled Nurse Psychiatry, Zambia Enrolled Nurse Midwife or Zambia Enrolled Mental Health Nurse. Crucially for the purposes of this study, the JICA survey also collected the GPS coordinates of each health facility.

The second data set is a de-identified individual-level extract from the health sector Zambia civil service payroll from 2014. This unique data set, made available by the Zambian government, contains information on the basic salary, allowances received and

specialisation of Zambian civil service employees.

Combining the payroll data set and the JICA survey permits us to generate two health worker related variables which serve as the main outcomes of this study. First, we simply compute the total number of health workers at each health facility regardless of type or specialisation. Second, we construct a quality-adjusted measure of health workers, the total facility wage bill, by summing up monthly basic salaries (excluding any allowances) received across all health workers within each facility.<sup>7</sup> Assuming that differences in basic salaries reflect differences in the quality of health workers, we can account for differences in the positions and specializations of health workers between facilities this way. Again, taking the case of nurses for illustrative purposes, a nurse listed as a Zambia Enrolled Nurse Psychiatry, a Zambia Enrolled Nurse Midwife or a Zambia Enrolled Mental Health Nurse has a higher qualification and earns a higher salary than a Zambia Enrolled Nurse. Similarly, a Registered Nurse is more qualified than an Enrolled Nurse or any of its specialisations. Among Registered Nurses, in turn, more specialised nurses such as Registered Diagnostic Nurses, Registered Ophthalmic Nurses, Registered Paediatric Nurses, Registered Theatre Nurses or Registered Nurse Midwives are more qualified than Registered Nurses and earn correspondingly higher monthly salaries. In this manner, we thus attach the number, type and specialisation of all health workers listed at each facility in the JICA survey to a monthly salary in the payroll data and compute the total monthly wage bill for each health facility in our data set.

The third data set utilised contains the GPS coordinates of all district capitals in Zambia. Combining the GPS coordinates of the district capitals and the GPS coordinates of each health facility from the JICA facility survey, we compute the running variable, distance between each facility and its respective district capital, measured in kilometres (km). This, in turn, allows us to determine whether or not a given health facility falls on the eligible side of the threshold to receive the rural hardship allowance, based on the eligibility criteria discussed in section 4.2.

The fourth data set utilised is a list of all health facilities actually receiving the rural hardship allowance in September 2020, obtained from the Payroll Management Establishment Control (PMEC) unit and the Zambia Ministry of Health (MoH).<sup>8</sup> Owing to the absence of unique common facility identifiers across this data set and the JICA facility survey, the two data sets are matched based on facility and district name. Following this procedure, a total of 2,340 health facilities, whose eligibility status for receipt of the rural hardship allowance is established based on GPS coordinates, are matched to the list of health facilities actually receiving the allowance.

---

<sup>7</sup>We measure the total facility wage bill in Zambian Kwacha (ZMW). 1 USD is approximately equal to 22 ZMW.

<sup>8</sup>PMEC is the government departmental unit responsible for maintaining the entire Zambian civil service payroll, while MoH is responsible for the recruitment, deployment and management of all public sector employed health workers in Zambia as discussed in section 4.2.

For our analysis, we restrict the resulting sample of health facilities as follows. First, we disregard all health facilities for which the JICA census is missing information on health workers, leaving us with 2,123 facilities. Second, we exclude all facilities in category-D districts since there is no variation in eligibility for the rural hardship allowance within these districts. Third, we exclude all hospitals as they are few, almost exclusively located in urban centres and provide more complex curative healthcare services than primary care facilities, based on the referral system of healthcare provision (see section 4.2). Fourth, we exclude health posts as they are rarely found in urban areas. This leaves a total of 755 health facilities comprised exclusively of health centres, the most common type of health facility in Zambia found on both the eligible and ineligible side of the threshold. Thus, in the discussion below, the term health facility will refer to health centres and our results will only directly apply to this type of facility. Finally, we zoom in on the sample of health centres located within 10km of the eligibility threshold on either side of the threshold owing to the fact that the identification rationale of the RDD framework only applies when comparing health facilities within a close neighbourhood of the eligibility threshold, which are most likely to be similar, apart from falling on the eligible or ineligible side of the threshold, respectively (Lee and Lemieux, 2010). This leaves us with a final sample of 171 health centres, utilised in the main analysis.

Table 4.3.1 below presents means of the outcome variables as well as characteristics of the health facilities included in the sample. Column 3 contains the difference in the means between eligible and ineligible facilities while the standard error in column 4 is the standard error of the difference in the means. In the full sample (Panel A), there is a significant difference in size between eligible and ineligible facilities, whether measured by staffing levels, catchment population or the number of maternal beds. Urban facilities on the ineligible side of the threshold are generally bigger. When restricting the sample to facilities in the neighborhood of the threshold (Panel B), these differences largely disappear, indicating that facilities in this sample are more similar. The likelihood of receiving the hardship allowance, on the other hand, is significantly higher among eligible facilities and the difference with ineligible facilities is actually larger in the restricted sample.

Table 4.3.1: Health Facility Characteristics

Variable	(1) Eligible	(2) Ineligible	(3) Difference	(4) SE	(4) p-value
<b>Panel A: Full sample</b>					
Health workers	4.32	14.97	10.65	1.22	0.00
Wage bill (ZMW)	13,993.39	50,002.87	36,009.48	4,215.18	0.00
Catchment population	9,512	20,407	10895	2,500	0.00
Maternal beds	2.68	3.30	0.61	0.37	0.10
In-patient beds	7.28	6.77	-0.51	1.60	0.75
Receive allowance	0.56	0.15	-0.41	0.03	0.00
Observations	481	274			
<b>Panel B: Within 10km</b>					
Health workers	4.20	4.90	0.70	0.69	0.31
Wage bill (ZMW)	13,224.16	15,892.74	2,668.58	2,348.63	0.26
Catchment population	11,963	8,368	-3,594	1814	0.05
Maternal beds	2.92	3.36	0.44	0.40	0.26
In-patient beds	8.58	6.90	-1.68	1.34	0.21
Receives allowance	0.69	0.20	-0.50	0.07	0.00
Observations	85	86			

*Notes:* The table shows means of health facility characteristics for health facilities that are eligible and ineligible to receive the rural hardship allowance, based on the distance-to-district-capital eligibility criteria, respectively. Columns 1 and 2 show the means of each respective variable, while column 3 shows the difference in the means. Columns 4 and 5 show the standard error and p-value of the difference in the various means, respectively. Health facility refers to a health centre. The sample is restricted to districts with health facilities on both the eligible and ineligible sides of the threshold. Panel A shows characteristics for all facilities in the sample, while panel B is restricted to facilities within a 20km bandwidth of the eligibility threshold (10km on each side of the threshold). A health worker is defined as either an Adherence Officer, Clinical Officer, Community Health Assistant, Dental Therapist, Elimination of Mother to Child Transmission (EMTCT) Officer, Environmental Health Technician, Laboratory Technologist, Medical Doctor, Medical Licentiate, Nurse, Nutritionist, Pharmacist, Pharmacy Technician, Physiotherapist, Psycho-social Councillor, Public Health Inspector or Radiographer. Health Facility wage bill refers to the total monthly wage bill per health facility, measured in Zambian Kwacha, ZMW; 1 USD approximately equals 22 ZMW. Catchment population refers to the headcount of the population in each health facility's catchment area. Receive allowance refers to the share of health facilities listed as actually receiving the allowance, based on the data from MoH/PMEC.

## 4.4 Empirical Strategy

### 4.4.1 Identification Strategy

Identifying the causal effect of rural hardship allowance schemes on staffing levels is challenging. Since hardship allowances are targeted to facilities in rural areas that lack basic infrastructure, facilities at which such allowances are paid are by construction different from facilities where such allowances are not offered. Hence, comparing staffing levels between facilities where the allowance is paid and where not confounds the causal effect of the allowance on staffing levels with other differences in determinants of staffing levels between these two types of facilities.

This study utilises a spatial fuzzy regression discontinuity design (RDD) to estimate the causal effect of the rural hardship allowance on health facility staffing in Zambia. The rural hardship allowance in Zambia is administered in such a manner that health workers stationed in health facilities outside a pre-specified threshold distance from their district capital are eligible to receive it. This permits the estimation of an RDD model within a close neighbourhood of the eligibility threshold, using the distance to the district capital as the running variable and eligibility to receive the rural hardship allowance as an instrument for receipt of the allowance (Hahn et al., 2001). A fuzzy as opposed to sharp RDD is necessitated because crossing the eligibility threshold for the receipt of the allowance increases the likelihood of receipt, but does not shift it from 0 to 1 for reasons discussed in section 4.2.

A key identifying assumption of the RDD framework is that of continuity of all potential determinants of the outcome of interest aside from the treatment studied within a close neighbourhood of the eligibility threshold (Hahn et al., 2001). This implies that healthcare facilities within a close neighbourhood of the threshold for receiving the allowance, but on the non-qualifying side, are sufficiently similar to those on the qualifying side and in this way, serve as a valid comparison group in identifying the causal effects of the allowance. If the continuity assumption is satisfied, assignment to treatment, within a close neighbourhood of the eligibility threshold, can be said to be as good as random (Lee and Lemieux, 2010). Thus, any discontinuous 'jumps' in outcome variables at the threshold can then be attributed to the treatment (Imbens and Lemieux, 2008).

While the continuity assumption is by definition not testable, two aspects can support its plausibility. The first is continuity of the density of health facilities along the running variable at the eligibility threshold, which implies the absence of an unusually large clustering of health facilities immediately to the right or left of the threshold for receipt of the allowance. Practically, this latter point requires that it be impossible for health facilities to precisely and systematically manipulate their GPS coordinates so as to fall immediately to the right of the qualifying side of the threshold, when they should in actual fact not (Lee and Lemieux, 2010). Precise and systematic manipulation of GPS coordinates is highly implausible in our setting because the coordinates are measured, not by the health

facilities themselves, but by the office of the Surveyor General under a different ministry, as noted in section 4.2. To assess this aspect, figure 4.4.1 below shows a histogram of the density of health facilities, plotted against the running variable, within a 20km bandwidth of the treatment threshold. As is evident from the figure, an unusual clustering of health facilities directly to the right of the threshold is absent. Additionally, figure 4.4.1 also shows local polynomial estimates of the density function of health facilities against the running variable, with bias-robust 95% confidence bands around the estimates, following the discontinuity in density testing approach<sup>9</sup> proposed by Cattaneo et al. (2020) and implemented using the procedure in Cattaneo et al. (2019). The local polynomial estimates of the density function reveal a smooth curve through the treatment threshold. Additionally, we employ the manipulation test based on Cattaneo et al. (2020, 2018), an extension of the McCrary test (McCrary, 2008), which yields a bias-robust t-statistic of 0.2760, with a p-value of 0.7524, implying that we fail to reject the null hypothesis of no manipulation of the running variable at the treatment threshold. The above thus further supports the plausibility of the continuity assumption and alleviates concerns regarding a systematic and precise manipulation of GPS coordinates or indeed even systematic health facility construction on either side of the threshold by MoH after the introduction of the allowance.

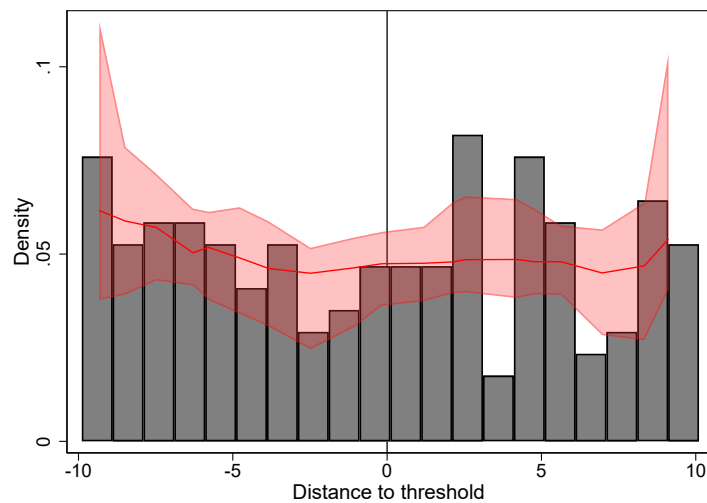


Figure 4.4.1: Health Facility Density around the Eligibility Threshold

*Notes:* This figure shows a histogram of the density of health centres, plotted against the running variable, distance-to-district-capital, within a 20km bandwidth of the eligibility threshold. The superimposed curve shows local polynomial estimates of the density function of health facilities against the running variable with bias-robust 95% confidence bands proposed in Cattaneo et al. (2020) and estimated using the procedure in Cattaneo et al. (2019).

<sup>9</sup>Manipulation testing procedures, such as McCrary (2008), require numerous data transformations such as pre-binning of the data and the choice of other 'tuning' parameters and might also suffer from bias arising from density estimation at boundary points (Hahn et al., 2001). The discontinuity in density testing approach proposed by Cattaneo et al. (2020), however, is robust to the above bias and allows for fully data-driven discontinuity in density testing, that does not require data transformations and the selection of tuning parameters, apart from the bandwidth.

The second aspect that can support the plausibility of the continuity assumption, is the continuity of observables unlikely to be affected by treatment at the eligibility threshold. To assess this aspect, we examine the continuity of catchment populations, maternal beds and in-patient beds at the threshold. Results from the estimation of RDD models, based on equation 4.3 discussed in section subsection 4.4.2 below, are shown in table 4.A.1 in annex 4.A.2. Results reveal no treatment effects of the rural hardship allowance on these variables in all specifications, further supporting the plausibility of continuity around the threshold in this setting.

#### 4.4.2 Estimation Strategy

In equation 4.1 below,  $y_i$  represents each respective outcome of interest; the stock of health workers and the total facility wage bill, for health facility  $i$ .  $allowance_i$  is an indicator for whether health facility  $i$  is listed as actually receiving the rural hardship allowance.  $distance_i$  represents the running variable, the distance of each health facility from it's respective district capital in kilometers (km). As per the literature, we specify this one-dimensional running variable such that  $distance_i > 0$  if health facility  $i$  falls on the eligible side of the threshold for receipt of the allowance.  $h(distance_i)$  is a continuous function of the running variable, estimated by a polynomial of order  $J$ .  $district_v$  are district fixed effects, included in all specifications and  $\epsilon_i$  is the error term. However, as noted above,  $\beta_2$  in equation 4.1 may potentially be biased owing to endogeneity arising from non-random selection effects. To control for this, eligibility to receive the rural hardship allowance, based on the distance-to-district-capital criteria and given by the indicator  $eligible_i$ , is used as an instrument for receipt of the rural hardship allowance  $allowance_i$ . Equation 4.2 shows the corresponding first stage of the fuzzy RDD model (Hahn et al., 2001).

$$y_i = \beta_0 + \beta_1 h(distance_i) + \beta_2 allowance_i + \beta_3 district_v + \epsilon_i; distance_i < |b| \quad (4.1)$$

$$allowance_i = \gamma_0 + \gamma_1 g(distance_i) + \gamma_2 eligible_i + \gamma_3 district_v + \omega_i; distance_i < |b| \quad (4.2)$$

$$y_i = \theta_0 + \theta_1 f(distance_i) + \theta_2 eligible_i + \theta_3 district_v + \tau_i; distance_i < |b| \quad (4.3)$$

Equation 4.3 represents the reduced form relationship between eligibility for the rural hardship allowance and each respective outcome of interest, with  $\theta_2$  being the coefficient capturing the causal effect of eligibility (Intention-to-Treat), localised to the eligibility threshold (Hahn et al., 2001). Equation 4.3 is estimated non-parametrically using local polynomial estimation and inference procedures proposed in Calonico et al. (2014, 2017), within a close neighbourhood of the eligibility threshold,  $distance < |b|$ .<sup>10</sup> Specifically,

---

<sup>10</sup>As opposed to parametric estimation procedures, such as Ordinary Least Squares (OLS), non-parametric bias-robust local polynomial regression estimation and inference procedures are argued to

a benchmark local linear regression specification is estimated, where the flexible function of the running variable  $f(\text{distance}_i)$ , is specified as  $f(\text{distance}_i) = \delta_1 \text{distance}_i + \delta_2 \text{distance}_i * \text{elig}_i$ , including an interaction term of the running variable and eligibility indicator. A second specification utilising a local polynomial regression specification is estimated as  $f(\text{distance}_i) = \sum_{k=1}^3 \delta_k \text{distance}_i^k$ , with a third-order polynomial of the running variable included to account for potential non-linearities in the underlying (unobservable) regression functions being estimated.

$$y_i = \beta_0 + \beta_1 h(\text{distance}_i) + \beta_2 \widehat{\text{allowance}}_i + \beta_3 \text{district}_v + \epsilon_i; \text{distance}_i < |b| \quad (4.4)$$

We follow the same non-parametric procedure to estimate the first stage (equation 4.2) and an instrumental variables specification as shown in equation 4.4, where  $\widehat{\text{allowance}}_i$  are residuals from equation 4.2. The resulting fuzzy RDD estimates from equation 4.4 are equivalent to Local Average Treatment Effect (LATE) estimates for compliers at the eligibility threshold, i.e. for those health facilities induced to receive the rural hardship by being located on the eligible side of the threshold. All local polynomial regression specifications utilise a triangular weighting kernel, which assigns a larger weight to observations closer to the threshold. The bandwidth  $|b|$  is chosen using two approaches. The benchmark specification uses a common bandwidth of 10km on each side of the threshold for all outcomes in order to ensure results are comparable across outcomes. A second approach involves estimating an optimal bandwidth for each regression specification using data-driven procedures proposed in Calonico et al. (2014) and Calonico et al. (2017). Unless otherwise stated, data-driven bandwidths are selected so as to minimise the mean squared error (mse), which has been noted to be an ideal procedure for both point estimation and inference purposes (Calonico et al., 2014, 2017). Another critical choice in RDD estimation relates to bin selection. Unless otherwise stated, we utilise quantile-spaced bins that mimic the variance of the underlying data. This bin selection procedure is recommended in the literature as it minimises the effects of differences in between-bin variance, by choosing the distance between bins such that an equal number of observations is included in each bin (Calonico et al., 2014, 2017).<sup>11</sup>

## 4.5 Results

In order to assess the existence of a first-stage relationship between eligibility to receive the rural hardship allowance, based on the distance-to-district-capital criteria and actual receipt of the allowance, table 4.5.1 below shows estimates based on the first stage equation 4.2 above. The existence of this first stage relationship between eligibility and actual receipt of the allowance, is an important first step in estimating the causal effect of the

---

be preferable in RDD settings, because they account for potential bias in coefficient estimates and their associated confidence intervals, arising from estimation of unknown regression functions on either side of the eligibility threshold (Calonico et al., 014a).

<sup>11</sup>Results presented below are robust to bin selection utilising an alternative, evenly spaced binning procedure. This approach is less recommended in the literature, owing to the effects of between-bin variance, given that the number of observations per bin varies under this procedure.

rural hardship allowance on our outcomes of interest. To this end, estimates in column 1 of the table utilise a specification with a 20km bandwidth (10km's on each side of the eligibility threshold), a first-order polynomial of the running variable, distance-to-district-capital, in kilometers (km) and an interaction term of the running variable and eligibility indicator. Similarly, column 2 of the table contains estimates from a specification utilising a 20km bandwidth, a third-order polynomial of the running variable and no interaction term. Column 3 shows estimates utilising a specification with a first-order polynomial, a data-driven 9.5km bandwidth on each side of the eligibility threshold and no interaction term.

Column 1 of table 4.5.1 reveals that crossing the eligibility threshold, increases the likelihood of actually receiving the rural hardship allowance by approximately 64 percentage points using a 10km bandwidth. Similarly, the more flexible specifications with a third order polynomial and a data driven bandwidth in columns 2 and 3, reveal between 73 and 71 percentage point increases in the likelihood of receiving the rural hardship allowance upon crossing the eligibility threshold, respectively. Table 4.5.1 also reveals sufficiently large F-statistics in all specifications, indicating that eligibility, based on the district-to-district-capital criteria, is a sufficiently strong instrument for receipt of the rural hardship allowance. Figure 4.5.1, graphically plots the first stage, visually depicting the increase in the likelihood of rural allowance receipt at the eligibility threshold.

With a sufficiently strong first stage relationship between eligibility and actual receipt of the rural hardship allowance established, table 4.5.2 below presents reduced-form, ITT estimates of the impact of rural hardship allowance eligibility, estimated using equation 4.3. Panel A of the table contains estimates of the impact of the rural hardship allowance eligibility on the stock of health workers, measured as the natural logarithm of the number of health workers per facility. Similarly, panel B presents reduced form estimates of the impact of the rural hardship allowance eligibility on the total monthly facility wage bill.<sup>12</sup> The estimates presented in column 1 of the table are based on a specification with a first-order polynomial of the running variable, distance-to-district-capital, in kilometers (km), a 20km bandwidth (10km on either side of the eligibility threshold) and an interaction term of the running variable and eligibility indicator.

The estimates in column 2 are based on a specification with a third-order polynomial of the running variable, a 20km bandwidth and no interaction term. Estimates in column 3 are based on a specification with a first-order polynomial of the running variable, different data-driven, mean squared error minimising bandwidths for each outcome, chosen using the procedure in Calonico et al. (2014, 2017) and no interaction term. Heteroscedasticity robust bias-corrected standard errors and coefficient estimates, which are estimated with the previously mentioned estimation procedures, are reported. All specifications include district fixed effects.

---

<sup>12</sup>The total monthly wage bill per health facility is measured in Zambian Kwacha (ZMW). 1 USD is approximately equal to 22 ZMW.

Table 4.5.1: Receipt of Hardship Allowance: First Stage

	Receives Hardship Allowance		
	(1)	(2)	(3)
Eligible	0.643*** (0.153)	0.727*** (0.214)	0.707*** (0.156)
Observations	171	171	157
Mean Ineligible	0.198	0.198	0.200
Bandwidth (km)	10	10	9.5
Polynomial	1	3	1
Interaction	Yes	No	No
F-Statistic	14.10	13.37	18.71

*Notes:* The table contains estimates of the first stage relationship between eligibility to receive the rural hardship allowance, based on the distance-to-district-capital eligibility criteria and the likelihood of actually receiving the rural hardship allowance as in equation 4.2 above. The dependant variable in each column is an indicator for actual receipt of the rural hardship allowance, while the regressors include an indicator for eligibility to receive the allowance, based on the district-to-district-capital criteria. Heteroscedasticity robust bias-corrected standard errors, in parentheses and coefficient estimates are shown. All specifications include district fixed effects. Column 1 utilises a specification with a first order polynomial of the running variable, a 20km bandwidth (10km on each side of the eligibility threshold) and an interaction term of the running variable and eligibility indicator. Column 2 utilises a specification with a third-order polynomial, a 20km bandwidth and no interaction term. Column 3 utilises a specification with a first-order polynomial a data-driven 9.514 km bandwidth on each side of the eligibility threshold, selected using the procedure in Calonico et al. (2014, 2017) and no interaction term;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Panel A of table 4.5.2 reveals that crossing the threshold for rural hardship allowance eligibility, increases the number of health workers by approximately 57% using a 10km bandwidth.<sup>13</sup> Similarly, the more flexible specifications with a third-order polynomial and a data-driven bandwidth in columns 2 and 3 respectively, reveal 41% and 51% increases in the number of health workers per facility at the eligibility threshold. However, only the coefficient estimates in columns 1 and 3 are statistically significant at the 5% level.

Panel B of table 4.5.2 reveals that crossing the eligibility threshold for the rural hardship allowance increases the monthly wage bill per health facility by approximately 60% using a 10km bandwidth. This same effect in the more flexible specifications ranges between

<sup>13</sup>To obtain this number, we transform the coefficient estimate  $\hat{\theta}_2$  using the expression  $\exp(\hat{\theta}_2) - 1$ . We adopt this convention for the rest of the paper.

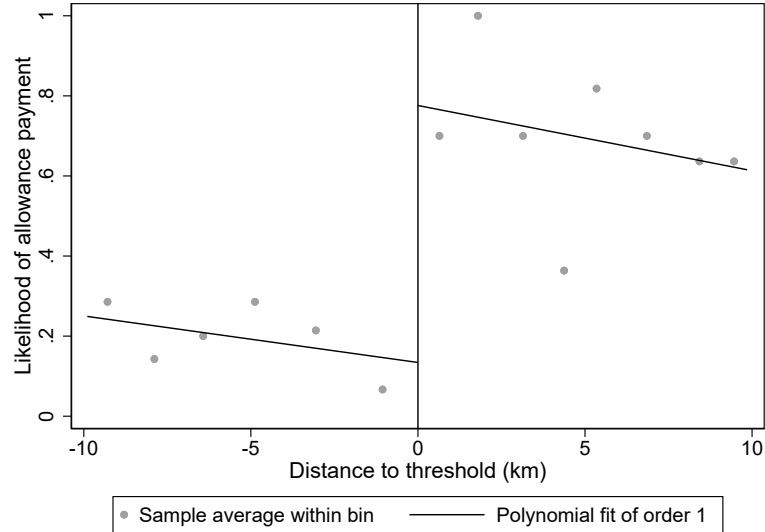


Figure 4.5.1: Receipt of Hardship Allowance: First Stage

*Notes:* This figure plots the first-stage relationship between eligibility to receive the rural hardship allowance, based on the distance-to-district-capital criteria (horizontal axis) and the likelihood of actually receiving the allowance (vertical axis). The specification utilises a first-order polynomial of the running variable with a 20km bandwidth (10km on each side of the eligibility threshold). All specifications utilise a triangular weighting kernel. Grey dots are quantile-spaced within-bin sample means, mimicking the variance of the underlying data.

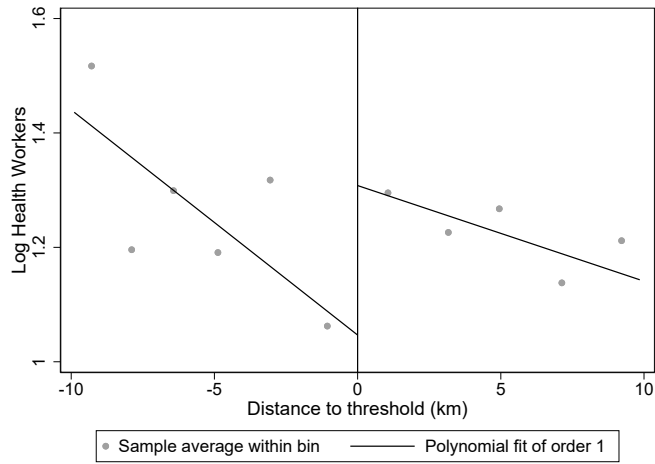
47% and 50% in columns 2 and 3, respectively. However, similar to the estimates in panel A, only the coefficient estimates in columns 1 and 3 are statistically significant at the 5% level.

One noteworthy observation from the results presented above, is that the increases in the quality-adjusted stock of health workers, represented by the monthly facility wage bill, seem to be slightly larger than the corresponding increases in the stock of health workers. This is especially evident when comparing the various sets of estimates in columns 1 and 2 of panels A and B, respectively, which are based on the same bandwidths and therefore, the same samples of facilities. Figure 4.5.2 graphically plots these results, using a 10km bandwidth and visually depicts the increase in the stock of health workers and the facility wage bill at the eligibility threshold.

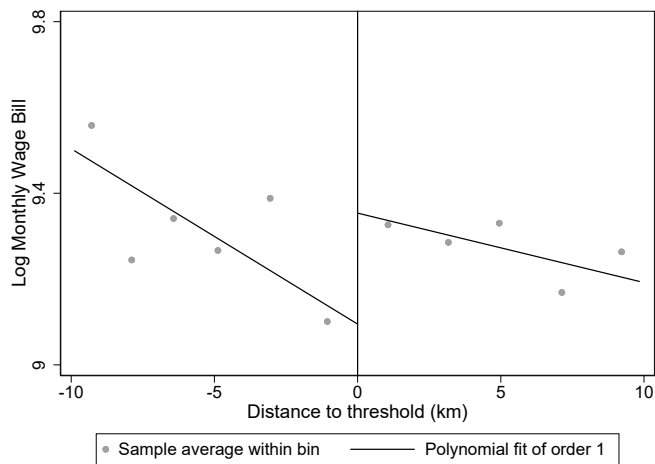
Table 4.5.2: Reduced Form Estimates of the Impact of the Rural Hardship Allowance

	(1)	(2)	(3)
<b>Panel A: Log Health Workers</b>			
Eligible	0.452**	0.340	0.413**
	(0.200)	(0.402)	(0.199)
	[0.571]	[0.405]	[0.511]
Observations	171	171	100
Mean Ineligible	1.26	1.26	1.21
Bandwidth (km)	10	10	7.2
<b>Panel B: Log Monthly Wage Bill</b>			
Eligible	0.468**	0.386	0.404**
	(0.196)	(0.393)	(0.198)
	[0.597]	[0.471]	[0.498]
Observations	171	171	97
Mean Ineligible	9.314	9.314	9.272
Bandwidth (km)	10	10	7.1
Polynomial	1	3	1
Interaction term	Yes	No	No

*Notes:* This table contains reduced form estimates of the impact of the rural hardship allowance eligibility on the stock of health workers, in panel A and the total wage bill per health facility, in panel B, respectively. The total wage bill per health facility is measured by the natural logarithm of total monthly wage bill per health facility in Zambian Kwacha ZMW; 1 USD approximately equals 22 ZMW. Heteroscedasticity robust bias-corrected standard errors, in parentheses and coefficient estimates are shown. Each column contains estimates from a different regression specification. Column 1 shows a specification with a first-order polynomial of the running variable, a 20km bandwidth (10km on each side of the eligibility threshold) and an interaction term of the running variable and eligibility indicator. Column 2 shows a specification with a third-order polynomial, a 20km bandwidth and no interaction term. Column 3 shows a specification with a first-order polynomial, different data-driven, mean squared error minimising bandwidths, each estimated using the procedure in Calonico et al. (2014, 2017) and no interaction term. The estimates in brackets are obtained by transforming the coefficient estimate  $\hat{\theta}_2$  using the following approximation  $\exp(\hat{\theta}_2) - 1$ . All specifications utilise a triangular weighting kernel and include district fixed effects;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .



(a) Number of health workers



(b) Facility wage bill

Figure 4.5.2: Health Facility Staffing (Reduced Form)

*Notes:* This figure shows the reduced form relationship between rural hardship allowance eligibility and staffing per health facility. The health facility wage bill is measured by the natural logarithm of total monthly wage bill per health facility (measured in Zambian Kwacha ZMW, with 1 USD approximately equals 22 ZMW). In panel A, log health workers per facility is the outcome variable, while in panel B, log monthly health facility wage bill is the outcome variable. Both figures display a specification with a first-order polynomial of the running variable, distance-to-district-capital in Kilometres (km), a 10km bandwidth on either side of the eligibility threshold and a triangular weighting kernel. Grey dots are quantile-spaced within bin sample means, mimicking the variance of the underlying data.

Table 4.5.3: Fuzzy RDD Estimates of the Impact of the Rural Hardship Allowance

	(1)	(2)	(3)
<b>Panel A: Log Health Workers</b>			
Rural	0.611*	0.825	0.590**
	(0.348)	(0.889)	(0.279)
	[0.842]	[1.282]	[0.804]
Observations	171	171	130
Mean Ineligible	1.262	1.262	1.262
Bandwidth (km)	10	10	10.002
<b>Panel B: Log Monthly Wage Bill</b>			
Rural	0.634*	0.919	0.619**
	(0.342)	(0.886)	(0.283)
	[0.885]	[1.507]	[0.857]
Observations	171	171	129
Mean Ineligible	9.314	9.314	9.314
Bandwidth (km)	10	10	10.026
Polynomial	1	3	1
Interaction term	Yes	No	No

*Notes:* This table contains Fuzzy RDD estimates of the impact of rural hardship allowance receipt at the eligibility threshold, on the stock of health workers, in panel A and the total wage bill per health facility, in panel B, respectively. The total wage bill per health facility is measured by the natural logarithm of total monthly wage bill per health facility, measured in Zambian Kwacha ZMW; 1 USD approximately equals 22 ZMW. Heteroscedasticity robust bias-corrected standard errors, in parentheses and coefficient estimates, are shown. Each column contains estimates from a different regression specification. Column 1 shows a specification with a first-order polynomial of the running variable, a 20km bandwidth (10km on each side of the eligibility threshold) and an interaction term of the running variable and eligibility indicator. Column 2 shows a specification with a third-order polynomial, a 20km bandwidth and no interaction term. Column 3 shows a specification with a first-order polynomial, different data-driven, mean squared error minimising bandwidths, each estimated using the procedure in Calonico et al. (2014, 2017) and no interaction term. The estimates in brackets are obtained by transforming the coefficient estimate  $\hat{\beta}_2$  using the following approximation  $\exp(\hat{\beta}_2) - 1$ . All specifications utilise a triangular weighting kernel and include district fixed effects;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Assessment of the impact of the rural hardship allowance continues in table 4.5.3. Specifically, the table contains Fuzzy RDD, LATE estimates of the impact of rural hardship allowance receipt at the eligibility threshold. Panel A of table 4.5.3 contains estimates of the impact of the rural hardship allowance receipt at the eligibility threshold on the number of health workers. Panel B presents estimates of the impact of the rural hard-

ship allowance receipt at the eligibility threshold on the total facility wage bill.<sup>14</sup> The estimates in column 1 of the table are based on a specification with a first-order polynomial of the running variable, distance-to-district-capital, in kilometers (km), a 20km bandwidth (10km on either side of the eligibility threshold) and an interaction term of the running variable and eligibility indicator. The estimates in column 2 are based on a specification with a third-order polynomial of the running variable, a 20km bandwidth and no interaction term. The estimates in column 3 are based on a specification with a first-order polynomial of the running variable, different data-driven, mean squared error minimising bandwidths for each outcome, chosen using the procedure in Calonico et al. (2014, 2017) and no interaction term. Heteroscedasticity robust bias-corrected standard errors and coefficient estimates, which are estimated with the previously mentioned estimation procedures, are reported. All specifications utilise a triangular weighting kernel and include district fixed effects.

Although qualitatively similar to the reduced form estimates presented above, the Fuzzy RDD, LATE estimates in table 4.5.3 are generally larger than the former estimates, as would be expected, given the need to account for the first stage in the Fuzzy RDD estimates. Specifically, column 1 panel A of table 4.5.3 reveals that receipt of the rural hardship allowance at the eligibility threshold, results in an 84% increase in the number of health workers. Similarly, the more flexible specifications in columns 2 and 3 respectively, reveal increases in the number of health workers between 128% and 80%, respectively. However, similar to the reduced form estimates shown above, only the Fuzzy RDD, LATE estimates in columns 1 and 3 are statistically significant at the 10 and 5% levels, respectively.

Panel B of table 4.5.3 reveals that receipt of the rural hardship allowance at the eligibility threshold, results in an 89% increase in the total wage bill per health facility, using a 10km bandwidth. The corresponding increases in the more flexible specifications in columns 2 and 3 are 151% and 86%, respectively. However, as with the estimates in panel A, only the coefficient estimates in columns 1 and 3 are statistically significant, at the same 10 and 5% levels, respectively.

## 4.6 Robustness checks

### 4.6.1 Bandwidth Choice

The choice of bandwidth is perhaps the most important single choice in RDD estimation. As noted above, the main analysis relies on a standard 10km bandwidth on each side of the eligibility threshold, which is held fixed across all outcomes to facilitate comparability, as well as different data-driven bandwidths chosen for each outcome, using the procedure in Calonico et al. (2014, 2017). The data-driven bandwidths are generally between 7 and 10km on either side of the eligibility threshold.

---

<sup>14</sup>The total wage bill per health facility is measured in Zambian Kwacha (ZMW). 1 USD is approximately equal to 22 ZMW.

To demonstrate that results are robust to the choice of bandwidth, table 4.A.2 in annex 4.A.2 shows reduced form estimates for all outcomes, for different bandwidths between 5 and 15km. Estimates for the stock of health workers in panel A and the monthly wage bill in panel B per facility all reveal qualitatively similar results as those in table 4.5.2. Specifically, the estimates reveal larger coefficient estimates for shorter bandwidths. At the same time, coefficient estimates become less statistically significant for shorter bandwidths as the number of facilities in the sample shrinks.

## 4.7 Conclusion

Self-reinforcing structural inequality between urban and rural areas is a major concern in many developing countries. Rural areas lack amenities valued by service delivery agents and rigid wage structures in the public sector limit the ability of governments to compensate these agents for the implied disutility from working in these areas. As a result, rural areas lack qualified personnel for service delivery and access to public services remains limited, thereby keeping the amenity value low and reinforcing existing rural-urban gaps in public service delivery and ultimately development levels.

A common policy tool employed by governments in developing countries to break this cycle of rural underdevelopment is the payment of salary supplements for working in areas defined as rural, also referred to as hardship allowances. The simple idea is that the allowance will compensate workers for the lack of access to amenities, such as basic infrastructure and services and opportunities for training and career development, in rural areas. The success of such schemes depends on the elasticity of labor supply in rural areas.

This paper evaluates the impact of paying a 20% salary supplement to health workers for working in rural areas on staffing levels at rural health facilities in Zambia. To identify the causal effect of the hardship allowance, we leverage a spatial discontinuity in the eligibility for the allowance. We employ a fuzzy regression discontinuity design, comparing staffing levels between nearby facilities on either side of the eligibility threshold. We find that the probability of payment of the hardship allowance increases by 71 percentage points at the eligibility threshold. At the same time, the number of health workers per health facility increases by 51%. Instrumenting the payment of the allowance with eligibility among facilities near the threshold, we estimate that the allowance increases staffing levels by approximately 80% in health facilities receiving it at the eligibility threshold. Hence, the allowance has a large effect on the availability of health workers in rural health facilities. Similarly, we find that the allowance also results in large increases in the quality-adjusted stock of health workers per facility. Assuming that differences in basic salaries reflect differences in the quality of health workers, we proxy health worker quality by computing the total facility wage bill received by health workers of various specialisations in each facility.<sup>15</sup> Our estimates reveal a 50% increase in the quality-adjusted stock of health

---

<sup>15</sup>We measure the total facility wage bill in Zambian Kwacha (ZMW). 1 USD is approximately equal to 22 ZMW.

workers per facility, with this estimate increasing to 86% once payment of the allowance is instrumented with eligibility to receive the allowance. Thus, the rural hardship allowance leads to increases in both the number of health workers per facility as well as their quality. Furthermore, we find that increases in the quality-adjusted stock of health workers, as measured by the monthly facility wage bill, are larger than the corresponding increases in the number of health workers.

With 60% of Zambia's population residing in rural areas, this finding is of particular relevance in the context of our study. It indicates that salary supplements are indeed a viable, although costly, option to ensure greater presence of high quality health workers in rural areas.

What remains unclear is to what extent greater presence of high quality health workers in rural areas can improve health service delivery and population health. If other complementary inputs to service provision and health production are absent or demand for health services is low, workforce increases may not translate into improvements in service delivery and health outcomes, or effects could be muted. In ongoing work, we examine this question by comparing service provision and recorded disease patterns between eligible and ineligible health facilities near the threshold.

## Bibliography

- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics* 97(4), 543–569.
- Alkema, L., D. Chou, D. Hogan, S. Zhang, A.-B. Moller, A. Gemmill, D. M. Fat, T. Boerma, M. Temmerman, C. Mathers, et al. (2016). Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the un maternal mortality estimation inter-agency group. *The Lancet* 387(10017), 462–474.
- Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011). Inside the war on poverty: The impact of food stamps on birth outcomes. *The Review of Economics and Statistics* 93(2), 387–403.
- Amarante, V., M. Manacorda, E. Miguel, and A. Vigorito (2016). Do cash transfers improve birth outcomes? evidence from matched vital statistics, program, and social security data. *American Economic Journal: Economic Policy* 8(2), 1–43.
- Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics* 10, 439–463.
- Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review* 110(5), 1355–94.
- Ashraf, N., O. Bandiera, and B. K. Jack (2014). No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics* 120, 1–17.
- Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21(1), 1–42.
- Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014). Conditional, unconditional and everything in between: A systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness* 6(1), 1–43.
- Barber, S. L. and P. J. Gertler (2008). The impact of mexico’s conditional cash transfer programme, oportunitades, on birthweight. *Tropical Medicine and International Health* 13(11), 1405–1414.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. Soucat, J. Sturdy, and C. M. Vermeersch (2011). Effect on maternal and child health services in rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet* 377(9775), 1421–1428.
- Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies* 70(3), 489–520.

- Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *American Economic Review* 95(3), 616–636.
- Besley, T. J. and M. Ghatak (2014). Solving agency problems: Intrinsic motivation, incentives, and productivity. Technical report.
- Bonfrer, I., E. Van de Poel, and E. Van Doorslaer (2014). The effects of performance incentives on the utilization and quality of maternal and child care in burundi. *Social Science & Medicine* 123, 96–104.
- Bowles, S. and S. Polania-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature* 50(2), 368–425.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies, and D. Wiesen (2013). *How Effective are Pay-for-Performance Incentives for Physicians?—A Laboratory Experiment*. Number 413. Ruhr Economic Papers.
- Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz, and D. Wiesen (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization* 131, 17–23.
- Cabrera, J. M. and D. Webbink (2020). Do higher salaries yield better teachers and better student outcomes? *Journal of Human Resources* 55(4), 1222–1257.
- Callaway, B. and P. H. Sant’Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.
- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal* 17(2), 372–404.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal* 14(4), 909–946.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6), 2295–2326.
- Cardona-Sosa, L. and C. Medina (2017). The effects of in utero programs on birth outcomes: The case of buen comienzo. *Journal of the Latin American and Caribbean Economic Association* 17(2), 93–134.
- Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal* 18(1), 234–261.
- Cattaneo, M. D., M. Jansson, and X. Ma (2019). lpdensity: Local polynomial density estimation and inference. *arXiv preprint arXiv:1906.06529*.
- Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531), 1449–1455.

- Chelwa, G., M. Pellicer, and M. Maboshe (2019). Teacher pay and educational outcomes: Evidence from the rural hardship allowance in zambia. *South African Journal of Economics* 87(3), 255–282.
- Chibuye, B., P. Hangoma, and M. Khadjavi (2021). The dark side of financial incentives in a health setting - experimental evidence from zambia. Working paper, Christian-Albrechts-University of Kiel.
- Chibwasha, C. J., A. Zanolini, M. Smid, B. Vwalika, M. P. Kasaro, M. Mwanahamuntu, J. S. Stringer, and E. M. Stringer (2016). Predictors and outcomes of low birth weight in lusaka, zambia. *International Journal of Gynecology & Obstetrics* 134(3), 309–314.
- Cross, S. S. and M. Baernholdt (2017). Identifying factors for worker motivation in zambia’s rural health facilities. *Journal of Health Care for the Poor and Underserved* 28(1), 63–70.
- Currie, J. and D. Almond (2011). Human capital development before age five. In *Handbook of Labor Economics*, Volume 4, pp. 1315–1486. Elsevier.
- Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics* 128(3), 1169–1218.
- Das, J. and J. Hammer (2014). Quality of primary care in low-income countries: Facts and economics. *Annual Review of Economics* 6(1), 525–553.
- de Bekker-Grob, E. W., M. Ryan, and K. Gerard (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21(2), 145–172.
- De Chaisemartin, C. and X. d’Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9), 2964–96.
- De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2018). Double for nothing? experimental evidence on an unconditional teacher salary increase in indonesia. *The Quarterly Journal of Economics* 133(2), 993–1039.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18(1), 105–128.
- Deci, E. L., R. Koestner, and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125(6), 627–693.
- Deserranno, E. (2019). Financial incentives as signals: Experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics* 11(1), 277–317.
- Díaz, J.-J. and V. Saldarriaga (2019). Encouraging use of prenatal care through conditional cash transfers: Evidence from juntos in peru. *Health Economics* 28(9), 1099–1113.

- Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics* 27(6), 1436–1450.
- Dussault, G. and M. C. Franceschini (2006). Not enough there, too many here: Understanding geographical imbalances in the distribution of the health workforce. *Human Resources for Health* 4(1), 1–16.
- Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics* 24(1), 211–223.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics* 108(2), 437–459.
- Fehr, E., G. Kirchsteiger, and A. Riedl (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review* 42(1), 1–34.
- Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy* 21(3), 673–705.
- Ferrinho, P., S. Siziya, F. Goma, and G. Dussault (2011). The human resource for health situation in zambia: Deficit and maldistribution. *Human Resources for Health* 9(1), 1–11.
- Figueiredo Walter, T. (2020). Misallocation in the public sector? cross-country evidence from two million primary schools. Technical report, STICERD Working Paper No. EOPP 70.
- Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. *Handbook of Economic Field Experiments* 2, 467–514.
- Friedman, J., J. N. Qamruddin, C. Chansa, and A. K. Das (2016). Impact evaluation of zambia's health results-based financing pilot project. Technical report, The World Bank.
- Gage, A. and S. Bauhoff (2021). The effects of performance-based financing on neonatal health outcomes in burundi, lesotho, senegal, zambia and zimbabwe. *Health Policy and Planning*.
- Gertler, P. and C. Vermeersch (2012). *Using Performance Incentives to Improve Health Outcomes*. The World Bank.
- Gneezy, U. and J. A. List (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5), 1365–1384.
- Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3), 791–810.
- Godager, G. and D. Wiesen (2013). Profit or patients' health benefit? exploring the heterogeneity in physician altruism. *Journal of Health Economics* 32(6), 1105–1116.

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.
- Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization* 106, 367–378.
- Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.
- Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature* 42(4), 1009–1055.
- Hennig-Schmidt, H., R. Selten, and D. Wiesen (2011). How payment systems affect physicians’ provision behaviour — an experimental investigation. *Journal of Health Economics* 30(4), 637–646.
- Herbst, C. H., M. Vledder, K. Campbell, M. Sjöblom, and A. Soucat (2011). The human resources for health crisis in zambia. *Washington (DC): World Bank*.
- Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.* 7, 24.
- Hoynes, H., D. Miller, and D. Simon (2015). Income, the earned income tax credit and infant health. *American Economic Journal: Economic Policy* 7(1), 172–211.
- Hoynes, H., M. Page, and A. H. Stevens (2011). Can targeted transfers improve birth outcomes?: Evidence from the introduction of the wic program. *Journal of Public Economics* 95(7-8), 813–827.
- Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635.
- James, N., K. Lawson, and Y. Acharya (2020). Evidence on result-based financing in maternal and child health in low-and middle-income countries: A systematic review. *Global Health Research and Policy* 5(1), 1–15.
- Kaarboe, O. and L. Siciliani (2011). Multi-tasking, quality and pay for performance. *Health Economics* 20(2), 225–238.
- Kesternich, I., H. Schumacher, and J. Winter (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? *Journal of Public Economics* 131, 1–11.
- Kramer, M. S. (1987). Determinants of low birth weight: Methodological assessment and meta-analysis. *Bulletin of the World Health Organization* 65(5), 663.
- Kramer, M. S. (2003). The epidemiology of adverse pregnancy outcomes: An overview. *The Journal of Nutrition* 133(5), 1592S–1596S.

- Lagarde, M. and D. Blaauw (2014). Pro-social preferences and self-selection into jobs: Evidence from south african nurses. *Journal of Economic Behavior & Organization* 107, 136–152.
- Lagarde, M. and D. Blaauw (2017). Physicians’ responses to financial and social incentives: A medically framed real effort experiment. *Social Science & Medicine* 179, 147–159.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review* 90(5), 1346–1361.
- Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281–355.
- Lemiere, C., C. Herbst, N. Jahanshahi, and E. Smith (2011). Reducing geographical imbalances of health workers in sub-saharan africa: A labor market perspective on what works, what does not, and why. world bank; 2011.
- Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2019). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly Journal of Economics* 134(3), 1627–1673.
- McCoy, D., S. Bennett, S. Witter, B. Pond, B. Baker, J. Gow, S. Chand, T. Ensor, and B. McPake (2008). Salaries and incomes of health workers in sub-saharan africa. *The Lancet* 371(9613), 675–681.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2), 698–714.
- Miller, G. and K. S. Babiarz (2013). Pay-for-performance incentives in low-and middle-income country health programs. Technical report, National Bureau of Economic Research.
- Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, and S. Rozelle (2012). Effectiveness of provider incentives for anaemia reduction in rural china: A cluster randomised trial. *Bmj* 345.
- Mullen, K. J., R. G. Frank, and M. B. Rosenthal (2010). Can you get what you pay for? pay-for-performance and the quality of healthcare providers. *The Rand Journal of Economics* 41(1), 64–91.
- Muralidharan, K. and N. Prakash (2017). Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics* 9(3), 321–50.
- Neilson, C., M. Bobba, T. Ederer, G. Leon-Ciliotta, and M. Nieddu (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in peru. Working paper, National Bureau of Economic Research.
- Ohlsson, A. and P. Shah (2008). *Determinants and Prevention of Low Birth Weight: A Synopsis of the Evidence*. Institute of Health Economics.

- Olken, B. A., J. Onishi, and S. Wong (2014). Should aid reward performance? evidence from a field experiment on health and education in indonesia. *American Economic Journal: Applied Economics* 6(4), 1–34.
- Paul, E., M. L. Dramé, J.-P. Kashala, A. E. Ndema, M. Kounnou, J. C. Aïssan, and K. Gyselincx (2018). Performance-based financing to strengthen the health system in benin: Challenging the mainstream approach. *International Journal of Health Policy and Management* 7(1), 35.
- Pega, F., S. Y. Liu, S. Walter, R. Pabayo, R. Saith, and S. K. Lhachimi (2017). Unconditional cash transfers for reducing poverty and vulnerabilities: Effect on use of health services and health outcomes in low-and middle-income countries. *Cochrane Database of Systematic Reviews* (11).
- Prust, M. L., A. Kamanga, L. Ngosa, C. McKay, C. M. Muzongwe, M. T. Mukubani, R. Chihinga, R. Misapa, J. W. van den Broek, and N. Wilmink (2019). Assessment of interventions to attract and retain health workers in rural zambia: A discrete choice experiment. *Human Resources for Health* 17(1), 26.
- Pugatch, T. and E. Schroeder (2014). Incentives for teacher relocation: Evidence from the gambian hardship allowance. *Economics of Education Review* 41, 120–136.
- Sappington, D. E. (1991). Incentives in principal-agent relationships. *Journal of Economic Perspectives* 5(2), 45–66.
- Shen, G. C., H. T. H. Nguyen, A. Das, N. Sachingongu, C. Chansa, J. Qamruddin, and J. Friedman (2017). Incentives to change: Effects of performance-based financing on health workers in zambia. *Human Resources for Health* 15(1), 20.
- Sherry, T. B. (2016). A note on the comparative statics of pay-for-performance in health care. *Health Economics* 25(5), 637–644.
- Sherry, T. B., S. Bauhoff, and M. Mohanan (2017). Multitasking and heterogeneous treatment effects in pay-for-performance in healthcare: Evidence from rwanda. *American Journal of Health Economics* 3(2), 192–226.
- Singh, P. (2015). Performance pay and information: Reducing child undernutrition in india. *Journal of Economic Behavior & Organization* 112, 141–163.
- Singh, P. and W. A. Masters (2020). Performance bonuses in the public sector: Winner-take-all prizes versus proportional payments to reduce child malnutrition in india. *Journal of Development Economics* 146, 102–295.
- United Nations, U. (2016). Transforming our world: The 2030 agenda for sustainable development.
- Zambia Ministry of Health, M. (2011). National health strategic plan (2011 - 2015). Technical report, Government of the Republic of Zambia.

Zambia Ministry of Health, M. (2017). National health sector strategic plan. Technical report, Government of the Republic of Zambia., Lusaka.

Zambia Ministry of Health, Z. S. A. and M. IFC (2019). Zambia demographic and health survey 2018. Technical report, Zambia Statistics Agency, Ministry of Health, and ICF, Lusaka, Zambia, and Rockville, Maryland, USA.

Zambia Ministry of National Development Planning, M. (2017). Seventh national development plan (sndp). Technical report, Zambia Ministry of National Development Planning.

# Appendices

## 4.A Appendices to Chapter Four

### 4.A.1 Assistance at Delivery

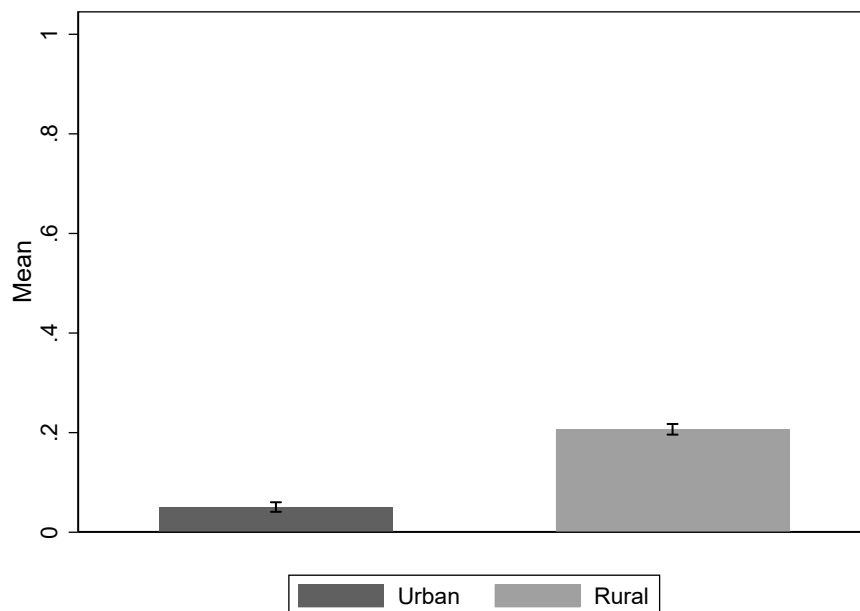


Figure 4.A.1: Assistance at Delivery: Friend, Relative or Traditional Birth Attendant.

*Notes:* This figure plots the proportion of women reporting being assisted by a friend, relative or traditional birth attendant during delivery in rural and urban areas respectively. The figures are the authors' calculations based on data from the 2018 Demographic and Health Survey.

## 4.A.2 Placebo Outcomes and Bandwidth

Table 4.A.1: Impact of the Rural Hardship Allowance on Placebo Outcomes

	Log Maternal Beds			Log In-patient Beds			Log Catchment Pop		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Eligible	0.236 (0.235)	0.475 (0.443)	0.300 (0.209)	0.0309 (0.260)	0.494 (0.481)	0.0657 (0.247)	0.126 (0.243)	-0.232 (0.383)	0.0942 (0.204)
Observations	165	165	98	154	154	90	167	167	96
Mean Ineligible	0.990	0.990	0.899	1.704	1.704	1.700	8.845	8.845	8.846
Bandwidth (km)	10	10	7.5	10	10	7.8	10	10	7.3
Polynomial	1	3	1	1	3	1	1	3	1
Interaction term	Yes	No	No	Yes	No	No	Yes	No	No

*Notes:* The table contains reduced form estimates of the impact of rural hardship allowance eligibility on placebo outcomes: Maternal Beds, In-patient Beds and Catchment Population per health facility, respectively. Heteroscedasticity robust bias-corrected coefficient estimates and standard errors, estimated using the procedure in Calonico et al. (2014, 2017), are shown (the standard errors are shown in parentheses). Within each panel, each column shows estimates utilising a different regression specification. Within each panel, columns 1, 4 and 7 show a specification with a first-order polynomial of the running variable, a 20km bandwidth (10km on each side of the eligibility threshold) and an interaction term of the running variable and eligibility indicator. Similarly, columns 2, 5 and 8 show a specification with a third-order polynomial, a 20km bandwidth and no interaction term. Columns 3, 6 and 9 show a specification with a first-order polynomial, different data-driven, coverage error rate minimising bandwidths, each estimated using the procedure in Calonico et al. (2014, 2017) and no interaction term. All specifications utilise a triangular weighting kernel and include district fixed effects;  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Table 4.A.2: Impact of Eligibility on Health Facility Staffing: Robustness to Bandwidth Selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
<b>Panel A: Log Health Workers</b>											
Eligible	0.441	0.405	0.423*	0.433*	0.446**	0.452**	0.456**	0.429**	0.407**	0.372*	0.346*
	(0.268)	(0.252)	(0.241)	(0.225)	(0.210)	(0.200)	(0.199)	(0.197)	(0.195)	(0.192)	(0.190)
<b>Panel B: Log Wage Bill</b>											
Eligible	0.430	0.426*	0.452*	0.459**	0.466**	0.468**	0.468**	0.438**	0.413**	0.375**	0.347*
	(0.269)	(0.252)	(0.238)	(0.220)	(0.206)	(0.196)	(0.195)	(0.194)	(0.192)	(0.190)	(0.190)
Polynomial	1	1	1	1	1	1	1	1	1	1	1
Observations	79	100	115	129	146	171	183	199	216	234	250
Bandwidth (km)	5	6	7	8	9	10	11	12	13	14	15

*Notes:* The table contains reduced form estimates of the impact of rural hardship allowance eligibility on the stock of health workers, in panel A and the total wage bill per health facility, in panel B, respectively. The total wage bill per health facility is measured by the natural logarithm of total monthly wage bill per health facility, measured in Zambian Kwacha ZMW; 1 USD approximately equals 22 ZMW. Each column contains estimates from a similar regression specification, with different bandwidths between 5km and 15km. Heteroscedasticity robust bias-corrected coefficient estimates and standard errors, estimated using the procedure in Calonico et al. (2014, 2017), are shown (the standard errors are shown in parentheses). All specifications include district fixed effects. The specification shown utilises a first-order polynomial of the running variable.  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\* $p < 0.001$ .

# Declaration

I hereby declare that I have produced my doctoral thesis, "*Essays on Financial Incentives to Healthcare Providers*," independently and without external assistance, and that I have made a significant contribution as co-author to other scientific articles. I have identified all word-for-word quotations of other authors, as well as comments based closely on other authors' ideas, and I have cited the sources according to the guidelines I received.

Date

Signature