

©Copyright 2025

Theradapuzha Varghese Ninan

Essays on the Drivers and Impacts of Clean Cooking Energy Transitions in India

Theradapuzha Varghese Ninan

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Rachel M. Heath, Chair

Alan Griffith

Emma Riley

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays on the Drivers and Impacts of Clean Cooking Energy Transitions in India

Theradapuzha Varghese Ninan

Chair of the Supervisory Committee:

Rachel M. Heath

Department of Economics

Household air pollution from cooking with solid fuels remains one of the most significant health hazards in low- and middle-income countries. This dissertation comprises three chapters that explore the underlying reasons behind the persistent use of solid fuels for cooking and examines the potential consequences of transitioning to cleaner alternatives.

In chapter 1, I study the impact of providing access to clean cooking technology on learning outcomes of children. About a third of the global population relies on solid fuels for cooking with significant known impacts on health and time use. However, there is limited evidence on the impact of this reliance on further downstream outcomes like children's educational attainment. This paper is the first to examine the impact of gaining access to liquefied petroleum gas (LPG) as a fuel for cooking on children's learning outcomes. Leveraging the variation in the district level intensity of the Pradhan Mantri Ujjwala Yojana (PMUY), one of the world's largest clean cooking transition program, and nationally representative data on foundational learning from India, I find that increased access to LPG leads to improved learning outcomes in both reading and math for primary school-aged children. Specifically, a 30 percentage-point improvement in PMUY coverage leads to an improvement in overall learning of approximately 0.1 standard deviations. This effect is comparable in magnitude to the median impact of some of the most well-known interventions that directly target learning outcomes in developing countries. I also find larger positive impacts on learning for girls, households in the middle of the wealth distribution, and older children. I provide evidence for the fact that these effects are driven by improved health of children, which may allow them to participate more in activities that aid learning like attending school more regularly.

In chapter 2, I explore the link between living in a multi-generational traditional household structure and use of solid fuels for cooking in a patriarchal society. I do this by looking at the effect of the father-in-law's death on the choice of household cooking fuel in India. Using a difference-in-differences model with household fixed effects, I find that the probability of using biomass for cooking is lower by about 6 percent in households where a co-residing father-in-law died compared to a household where they did not. I also find that the probability of collecting every major solid fuel is lower as well. I provide evidence for this effect to be driven primarily by the father-in-law's preference for food that is cooked in the traditional way without the use of modern equipment.

In chapter 3, I examine whether providing information about the health risks of cooking with solid fuels for primary cooks and children can change beliefs and behaviors, and whether the identity of the information recipient within the household matters. We test this using a cluster randomized controlled trial with 2000 households in rural India where we inform either the (male) household head or the (female) primary cook about the actual health risks of cooking with solid fuels. An endline survey conducted 16 weeks later shows that information provision leads to belief updating, especially when the primary cook is informed. Information also transmits within households, both from the primary cook to the household head and vice-versa. However, despite these changes in beliefs, we find no evidence of increased adoption of clean cooking fuels or other mitigating behavior. One likely reason that we find in our analysis is that the changes in beliefs are concentrated among primary cooks with low intrahousehold bargaining power at baseline, who are unable to effect change.

Together, the three chapters of this dissertation offer novel and important evidence on the drivers and consequences of cooking with solid fuels. They also make significant contributions to the literature on children's learning, intrahousehold dynamics and decision-making and the role of beliefs in shaping behavior in the context of developing countries.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	v
Chapter 1: Cooking Up Success: The Impact of Access to Clean Stoves on Children’s Learning	1
1.1 Introduction	1
1.2 Background	6
1.3 Data	8
1.4 Empirical Strategy	14
1.5 Results	17
1.6 Mechanisms	29
1.7 Conclusion	32
Chapter 2: When Tradition Pollutes: Health and Environmental Costs of Multi-Generational Living in India	39
2.1 Introduction	39
2.2 Conceptual Framework	43
2.3 Data	46
2.4 Empirical Strategy	48
2.5 Results	50
2.6 Robustness	54
2.7 Mechanisms	60
2.8 Conclusion	64
Chapter 3: Intrahousehold Differences in Perceived Health Risks of Using Solid Fuels for Cooking	70
3.1 Introduction	70
3.2 Experimental Design	74
3.3 Data and Empirical Strategy	80
3.4 Results	81

3.5 Conclusion	91
Appendix A: Appendix for Chapter 1	96
Appendix B: Appendix for Chapter 2	117
Appendix C: Appendix for chapter 3	127
C.1 Information Script:	129

LIST OF FIGURES

Figure Number	Page
1.1 Number of districts that received various amounts of PMUY connection per household	9
1.2 Trends of learning ability between 2007 and 2022 for primary school children in India. Source: ASER Dataset	12
1.3 Event study style plot for overall learning	19
1.4 Event study style plot for reading ability	20
1.5 Event study style plot for math ability	20
1.6 Event Study style graph for overall learning ability for a subset of states for which parallel pre-trends hold for all years	22
1.7 Event study style graph for reading ability for a subset of states for which parallel pre-trends hold for all years	22
1.8 Event Study style graphs for Math ability for a subset of states for which parallel pre-trends hold for all years	23
1.9 Differential impact on girls and boys	27
1.10 Differential impact by wealth of the household	27
1.11 Heterogeneity by Age	29
3.1 Cumulative probability of the belief index for the primary cook and household head	78
A.1 Pre-treatment coefficient for overall learning	101
A.2 Pre-treatment coefficient for reading ability	102
A.3 Pre-treatment coefficient for Math ability	102
A.4 Event Study style graph for overall learning ability: Binary treatment	106
A.5 Event study style graph for reading ability: Binary treatment	107
A.6 Event Study style graphs for Math ability: Binary treatment	107
A.7 Event study style plot for overall learning with additional time varying controls	108
A.8 Event study style plot for reading ability with additional time varying controls	108
A.9 Event study style plot for Math ability with additional time varying controls	109
A.10 Synthetic difference in differences unit weights, time weights and parallel trends for the learning index	110
A.11 Unit weights, time weights and parallel trends for the reading index	110
A.12 Synthetic difference in differences unit weights, time weights and parallel trends for math index	111

A.13 Synthetic difference in differences unit weights, time weights and parallel trends for grade 5 reading	111
A.14 Synthetic difference in differences unit weights, time weights and parallel trends for grade 5 math	112
B.1 Bias in covariates in the original sample versus the matched sample	119
C.1 Number of diseases out of 6 correctly identified by PC and HH	128
C.2 Number of factors correctly identified as being less risky than solid fuels for mortality and morbidity	128

LIST OF TABLES

Table Number	Page
1.1 Comparison of high intensity PMUY districts and low intensity PMUY districts before and after the introduction of PMUY	13
1.2 Description of how the outcome variables used in the analysis were constructed . . .	15
1.3 Parallel Pre-Trends test using data from 2011-2014	17
1.4 Impact of PMUY coverage on learning	18
1.5 Parallel trends with a subset of 17 states	21
1.6 Impact of PMUY connections on learning for subset of 17 states	21
1.7 Synthetic difference-in-differences	24
1.8 Sensitivity analysis based on Kahn-lang and Lang using all pre 2016 data	25
1.9 Sensitivity analysis based on Kahn-lang and Lang using data from 2011-2014 and 2022	26
1.10 Impact of PMUY on school attendance	30
2.1 Summary Statistics: Households with FIL dead vs FIL not dead	47
2.2 Impact of the death of the FIL on solid fuel use	50
2.3 Impact of the death of the FIL on solid fuel collection	52
2.4 Impact of the death of the FIL on clean fuel use	52
2.5 Impact on solid fuel use with households where MIL died as the control group . . .	55
2.6 Impact on solid fuel collection with households where MIL died as the control group	56
2.7 Impact on solid fuel use with MIL dead as the treatment group	57
2.8 Impact on solid fuel collection with MIL dead as the treatment group	57
2.9 Impact on clean fuel use with MIL dead as the treatment group	58
2.10 Impact on solid fuel use: Matched difference-in-differences	59
2.11 Impact on solid fuel collection: Matched difference-in-differences	60
2.12 Impact on clean fuel use: Matched difference-in-differences	61
2.13 Impact of FIL death on other kitchen labor saving technology	62
2.14 Impact of FIL death on technology not used by adult women	62
3.1 Intrahousehold Difference in Belief Index	77
3.2 Effects of Information Intervention	82
3.3 Effect of Information Intervention on PC's Beliefs	83
3.4 Effect of Information Intervention on PC's Beliefs	84

3.5	Effect of Information Intervention on HH's Beliefs	85
3.6	Effect of Information Intervention on HH's Beliefs	85
3.7	Effect of Information Intervention on Beliefs about Household Head	88
3.8	Effect of Information Intervention (Secondary Outcomes)	88
3.9	Heterogeneity in beliefs about risks to PCs and Children by PC's baseline empowerment index	90
3.10	Heterogeneity in beliefs about risks to Household Head by PC's baseline empowerment index	90
3.11	Heterogeneity by PC's baseline empowerment index	91
A.1	Parallel Pre Trends test using data from 2007-2014	97
A.2	Parallel Pre Trends test using data from 2007-2014	98
A.3	Parallel Pre Trends test using data from 2011-2014	99
A.4	Impact of PMUY connections on learning for years between 2011-2016 only	100
A.5	Impact of PMUY connections on learning	103
A.6	Impact of PMUY connections on learning: Binary Treatment	104
A.7	Impact of PMUY connections on learning: Binary Treatment	105
A.8	Heterogenous impact of PMUY connections on learning by gender: Binary treatment	113
A.9	Heterogenous impact of PMUY connections on learning by wealth: Binary treatment	114
A.10	Impact of PMUY connections on attendance: continuous treatment	115
A.11	Parallel pre-trends test for out of school children using data from 2007-2014	115
A.12	Impact of PMUY connections on learning on out-of-school children	116
A.13	Impact of PMUY connections on Enrollment (=1 if child enrolled in school)	116
B.1	Impact of the death of the FIL on solid fuel use with additional time varying controls	120
B.2	Impact of the death of the FIL on clean fuel use with additional time varying controls	120
B.3	Impact of the death of the FIL on solid fuel collection with additional time varying controls	121
B.4	Impact of the death of only the FIL on solid fuel use	121
B.5	Impact of the death of only the FIL on clean fuel use	122
B.6	Impact of the death of only the FIL on solid fuel collection	122
B.7	Summary Statistics: Households with FIL dead vs MIL dead	123
B.8	Impact of the death of the FIL on income and household size	124
B.9	Impact of the death of the FIL on household splitting	125
B.10	Impact of household splitting on solid fuel use	125
B.11	Impact of the death of the FIL on solid fuel collection	126
B.12	Impact of the death of FIL on DIL's autonomy	126

C.1 Underestimation of Primary Cook's Symptoms	127
C.2 Underestimation of Children's Symptoms	127
C.3 Summary Statistics and Balance Test	139
C.4 Belief Index Components	140
C.5 Solid Fuel Use Index Components	141
C.6 Clean Fuel Use Index Components	141
C.7 Effect of Information Intervention on PC's Beliefs	142
C.8 Effect of Information Intervention on PC's Beliefs (PC's Symptoms)	143
C.9 Effect of Information Intervention on PC's Beliefs(Children's Symptoms)	143
C.10 Effect of Information Intervention on HH's Beliefs	144
C.11 Effect of Information Intervention on HH's Beliefs (PC's Symptoms)	145
C.12 Effect of Information Intervention on HH's Beliefs(Children's Symptoms)	146
C.13 Effect of Information Intervention on Solid Fuel Use	147
C.14 Effect of Information Intervention on Clean Fuel Use	148
C.15 Heterogeneity by HH's baseline belief index	149
C.16 Heterogeneity by PC's baseline belief index	150
C.17 Heterogeneity by PC's education	150
C.18 Heterogeneity by PC's education	151
C.19 Heterogeneity by Child's Gender	151
C.20 Heterogeneity by Child's Gender	152

ACKNOWLEDGMENTS

I am deeply grateful to my chair, Rachel Heath for not just her outstanding academic guidance but for also being an exceptionally kind and supportive advisor. Rachel, your mentorship was critical to the completion of this dissertation- both for the depth of your feedback and for the generous support you offered during challenging times. Your firm yet gentle encouragement to continually improve my research is the foundation upon which this work rests.

I also want to thank my committee members Alan Griffith and Emma Riley. Alan, thank you for being so generous with your time, your invaluable technical feedback, and for encouraging me to showcase my work to a broader audience. Emma, I am especially thankful for all that you taught me about designing and running field experiments. I would have never had the confidence to co-lead one without your guidance. I also want to thank Isabelle Cohen, for her feedback on my presentations and for being a wonderful mentor and collaborator. I am grateful to Michelle Foshee for her kindness and the generous time she devoted to supporting me. I owe a special thanks to Professor Krishnendu Gosh Dastidar and Professor Subrata Guha - without your early encouragement, I would have never pursued a PhD.

None of this would have been possible without my wife Ananya, who is not just my life partner but also my closest intellectual companion. Your ability to challenge me while maintaining unwavering belief in my work, your strength during our hardest moments, and your willingness to make space for my work, even when it meant taking on more for yourself, sustained me throughout this journey. Your support has been transformative, selfless and unconditional.

I am grateful to my parents for their countless sacrifices and for supporting me unconditionally, always encouraging me to follow my own path without pressure. To my nephew, Tarun, and my niece, Diveesha, the joy that you brought into my life made even the toughest days easier.

Finally, I want to thank sister and brother-in-law, Ramiya and Arun, Ananya's parents; my India family- Balan Maman, Maami, Bharat, Bharathi, Nikhil, Pranav and Aakash and my Seattle

family - Deepesh, Prerna, Shashank, Siva, Gayatri, Amandeep, and Debasmita, for your unwavering support throughout this long and often trying journey.

DEDICATION

To my dear wife, Ananya, and my beloved parents, Shibu and Mini without whom this dream
would never have found its wings

Chapter 1

COOKING UP SUCCESS: THE IMPACT OF ACCESS TO CLEAN STOVES ON CHILDREN'S LEARNING

1.1 Introduction

Understanding the determinants of early life human capital formation is a key question in development economics. Apart from traditional inputs to education, such as infrastructure (Duflo, 2001, 2004), teacher quality (Rockoff, 2004; Rivkin et al., 2005), class size (Hanushek, 1998; Ehrenberg et al., 2001) and technology (Ball et al., 2006), there is an emerging consensus on the role of environmental factors like climate change (Prentice et al., 2024) and the accompanying abnormal rainfall, extreme temperatures (Park et al., 2020) and pollution (Balakrishnan and Tsaneva, 2021; Wen and Burke, 2022). However, evidence on the impact of indoor/household air pollution on educational achievement is lacking, especially in developing countries. This paper fills this gap by examining the impact of a program in India, the Pradhan Mantri Ujjwala Yojana (PMUY) that dramatically expanded access to non-polluting sources of cooking energy and technology, on the learning outcomes of primary school children in India.

Cooking with solid fuels like firewood, dung and agricultural residue is one of the biggest health hazards in developing countries like India (Bennitt et al., 2021; Naghavi et al., 2024). Shifting to non-polluting sources of cooking, either exclusively or even partially, could improve children's learning outcomes. For instance, if children are less sick, they will have more time to focus on school work and attend school more regularly. If mothers are less sick and have more time, then they will be able to better aid their children's learning. In May 2016, the Indian government launched

PMUY with the aim to provide subsidized connections of Liquefied Petroleum Gas (LPG)- the most commonly used non-polluting cooking fuel in India, to socio-economically marginalized households. This program is unique because most other large-scale programs in developing countries that have encouraged households to switch from traditional cooking methods have been to other polluting technologies like improved cookstoves or kerosene stoves.¹ The few programs that have focused on non-polluting technology like gas and electricity (for example, a program in Peru introduced in 2009) have been at a much smaller scale in terms of number of beneficiaries, and provide lower financial support than PMUY. More importantly, none of these programs have been evaluated for their impact on educational outcomes.

This paper uses the variation in the intensity of the PMUY program across districts in India, combined with large, nationally and district-representative repeated cross-sectional surveys on children's learning, to estimate plausibly causal effects of increase in LPG access on learning outcomes of children. I use a difference-in-differences (DiD) strategy with two-way (year and district) fixed effects to assess this causal impact. I find a positive impact of the program on learning outcomes of children. Specifically, I find that a 10 percentage-point increase in PMUY LPG connections in a district leads to a significant 0.032 standard deviation improvement in an index of children's learning. I also find almost equivalent increases in individual math and reading skills. For example, the math and reading indices show an improvement of 0.032 standard deviations and 0.027 standard deviations respectively, for a 10-percentage-point increase in PMUY LPG connections. In addition, the proportion of class 5 students that can perform division and read at the grade 2 level each show an improvement of 2.6 percentage points. Given that the median district had a PMUY coverage of 30% , the effects observed in this study are as large as the median of those observed in interventions that are directly targeted at improving learning in developing countries ([Evans and Yuan, 2022](#)).

A key assumption for unbiased estimation of the DiD parameter is the parallel trends assumption.

¹Indonesia had a large scale program to encourage transition from kerosene to LPG. Kerosene however, emits much less PM2.5 compared to solid fuels which is the strongest predictor for mortality and morbidity caused due air pollution

Even though the assumption cannot be directly tested, a standard way to show that it is plausible would be to test whether trends were parallel for districts with different intensities of PMUY coverage, in the periods before treatment occurred. I find that trends between the ‘treatment’ and ‘control’ groups are parallel for the years immediately before the treatment but not parallel for the earliest years for which data is available. I offer a plausible explanation for why this might be the case and contend that it does not pose a threat to the identification strategy. Specifically, I argue that the Right to Education act, implemented in 2009 and 2010, was a massive shock to the Indian education sector which in (only) the years of its implementation caused differential trends between high-intensity and low-intensity PMUY districts. However, after this initial shock, the groups transitioned to parallel trends as PMUY approached its launch and that there are no shocks other than PMUY that would deviate the education outcomes from parallel trends.

In addition, I conduct other analyses to verify this claim and confirm the robustness of my results. These include (i) looking at a subset of states for which the trends are parallel in all of the pre-period, (ii) a synthetic difference-in-differences approach as prescribed by [Arkhangelsky et al. \(2021\)](#), (iii) Adding group trends to extrapolate pre-trends of the treatment group to the post period ([Kahn-Lang and Lang, 2020](#)) and (iv) Estimating other specifications with time-varying controls which rely on a conditional parallel trends assumption. All these analyses give results that are qualitatively similar to my main results. The effect sizes are also similar to my main DiD specification (they are slightly lower for synthetic DiD)

I also look at heterogeneous impacts of the program by wealth, gender and age. I find that the program has a larger positive impact on children in those households that are in the middle of the wealth distribution, on girls and older children. Higher impact on children belonging to households in the middle of the wealth distribution is consistent with the fact that not only were these the households that were targeted under PMUY but are also those households that could afford to purchase regular LPG refills. Higher impact on girls is consistent with the fact that in a largely patriarchal society like India, girls are likely to be indoors more and thus would benefit more from

a reduction in household air pollution.

In terms of mechanism, I provide suggestive evidence for why the dominant channel is likely to be an improvement in health outcomes of children leading to an increase in activity that is complimentary to learning such as attending school instead of channels like a change in the child’s or mother’s time use, improvement in their cognition or improvement in teachers’ health. First, our main specification focuses on enrolled children aged 10 and under. Children aged 10 and under are less likely to be helping with household chores like collecting and preparing fuel or cooking. Second, I find an increase in school attendance of children and no impact on teachers’ attendance. Third, I find no impact on the learning of children who are not enrolled in school. If higher LPG worked through improved cognition only, we should have seen a positive impact in these students as well. I also argue that the effect due to primary cook’s time use is probably secondary due to evidence that cooking with clean fuels saves only about 5 minutes per day in India for the kind of households we are interested in ([Afridi et al., 2023](#)).

Recent and quickly growing literature has evaluated the impact of air pollution on learning and cognition among children in both developed and developing countries. For example, [Grineski et al. \(2020\)](#) find that kindergarten exposure to air pollution is linked to lower reading, math, and science competencies in the US, while [Palacios et al. \(2022\)](#) show that indoor air pollution in schools in the Netherlands was associated with lower test scores. In developing countries, [Balakrishnan and Tsaneva \(2021\)](#) find that high ambient air pollution causes lower learning in India by up to 2.4 percentage points and [Requia et al. \(2022\)](#) report that a $10 \mu\text{g}/\text{m}^3$ increase in the long-term average PM2.5 around Brazilian schools is associated with a 3 point reduction in test scores. However, no study has yet examined the impact of *household* air pollution on children’s learning, particularly in the context of developing countries. This paper is the first to address this gap, as cooking with solid fuels is the largest contributor of *household* air pollution in low and middle income countries ([Shupler et al., 2018](#); [Somanathan et al., 2022](#); [Berkouwer and Dean, 2024](#)).

The health effects of *household* air pollution might differ from those of ambient air pollution due to significantly higher exposure to harmful pollutants, such as PM2.5 particles, associated with it. For example, [Somanathan et al. \(2022\)](#) find that during cooking hours, PM2.5 concentrations in the kitchen could be as large as 4 times the ambient concentration and around 25 times above the WHO prescribed limit. [Berkouwer and Dean \(2024\)](#) suggest that average pollutant concentrations are associated with negative health effects rather than short-term spikes. However, their results pertain to adults and the effects on children, the focus of this paper, may be different. Furthermore, their paper examines the switch from more polluting to less polluting cooking technology using the same fuel. In contrast, this paper studies a shift from a polluting to non-polluting technology which is likely to result in a more substantial reduction in pollution spikes. Importantly, their paper is set in urban Kenya, whereas ours looks at rural India, where cooking with solid fuels is expected to be one of the largest contributors to ambient air pollution as well. This paper thus contributes to the literature documenting the negative impacts of cooking with solid fuels. While most studies in this area have focused on the impacts on health and time use ([Smith et al., 2000](#); [Smith, 2000, 2013](#); [Imelda, 2020](#); [Yun et al., 2020](#); [Krishnapriya et al., 2021](#); [Frostad et al., 2022](#); [Afridi et al., 2023](#); [Chowdhury et al., 2023](#)), there is very limited evidence, however, on its effects on further downstream outcomes like learning which this paper explores.

This paper also adds to the literature that identifies the factors for low academic achievement in low-income countries ([Glewwe and Muralidharan, 2016](#)). Such low achievement has been linked to slower GDP growth and human development ([Hanushek and Woessmann, 2012](#); [Hanushek, 2012](#)). Approximately one-third of the world’s population still relies on solid fuel for cooking, with most of these households located in the poorest regions of the world, like Sub-Saharan Africa and South Asia. This paper suggests that the use of solid fuels for cooking in poor regions contributes to lower learning outcomes, which in turn perpetuates poverty. Thus, a policy that improves access to clean cooking technology has the added benefit of improving human capital accumulation.

The rest of the paper is organized as follows. Section [1.2](#) outlines the background, in particular, it

describes cooking technology in the Indian context and the PMUY scheme. Section 1.3 describes the data used in the paper and provides important summary statistics with respect to the PMUY scheme and learning in India over the years. Section 1.4 outlines the identification strategy and its limitations. Section 1.5 discusses the main findings, robustness checks, and important treatment effect heterogeneity, and how that renders support to our findings. Section 1.6 explores potential mechanisms, and Section 1.7 concludes.

1.2 Background

1.2.1 Cooking Energy in India

Historically, India has had one of the largest populations reliant on solid fuels for cooking world-wide. The most commonly used solid fuels in India are firewood, animal dung and agricultural residue/twigs. In 2015, it had more than 650 million individuals (or about 50 % of the population) who were primarily reliant on solid fuel for cooking. Out of these more than 90 % lived in rural areas (or about 75 % of the rural population) (Vardell, 2020). Even though successive governments recognized this as a serious health hazard, the focus until the late 2000s was to encourage the use of *Improved Cookstoves* and *Biogas*. These initiatives hardly took off and had very limited success in reducing the dependence on traditional cooking (Jain et al., 2018).

Since the mid 2010's however, there has been an enthusiastic push to encourage the adoption of gas (mainly Liquefied Petroleum Gas (LPG)) to replace biomass. The government introduced several schemes like the *PAHAL* scheme, the *Smokeless Villages* initiative, the *Give it up* initiative leading up to the *Pradhan Mantri Ujjwala Yojana*. In 2022, the proportion of households using biomass as primary cooking fuel had halved in comparison to 2015 to 24.7% and the number of people using LPG had almost doubled, mostly driven by the rural population (Vardell, 2020). About 62% of the rural population was primarily using gas and the percent using biomass had gone down to

37%. Overall, the steps undertaken by the government since 2014, particularly the *Pradhan Mantri Ujjwala Yojana*- which is described in more detail in the next subsection has significantly changed the landscape of cooking fuel energy access and availability in India.

1.2.2 *Pradhan Mantri Ujjwala Yojana*

The Government of India launched the *Pradhan Mantri Ujjwala Yojana (PMUY)* in May 2016 to improve access to and encourage the use of Liquefied Petroleum Gas (LPG), the most commonly used non-polluting cooking fuel in India. The scheme aimed to provide deposit free LPG connections to adult women belonging to below poverty line (BPL) households as identified by the Socio- Economic caste census (SECC) of 2011. The scheme provided a one-time assistance towards the security deposit of the LPG cylinder, cost of the pressure regulator and hose and waived off installation charges. This amounted to Rs. 1600 in 2019 (roughly \$20). In addition, Oil Marketing Companies provided an interest free loan for the cost of the stove and the first LPG refill, which was to be recovered from the subsidy that was provided on LPG refills. The amount of the loan was Rs. 1507 (approx \$18) in 2019. In order to get the LPG connection, the beneficiary from the eligible BPL family had to fill up a prescribed form, apply to the nearest LPG distributor, complete the *Know your Customer* requirements and provide other documentation. On receipt of the application, the LPG distributor was required to verify the application details against the SECC database and carry out field verification of the information. This was done in order to ensure the genuineness of the beneficiary and also to ensure that the family of the applicant does not already possess an LPG connection (CAG, 2019).

Another component of the program was to enhance the LPG supply infrastructure, such as expanding the network of LPG distributors and establishing new LPG bottling facilities, particularly in regions with low LPG penetration and rural areas. ² Although the goal of 10,000 new distribution

²Refer to the case study by *Social Cops*: <https://socialcops.com/case-studies/finding-the-best-locations-for-10000->

centers by 2019 was not met, approximately 6,100 new distribution centers were commissioned. Since, LPG distributors play a crucial role in determining the eligibility of households to get a PMUY connection and distribute regular refills- their location and accessibility, in addition to SECC status of the household, their awareness about the program and their ability to provide documentation determined whether the household was able to get a PMUY connection ([CAG, 2019](#)).

Initially, the target was to provide 50 million LPG connections under the program, which was later revised to 80 million households in 2018 ([CAG, 2019](#)). The target of 80 million was achieved in September 2019, seven months ahead of schedule ([Utpal Bhaskar, livemint](#)). This makes PMUY the world’s largest clean cooking transition program. In August 2021, the target was again revised to cover 10 million more BPL households and now the government would also cover the cost of the first refill and hotplate/stove (the program was renamed Ujjwala 2.0).

Overall the program has been very successful in increasing the access of LPG in India. LPG coverage was about 56% percent in 2015 before the program and had increased to more than 99% in 2022 according to the government’s estimates. Out of all the active customers in 2022, around 29% were PMUY customers ([Agarwal et al., 2023](#)).

1.3 Data

1.3.1 PMUY data

The data on the number of PMUY LPG connections released per district is obtained from the PMUY website maintained by the government of India. ³. This data is available for all districts

new-lpg-centers/

³One can access the data from [here](#). In order to download the data one needs to receive a one time password on an Indian cell phone number. The data shows number of PMUY connections released with a lag on approximately

of India. In order to estimate the number of connections released per household in a district, I use the National Family Health Survey conducted in 2019-2021 (NFHS-5) to get the average household size in a district and use estimates of district population from [Wang et al. \(2021\)](#).⁴

Figure 1.1 depicts the number of PMUY connections released per 100 households across districts, with data up to August 2021. We see that there is a large variation in the proportion of households that received a PMUY connection, with Karaikal in Puducherry receiving only 0.6 connections per 100 households and Jaisalmer in Rajasthan receiving about 79 connections per 100 households. The median district received about 29 connections per 100 households.

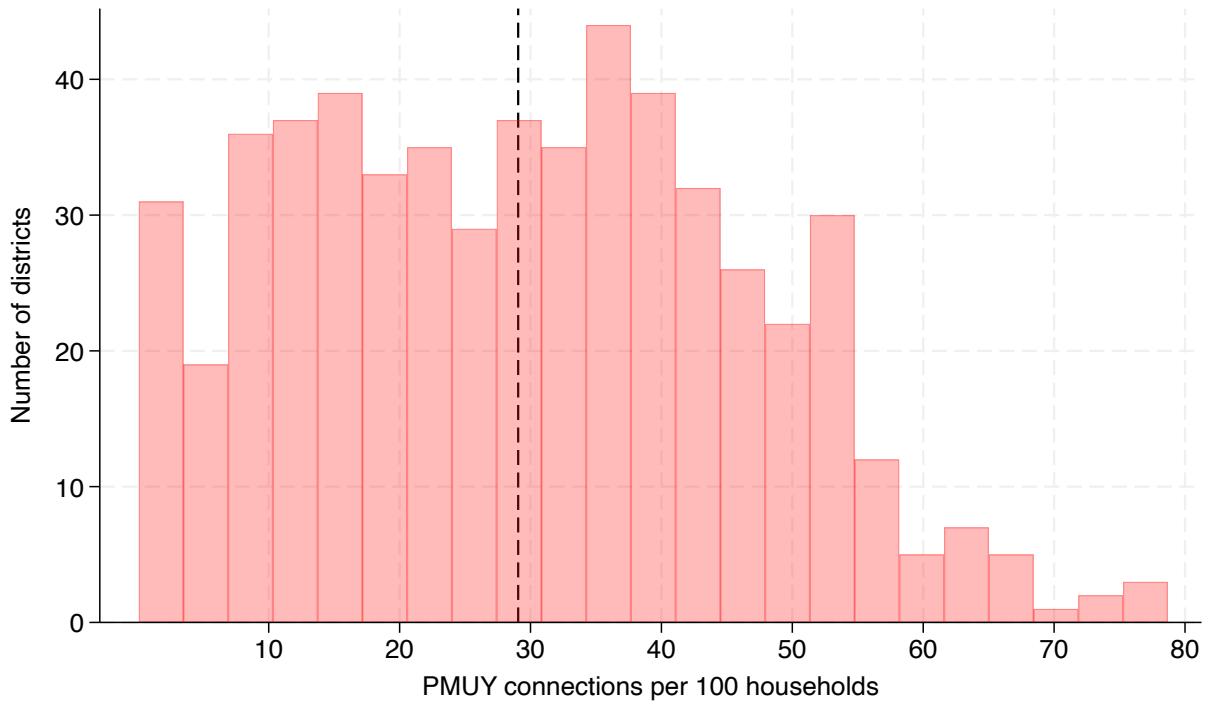


Figure 1.1: Number of districts that received various amounts of PMUY connection per household

a month. I use the for the number of connections released until the end of the first phase of the program

⁴Results are very similar if use the number of connections released per person instead of per household

1.3.2 Learning Data

I use data on learning ability from the Annual Status of Education Report (ASER), which tests children between ages 5 and 15 on math, reading and English ability. I use the data from 2007 to 2022.⁵ The ASER survey is representative at the district level but is conducted only in rural areas.⁶ It is conducted at home and thus includes both in-school, out-of-school and never enrolled children. It is a ‘floor test’ which tests children on foundational ability and all children are assigned the same test regardless of their age or class. For example, for math, students are tested on recognizing single digit numbers, two digit numbers, subtraction and three by one division. Children in India are expected to attain these skills by class 3 (age 8 if they start school on time and are not held back). I look at only children between ages 5 and 10 since the focus of this paper is on early life human capital accumulation. In addition, as stated above the tests assigned in ASER are not meant to evaluate grade specific competencies and are not appropriate to test the learning outcomes of older children.

Figure 1.2a and figure 1.2b show the proportion of children aged 5-10 who can perform various tasks in reading and mathematics. Notably, there is a steep decline in learning after 2010. As has been documented in [Shah and Steinberg \(2019\)](#) one particular hypothesis explaining this decline is the implementation of the Right to Education Act (RTE), which was passed in the Indian parliament in August, 2009 and implemented across India from the 2010 academic year. RTE made education for children 6-14 a fundamental right, thus mandating free and compulsory education. [Shah and Steinberg \(2019\)](#) find that overall learning ability dropped significantly after the implementation

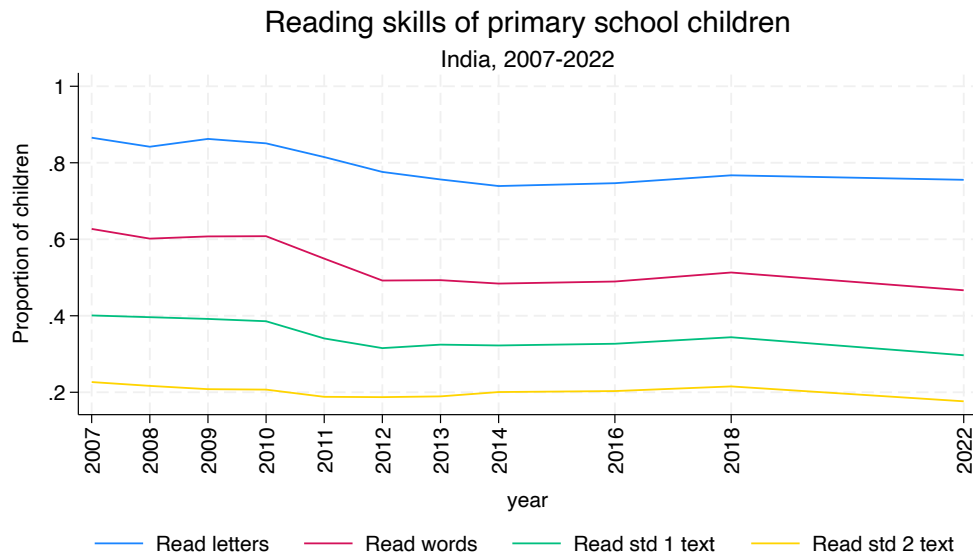
⁵ASER data is available publicly for all years between 2007 and 2014, and for 2016, 2018 and 2022. The English ability tests were conducted only in 2007, 2009, 2012, 2014, 2016 and 2022

⁶The PMUY data follows the most recent district definitions, while ASER data from 2007 to 2014 uses the 2001 Census district definitions and sampling frame. Starting in 2016, ASER adopted the 2011 Census district definitions. To maintain consistency, I assign newer districts to their corresponding parent districts based on the 2001 Census definitions. For districts that were formed from two or three older districts, I assign them to the parent district that contributed the most *Tehsils* (administrative sub-units). I drop the states of Andhra Pradesh, Arunachal Pradesh and some other districts from the analysis since the rule described above could not be applied given that there were cases where parts of the same *Tehsils* contributed to multiple newer districts. I use 564 out of 601 districts for the final analysis.

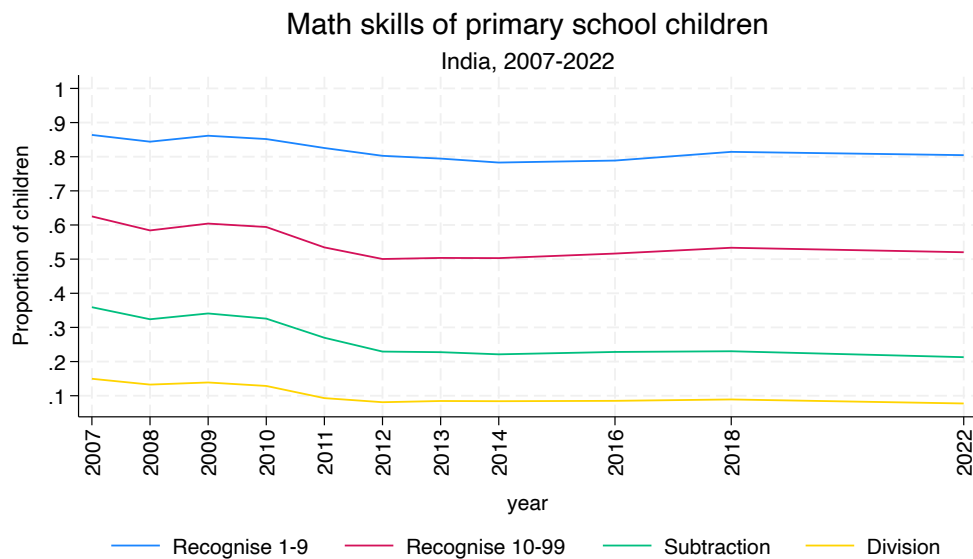
of RTE across India despite an increase in enrollment, and an increase in inputs such as improved student-teacher ratios, teachers' education levels and toilets for girls (a proxy for school infrastructure). They hypothesize that negative peer effects due to increased enrollment, the 'no-detention' policy of the RTE and the way students were evaluated, might have caused this decrease in learning. This is an important point that I will return to later while discussing the identification strategy. Importantly, learning post-2010 had not recovered to pre-RTE levels until 2022 although there has been some improvement.

1.3.3 Summary Statistics

Table 1.1 shows summary statistics and differences between what I define as *high intensity* and *low intensity* PMUY districts before and after PMUY. Since PMUY was a national program, there were no districts that did not receive any connections. Thus, I divide the districts into two groups: *high intensity* districts are the ones that received more connections per household than the median district and the *low intensity* districts are the ones that received less connections per household than the median district. Table 1.1 shows that in *high intensity* districts children are younger and thus, enrolled in lower grades. This could be as a result of higher fertility rates in these districts. Additionally, there was a lower likelihood of female children in the pre-period in *high intensity* districts, potentially due to son preference. They are also more likely to not be enrolled in school and to have parents who have not gone to school. These districts also tend to be poorer in terms of assets owned and services available to them like television, mobile/phone, toilet and an electricity connection. This is not surprising, since PMUY was targeted towards households that were below the poverty line or belonging to 'backward' castes and other marginalized groups. Thus, since the distribution of PMUY connections was not random, I cannot simply compare households in *high intensity* and *low intensity* districts in the post period to identify the causal impact of PMUY on learning. I will use a difference-in-differences strategy with district fixed effects to identify this impact.



(a) The proportion of students in grades 1 to 5 who can perform readings tasks like recognising letters and words, reading a grade 1 level text and reading a grade 2 level text.



(b) The proportion of students in grades 1 to 5 who can perform various mathematics tasks like recognizing single digit and two digit numbers, performing a 2 digit by 2 digit subtraction and performing a 3 digit by 1 digit division.

Figure 1.2: Trends of learning ability between 2007 and 2022 for primary school children in India.
Source: ASER Dataset

Table 1.1: Comparison of high intensity PMUY districts and low intensity PMUY districts before and after the introduction of PMUY

Variable	(1) Pre PMUY High Intensity	(2) Pre PMUY Low Intensity	(3) Post PMUY High Intensity	(4) Post PMUY Low Intensity	(5) (1)-(2)	(6) (3)-(4)
Female	0.461 (0.498)	0.471 (0.499)	0.488 (0.500)	0.489 (0.500)	-0.011*** (0.002)	-0.001 (0.002)
Age of Child	9.536 (3.727)	9.679 (3.715)	9.518 (3.797)	9.571 (3.767)	-0.144*** (0.025)	-0.053** (0.026)
Class of Child	4.919 (3.192)	5.257 (3.277)	5.209 (2.935)	5.409 (2.966)	-0.338*** (0.035)	-0.200*** (0.032)
Not Enrolled	0.021 (0.142)	0.007 (0.085)	0.029 (0.168)	0.013 (0.112)	0.013*** (0.001)	0.016*** (0.002)
Mother School	0.398 (0.489)	0.623 (0.485)	0.516 (0.500)	0.756 (0.430)	-0.225*** (0.014)	-0.240*** (0.014)
Father School	0.669 (0.471)	0.795 (0.403)	0.721 (0.449)	0.846 (0.361)	-0.126*** (0.009)	-0.125*** (0.009)
House has Electricity	0.598 (0.490)	0.871 (0.335)	0.851 (0.357)	0.953 (0.211)	-0.272*** (0.017)	-0.103*** (0.008)
Toilet	0.279 (0.449)	0.571 (0.495)	0.602 (0.489)	0.802 (0.399)	-0.292*** (0.019)	-0.199*** (0.014)
T V	0.339 (0.473)	0.661 (0.473)	0.463 (0.499)	0.775 (0.418)	-0.322*** (0.017)	-0.312*** (0.017)
Mobile/phone	0.629 (0.483)	0.744 (0.436)	0.860 (0.347)	0.911 (0.285)	-0.116*** (0.011)	-0.050*** (0.007)
Observations	2,841,697	2,394,905	1,081,859	841,621	5,236,602	1,923,480

The data is from ASER 2007-2022. Pre PMUY are the years from 2007 to 2014, Post PMUY are the years 2016, 2018 and 2022. Differences are obtained by a regression of the variable on the binary variable high intensity district=1 if the district had more PMUY coverage than the median district. Heteroskedasticity-robust standard errors clustered at the district level are in the parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1.4 Empirical Strategy

1.4.1 Main Specification

I use a difference-in-differences design to estimate plausibly causal impact of PMUY on the learning ability of students. Specifically, I estimate the following equation:

$$y_{idt} = \beta_1 PMUYconnec_d * Post_t + \theta_2 X_{dt} + \delta_t + \gamma_d + \epsilon_{id} \quad (1.1)$$

Where $PMUYconnec_d$ is the number of PMUY connections released per 10 households in the district (i.e. the intensity of treatment) until 2021, $Post_t$ is equal to 1 if year= 2022 and 0 if year = 2011, 2012, 2013 or 2014. X_{dt} is a vector of time-varying controls at the district level that include the proportion of repeaters and the proportion of children not enrolled in 2009 interacted with a year dummy obtained from District Information on School Education (DISE) data. These covariates allow for linear time trends for districts that had different educational attainment before the introduction of RTE.⁷ γ_d is the district fixed effect, δ_t is the year fixed effect and ϵ_d is the district clustered error. y_{idt} is the measured learning outcome. Table 1.2 describes how the outcome variables used in the analysis were constructed. β_1 is the coefficient of interest.⁸ It is important to note that merely getting a PMUY connection would not imply an improvement in learning outcomes. As noted earlier, LPG in India is sold in cylinders that would typically last a household of four between 4-5 weeks if used exclusively. Therefore, PMUY is expected to have a positive impact only if households get refills regularly. In that sense, β_1 is an intention-to-treat estimate.

⁷I also estimate equation 1.1 with additional time varying controls namely, the age of the child, the gender of the child, and state times a linear time trend. The results are very similar to the main specification.

⁸I also dichotomize the treatment variable $PMUYconnec_d$ into high intensity and low intensity districts (as defined before) and estimate equation 1.1 where treated districts are the high intensity districts.

Table 1.2: Description of how the outcome variables used in the analysis were constructed

Variable	Description
Math Index	An Anderson (2008) index of the indicator variables describing whether the child can recognise single and two digit numbers, do a two by two subtraction and a three by one division.
Read Index	An Anderson (2008) index of the indicator variables describing whether the child can recognise letters, read words, grade 1 level text and grade 2 level text.
Learning Index	An Anderson (2008) index of all the math and reading indicator variables described above.
Reading Grade 5	An indicator variable for whether a class V student can read a standard 2 level text.
Division Grade 5	An indicator variable for whether a class V students can do a three by one division.

1.4.2 Identification

For β_1 to be a causal estimate, the parallel trends assumption needs to hold. The assumption states that learning outcomes in districts with high PMUY coverage should trend identically to learning in districts with low intensity of PMUY in the post period, in the absence of treatment. This is an assumption about a counterfactual and therefore cannot be directly tested. A common approach to assess the plausibility of this assumption is to check whether the trends were parallel before treatment occurred. This can be done using ASER data from 2007 to 2014. As seen in Appendix [A.1](#) the coefficient of *High Intensity X Year* is negative and significant. This indicates that pre-trends are not parallel for all these years, with *high intensity* districts experiencing a larger decline in learning ability during the pre-PMUY periods. Appendix table [A.2](#) shows that this is true for the individual outcomes that constitute the index as well.

Appendix figures A.1, A.2 and A.3 show the pre treatment coefficients for math and reading outcomes with 2014 as the base year. The point estimate of the coefficients as well as the 95% confidence interval are reported. As can be seen from these plots, trends are parallel between 2007 and 2009 and between 2011 and 2014. However, learning declines more in the *high intensity* districts as compared to the low intensity districts between 2009 and 2011.

A plausible reason for this deviation from parallel trends could be the implementation of the RTE Act that was described in section 3. Learning had dipped across the country after the implementation of the Act. Given that the treatment districts were poorer and had a larger proportion of households belonging to marginalized groups (Scheduled Caste, Scheduled Tribes and Other Backward Class), it is plausible that the immediate adverse impact of the RTE was more pronounced in these districts. For example, the peer effects would be larger given that they started with higher out-of-school children, or that schools in these districts were less able to implement the new continuous evaluation system enshrined in the RTE. This could account for the deviation from parallel pre-trends immediately after the implementation of RTE.

Pre-trends are parallel for the outcomes if we only look at learning outcomes between 2011 and 2014 as seen in Table 1.3. It is thus important to note that if the new trend that we observe between 2011- 2014 holds in the counterfactual state of no PMUY, then the parallel trends assumption is a valid one and the DID estimate would be unbiased. On the other hand, if the negative trend in A.1 continued, then one would underestimate the actual effect. Both of these situations do not invalidate our analysis and the results presented.

The concern however, is if the differential decline in scores is followed by a situation where the differences in trends are reversed. For example, due to government action if the high intensity regions recover faster than the low intensity regions. It is unlikely that there was any such differential recovery. Firstly, there were no specific national government programs aimed at improving learning after 2010. Moreover, we cannot reject the hypothesis that DiD estimates are not different from

Table 1.3: Parallel Pre-Trends test using data from 2011-2014

	(1) Learning Index	(2) Math Index	(3) Read Index	(4) Division:Grade 5	(5) Read:Grade 5
Year (0-3)	-0.032*** (0.006)	-0.029*** (0.006)	-0.035*** (0.006)	0.002 (0.004)	0.003 (0.004)
High intensity x Year	0.009 (0.007)	0.012* (0.007)	0.003 (0.007)	0.006 (0.005)	0.005 (0.005)
N	907317	902482	905549	96710	96952
R2	0.08	0.08	0.08	0.11	0.10

This data only uses ASER data for years 2011- 2014. High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Year is a linear variable with year=0 for year 2007 and year=7 for year 2014. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * $p < 0.001$, ** $p < 0.05$, *** $p < 0.01$

zero between 2013 and 2016 for all outcomes as seen in table A.4. In addition, to provide further robustness of my results, in Section 1.5.2, I perform additional analyses to address these concerns. These involve using some of the most recent advancements in causal analysis like synthetic DiD, looking at the estimates for a subset of states for which pre-trends are parallel even from 2007-2010 and extrapolating the pre-treatment trend of the treated group into the post period.

1.5 Results

1.5.1 Impact on learning

Table 1.4 shows the impact of PMUY on learning in a DiD framework with years 2011 to 2014 (where pre-trends are parallel) as the pre-period and 2022 as the post period. We see improvements in both math and reading ability.

For mathematics, an increase of 1 PMUY connection per 10 households (or 10-percentage-points) corresponds to an increase of 0.032 standard deviations in the math index. The ability of class 5 students to perform division increases by 2.6 percentage points (or about 10% on a pre intervention average of 26%). For reading, a 10 percentage-point increase in PMUY connections leads to an

increase of 0.027 standard deviations in the reading index. The probability of a class 5 student being able to read a grade 2 text increases by 2.6 percentage points (or about 5.5% on a pre intervention average of 47%). Overall, the learning index increases by 0.032 standard deviations. All coefficients are significant at the 1% level. In the median district, which receives about 2.9 connections per 10 households, this translates to an effect size of roughly 0.09 standard deviations on the indices of learning and about 8 percentage points improvement in grade 5 reading and math.⁹ Appendix table A.5 shows the effect on the components of the indices. All outcomes are positive and significant with effect sizes ranging from about 1 % for recognising single digit numbers to about 10% for performing a three by one division. Figures 1.3, 1.4 and 1.5 show these results in event study style plots with 2014 as the base year along with a 95 % confidence interval.

Table 1.4: Impact of PMUY coverage on learning

	(1) Learning Index	(2) Math Index	(3) Read Index	(4) Division:Grade 5	(5) Read:Grade 5
Post	-0.080*** (0.016)	-0.049*** (0.015)	-0.112*** (0.016)	-0.056*** (0.010)	-0.114*** (0.013)
Post X PMUY connections	0.032*** (0.004)	0.032*** (0.004)	0.027*** (0.005)	0.026*** (0.003)	0.026*** (0.003)
Mean Dependent	-0.00	0.00	-0.00	0.31	0.50
Mean Dependent:Pre	-0.10	-0.10	-0.08	0.26	0.46
N	1,175,618	1,170,387	1,173,418	127,086	127,313

Treatment variable is the number of PMUY connections released in each district per 10 households until August 2021. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Appendix tables A.6 and A.7 where the treatment variables has been dichotomised. *high intensity* districts are defined as districts that had a higher PMUY connection per thousand households than the median district (i.e more than 2.9 PMUY connections per 10 households). The results show that learning in high intensity districts was higher than learning in low intensity districts for overall learning, math and english as before. Learning and Math indices were higher by about 0.1 standard deviations whereas the reading index was higher by about 0.9 standard deviation. Effects on other outcomes ranged from 3 percent for recognising single digit numbers to 30 percent for division.

⁹The average PMUY connections in *high intensity districts* is 4.38 per 10 households, this translates to an effect size of 0.14 standard deviations for the learning index

Appendix Figures A.4, A.5 and A.6 show these results in event study style plots with 2014 as the base year along with a 95 % confidence interval. Underscoring the point made in table 1.4 we see that overall learning, math and reading ability improved in 2022 vis-a-vis 2014 and is significant at the 95% confidence level. This analysis has the advantage of clearly classifying districts as ‘treatment’ or ‘control’ and depends on the parallel trends assumption for binary treatments over which there is a much larger consensus of what causal parameter is identified as opposed to the continuous case (Callaway et al., 2024). Appendix figures A.7, A.8 and A.9 show the effects with additional time varying controls. The effects are very similar to our main specification.

It is noteworthy that the median effect size of 234 studies, including randomized control trials and quasi-experimental studies evaluating educational interventions in low- and middle-income countries, is 0.1 standard deviations (Evans and Yuan, 2022)- closely aligning with the effect observed in this study. In fact, the effect sizes observed here are approximately 0.02 standard deviations higher than the median effect size of previous learning interventions in South Asia. This means that PMUY achieved outcomes comparable to those of targeted educational interventions, despite not being explicitly focused on education. Furthermore, the improvements in reading ability observed in this study translate to 0.47 to 0.62 equivalent years of education in low- and middle-income countries, based on the methodology described in Evans and Yuan (2019).

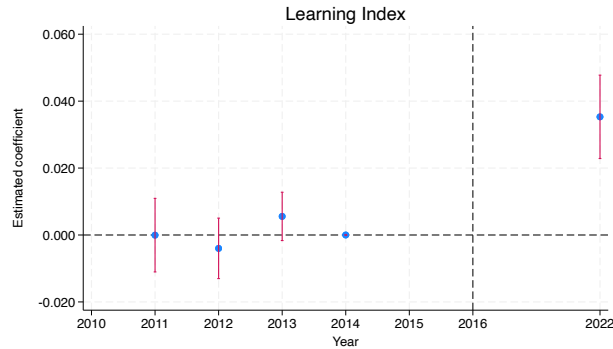


Figure 1.3: Event study style plot for overall learning

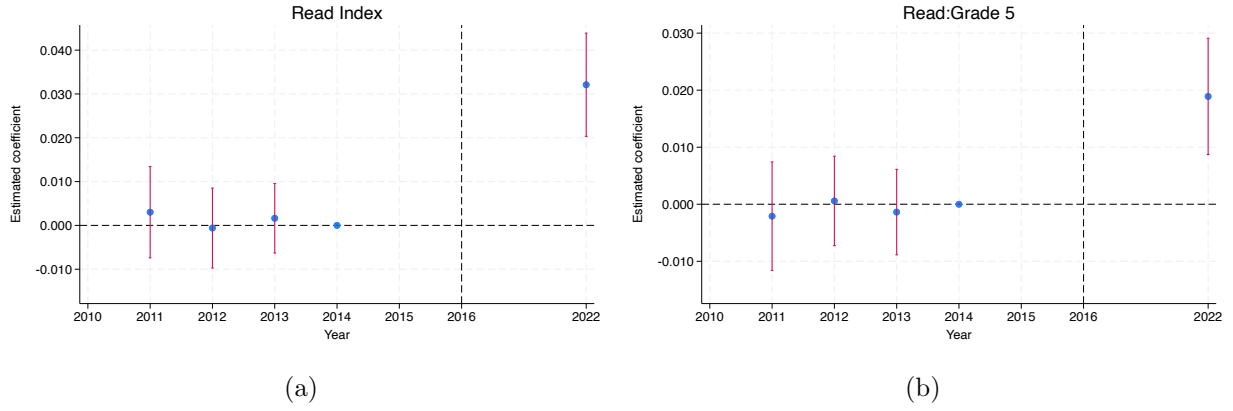


Figure 1.4: Event study style plot for reading ability

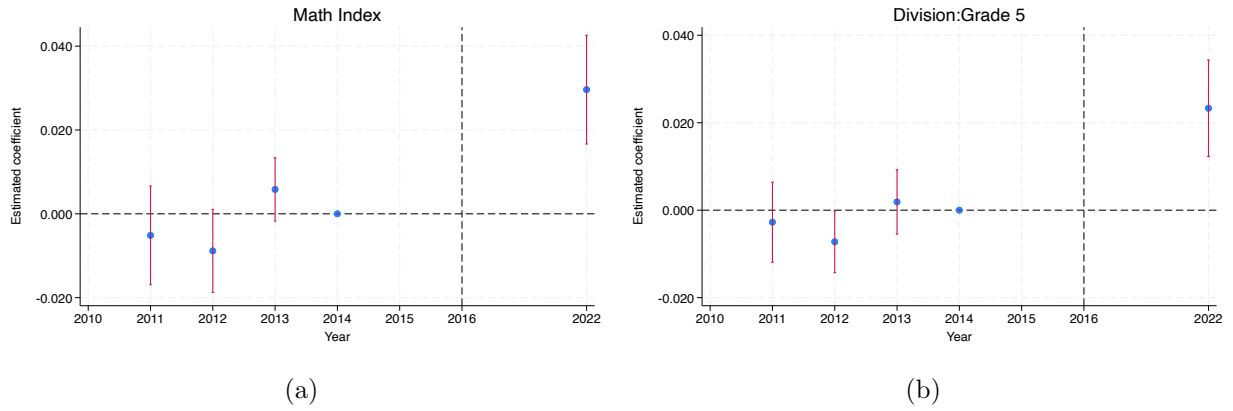


Figure 1.5: Event study style plot for math ability

1.5.2 Robustness

As noted above, the trends are not parallel if I include all the years for which I have data. If this implies that the trends would also not be parallel post-treatment absent the treatment then the estimates presented in the previous section would be biased. One needs to be particularly concerned about these estimates being biased upward. In what follows, I conduct some tests to convince the reader that such a positive bias is unlikely to be a significant concern for this paper's analysis and

thus render support to the identification strategy used.

A subset of states where parallel pre-trends hold

Table 1.5: Parallel trends with a subset of 17 states

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Year (0-7)	-0.032*** (0.003)	-0.031*** (0.004)	-0.029*** (0.003)	-0.006** (0.002)	-0.000 (0.003)
High intensity x Year	-0.002 (0.004)	-0.001 (0.004)	-0.004 (0.004)	-0.001 (0.003)	0.001 (0.003)
N	1,052,428	1,044,864	1,049,411	110,618	110,943

High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Year is a linear variable with year=0 for year 2007 and year=7 for year 2014. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Table 1.6: Impact of PMUY connections on learning for subset of 17 states

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Post	-0.144*** (0.024)	-0.115*** (0.024)	-0.168*** (0.024)	-0.074*** (0.016)	-0.116*** (0.017)
Post X PMUY connections	0.035*** (0.008)	0.035*** (0.008)	0.027*** (0.007)	0.030*** (0.005)	0.026*** (0.005)
N	1,204,048	1,196,360	1,200,942	128,178	128,511

Treatment variable is the number of PMUY connections released in each district per 10 households until August 2021. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

First, I consider a subset of 17 Indian states and territories where parallel pre-trends cannot be rejected for all the pre-2016 years for which I have data. ¹⁰ These states include 296 districts with the median district having 2.5 connections per 10 households. Table 1.5 shows that for these states, the hypothesis that the trends from 2007 to 2014 are parallel cannot be rejected. This can also be seen in the figures 1.6, 1.7 and 1.8 which shows event study plot for only these states- the 95%

¹⁰These states are Haryana, Jammu and Kashmir, Karnataka, Tamil Nadu, Punjab, Sikkim, Uttar Pradesh, Odisha, Meghalaya, Mizoram, Manipur, Nagaland and Tripura, Maharashtra, Dadra and Nagar Haveli and Daman and Diu. These states had 45 % of India's population according to the 2011 Census or roughly 540 million people.

confidence interval for all years before 2016 include 0. These states also saw a decrease in learning post Right to Education, 2009, but here the high intensity and low intensity districts did not see differential impacts on learning.

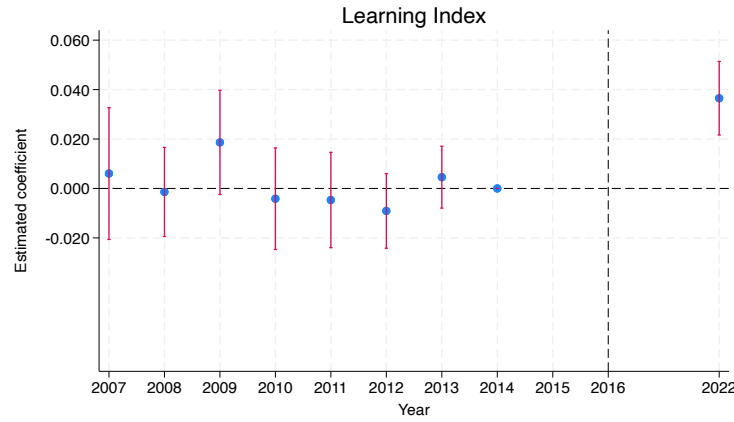


Figure 1.6: Event Study style graph for overall learning ability for a subset of states for which parallel pre-trends hold for all years

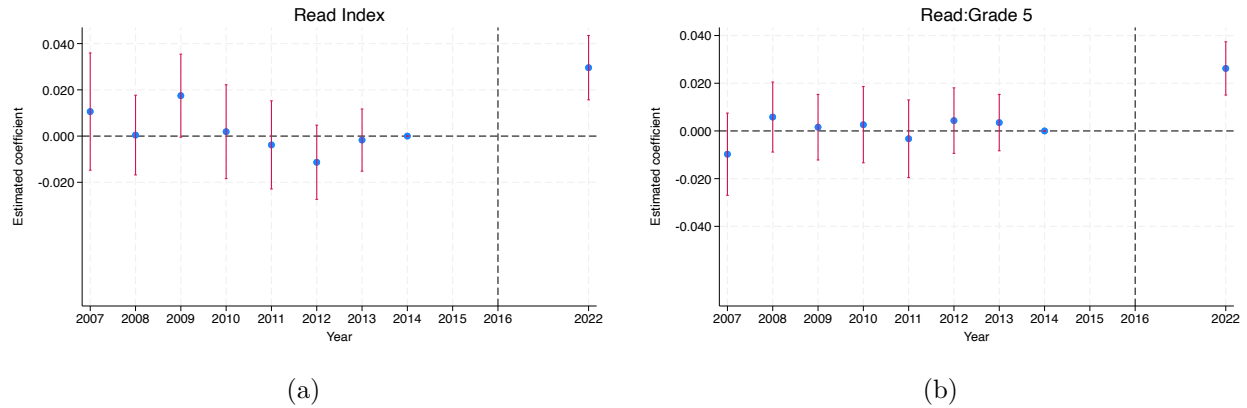


Figure 1.7: Event study style graph for reading ability for a subset of states for which parallel pre-trends hold for all years

Table 1.6 shows the DiD estimates for only these states taking all the data from 2007-2014 as the pre-period. We see that not only are the results in the same direction as before, but the effect sizes are between 15% to 30% larger. These findings are reiterated in the event study type graphs in

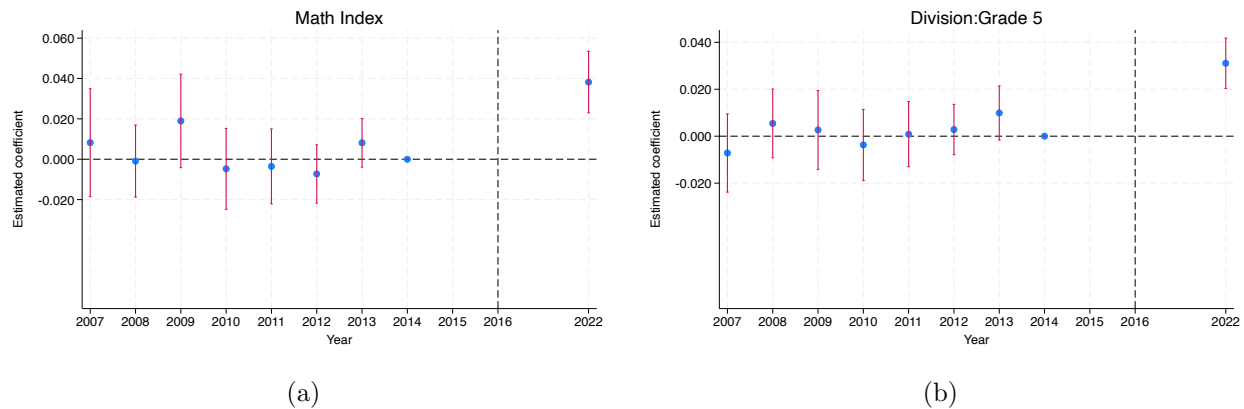


Figure 1.8: Event Study style graphs for Math ability for a subset of states for which parallel pre-trends hold for all years

figure 1.6- 1.8. We see that null hypothesis of no effect cannot be rejected for the years 2007-2014, but the effect is positive and significant at the 95% level in 2022.

Synthetic Difference-in-Differences

[Arkhangelsky et al. \(2021\)](#) have introduced a new method for causal analysis called synthetic difference-in-differences (synthetic DiD) which combines the features of the traditional difference-in-differences and synthetic controls methods. Synthetic DID finds optimal unit weights that align pre-exposure trends in the control units and the treatment units. It also finds time weights that balance pre-exposure time periods with post-exposure ones. [Arkhangelsky et al. \(2021\)](#) find that that synthetic DID method “is competitive with (or dominates) DiD in applications where DiD methods have been used in the past, and likewise is competitive with (or dominates) synthetic control (SC) in applications where SC methods have been used in the past” in terms of bias and root-mean-squared error. Inference is done with bootstrap or permutation methods.

Synthetic DiD requires a balanced panel and can be used only when the treatment is binary. Since

ASER is a repeated cross section, I collapsed the data at the district level and used only those districts in the analysis that had data for all years from 2007-2014 and 2022. The outcome variable is now the district average of the learning, math and reading index as well as proportion of class 5 children that can perform division and read a standard 2 text in that district. Table 1.7 shows the average treatment effect on the treated (ATT) from the synthetic difference-in-differences. The effects on both math and reading and thus, overall learning point in the same direction as the main results (albeit a little smaller) and significant at 1% level.

Table 1.7: Synthetic difference-in-differences

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Post X High	0.050***	0.043**	0.047**	0.060***	0.046***
Intensity	(0.018)	(0.017)	(0.020)	(0.011)	(0.018)
N	4,347	4,347	4,347	4,347	4,347

High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Post=1 for Year =2022 and 0 otherwise. All regressions include district fixed effects. Bootstrapped standard errors are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Robustness to non-parallel trends

Here, I use the sensitivity analysis suggested by [Kahn-Lang and Lang \(2020\)](#) and increasingly used in recent literature where parallel pre-trends are violated (for example, see [Griffith and Noonan \(2022\)](#)) to show that our results are robust qualitatively to non-parallel trends. I estimate equation 1.2 and it differs from the binary treatment version of equation 1.1 in that I add a time trend and a group trend. The identification assumption is that the high intensity districts would continue to trend the way they did before 2016 in the post period. I do two things: First, I estimate equation 1.2 for the entire pre-period that I have the data for. The results are in table 1.8. Here as we can see the coefficient on *High Intensity District X Year=2022* is positive, significant and much higher than in table 1.4. This is because the overall pre-trend is downward sloping for high intensity

districts. That is if this were true then we are underestimating the treatment effect.

$$y_{idt} = \delta_t * year + \beta_1 \text{High intensity district}_d * Post_t + \beta_1 \text{High intensity district}_d * Year_t + \theta_2 X_{it} + \gamma_d + \epsilon_{id} \quad (1.2)$$

Table 1.8: Sensitivity analysis based on Kahn-lang and Lang using all pre 2016 data

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Year	-0.013*** (0.002)	-0.011*** (0.002)	-0.014*** (0.002)	-0.005*** (0.001)	-0.005*** (0.001)
High Intensity District X Year	-0.031*** (0.004)	-0.031*** (0.004)	-0.029*** (0.004)	-0.017*** (0.003)	-0.013*** (0.003)
High Intensity District X Year=2022	0.423*** (0.040)	0.432*** (0.041)	0.383*** (0.037)	0.239*** (0.031)	0.181*** (0.033)
N	2,368,665	2,351,004	2,361,432	262,102	262,860

Includes data from all years (2007-2014 and 2022) High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

The more likely case however, is that after the RTE shock in 2010, *high intensity* districts recover faster and thus, we are capturing this effect instead that of PMUY. In order to test this hypothesis, I estimate 1.2 only using the data from 2011-2016 assuming that correct counterfactual is the trend in the treatment groups during this period.¹¹ The results shown in 1.9, show that the indices for learning are still qualitatively similar to what we saw in earlier. The effects on grade 5 reading and math however disappear.

Heterogeneity

In this section we show some interesting heterogeneity in our treatment that also support the fact that we are indeed capturing the PMUY on learning and not the impact of some other confounding factor.

¹¹PMUY was launched in May 2016 and connections were rolled out rather slowly at first. ASER was conducted in September 2016, and we can assume that we would not see much of an impact of the program in ASER 2016

Table 1.9: Sensitivity analysis based on Kahn-lang and Lang using data from 2011-2014 and 2022

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Year	-0.006*** (0.001)	-0.003** (0.001)	-0.010*** (0.002)	-0.003*** (0.001)	-0.008*** (0.001)
High Intensity	0.002 (0.004)	0.003 (0.004)	-0.001 (0.004)	0.009*** (0.002)	0.009*** (0.003)
District X Year	0.090*** (0.029)	0.080*** (0.029)	0.105*** (0.030)	-0.008 (0.022)	-0.009 (0.024)
N	1,376,158	1,370,505	1,373,728	147,842	148,092

Includes data from 2011-2016 and 2022 High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Impact of PMUY by gender

Figure 1.9 shows heterogeneous impact of PMUY by gender. While the impact of PMUY is positive and significant for boys for all our outcomes, we see that the impact is higher for girls (this is the coefficient *Post X Female X PMUY connections*). While such heterogeneity is interesting by itself, this is also what we would expect if we were indeed capturing the impact of PMUY. Girls in India are more likely to be at home as compared to boys because of son preference and consequent gender bias. Girls are much less likely to attend school on a day-to-day basis as compared to boys and are also more likely to help with household chores and less likely to spend time outdoors. Results using binary treatment is given in Appendix table A.9

Impact of PMUY by wealth

As mentioned above, PMUY has been targeted towards BPL households and those who do not already have an LPG connection. By this logic, the impact of PMUY should be larger on poorer households and there should be almost no impact on the richest households. However, PMUY would lead to learning gains in poor households only if they use LPG regularly. Even though the

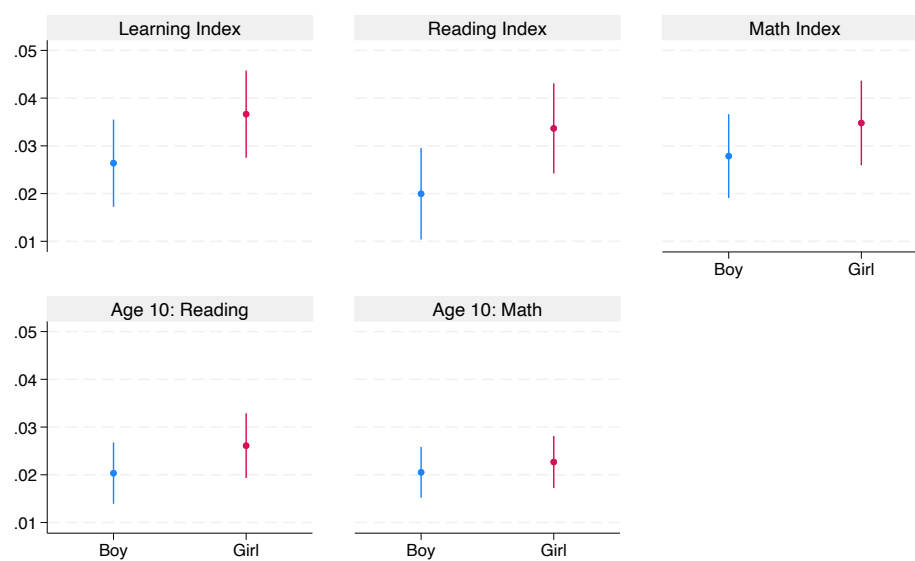


Figure 1.9: Differential impact on girls and boys

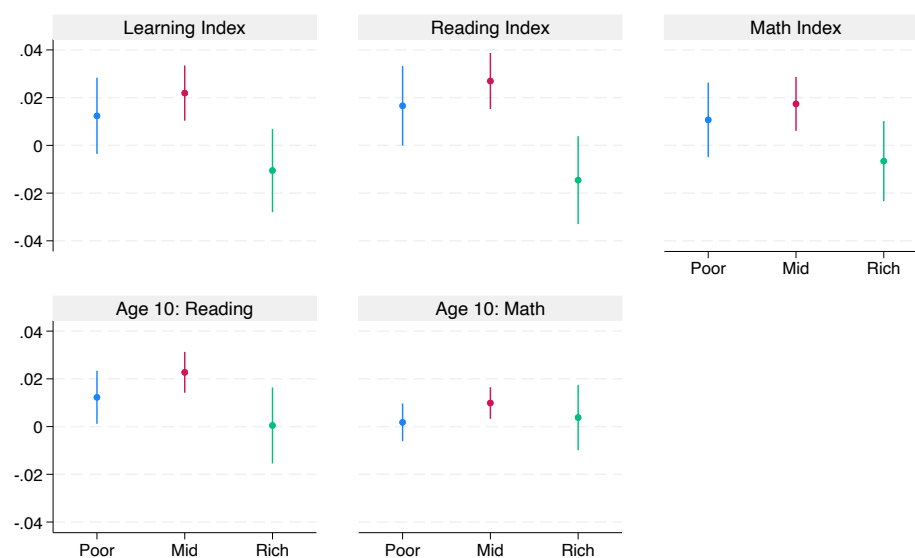


Figure 1.10: Differential impact by wealth of the household

price of an LPG refill is subsidized, it is still expensive for the poorest households. Thus, the impact of PMUY should be largest on those who are poor enough to be targeted by PMUY but rich enough to be able to afford LPG refills.

In order to construct a measure of household wealth, I sum up the number of assets that are owned by the household collected in the ASER survey. These range from 1-6 for 2022 and 1-5 for all other years. I then construct three binary variables for the poorest, middle wealth and richest households. The poorest households are in the lowest quartile, the richest ones are in the top decile and everyone else is considered as being in the middle wealth group.

We see an impact in line with our expectation in Figure 1.10. The impact of PMUY on the richest group for different learning outcomes is slightly negative and not statistically different from zero. The impact on the poorest households are positive and significant for some outcomes, but the impact on households in the middle of the wealth distribution is the highest and significant for all outcomes. The households in the middle are most likely the ones targeted by PMUY and can also afford to get refills while the poorest aren't able to.

Impact of PMUY by age

Figure 1.11 shows the impact of PMUY by age of the child. Among school aged kids (aged 6 or above), older kids (aged 9 and 10) seem to have a higher impact of the program than younger kids (aged 6, 7 and 8). This is in line with what we find in the 1.6 section below. We find evidence for increase in school attendance because of PMUY. Older children may struggle more to learn without attending school, as the skills they learn become more complex (for example, division as opposed to addition), and fewer family members at home may have the ability/expertise to help them.

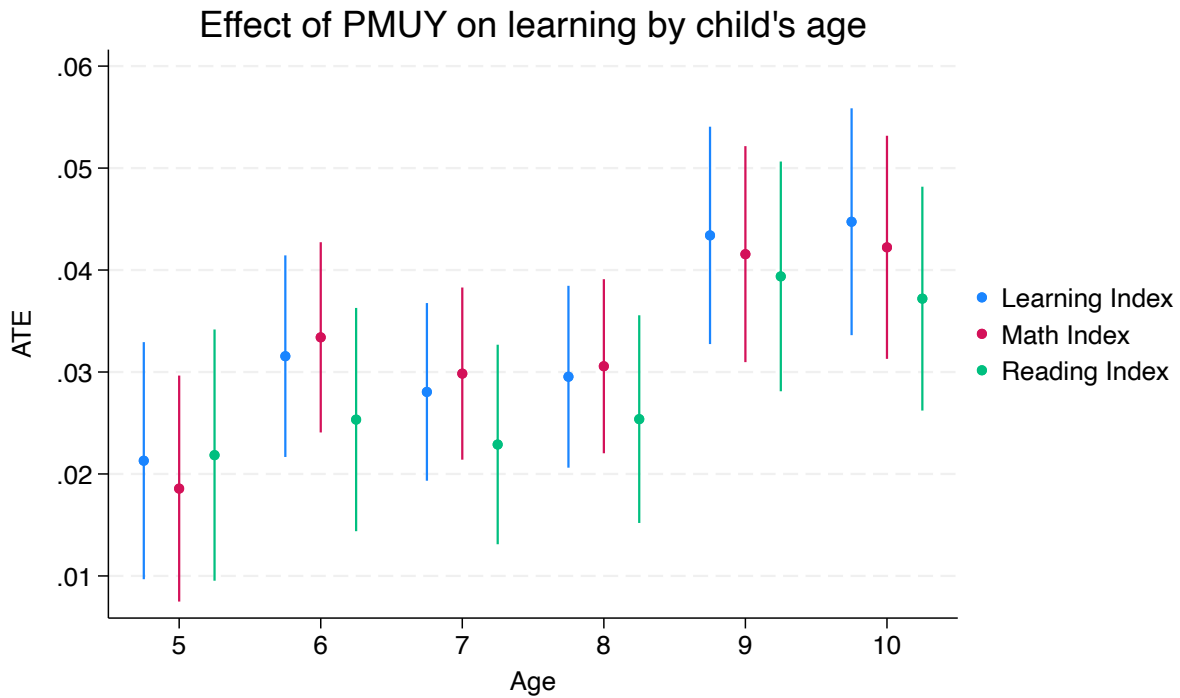


Figure 1.11: Heterogeneity by Age

1.6 Mechanisms

There are five possible ways in which access to clean cookstoves can improve learning outcomes of children. First, the reduction in household air pollution could improve the health of children. Healthier children are then able to attend school more and spend more time studying. The link between improved health and reduced school absenteeism is well documented (for example, see the classic study on deworming by [Miguel and Kremer \(2004\)](#)) Second, the reduction in household air pollution leads to an improvement in children's cognition. "Cognition refers to all the activities and processes concerned with the acquisition, storage, retrieval and processing of information - regardless of whether these processes are explicit or conscious" ([Bayne et al., 2019](#)). Improved cognition would lead to improved learning even without spending increased time attending school or studying. Third, switching to more efficient cooking fuels like gas could save children's time if

Table 1.10: Impact of PMUY on school attendance

	(1) Total Class 1-5	(2) Class 1	(3) Class 2	(4) Class 3	(5) Class 4	(6) Class 5
Post	-0.009 (0.006)	-0.008 (0.006)	-0.009 (0.007)	-0.005 (0.006)	-0.010* (0.006)	-0.003 (0.006)
Post X PMUY connections	0.007*** (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.007*** (0.002)	0.005** (0.002)
Mean Dependent	0.72	0.71	0.72	0.73	0.73	0.73
Mean Dependent:Pre	0.71	0.71	0.71	0.71	0.72	0.72
N	73,593	72,316	72,276	72,328	72,233	66,621

Treatment variable is the number of PMUY connections released in each district per household until August 2021. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * $p < 0.001$, ** $p < 0.05$, *** $p < 0.01$

they help with household chores. This would be more likely for older children. Fourth, time could also be saved for the primary cook and this might lead to her contributing more to the child's learning, for instance, by helping with schoolwork. Fifth, an improvement in teachers health could lead to increased teacher attendance and aid child learning.

Since we focus on children aged 10 and under in this paper, the channel through their time use seems very unlikely to be the dominant one. India's 2019 Time Use Survey shows that children between age 5 and 10 spend only about 3 minutes on cooking and related activities, such as collecting firewood. This is expected because cooking on an open fire is considered dangerous for a child. Additionally splitting, chopping and carrying firewood require considerable strength. Time savings of the primary cook are also not expected to be a dominant channel, since [Afridi et al. \(2023\)](#) show that, in practice, time savings are minimal- only about 5 minutes per day- when PMUY households start using non-polluting cooking fuels. They provide evidence that this is due to fuel stacking in rural households, even after they begin using clean fuels. Their finding indicate that most of the transition to clean cooking happens at the intensive margin with very few household switching exclusively to clean cookstoves. Most PMUY customers nationally do not use [LPG exclusively](#), and thus the time savings channel of the primary cook is also unlikely. I also find ([Appendix table A.10](#)) that there is no impact of PMUY on teacher's attendance, ruling out this factor that as the

dominant channel.

I provide suggestive evidence for the fact that improvement in children’s health leading to increased attendance is the dominant channel. The negative impact of air pollution on school attendance has been documented in both developing and developed countries ([Currie et al., 2009](#); [Chen et al., 2018](#)). Table 1.10 shows the impact of PMUY on attendance which is defined as the proportion of students who were attending school on the day the data was collected out of the total enrollment. The attendance comes from the school module of ASER- and is collected from one random school in each village that is visited as part of the household module. Column 1 of table 1.10 shows that overall primary school attendance goes up by 0.6 percentage points when the PMUY connections increase by 10 percentage points. This means that for the median district, the effect size is about 1.8 percentage points, a modest but significant impact. Columns 2-6 show that this increase is consistent across all classes from 1 to 5.

In addition, Appendix tables A.11 and A.12 show the impact on out-of-school children. One cannot reject the hypothesis that there was no impact of PMUY on the learning of these children. This is despite the fact that enrollment shows a slight decline in *higher intensity districts* which means our coefficient might be positively biased (see appendix table A.13). The fact that there is no impact on out-of-school children, suggests that cognition alone is perhaps not an important channel. In addition, the fact that we see the largest effect sizes on the hardest skills and on older children suggests that school attendance is the dominant mechanism- if older children don’t go to school, it is harder to make up the learning losses. [Balakrishnan and Tsaneva \(2021\)](#) also find that school attendance is an important mechanism through which ambient air pollution impacts children’s learning.

1.7 Conclusion

PMUY was launched in 2016 with an aim to reduce the reliance of poor households in India on solid cooking fuels, which generate harmful indoor air pollutants. These pollutants pose significant health risks, particularly to women and children, who are most vulnerable to the adverse effects of indoor air pollution because of the way household work in most Indian households is distributed. This paper shows that this program had the added benefit of improving educational outcomes of primary school children. The children in districts that received a higher proportion of PMUY LPG connections had larger learning improvements in both reading and mathematics. A 30 percentage-point increase in LPG connections leads to an improvement in learning of 0.1 standard deviations which is close to some of the most successful learning interventions conducted in low- and middle-income countries. It is important to note that most PMUY customers do not use LPG exclusively due to the high cost of refills and the average LPG refills for PMUY customers is close to only 4 cylinders per year. For context, a family of four is expected to use one cylinder for 4-5 weeks if they only use LPG for cooking. Therefore, the impacts that we have noted are conservative relative to what they could be if there were no stacking of LPG with solid fuels.

This study uses a difference-in-differences approach with district and year fixed effects to identify the causal impact of PMUY on learning. However, the study is not without limitations. In particular there are three major concerns. First, I have access to only aggregate PMUY connections released at the end of the first phase of the scheme in 2021. Ideally, the paper could have used the intensity of treatment coinciding with all the ASER surveys done after the launch of PMUY, that is 2016 and 2018 in addition to 2022. Second, there could be technical concerns because pre-trends are not parallel in the first four years for which we have data (2007-2010). This I hypothesize is because there was a massive shock to the Indian education sector due to the introduction of the Right to Education Act in 2009. I have tried to alleviate these concerns with a variety of robustness tests as well as time varying controls in our main specification but ideally, we should have controlled for all aspects of our districts that are correlated with PMUY intensity and effectiveness of RTE

implementation. Unfortunately, such measures are unavailable publicly. Third, even though I provide suggestive evidence for the fact that the improvement in learning occurred due to health improvement of children and resulting change in their behavior, more research is needed to exactly pin down the mechanism with better data collection. Despite these limitations, this paper is the first to demonstrate the relationship between access to clean cooking fuels and learning ability and provide plausibly causal estimates for such a relationship in low- and middle-income countries. It provides evidence for how household air pollution, as part of children's home environment, affects their learning and subsequently influences their educational achievement.

Overall, this study has many implications for policy. This study makes a case for providing subsidized connections to clean cooking technology like gas and electric stoves. While evaluating the costs and benefits of such an incentive, policymakers should, in addition to, the impact on health and time savings, also consider the benefits to children's learning and consequently their adult outcomes. Such an analysis is beyond the scope of this paper and is an interesting avenue of future research. India, which has for a large part of the last 3 decades been the largest solid fuel using country in the world has done well in recent years in increasing access to clean cooking, but the refill rates are abysmally low. This paper suggests that subsidizing LPG refills could have long lasting impacts on learning and thus, adult outcomes.

BIBLIOGRAPHY

- Afridi, F., S. Debnath, T. Dinkelman, and K. Sareen (2023). Time for clean energy? cleaner fuels and women’s time in home production. *The World Bank Economic Review* 37(2), 283–304.
- Agarwal, M. K., G. Mishra, and S. Ansari (2023). Attaining social inclusion through pradhan mantri ujjwala yojana (pmuy) in india.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association* 103(484), 1481–1495.
- Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021). Synthetic difference-in-differences. *American Economic Review* 111(12), 4088–4118.
- Balakrishnan, U. and M. Tsaneva (2021). Air pollution and academic performance: Evidence from india. *World Development* 146, 105553.
- Ball, S. B., C. Eckel, and C. Rojas (2006). Technology improves learning in large principles of economics classes: Using our wits. *American Economic Review* 96(2), 442–446.
- Bayne, T., D. Brainard, R. W. Byrne, L. Chittka, N. Clayton, C. Heyes, J. Mather, B. Ölveczky, M. Shadlen, T. Suddendorf, et al. (2019). What is cognition? *Current Biology* 29(13), R608–R615.
- Bennitt, F., S. Wozniak, K. Causey, K. Burkart, and M. Brauer (2021). Estimating disease burden attributable to household air pollution: new methods within the global burden of disease study. *The Lancet Global Health* 9, S18.

- Berkouwer, S. B. and J. T. Dean (2024). Cooking, health, and daily exposure to transient air pollution peaks.
- CAG (2019). Report on pradhan mantri ujjwala yojana. Technical report, Comptroller and Auditor General of India.
- Callaway, B., A. Goodman-Bacon, and P. H. Sant’Anna (2024). Difference-in-differences with a continuous treatment. Technical report, National Bureau of Economic Research.
- Chen, S., C. Guo, and X. Huang (2018). Air pollution, student health, and school absences: Evidence from china. *Journal of Environmental Economics and Management* 92, 465–497.
- Chowdhury, S., A. Pillarisetti, A. Oberholzer, J. Jetter, J. Mitchell, E. Cappuccilli, B. Aamaas, K. Aunan, A. Pozzer, and D. Alexander (2023). A global review of the state of the evidence of household air pollution’s contribution to ambient fine particulate matter and their related health impacts. *Environment international* 173, 107835.
- Currie, J., E. A. Hanushek, E. M. Kahn, M. Neidell, and S. G. Rivkin (2009). Does pollution increase school absences? *The Review of Economics and Statistics* 91(4), 682–694.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review* 91(4), 795–813.
- Duflo, E. (2004). The medium run effects of educational expansion: Evidence from a large school construction program in indonesia. *Journal of Development Economics* 74(1), 163–197.
- Ehrenberg, R. G., D. J. Brewer, A. Gamoran, and J. D. Willms (2001). Class size and student achievement. *Psychological science in the public interest* 2(1), 1–30.
- Evans, D. and F. Yuan (2019). Equivalent years of schooling: A metric to communicate learning gains in concrete terms. *World Bank Policy Research Working Paper* (8752).
- Evans, D. K. and F. Yuan (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis* 44(3), 532–540.

- Frostad, J. J., Q. P. Nguyen, M. M. Baumann, B. F. Blacker, L. B. Marczak, A. Deshpande, K. E. Wiens, K. E. LeGrand, K. B. Johnson, M. Abbasi-Kangevari, et al. (2022). Mapping development and health effects of cooking with solid fuels in low-income and middle-income countries, 2000–18: a geospatial modelling study. *The Lancet Global Health* 10(10), e1395–e1411.
- Glewwe, P. and K. Muralidharan (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education*, Volume 5, pp. 653–743. Elsevier.
- Griffith, A. and T. Noonan (2022). The effects of public campaign funding: Evidence from seattle’s democracy voucher program. *Journal of Public Economics* 211, 104676.
- Grineski, S. E., T. W. Collins, and D. E. Adkins (2020). Hazardous air pollutants are associated with worse performance in reading, math, and science among us primary schoolchildren. *Environmental research* 181, 108925.
- Hanushek, E. (1998). The evidence on class size.
- Hanushek, E. A. (2012). The economic value of education and cognitive skills. *Handbook of education policy research*, 39–56.
- Hanushek, E. A. and L. Woessmann (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. *Journal of economic growth* 17, 267–321.
- Imelda (2020). Cooking that kills: cleaner energy access, indoor air pollution, and health.
- Jain, A., S. Tripathi, S. Mani, S. Patnaik, T. Shahidi, and K. Ganesan (2018). Access to clean cooking energy and electricity: Survey of states 2018. *CEEW Report, Council on Energy, Environment and Water (CEEW), New Delhi, India*.
- Kahn-Lang, A. and K. Lang (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics* 38(3), 613–620.

- Krishnapriya, P., M. Chandrasekaran, M. Jeuland, and S. K. Pattanayak (2021). Do improved cookstoves save time and improve gender outcomes? evidence from six developing countries. *Energy Economics* 102, 105456.
- Miguel, E. and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72(1), 159–217.
- Naghavi, M., K. L. Ong, A. Aali, H. S. Ababneh, Y. H. Abate, C. Abbafati, R. Abbasgholizadeh, M. Abbasian, M. Abbasi-Kangevari, H. Abbastabar, et al. (2024). Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. *The Lancet* 403(10440), 2100–2132.
- Palacios, J., P. Eichholtz, N. Kok, and N. Duran (2022). Indoor air quality and learning: evidence from a large field study in primary schools. *MIT Center for Real Estate Research Paper* (22/13).
- Park, R. J., J. Goodman, M. Hurwitz, and J. Smith (2020). Heat and learning. *American Economic Journal: Economic Policy* 12(2), 306–39.
- Prentice, C. M., F. Vergunst, K. Minor, and H. L. Berry (2024). Education outcomes in the era of global climate change. *Nature Climate Change* 14(3), 214–224.
- Requia, W. J., C. C. Saenger, R. E. Cicerelli, L. M. de Abreu, and V. R. Cruvinel (2022). Air quality around schools and school-level academic performance in brazil. *Atmospheric Environment* 279, 119125.
- Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *econometrica* 73(2), 417–458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American economic review* 94(2), 247–252.
- Shah, M. and B. Steinberg (2019). The right to education act: Trends in enrollment, test scores, and

- school quality. In *AEA Papers and Proceedings*, Volume 109, pp. 232–238. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Shupler, M., W. Godwin, J. Frostad, P. Gustafson, R. E. Arku, and M. Brauer (2018). Global estimation of exposure to fine particulate matter (pm_{2.5}) from household air pollution. *Environment international* 120, 354–363.
- Smith, K. R. (2000). National burden of disease in india from indoor air pollution. *Proceedings of the National Academy of Sciences* 97(24), 13286–13293.
- Smith, K. R. (2013). Biofuels, air pollution, and health: a global review.
- Smith, K. R., J. M. Samet, I. Romieu, and N. Bruce (2000). Indoor air pollution in developing countries and acute lower respiratory infections in children. *Thorax* 55(6), 518–532.
- Somanathan, E., T. Ninan, M. H. Bergin, M. Jeuland, E. Gupta, U. Kumar, R. Kamdar, V. Chowdhury, S. Chandna, K. Barkjohn, et al. (2022). *Electric stoves as a solution for household air pollution: Evidence from rural India*. Indian Statistical Institute, Delhi, Economics and Planning Unit.
- Vardell, E. (2020). Global health observatory data repository. *Medical reference services quarterly* 39(1), 67–74.
- Wang, W., R. Kim, and S. V. Subramanian (2021). Population Estimates for Districts and Parliamentary Constituencies in India, 2020.
- Wen, J. and M. Burke (2022). Lower test scores from wildfire smoke exposure. *Nature Sustainability* 5(11), 947–955.
- Yun, X., G. Shen, H. Shen, W. Meng, Y. Chen, H. Xu, Y. Ren, Q. Zhong, W. Du, J. Ma, et al. (2020). Residential solid fuel emissions contribute significantly to air pollution and associated health impacts in china. *Science advances* 6(44), eaba7621.

Chapter 2

WHEN TRADITION POLLUTES: HEALTH AND ENVIRONMENTAL COSTS OF MULTI-GENERATIONAL LIVING IN INDIA

2.1 *Introduction*

Household air pollution from solid fuels is one of the biggest health hazards in the world ([Murray, 2024](#)) and was responsible for an estimated 3.2 million deaths in 2020, including over 237,000 children under the age of 5 ([WHO 2022](#)). Roughly 30% of the world’s population relies on biomass-fueled stoves or open fires for cooking ([Shupler et al., 2018](#)). The incomplete combustion of these fuels, which include firewood, animal dung, crop residue and charcoal, emit toxic pollutants like particulate matter, carbon monoxide, and nitrogen dioxide. Exposure to these pollutants, especially high concentrations of PM2.5 has been linked to premature death as well as chronic diseases. In developing countries like India, where over 40% of households still do not use non-polluting fuels like gas and electricity as their primary cooking fuel ([International Institute for Population Sciences \(IIPS\), 2021](#)), women, and young children in their care face a disproportionate disease burden as a consequence of using these ‘dirty’ fuels. This is due to the primary role of girls and women in cooking, fuel collection and care-giving ([WHO 2022](#)). Average PM2.5 concentrations in rural kitchens using solid fuels have been noted to be more than 20 times the WHO safe limit and more than 2.5 times the ambient concentration ([Somanathan et al., 2022](#)).

This paper estimates the causal impact of a multi-generational traditional household structure on the household’s cooking fuel choice in a society like India, where patriarchy and patrilocality are prevalent. In particular, I look at the impact of the death of a co-residing father-in-law of pre-

sumably the primary cook(s) on the household's cooking fuel choice. Older males in the household are expected to have a lot of say in major household decisions in a patriarchal society like rural India, including in things like which fuel to use for cooking. However, these males would generally be minimally involved in the cooking process. Hence, if the older male like the father-in-law has a preference for food cooked using traditional stoves or underestimates the health impacts on the primary cooks, then their death might encourage the primary cooks to adopt cleaner cooking technologies. I use a nationally representative two wave panel survey of households conducted in 2005 and 2012 and employ a difference-in-differences (DiD) design with household fixed effects as my main specification to identify this effect. I restrict my analysis to households where the father-in-law and daughter-in-law co-resided in the same household during the first wave of the survey and then compare households where the father-in-law had died by the second wave (treated) with households where they had not (control).

The estimated DiD coefficient would be unbiased only if the parallel trends assumption is valid, that is only if the trends in cooking fuel use would be parallel in households where the FIL died compared to households where they didn't. The use of household fixed effects means that any time-invariant variables at the household level will not bias the estimates. Thus, we need to be worried about time-varying household-level confounders such as changing socioeconomic status. To address such concerns, I use households where only the mother-in-law died as the control group instead of households where the father-in-law did not die. Such a comparison eases concerns about endogeneity since we see that the two groups of households are similar on observable variables. In addition, I also use households where the MIL died as a placebo; if we are capturing the effect of time-varying variables that are correlated with both the death of the father-in-law and mother-in-law, then this regression should pick that up too. I find no effect in such a specification. I also consider a matched difference-in-differences strategy where control households are chosen using propensity score matching and are matched on variables like income, education and caste. I argue that households that are similar to each other in these respects are more likely to fulfill the parallel trends assumption. All these tests point qualitatively in the same direction as the effect in my main

specification. In fact, the estimated effect sizes are very similar as well.

Specifically, I find that the death of the father-in-law reduces the probability of using solid fuel for cooking by a statistically significant 3.3 percentage points or 6 percent. The probability of using every commonly used solid fuel like firewood, animal dung and crop residue is also significantly lower with effect sizes ranging from 6 percent to 33 percent. I also find that households where the father-in-law had died had a significantly lower probability of collecting each of these solid fuels (which is more cost-effective than purchasing these fuels) despite the fact that household incomes decrease after the father-in-law's death. The effect sizes range from 5 percent for firewood to 25 percent for crop residue.

I evaluate a variety of different mechanisms that could be driving these results. For example, I examine whether the preference of the father-in-law for food cooked in the traditional way, which includes using biomass, drives these results. I also study the impact of household splitting due to the death of the patriarch, income effects due to the death of a productive member, and impact on the daughter-in-law's autonomy due to the father-in-law's death. Here, I find that the preference of the father-in-law for food cooked in the traditional way without using modern mechanical equipment is the most plausible explanation.

This paper contributes to two different strands of literature. First, it contributes to understanding the sticky nature of cooking energy use especially in the South Asian context which has one of the largest numbers of solid fuel users and is simultaneously one of the most polluted regions in the world. Despite extensive attention from researchers and policymakers, household air pollution from cooking with solid fuels remains an intractable problem in all but the richest countries. Some reasons that have been highlighted in the literature include households' low purchasing power and liquidity constraints ([Jeuland et al., 2015, 2023](#); [Afridi et al., 2021](#)), unreliable cooking fuel supply ([Somanathan et al., 2022](#)), maintenance needs of 'improved cookstoves (ICS)' ([Hanna et al., 2016](#)), women's low bargaining power in the household ([Miller and Mobarak, 2013](#); [Krishnapriya](#)

et al., 2022; Kishore and Spears, 2014) and cultural preference for food cooked using solid fuels (Pattanayak et al., 2019; Gupta and Köhlin, 2006). One aspect that has received no attention in the existing literature is the significance of traditional family structure and the pressure it exerts in the choice of cooking fuel. This is the first paper to do so and finds that it is plausible that family structure has an important role to play in such settings.

Second, this study contributes to the scant literature that studies intra-household dynamics in patriarchal multi-generational households in developing countries. In countries like India where patrilocality is the norm, it is not uncommon to see multiple generations (including siblings, spouses and kids) cohabiting, sharing the same kitchen and pooling resources for consumption. In such settings, senior or male members of the family can exert significant influence on all aspects of household decision making including large household purchases and technology adoption including the choice of cooking fuel and stove. This setting also diverges from conventional household models, where apart from herself, only a married woman’s husband primarily makes choices that affect her well-being. In a multigenerational setting, in addition to the husband, other members of the household can substantially affect a women’s well-being. For example, her father-in-law, mother-in-law, or brother-in-law can significantly impact her mobility, social connections, health and employment decisions. Even though other studies (Anukriti et al., 2020, 2022; Khanna and Pandey, 2021) have looked at the mother-in-law’s role in the daughter-in-law’s formation of social networks, fertility and employment decisions, to the best of my knowledge, this is the first study that looks at the impact of the pressure that the father-in-law exerts on household choices that affect health and time use of daughter-in-laws.

The results in this paper can be contextualised in a setting where the father-in-law in the household who has the most bargaining power does not find it beneficial to adopt the use of clean cookstoves either due to inaccurate information or due to apathy towards the daughter-in-law. In such a case, his death could change incentives such that adoption of clean cookstoves becomes more likely. Ethnographic studies like Vyas et al. (2021) and Palriwala (1993) have shown the prevalence

of norms in the Indian context where a skilful daughter-in-law is expected to preserve household resources and utilize the household budget in a way that prioritises her family’s well being over her own comfort. This paper shows that in a ‘joint’ family these pressures could have perilous impacts on the daughter-in-law’s health and productivity.

The rest of the paper is organised as follows: Section 2.2 discusses the conceptual framework and theoretical considerations, Section 2.3 describes the data that is used in the analysis, Section 2.4 presents the main empirical strategy and discusses the limitations of the analysis, Section 2.5 presents the main results, Section 2.6 discusses the robustness of the results. Section 2.7 discusses the plausible mechanisms driving the results in this paper and Section 2.8 concludes.

2.2 Conceptual Framework

India is a largely patriarchal and patrilocal society, where in many households two or more adult generations live in the same household. In NFHS-5, about 4 in 10 households live in these so-called ‘joint families’. In such families, the primary responsibility of household chores including cooking and collecting fuel is on the daughter-in-law(s) of the household. Hence, the daughter-in-law(s) has the most to gain if households shift from using solid fuels to clean cooking fuels. These benefits would primarily be in the form of health benefits, as switching to clean fuels would drastically reduce their average exposure to harmful pollutants like particulate matter. Particulate matter, especially PM2.5 from household air pollution has been linked to respiratory illness, cancer, heart disease and consequently lower life expectancy (Lelieveld et al., 2020; Cohen et al., 2017). In addition to health benefits, since many of these solid fuels are collected and not purchased from the market, reducing their consumption could also provide time-use benefits by allowing the woman to allocate time to other activities.

However, the daughter-in-law(s) by virtue of being an outsider to the household as well as a young

woman often has the least bargaining power and say in household decisions ([Anukriti et al., 2020](#)). Her father-in-law, on the other hand, as one of the older males generally has a lot of say in most decisions in the household. Thus, if the overall utility from switching from ‘dirty’ to ‘clean’ cookstoves is lower than not switching for the father-in-law, the household might not switch to clean fuels.

The father-in-law might not find it valuable to switch because, first, he might have a preference for food that is cooked in the traditional way, that is without using any modern equipment like mechanized grinders, pressure cookers and refrigerators. This is likely because he is used to food prepared in this way. Second, clean cooking is more expensive given how clean fuels can not be collected freely unlike solid fuels like wood, crop residue and animal dung and need to be purchased. They are much more expensive than solid fuels, and clean stoves themselves cost more than the traditional ones. This means that the household might have to transfer household resources towards cooking from things that the father-in-law values more.

In addition, the father-in-law (i) might not care enough about the welfare of the primary cook or (ii) might underestimate the extent to which the use of solid fuels causes discomfort and health issues to the primary cook and other members. To see why (i) could be the case, it is important to note that in the traditional Indian setting, males and especially older males are not very involved in the cooking and fuel collection process ([National Statistical Office \(NSO\), 2020](#)). Thus, the father-in-law does not directly benefit much from switching to cleaner cooking practices since his exposure to air pollution is not that much higher than what he is already exposed to vis-à-vis ambient air pollution. In addition, he also does not bear the burden of collecting and preparing these fuels,, which is also primarily the women’s responsibility. On the other hand, (ii) could be true because the head does not spend a lot of time in the kitchen and thus is exposed to much lower household air pollution compared to the women. Also, in a patriarchal and conservative society like India, there isn’t much communication between the father-in-law and daughter-in-law. In fact, in such a setting it might even be considered inappropriate for the daughter-in-law to make demands for

improved cook stoves from the household head as it might not be line with the view of a ‘good’ daughter-in-law in the traditional Indian context.

Because of these reasons, if the father-in-law has a lot of weight in household decision making (an assumption that is not unwarranted)- the optimal choice for the household might be to continue using solid fuels.

However, if the father-in-law dies in a household where the father-in-law and daughter-in-law co-reside, the constraints described above will start to loosen. The daughter-in-law might have a larger weight in household decisions, there might be lower preference in the household for food cooked in the traditional way and also there might be better communication between the primary cook and the new household head (which could be either her husband, her MIL or even her husband’s brother). All these individually or taken together might encourage the use of clean fuels vis-à-vis dirty fuels.

On the other hand, if the father-in-law contributed to the household income, then his death might result in lower household income. Combining this with the fact that clean fuels and stoves are more expensive than solid fuels, the death of the father-in-law might actually encourage the use of solid fuels. Additionally, the death of the father-in-law, mechanically reduces the number of adult people for whom food needs to be prepared. This might also make the use of solid fuels easier for the primary cook.

This paper tries to causally identify the effect of the death of the father-in-law on the use of solid fuels in the household. The direction of the effect will depend on which of the above effects dominate. Even though I am not able to disentangle every mechanism, I will provide some suggestive evidence on the plausible ones.

2.3 Data

The primary data for this paper comes from the Indian Human Development Survey (IHDS), a nationally representative two-wave panel survey of 42,152 households in 1,420 villages and 1,042 urban neighbourhoods in India. The first wave was conducted in 2005-2006, and the second wave was conducted in 2011-2012, with an 85 % re-interview rate between the two surveys.

The panel structure of the data as well as detailed information collected on household cooking fuel use and collection, wealth, incomes and household rosters and demographic information including information on household members who had died between the two waves makes the IHDS appropriate for my analysis. I am able to identify households where father-in-laws and daughter-in-laws cohabit as well as those households where the father-in-law had died between the first and second wave. I am also able to use the information on incomes, assets and gender relations to explore different mechanisms as well as test alternate specifications.

Because the choice of the daughter-in-law to live with her father-in-law is endogenous, I first subset the data for only those households where the daughter-in-law is co-residing with the father-in-law in the first wave of the IHDS (henceforth, the baseline). There are 8,962 such households where 11,312 daughter-in-laws live with 9,056 father-in-laws. Details on how the co-residing daughter-in-law, father-in-law pair and their respective households were identified is given in Appendix B. Given the panel nature of the data and the tracking sheet from the second wave, I then tag father-in-laws that had died between the first and second wave. 1,997 father-in-laws in 1,934 baseline households had died by the second wave.

The data for cooking fuel use and collection comes from the household module of the same survey. In particular, households are asked whether they use firewood/twigs, dung cake, crop residue/by-product, charcoal, kerosene and Liquefied Petroleum Gas (LPG). They are also asked what purpose these fuels are mainly used for. I code the household as having used a particular fuel for cooking

if they report having used it for cooking only.

Table 2.1: Summary Statistics: Households with FIL dead vs FIL not dead

	(1) FIL alive	(2) FIL dead	(3) Difference	(4) N
Income	136,102.27	128,060.95	-8,041.31	8,962
Household Consumption	129,393.61	119,492.37	-9,901.24***	8,952
Household Assets- quintile	3.42	3.29	-0.13***	8,962
Below poverty line	0.22	0.26	0.04***	8,962
Urban	0.27	0.24	-0.04***	8,962
Highest adult education	9.07	8.40	-0.67***	8,952
Highest female education	5.65	5.19	-0.46***	8,886
Household size	7.49	7.43	-0.06	8,962
Number of adult males	2.41	2.42	0.01	8,962
Number of adult females	2.04	2.01	-0.03	8,962
Firewood used for cooking	0.49	0.52	0.03**	8,897
Dung used for cooking	0.34	0.35	0.01	8,792
Biomass used for cooking	0.54	0.58	0.04***	8,920
Uses dirty stove	0.76	0.80	0.04***	8,962
Collects firewood	0.59	0.62	0.04***	8,897
Collects dung	0.44	0.44	-0.00	8,792
Collects any fuel	0.52	0.57	0.05***	8,920
HH uses LPG	0.41	0.37	-0.04***	8,575

This tables compares households where a coresiding FIL had died between IHDS 1 and IHDS 2 and households where they had not. It is obtained by regression of an indicator variable equal to 1 if the FIL is dead on various variable obtained from the 2005 survey

All values are in 2005 Rupees

Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Households are also asked which type of stove they use for cooking and I code a household as having used a ‘dirty stove’ if they use a stove that uses biomass as fuel.¹ Similarly, households are asked where they get most of these fuels from. I code them as having collected a particular fuel if they collect at least some of their fuel, that is, they use that particular fuel but do not exclusively purchase it.

For some outcomes where several different measures need to be combined (for example, women’s autonomy), I group these related variables into index variables following (Anderson, 2008) to mit-

¹The survey reports only one stove for each household, presumably the primary one. However, there is evidence that ‘fuel stacking’ (Cheng and Urpelainen, 2014) of LPG and biomass had actually increased at least till 2010 and so it is possible that some ‘dirty stoves’ users had ‘clean’ LPG stoves as well and vice versa.

igate the multiple testing problem. I construct the indices in three steps. First, I re-code all contributing variables so that higher values correspond to ‘better’ outcomes. Second, I standardize the individual outcomes using the baseline control group mean and standard deviation for that outcome. Third, I calculate the average of the standardized constituent outcomes, weighted by the inverse covariance matrix.²

Table 2.1 compares baseline information for households where the FIL had died by the second wave to households where the FIL was alive in the second wave. In 2005, households where the FIL died by the second wave have lower income, consumption, wealth and are more likely to be below the poverty line but have similar household size and household compositions. They also have lower levels of education. Given that they are poorer and less educated, it is not surprising that these households are also more likely to use solid fuels for cooking as can be seen from the proportion that uses firewood, biomass and LPG for cooking as well the proportion of those that use dirty stoves as their primary stove. As a result, these households are also more likely to collect firewood and dung.

Given these baseline differences, we cannot simply compare the households where the FIL died to households where they did not, as the FIL’s death is likely to be correlated with other factors that affect household fuel choice and usage. Thus, we will employ a difference-in-differences strategy which is described below.

2.4 Empirical Strategy

As described earlier, the choice to live with your father-in-law is endogenous and therefore we cannot simply compare households where a daughter-in-law lives with an father-in-law and where

²I will estimate the covariance matrix and hence the weights using only observations that have non-missing values for all outcomes in the index. The third step is implemented with the user written Stata command *icw_index* (Bouguen and Varejkova, 2020)

she does not. We also cannot do a cross-sectional comparison of households where the father-in-law died between IHDS 1 and IHDS 2 with households where they did as father-in-law's death is likely to be non-random. Thus, in order to estimate the causal effect of the daughter-in-law co-residing with her father-in-law on household fuel choice, I use a difference-in-differences strategy where I compare treatment (where the father-in-law died by IHDS 2) and control (where the father-in-law did not die in IHDS 2) households, pre and post treatment. I also include household fixed effects in order to account for time-invariant factors that differ across different types of households and could bias the estimates. This strategy is similar to [Khanna and Pandey \(2021\)](#) and [Evans and Miguel \(2007\)](#) except that the data are at the household level instead of the individual level and 'treatment' is defined differently. In other words, I estimate the equation:

$$y_{it} = \beta_0 + \beta_1 \text{FIL died}_i * \text{Post}_{it} + \beta_2 \text{Post}_{it} + \delta_i + X'_{it}\gamma + \epsilon_{it} \quad (2.1)$$

where y_{it} is the outcome variable for household i , such as whether the household uses solid fuels for cooking. $FILdied_i$ is an indicator for whether the co-residing father-in-law died between the two waves of the survey. $Post_i$ is an indicator for the second wave of the survey. δ_i are household fixed effects and ϵ_{it} is the idiosyncratic error term clustered at the primary sampling unit (PSU) level. β_1 is the causal coefficient of interest. Here, the first difference is the difference in y_{it} in the pre period and the second difference is the difference in y_{it} in the post period.

My preferred specification is one that is the most parsimonious (without any controls) because including time varying covariates that are affected by treatment could introduce bias. However, I also estimate the above equation with a vector of time varying control variables X_{it} in order to support the mechanisms and show robustness. The controls that I use in different specifications include inflation adjusted income, the wealth quintile of the household, and the size and composition of the household. The results are very similar to my preferred specification, hence providing evidence that the results would hold even if you consider these variables as potential confounders.

A crucial assumption for equation 2.1 to provide an unbiased causal estimate is for the parallel trends assumption to hold. The assumption states that the treatment and control households would have followed the same trend in absence of the ‘treatment’. I perform a variety of robustness checks described in section 2.6 to show why this is a plausible assumption in this setting.

2.5 Results

Table 2.2: Impact of the death of the FIL on solid fuel use

	(1) Any Dirty Fuel	(2) Firewood	(3) Dung	(4) Crop Residue	(5) Charcoal
Post	0.054*** (0.012)	0.051*** (0.012)	0.051*** (0.011)	0.110*** (0.010)	-0.005** (0.003)
Post x FIL dead	-0.033** (0.015)	-0.029* (0.015)	-0.025* (0.013)	-0.033*** (0.013)	0.001 (0.004)
Constant	0.573*** (0.006)	0.515*** (0.007)	0.362*** (0.006)	0.111*** (0.006)	0.029*** (0.001)
Baseline Mean	0.54	0.49	0.34	0.10	0.03
N	20289	20266	20161	19889	19630

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 if the household is one where the FIL had died by the second wave. All regressions include household fixed effects. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where FIL coresides with either MIL or FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.2 shows the impact of father-in-law’s death on the use of biomass for cooking. First, interestingly the use of biomass for cooking had increased in 2012 by more than 5 percentage points for households where the father-in-law had not died (the coefficient of Post). This could be because of the shift away from kerosene as cooking fuel. During the end of the 2000s, there has been a concerted effort to improve targeting of kerosene subsidies that are available through the public distribution system. Alternatively, this could also indicate a rise in ‘fuel stacking’- a practice where multiple sources of energy are used for the same purpose like cooking.

The causal parameter of interest is however the coefficient of Post x FIL dead which is the estimated difference-in-difference parameter. We observe in column 1 of Table 2.2 that the use of ‘any dirty

fuel' is lower by 3.3 percentage points which is about 6% on a baseline control mean of 54%. This is significant at the 5 % level. The variable 'any dirty fuel' is 1 if the household uses firewood, dung, crop residue or charcoal for cooking only and 0 otherwise. Columns 2 to 5 look at the impact on each of these fuels separately. Firewood and animal dung are the most commonly used and see a 3 percentage point (about 6% on a baseline control mean of 49 %) and 2.5 percentage point (7% on a baseline control mean of 34 %) decline respectively. Crop and agricultural residue, which rounds up the top three most commonly used solid fuels for cooking in India, also declines by 3.3 percentage points (33% on a baseline control mean of 10 %). We see no decline in the use of charcoal but that is because it is hardly used as a cooking fuel in India (only about 2% in India and 3% in the sample in this paper) .

Solid fuels can be freely collected from the surroundings whereas clean fuels need to be purchased either from government licensed retail stores where they are subsidized or from private stores where they are often sold at a markup. For example, firewood can be collected from nearby forests and agriculture residue and dung can be obtained as a byproduct from cultivation and livestock rearing that a majority of rural households are engaged in. Table 2.3 shows that the death of an father-in-law reduces the probability of collecting any solid fuel by 1.3 percentage point (around 1.7 % on a baseline control mean of 69%). Individually, the probability of collecting firewood is lower by 2.6 percentage points (around 4.5 % on a baseline control mean of 59%), of collecting dung by 2.7 percentage points (around 6% on a baseline control mean of 44%) and of collecting crop residue by about 4 percentage points (around 23% on a baseline control mean of 17%).

Table 2.4 shows the impact of father-in-law's death on the use of clean fuel for cooking. The coefficient of Post in column 1 shows that the use of 'any clean fuel' for cooking has increased between 2005 and 2012 by 3.3 percentage points. Combined with the fact that use of dirty fuel was increasing during this period, this indicates higher 'fuel stacking' among households in the control group. However, this masks the fact that kerosene is being used 5.5 percentage points less compared

Table 2.3: Impact of the death of the FIL on solid fuel collection

	(1) Collects any fuel	(2) Collects Firewood	(3) Collects Dung	(4) Collects Crop Residue
Post	-0.028*** (0.006)	-0.023*** (0.008)	-0.042*** (0.007)	0.094*** (0.011)
Post x FIL dead	-0.013 (0.011)	-0.026** (0.013)	-0.027** (0.012)	-0.038*** (0.014)
Constant	0.707*** (0.003)	0.612*** (0.004)	0.467*** (0.004)	0.182*** (0.006)
Baseline Mean	0.68	0.59	0.44	0.17
N	20285	20265	20154	19886

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 if the household is one where the FIL had. died by the second wave. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with either MIL or FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Impact of the death of the FIL on clean fuel use

	(1) Any Clean Fuel	(2) Kerosene	(3) LPG
Post	0.033*** (0.010)	-0.055*** (0.008)	0.092*** (0.007)
Post x FIL dead	0.012 (0.014)	0.008 (0.012)	0.006 (0.012)
Constant	0.399*** (0.005)	0.146*** (0.004)	0.289*** (0.004)
Baseline Mean	0.42	0.14	0.32
N	20290	20382	20382

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with either MIL or FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

to the baseline whereas LPG is being used about 9.2 percentage points more.³ The coefficient of interest on Post x FIL dead is positive in all three columns which means that the use of ‘clean fuels’ for cooking is higher in houses where the father-in-law died but these are not statistically significant. There could be a couple of reasons for why the decrease in cooking with biomass is not completely offset by an increase in clean fuel. First, some clean fuels are not captured by the IHDS surveys, for example biogas or electricity. Second, and this seems to be the more likely explanation is that among treatment households those that were using both solid fuels as well as LPG or kerosene for cooking have now switched to using mainly one of the cleaner fuels. This would then show up as a negative coefficient for Post x FIL dead in table 2.2 but a positive but not significant coefficient in table 2.4. This is expected as households using both LPG and firewood will have fewer constraints than those exclusively using solid fuels (for example, liquidity constraints or unfamiliarity with the technology) that might be preventing the switch.

Appendix tables B.1 to B.3 show the results described above with time varying controls like inflation adjusted income, wealth and number of adults in the household. The coefficients are quite similar to what we saw above.

The results described above are on the sample of households where the daughter-in-law co-resides with the father-in-law in the baseline. Here, she could or could not be co-residing with the MIL in the baseline. In addition, treated households are defined as those where the father-in-law has died between wave 1 and 2. It is possible that in these households the MIL had died by the second wave as well. In order to eliminate the effect of the MIL dying, appendix tables B.4 -B.6 present the effect on cooking fuel choice as well as collection of fuels on a subsample where the daughter-in-law was co-residing with both the MIL and father-in-law in baseline. In addition, a household is ‘treated’ when only the father-in-law dies.

³An important caveat is that kerosene is not strictly a ‘clean’ fuel and elevates the level of many pollutants like $PM_{2.5}$, PM_{10} , SO_2 and NO_2 . However, kitchen concentrations of Particulate Matter (PM_5) have been reported to be twice in households using wood as compared to kerosene. Moreover, kerosene is often considered a step up in the energy ladder (Lam et al., 2012)

The point estimates of the coefficients of interest on *Only FIL dead X Post* are in the same direction as those in our main subsample, with some of these effects even larger than before. However, we are underpowered (due to the smaller sample size) to reject the null of no effect.

2.6 Robustness

2.6.1 Parallel Trends

For the DiD estimates to be unbiased, the parallel trends assumption needs to hold. My study has limitations vis-a-vis the standard procedures to provide this evidence. Since the IHDS is only a two-wave panel, I cannot do a standard pre-trends analysis or placebo tests. The IHDS sample was partly (randomly) selected from a previous survey, the HDPI survey conducted in 1993-94. The HDPI however, has no questions about the household's fuel use or cooking behavior. Thus, we are unable to use this data to check for parallel pre-trends.

Specifically, for parallel trends to hold we need to be worried about time-varying unobservables that affect the treatment and control households differently. Two of these in particular are (i) preference towards health and (ii) since the households where father-in-law died are poorer, any government schemes that encouraged poorer households to adopt cleaner cooking fuels.

(i) is less of a concern because it would bias the estimates to be less negative given that the correlation between higher health preference and father-in-law's death as well as the adoption of clean cook stoves is negative. For (ii) between 2005 and 2012, the Government of India or any other organization, to the best of my knowledge, did not introduce any new program to discourage the use of solid fuels for cooking nationally. In addition, below I conduct robustness checks by varying the definition of the control group as well as conducting a placebo analysis. The robustness of the results to these specifications help assuage concerns about the non-validity of parallel trends.

2.6.2 Households with MIL dead as control

A reason to suspect the validity of parallel trends is that the treatment and control households at baseline are quite different in levels of observable variables, and thus one might suspect that these would be trending differently as well. In order to assuage this concern, I use the co-residing households where the MIL only died as the control group instead of all households where the father-in-law did not die. These households, as can be seen from appendix table B.7 are quite similar on most observable characteristics (except the household size, number of adults and probability of collecting dung) and therefore it is plausible that these groups would not be trending differently on use of solid fuels either. The results of the regression are presented in Tables 2.5 and 2.6. We see that the effect size on the use of dirty fuels is similar (albeit not significant due to being under-powered with a smaller sample size).

Table 2.5: Impact on solid fuel use with households where MIL died as the control group

	(1) Any Dirty Fuel	(2) Firewood	(3) Dung	(4) Crop Residue	(5) Charcoal
Post	0.075*** (0.026)	0.055** (0.027)	0.061** (0.026)	0.125*** (0.022)	-0.001 (0.007)
Post x FIL dead	-0.038 (0.029)	-0.023 (0.031)	-0.024 (0.027)	-0.041 (0.025)	-0.005 (0.007)
Constant	0.577*** (0.010)	0.526*** (0.010)	0.355*** (0.009)	0.108*** (0.008)	0.026*** (0.002)
Baseline Mean	0.56	0.52	0.38	0.10	0.03
N	4180	4176	4163	4106	4044

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 for households where the FIL died between baseline and endline. FIL dead=0 for households where the MIL died between baseline and endline Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Impact on solid fuel collection with households where MIL died as the control group

	(1) Collects Firewood	(2) Collects Dung	(3) Collects Crop Residue
Post	-0.013 (0.022)	-0.032 (0.021)	0.100*** (0.026)
Post x FIL dead	-0.042 (0.026)	-0.044* (0.025)	-0.032 (0.029)
Constant	0.628*** (0.007)	0.467*** (0.007)	0.182*** (0.009)
Baseline Mean	0.63	0.51	0.18
N	4176	4161	4106

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 for households where the FIL died between baseline and endline. FIL dead=0 for households where the MIL died between baseline and endline Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6.3 Households where Mother-in-Law died as placebo

I also use households where the mother-in-law died as a placebo. Given that I have used household fixed effects and time fixed effects, the coefficients could be biased if there are time-varying variables correlated with Post x FIL dead. However, most of these variables should also be correlated with Post x MIL dead and the coefficients of a regression on this therefore should be similarly biased. However, we observe in tables 2.7, 2.8 and 2.9 that the coefficient on Post x Only MIL dead⁴ are opposite in sign to what we observed with father-in-law dead as well as insignificant. Table 2.7 for example, shows that the probability of using any solid fuel for cooking as well as the probability of using each solid fuel for cooking is actually higher due to the mother-in-law's death. So is the probability of collecting different solid fuels as seen in table 2.8. This further provides proof that my identification is robust.

⁴I exclude households where the father-in-law died from the analysis. Hence the 'Only MIL dead'

Table 2.7: Impact on solid fuel use with MIL dead as the treatment group

	(1) Any Dirty Fuel	(2) Firewood	(3) Dung	(4) Crop Residue	(5) Charcoal
Post	0.046*** (0.011)	0.046*** (0.011)	0.048*** (0.009)	0.096*** (0.009)	-0.004** (0.002)
Post x Only MIL dead	0.029 (0.024)	0.009 (0.026)	0.013 (0.024)	0.029 (0.022)	0.004 (0.006)
Constant	0.555*** (0.006)	0.501*** (0.006)	0.324*** (0.005)	0.099*** (0.005)	0.031*** (0.001)
Baseline Mean	0.53	0.48	0.30	0.09	0.03
N	23622	23594	23473	23169	22878

Post is an indicator for the second wave of the IHDS survey. Only MIL dead=1 if only the MIL has died in the household between the two waves and 0 otherwise. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with MIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Impact on solid fuel collection with MIL dead as the treatment group

	(1) Collects Firewood	(2) Collects Dung	(3) Collects Crop Residue
Post	-0.024*** (0.006)	-0.039*** (0.006)	0.082*** (0.010)
Post x Only MIL dead	0.011 (0.022)	0.007 (0.021)	0.018 (0.025)
Constant	0.589*** (0.004)	0.424*** (0.003)	0.167*** (0.005)
Baseline Mean	0.57	0.40	0.16
N	23590	23463	23163

Post is an indicator for the second wave of the IHDS survey. Only MIL dead=1 if only the MIL has died in the household between the two waves and 0 otherwise. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with MIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Impact on clean fuel use with MIL dead as the treatment group

	(1) Any Clean Fuel	(2) Kerosene	(3) LPG
Post	0.044*** (0.008)	-0.054*** (0.007)	0.103*** (0.006)
Post x Only MIL dead	-0.031 (0.026)	-0.025 (0.019)	-0.023 (0.021)
Constant	0.407*** (0.005)	0.151*** (0.004)	0.293*** (0.003)
Baseline Mean	0.43	0.15	0.32
N	23635	23731	23731

Post is an indicator for the second wave of the IHDS survey. Only MIL dead=1 if only the MIL has died in the household between the two waves and 0 otherwise. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with MIL at baseline.

0.8* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6.4 Matched difference-in-differences

Another way to strengthen the credibility of the parallel trends assumption is to enhance the comparability of the treatment and control groups by matching on observable baseline variables. I use propensity score matching and inverse probability weighting.⁵ I match on five variables: Inflation adjusted annual household income, asset quintile that the household is in, highest female education level in the household, highest education level in the household and whether the household belongs to the backward castes or not.

Matching on these observable household variables could help address endogeneity from unobservable variables as well. For example, these variables could determine the health status as well as ability to seek medical attention of individuals in the household and therefore be correlated with the death of the father-in-law. Table 2.1 shows that households where father-in-law dies by the second wave tend to be poorer and less educated. These two variables could also put the treatment and control

⁵Used the user written psmatch2 on Stata to get propensity scores. I also use kernel matching and nearest neighbour matching and the results are similar. Results are available on request.

households on different trends because, first, eligibility to various social schemes in India at the local as well as national level are determined generally on the basis of household wealth, income, caste or a combination of all three. Second, household education may determine how much information they have about different government schemes as well as their knowledge about the harms of using solid fuels for cooking. Overall, this matching process ensures we are only comparing households that were similar on many important characteristics to begin with and thus for them, the parallel trends assumption seems more plausible.

Tables 2.10- 2.12 show the regression estimated by equation 2.1 for the matched sample. The probability of using any dirty fuel for cooking is lower by 4.6 percentage points and significant at the 1% level. The probability of using every commonly used solid fuel for cooking is also significantly lower just as our main results. In fact, the effect sizes in this matched sample are slightly larger than in our main results (Table 2.2)

Table 2.10: Impact on solid fuel use: Matched difference-in-differences

	(1) Any Dirty Fuel	(2) Firewood	(3) Dung	(4) Crop Residue	(5) Charcoal
Post	0.058*** (0.012)	0.054*** (0.013)	0.052*** (0.011)	0.112*** (0.011)	-0.005** (0.003)
Post x FIL dead	-0.036** (0.015)	-0.034** (0.016)	-0.026** (0.013)	-0.038*** (0.013)	0.001 (0.004)
Constant	0.575*** (0.007)	0.517*** (0.007)	0.360*** (0.006)	0.115*** (0.006)	0.027*** (0.001)
Baseline Mean	0.55	0.49	0.34	0.10	0.03
N	19853	19833	19760	19489	19233

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 if the household is one where the FIL had. died by the second wave. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample is obtained from an 1 to m propensity score matching procedure

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We also see similar direction and magnitude of effect on the usage of clean fuels in table 2.12 as the main result. Table 2.11 shows that the probability of collecting the three main solid fuels is also lower and is in the range of 2 percentage points to 4.2 percentage points.

Overall, all the tests in the last three subsections point in the same direction and thus can be seen as evidence for the robustness of our results. The next section talks about the plausible mechanisms that could be driving our results.

Table 2.11: Impact on solid fuel collection: Matched difference-in-differences

	(1) Collect Any Fuel	(2) Collects Firewood	(3) Collects Dung	(4) Collects Crop Residue
Post	-0.026*** (0.006)	-0.022*** (0.008)	-0.043*** (0.007)	0.096*** (0.011)
Post x FIL dead	-0.016 (0.011)	-0.030** (0.013)	-0.026** (0.012)	-0.042*** (0.014)
Constant	0.708*** (0.004)	0.615*** (0.005)	0.459*** (0.004)	0.187*** (0.006)
Baseline Mean	0.69	0.60	0.45	0.17
N	19852	19832	19753	19486

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 if the household is one where the FIL had. died by the second wave. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample is obtained from an 1 to m propensity score matching procedure

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.7 Mechanisms

2.7.1 Preference for food cooked in the traditional way

As discussed earlier, the biggest beneficiary of cooking with clean cook stoves is the primary cook and the father-in-law either is unaware of the potential risks of cooking with solid fuels or is aware of the risks but still prefers that the primary cook cooks with it. In either case, one reason why the father-in-law might hinder switching to cleaner cooking practices is if he has a preference for food that is cooked in the traditional way.

In India, food until very recently was cooked in a labor intensive manner, without using any modern equipment. This includes preparing *rotis* (Indian flatbread) on an open flame over solid fuel, cooking staples such as rice and lentils by boiling it in a pot, hand grinding spices and herbs

Table 2.12: Impact on clean fuel use: Matched difference-in-differences

	(1) Any Clean Fuel	(2) Kerosene	(3) LPG
Post	0.031*** (0.010)	-0.057*** (0.008)	0.091*** (0.007)
Post x FIL dead	0.018 (0.014)	0.011 (0.012)	0.010 (0.012)
Constant	0.405*** (0.005)	0.148*** (0.004)	0.294*** (0.004)
Baseline Mean	0.42	0.15	0.31
N	19847	19936	19936

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 if the household is one where the FIL had. died by the second wave. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample is obtained from an 1 to m propensity score matching procedure

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

as well as preparing every meal ‘fresh’ with ‘fresh’ ingredients, that is ingredients and cooked food were seldom stored in refrigerators. On the other hand, food can be cooked with less effort and rather quickly using clean cook stoves (which burn more efficiently), pressure cooking lentils and rice, using an electric mixer and grinder, as well as eating refrigerated meals.

The preference of the household head combined with his apathy towards the primary cook or lack of awareness (or both) could propagate the use of traditional cooking practices even if it is detrimental for the primary cook. Table 2.13 shows that the death of the FIL leads to an increase in the adoption of pressure cooker by almost 4 percentage points (about 8 % on a baseline control mean of 50 %), of refrigerator by about 2.5 percentage points (about 10 % on a baseline control mean of 22 %) and of mixer/grinder by 1.5 percentage points (about 6 % on a baseline control mean of 29 %). This hypothesis is also supported by table 2.14 which shows that technology that is primarily used by men like motorcycle, tractor and car, shows no such increase post the FILs death.

Table 2.13: Impact of FIL death on other kitchen labor saving technology

	(1) Pressure Cooker	(2) Refrigerator	(3) Mixer/Grinder
Post	0.074*** (0.007)	0.087*** (0.005)	0.065*** (0.006)
Post x FIL dead	0.039*** (0.012)	0.024** (0.010)	0.016 (0.011)
Constant	0.470*** (0.003)	0.189*** (0.003)	0.251*** (0.003)
Baseline Mean	0.50	0.22	0.29
N	20333	20349	20356

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with either MIL or FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.14: Impact of FIL death on technology not used by adult women

	(1) Motor Vehicle	(2) Motorcycle	(3) Car	(4) Tractor
Post	0.093*** (0.006)	0.087*** (0.006)	0.031*** (0.003)	0.020*** (0.006)
Post x FIL dead	-0.010 (0.012)	-0.009 (0.012)	-0.003 (0.006)	-0.009 (0.012)
Constant	0.248*** (0.003)	0.243*** (0.003)	0.025*** (0.002)	0.076*** (0.003)
Baseline Mean	0.28	0.28	0.04	0.07
N	20358	20356	19724	11192

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with either MIL or FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.7.2 *Other mechanisms*

Another way in which the death of a father-in-law could affect the use of solid fuels is through its effect on household income and the household size. However, as shown in appendix table [B.8](#) , the effect on annual household income as well as the number of persons (and adults) in the household is negative. If anything, this should drive the household to adopt more solid cooking fuel as these are relatively less expensive (and could even be collected freely) compared to clean cooking fuel.

[Debnath \(2015\)](#) has shown that the death of the patriarch could lead to the remaining household splitting into two or more. Split households might in turn use higher cleaner fuels because the primary cooks are free from the pressure of other family members. Appendix Tables [B.9](#) -[B.11](#) show why this is not likely. First, in this sample, the probability of a household splitting is about 6 percentage points lower in treatment households (where father-in-law dies) vs control. This might be the case because we are looking at a much shorter time frame than [Debnath \(2015\)](#) is. Additionally, using split household as the treatment, we find that they are not likely to use less dirty fuel for cooking as compared to households that are not split. They are also not less likely to collect fuel post the split.

Another possible pathway is that when the father-in-law dies, the daughter-in-law's autonomy within the household increases and she is able to play a larger part in decision making within the household. The parent-in-law has been shown to exert significant influence on the life of a daughter-in-law in this context ([Khanna and Pandey, 2021](#); [Anukriti et al., 2020](#)). To investigate this I first construct an index of autonomy and an index of financial autonomy following [Anderson \(2008\)](#). The autonomy index is constructed from binary indicators for the daughter-in-law which include whether she has a say in cooking decisions, in household purchases, in the number of children, what to do if the child is ill and in their children's wedding. It also includes questions about whether the daughter-in-law needs permission to go to the health centre, to a friend's home or the grocery shop. I also include whether the daughter-in-law practices purdah and whether she eats after others in

the household. The financial autonomy index includes an indicator variable that includes whether the daughter-in-law has her own bank account and whether she has cash in hand that she can use.

Appendix table [B.12](#) shows the impact of father-in-law's death on these indices. Even though the effect is positive, these are not statistically significant. This provides further evidence that the death of the father-in-law encourages clean fuel use via a reduction in food cooked in the traditional way.

2.8 Conclusion

In this paper, I demonstrate that in a multi-generational household within patriarchal societies like India, the father-in-law of the household exerts significant influence over the choice of cooking fuel, despite not directly participating in cooking or related activities like fuel preparation and collection. Using nationally representative panel data from India, I show that when a co-residing father-in-law dies, the probability of using any solid fuel for cooking significantly decreases. The probability of using each commonly used solid fuel like firewood, dung and crop residue also decreases. In addition, it is also less likely that households are collecting any of these fuels. All these point in the direction of switching away from solid fuels.

I also find that the primary mechanism driving these results is the father-in-law's preference for food cooked in the traditional way, that is without any modern equipment. I find that not only does the use of solid fuels go down, the use of mixer-grinder, refrigerators and pressure-cookers goes up. Other potential mechanisms like household splitting, increase in the autonomy of the daughter-in-law and the change in the composition of the household are ruled out as dominant mechanisms. These findings highlight the substantial role of intra-household dynamics in shaping gendered technology adoption and resource use.

The paper however, is not without limitations. First, the data that I use has only two waves which precludes me from performing standard pre-treatment parallel trends tests. Even though I perform matched difference-in-differences, use households where the mother-in-law (only) died as the control group, as well as conduct placebo tests, there may still be concerns. A third wave of the IHDS (which is underway) will alleviate some of these concerns. Second, the paper does not fully show the mechanics of reduction in solid fuels. For example, which clean fuel do households switch to? On the intensive margin, how much is the switch from say firewood to LPG. More data on the quantity of fuel used for cooking is required to answer this question- something that future research can do when such data becomes available.

This paper suggests that women that do not co-reside with their father-in-law would be less likely to use solid fuels for cooking. The paper also provides evidence for the preference of older males for cooked in the traditional way. If such preference is decreasing in age, then we can expect households to automatically shift to sources of cleaner fuel as the older generation passes way. However, policymakers can also take additional steps to accelerate the process. For example, informing older male household members about the costs of using solid fuels for cooking as well as finding ways to decrease their aversion to food cooked using modern technology, either by sharing information about their benefits or subsidizing these technologies could nudge the head towards a mix of fuel choice that leans towards non-polluting fuels.

BIBLIOGRAPHY

- Afridi, F., S. Debnath, and E. Somanathan (2021). A breath of fresh air: Raising awareness for clean fuel adoption. *Journal of Development Economics* 151, 102674.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association* 103(484), 1481–1495.
- Anukriti, S., C. Herrera-Almanza, M. Karra, and R. Valdebenito (2022). Convincing the mummy-ji: improving mother-in-law approval of family planning in india. In *AEA Papers and Proceedings*, Volume 112, pp. 568–72. American Economic Association.
- Anukriti, S., C. Herrera-Almanza, P. K. Pathak, and M. Karra (2020). Curse of the mummy-ji: the influence of mothers-in-law on women in india. *American Journal of Agricultural Economics* 102(5), 1328–1351.
- Bouguen, A. and T. Varejkova (2020). Icw_index: Stata module to aggregate the variables included in the varlist into an index.
- Cheng, C.-y. and J. Urpelainen (2014). Fuel stacking in india: Changes in the cooking and lighting mix, 1987–2010. *Energy* 76, 306–317.
- Cohen, A. J., M. Brauer, R. Burnett, H. R. Anderson, J. Frostad, K. Estep, K. Balakrishnan, B. Brunekreef, L. Dandona, R. Dandona, et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet* 389(10082), 1907–1918.

- Debnath, S. (2015). The impact of household structure on female autonomy in developing countries. *The Journal of Development Studies* 51(5), 485–502.
- Evans, D. K. and E. Miguel (2007). Orphans and schooling in africa: A longitudinal analysis. *Demography* 44, 35–57.
- Gupta, G. and G. Köhlin (2006). Preferences for domestic fuel: analysis with socio-economic factors and rankings in kolkata, india. *Ecological Economics* 57(1), 107–121.
- Hanna, R., E. Duflo, and M. Greenstone (2016). Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves. *American Economic Journal: Economic Policy* 8(1), 80–114.
- International Institute for Population Sciences (IIPS) (2021). National family health survey (nfhs-5), 2019-21: India. Mumbai: IIPS.
- Jeuland, M., V. Bhojvaid, A. Kar, J. Lewis, O. Patange, S. Pattanayak, N. Ramanathan, I. Rehman, J. T. Soo, and V. Ramanathan (2015). Preferences for improved cook stoves: Evidence from rural villages in north india. *Energy Economics* 52, 287–298.
- Jeuland, M., M. A. Desai, E. F. Bair, N. M. A. Cader, D. Natesan, W. J. Isaac, S. Sambandam, K. Balakrishnan, G. Thangavel, and H. Thirumurthy (2023). A randomized trial of price subsidies for liquefied petroleum cooking gas among low-income households in rural india. *World Development Perspectives* 30, 100490.
- Khanna, M. and D. Pandey (2021). Reinforcing gender norms or easing housework burdens? the role of mothers-in-law in determining women’s labour force participation’. published online.
- Kishore, A. and D. Spears (2014). Having a son promotes clean cooking fuel use in urban india: Women’s status and son preference. *Economic Development and Cultural Change* 62(4), 673–699.
- Krishnapriya, P., M. Jeuland, J. Orgill-Meyer, and S. K. Pattanayak (2022). Gendered demand for environmental health: Evidence of complementarities from stove auctions in india. *Available at SSRN* 4216609.

- Lam, N. L., K. R. Smith, A. Gauthier, and M. N. Bates (2012). Kerosene: a review of household uses and their hazards in low-and middle-income countries. *Journal of Toxicology and Environmental Health, Part B* 15(6), 396–432.
- Lelieveld, J., A. Pozzer, U. Pöschl, M. Fnais, A. Haines, and T. Münzel (2020). Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective. *Cardiovascular Research*.
- Miller, G. and A. M. Mobarak (2013). Gender differences in preferences, intra-household externalities, and low demand for improved cookstoves. Technical report, National Bureau of Economic Research.
- Murray, C. J. (2024). Findings from the global burden of disease study 2021. *The Lancet* 403(10440), 2259–2262.
- National Statistical Office (NSO) (2020). Time use in india – 2019. https://www.mospi.gov.in/sites/default/files/publication_reports/Report_TUS_2019_0.pdf. Ministry of Statistics and Programme Implementation, Government of India.
- Palriwala, R. (1993). Economics and patriliney: Consumption and authority within the household. *Social scientist*, 47–73.
- Pattanayak, S. K., M. Jeuland, J. J. Lewis, F. Usmani, N. Brooks, V. Bhojvaid, A. Kar, L. Lipinski, L. Morrison, O. Patange, et al. (2019). Experimental evidence on promotion of electric and improved biomass cookstoves. *Proceedings of the national Academy of Sciences* 116(27), 13282–13287.
- Shupler, M., W. Godwin, J. Frostad, P. Gustafson, R. E. Arku, and M. Brauer (2018). Global estimation of exposure to fine particulate matter (pm_{2.5}) from household air pollution. *Environment international* 120, 354–363.
- Somanathan, E., T. Ninan, M. H. Bergin, M. Jeuland, E. Gupta, U. Kumar, R. Kamdar, V. Chowdhury, S. Chandna, K. Barkjohn, et al. (2022). *Electric stoves as a solution for household air pol-*

lution: Evidence from rural India. Indian Statistical Institute, Delhi, Economics and Planning Unit.

Vyas, S., A. Gupta, and N. Khalid (2021). Gender and lpg use after government intervention in rural north india. *World Development* 148, 105682.

Chapter 3

INTRAHOUSEHOLD DIFFERENCES IN PERCEIVED HEALTH RISKS OF USING SOLID FUELS FOR COOKING¹

3.1 Introduction

Household air pollution (HAP) is one of the biggest health hazards in developing countries. Despite extensive attention from researchers and policymakers, HAP from cooking with solid fuels remains an intractable problem. Some reasons identified in the literature include households' low purchasing power ([Jeuland et al., 2015, 2023](#); [Afridi et al., 2021](#); [Berkouwer and Dean, 2022](#)), unreliable fuel supply ([Somanathan et al., 2022](#)), maintenance needs of 'improved cookstoves' ([Hanna et al., 2016](#)), low intrahousehold bargaining power of women ([Miller and Mobarak, 2013](#); [Krishnapriya, 2022](#)) and cultural preference for food cooked using traditional stoves ([Pattanayak et al., 2019](#); [Gupta and Köhlin, 2006](#)).

Lack of awareness about health risks could be another driver of solid fuel usage ([Afridi et al., 2021](#); [Yokoo et al., 2023](#)). Additionally, women in low-and middle-income countries like India have a much higher exposure to HAP from using solid fuels as compared to men because they are primarily responsible for cooking and related activities in the household ([National Statistical Office, Ministry of Statistics and Programme Implementation, 2020](#)). It is possible therefore, that the lack of awareness about health risks is more acute among men. Such intrahousehold differences in beliefs have been observed to be the case in other contexts. For instance, [Ashraf et al. \(2022\)](#) show that

¹This chapter is based on a joint paper with Ananya Diwakant, a Ph.D student at the University of Washington.

husbands in Zambia have lower awareness of maternal risk factors compared to wives. In addition, an asymmetry in beliefs may be critical in contexts where the primary cook lacks the bargaining power to advocate for technology that would primarily be to their benefit. This is especially true in patriarchal societies where women have low bargaining power in the household and where cultural norms might hinder communication between the primary cook and household head (Vyas et al., 2021; Anukriti et al., 2022).

This paper first examines whether there is a lack of awareness about health risks of cooking with solid fuels and if beliefs about these risks differ between the household head and the primary cook. To do this, we conduct a baseline survey and elicit beliefs about the health risks of using solid fuels specifically for the primary cook and children, both short-term risks (symptoms such as sore eyes, difficulty breathing, dry cough, headaches, skin rashes) and long-term risks (lung cancer, bronchitis, neonatal disorders, shortened lifespan, perception of the relative risk of solid fuel use). These beliefs are elicited separately from both the (male) household head and the (female) primary cook. We find that while both primary cooks and household heads underestimate women and children's health risks of using solid fuels for cooking, there is no statistically significant difference between the household members. Therefore, despite the gendered nature of cooking and related activities, it is not the case that the ones facing higher exposure are more aware of the health risks. This could be because the primary cook, while acknowledging the risk to herself, underestimates the prevalence of the symptoms in the broader population. Moreover, both the household head and the primary cook may be unaware of the link between solid fuel use and long-term health risks, as these consequences are not immediately visible or salient.

We then conduct a cluster randomized controlled trial to examine whether providing accurate information about health risks leads to belief updation and subsequent change in fuel use. We randomly assign villages into one of four equal sized groups and vary two things: i) identity of the information recipient - primary cook or household head. This allows us to test whether the information affects their beliefs and household's behavior differently ii) content of the information

given to household heads - about risks to primary cooks or about risks to both primary cooks and children. This allows us to test whether providing additional information about children who the head might care about more has a larger effect on the household heads' beliefs and behavior. These make up the three treatment groups and the fourth is a pure control group where no information was given.

We find that receiving information directly increases an index of the primary cooks' beliefs about the aforementioned short-term and long-term health risks of using solid fuels by a significant 0.22 standard deviations and household heads' by only around 0.09 standard deviations, which is not significant at 10%. Additionally, to examine the existence and direction of transmission of information within the household, we test if the non-targeted household member also updates their beliefs. We find that in households where the primary cook was informed, household heads' beliefs about long-term health risks also increase by 0.20 standard deviations, and this is not statistically different from the effect when they were directly informed. The effects on primary cooks' beliefs in households where the heads were informed also change almost as much as heads' own beliefs but much lower than when they are directly informed. This suggests transmission of information within the household by both household members but informing the primary cook about the risks to dominate informing the head. This is consistent with the traditional collective household model's assumption of information flowing freely within the household, since neither party has a clear incentive to hide information in this context.

We find, however, that the updation of beliefs does not translate to any meaningful change in cooking fuel use and other mitigating behavior to reduce the negative health consequences of cooking with solid fuels in any treatment group. The primary cook is plausibly the biggest beneficiary of switching to clean cooking fuels but may not control the resources in the household required to make that switch ([Miller and Mobarak, 2013](#)). In this paper, we find that informing the primary cook leads to a larger upward revision in beliefs of health risks to the primary cook compared to household heads, but this does not lead to behavioral change. We find evidence for the fact that this

is because the revision in beliefs is concentrated in those primary cooks who have low bargaining power in household decision making and therefore cannot effect any change in household behavior that is primarily to their benefit. Even though household heads also revise beliefs about the risks to primary cooks in households where the primary cook is informed- specifically in the subgroup where women have low bargaining power, there is no effect on their beliefs about risks to themselves. This may be why households do not significantly change fuel use.

This study contributes to three strands of literature. It adds to the literature on the adoption of clean cooking technology and is the first to examine the intrahousehold differences in perceived health risks of cooking with solid fuels. It also adds to the new and growing literature that looks at the transmission of information within the household ([Conlon et al., 2021](#); [Ziparo, 2016](#); [Ashraf et al., 2022](#)), being the first study to look at such transmission in the context of gendered technology. In addition, this study adds to the literature on information interventions and underscores the importance of intrahousehold dynamics as a reason for why these might fail in contrast to the frictionless household models described [Chiappori \(1992\)](#) and [Lundberg and Pollak \(1996\)](#) where the identity of the information recipient would not matter. A study that is in principle close to ours is [Krishnapriya \(2022\)](#) which provides information about improved cookstoves to men and women separately. However, the author does not examine the impact of the information on beliefs or the intrahousehold differences in them. They also do not examine the importance of information flow. Moreover, their use of pamphlets and posters displayed on the walls of dwellings and in villages may have led to treatment contamination, as other household members could also have been exposed to the information.

Our study also contributes to the literature on why information interventions often succeed in changing beliefs but fail to induce behavioral change (see [Haaland et al. \(2023\)](#) for a great review). We provide the first evidence that such interventions may be ineffective when they are most effective at shifting the beliefs of only a subgroup with limited bargaining power. Even though our setting focuses on a rural household in a developing country, this could be important in other settings

like developed countries, corporate organizations and other institutional contexts where decision making power is concentrated among few individuals or groups but benefits of any intervention accrue to others.

The rest of the chapter is organized as follows: Section 3.2 discusses the setting and experimental design, Section 3.3 describes the main empirical strategy and data used in the analysis, Section 3.4 presents the results, and Section 3.5 concludes.

3.2 Experimental Design

3.2.1 Sample

The study was conducted in the state of Odisha in India. Odisha has one of the highest proportions and absolute numbers of solid fuel users in India. 65.3% of households do not primarily use clean fuel for cooking as compared to 41.4% across India. In rural Odisha, this number is higher at 73.9% compared to 56.8% in rural India ([International Institute for Population Sciences \(IIPS\), 2021](#)). Koraput, our study district, has an even higher proportion of households primarily not using clean fuels for cooking (68.5%). One reason for this is that the district is heavily forested with firewood available freely and abundantly.

Our sample consists of 20 households each in 100 villages in this district. The 100 villages were selected based on a combination of two factors: first, where prior studies were conducted by our survey partner, ensuring representation across all the administrative blocks of the district, and second, logistical feasibility. To be able to get enough households that satisfy our inclusion criteria, we only sampled from villages with more than 80 households as recorded in the 2011 census. Once villages which were also a part of previous studies ² were exhausted, we sampled nearby villages.

²All of these studies were descriptive surveys and none had any interventions.

Households in each village were selected based on systematic random sampling proportionate to the village size, with every n th household chosen from a village landmark. We only included households that primarily used solid fuels for cooking during our baseline and also had a Liquefied Petroleum Gas (LPG) connection. This was to ensure that some of the other constraints of low usage of LPG like unfamiliarity with the technology and unavailability of refills were not strictly binding in our sample. LPG is the most commonly used clean cooking fuel in India and has been the focus of a large government scheme to increase access to clean cooking fuels. We also only included households with a female primary cook, a male household head and at least one child under the age of 6 so that the intrahousehold dynamics relevant to the aforementioned hypotheses are applicable to the sample.³ In every household, we surveyed the female primary cook and the male household head. We were able to reach out to 1995 households out of 2000 during the intervention stage, and were able to survey 1996 households during the endline.⁴

3.2.2 Baseline Survey

The baseline survey elicited two types of beliefs from both the primary cook and the household head. First, we elicited beliefs about long-term health consequence of solid fuel use. Here, as in [Afridi et al. \(2021\)](#), we described different diseases, some of which are associated with solid fuel use and others that are not. We asked whether the individual thought these could be caused or aggravated by using solid fuels. We also asked their beliefs about whether using solid fuels affects lifespan of people aged 60 or above, young adults aged 15 to 49 and, children under the age of 5. We also measured their perception of the disease burden of solid fuel use relative to other diseases and health risks like alcohol use, smoking, dietary risks, very high or low temperatures and unsafe sanitation. We use data from the 2021 Global Burden of Disease data for Odisha ([Murray, 2024](#))

³We randomly chose 20 households in each village that satisfied the inclusion criteria mentioned previously. 11 villages had fewer than 20 eligible households (ranging between 12 to 18), in which case we surveyed the remaining households from an adjacent village. For the analysis, we combine such adjacent villages into one.

⁴In one household, a different household head was surveyed at endline as compared to the baseline and in 4 households, a different primary cook was surveyed.

to compare these beliefs with the actual risks.

Second, we elicited their beliefs about primary cooks, young children and male household heads experiencing symptoms that are associated with using solid fuels like sore/runny eyes or redness, difficulty breathing, dry cough/wheezing, headaches/dizziness and, skin rashes/itching in the preceding 30 days, given that they use dirty/clean fuels similar to [Chattopadhyay et al. \(2021\)](#). The questions were framed as follows: “*Out of 100 primary cooks in households similar to yours in your district that primarily use (solid fuels/not solid fuels) for cooking, how many primary cooks do you think would have had (sore or runny eyes or redness/difficulty breathing/dry cough or wheezing/headaches or dizziness/skin rashes or itching) in the last 30 days?*” The same question was asked about children under age 6 and male household heads.

We also collected information about the prevalence and frequency of the aforementioned symptoms for the primary cook and children under the age of 6 in the household in the last 30 days. This information was key, since we used it to conduct the intervention. In addition to eliciting beliefs and health status, we also collected detailed information about the households’ fuel use and collection patterns, preferences for food cooked with traditional methods, income, consumption as well as measures of autonomy and time use of the primary cook.

Underestimation of health risks at Baseline

We find evidence of significant underestimation by both primary cooks and household heads of the long-term health risks and the short-term health risks especially to primary cooks. About 80% of primary cooks and heads are able to identify only 3 out 7 diseases associated with solid fuel use correctly (Appendix figure [C.1](#)) and more than 65% of heads and primary cooks think use of solid fuels does not affect life expectancy. For short term risks, more than 90% of the primary cooks and household heads underestimate the prevalence of dry cough, sore eyes and headaches for primary

cooks (Appendix table C.1). The under-estimation for short-term risk exists but is less serious for children under the age of 6 (Appendix table C.2). This stems from accurate beliefs about low prevalence of symptoms among children as reported by the primary cooks, rather than accurate beliefs about high prevalence. One reason for this could be that in our sample children do not spend a lot of time with the primary cooks while they are cooking. On average, primary cooks report that children spend around 25 minutes a day with her while she is cooking, with approximately 63% children spending no time at all.

Intrahousehold differences in beliefs at Baseline

To test whether there are intrahousehold differences in beliefs about health risks of cooking with solid fuels, we use the baseline survey data to estimate following equation:

$$y_{ih} = \delta_h + \beta_1 * Head_{ih} + \epsilon_h \quad (3.1)$$

where, y_{ih} is an inverse-covariance weighted index of beliefs of individual i in household h about short-term and long-term health risks to primary cooks and children, δ_h is household fixed effect and ϵ_h is the standard error clustered at the household level. $Head_{ih} = 1$ if we are considering the head's belief. The results below show that there is no statistically significant difference between the beliefs of primary cooks and household heads with the household.

Table 3.1: Intrahousehold Difference in Belief Index

	(1)
	Belief Index
Household Head	0.000
	[-0.042,0.042]
Observations	4000

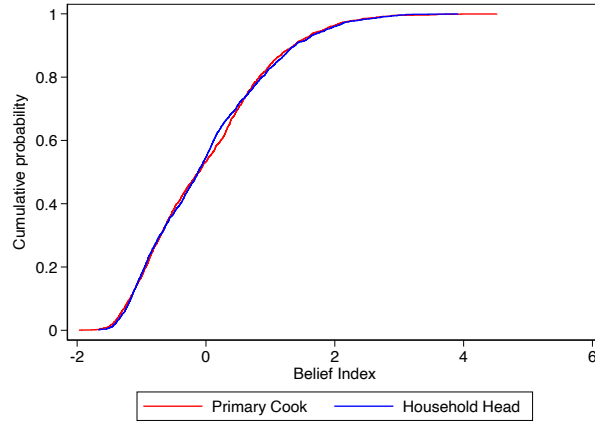


Figure 3.1: Cumulative probability of the belief index for the primary cook and household head

3.2.3 Randomization and Intervention

Since there is evidence for significant underestimation for health risks, we conducted an information intervention. After the baseline, we randomly assigned villages into one of four groups of the same size (i.e., 25 villages in each group), stratified by two variables at baseline: a dummy variable for above median value of an inverse-covariance weighted index of access to LPG ⁵ and a dummy variable for above median value of an inverse-covariance weighted index of underestimation of health risks of solid fuel ⁶. We use these strata as villages with better access to LPG or higher magnitude of underestimation of risks might see a higher impact of the information intervention. We use a cluster randomized design, clusters being villages, to prevent spillover and risk of contamination of treatment as villages in this setting are geographically spread out.

In treatment group 1 (T1), we shared information with the household head about the long-term

⁵The index includes villages' average of time taken on a trip to refill an LPG cylinder, proportion of households that do not receive a subsidy for LPG refills, and 12 minus the number of LPG refills in the last 12 months

⁶The index includes villages' average of magnitude of underestimation of primary cooks experiencing the five symptoms by primary cooks and households heads, number of diseases incorrectly not identified as being associated with solid fuel use, number of risks incorrectly identified as being smaller than solid fuel use and 2 minus the sum of indicator variables of the questions on whether lifespan of young children and older adults are adversely affected

disease burdens and relative risk of using solid fuels for primary cooks as well as children below the age of 6. We also shared descriptive statistics from our baseline survey about the prevalence of various symptoms for the primary cooks and children in our study sample. This included data from 2000 households that primarily use solid fuels (our main sample) as well as data from 245 households in the same villages that primarily use LPG. The intervention for treatment group 2 (T2) was similar to T1, except we did not mention the health risks for children. Treatment group 3 (T3) was similar to T1, except the information was given to the primary cook. We also had a pure control group where no information was shared. The full information script and answers to be given by the enumerators for potential questions asked by the respondents is given in [Appendix C.1](#).

In the treated households, the information was shared by the enumerator in a one-on-one session privately with the targeted individual. Additionally, we also measured the weight of LPG cylinders that were present in the households after the intervention.

Some summary statistics and the validity of our randomization are shown in [Appendix Table C.3](#) . We see that out of 14 variables, only 2 are statistically significant at the 5% level. The household size in the control group is smaller by 0.29 than T1 on a control mean of 5.46 and the average time spent by children in the kitchen with the primary cook is greater in the control group by 0.17 hours on a control mean of 0.35 hours. We control for both of these variables in all of our analyses, and the results are similar if we do not explicitly control but include them in the variables chosen from in PDS LASSO.

3.3 Data and Empirical Strategy

3.3.1 Data

We use data from the baseline, intervention and endline surveys. The endline survey was conducted 16 weeks after intervention and a pre-analysis plan was lodged before the survey started. In the endline, we again elicited beliefs held by both the household head and the primary cook as in the baseline, as well as collected data on their use and collection of different fuels and a few secondary outcomes to measure downstream effects of the interventions like self-reported health outcomes of primary cooks and young children, time use of the primary cooks and self-reported mental health outcomes of the primary cooks and the household heads.

3.3.2 Outcomes

We look at four primary outcomes as outlined in our pre-analysis plan: an index of primary cooks' beliefs about health risks to women and children, an index of household heads' beliefs about health risks to women and children, an index of solid fuel used since the intervention, and an index of clean fuel used since intervention. The belief indices are inverse-covariance weighted indices ([Anderson, 2008](#)) of the household member's beliefs about the long-term health consequences and incidence of short-term symptoms in households that primarily use solid fuels for cooking, and are similar to them in other respects, in their district (all individual components are listed in Appendix Table [C.4](#)). The solid fuel use index and clean fuel use index are also inverse-covariance weighted, and include fuel use on both intensive and extensive margins (Appendix Table [C.5](#) and [C.6](#)).

3.3.3 Empirical Strategy

To estimate the intent-to-treat effects of the information intervention, we use the following ANCOVA model:

$$Y_{iv} = \alpha + \beta_1 T_{1v} + \beta_2 T_{2v} + \beta_3 T_{3v} + Y_{0iv} + X_{0iv} + \gamma_s + \epsilon_{iv} \quad (3.2)$$

where, Y_{iv} is the outcome variable of interest, T_{1v} is a dummy variable equal to 1 if village v belongs to T1, T_{2v} is a dummy variable equal to 1 if village v belongs to T2, T_{3v} is a dummy variable equal to 1 if village v belongs to T3, Y_{0iv} is the equivalent measure, or a close proxy, if available in the baseline survey, X_{0iv} is a vector of controls selected using post double selection LASSO from the variables included in the baseline survey for each outcome, γ_s are strata fixed effects and ϵ_{iv} is the standard error clustered at the village level. The coefficients of interest are β_1 , β_2 and β_3 which are intent-to-treat estimates of being assigned to each treatment arm. When analyzing treatment effects on the primary outcomes, we also account for multiple testing following [Benjamini et al. \(2006\)](#) and report sharpened FDR q values in addition to standard p -values.

3.4 Results

We look at the effect of the information intervention on the four pre-specified outcomes - household head and primary cook's beliefs about health risks to primary cooks and children, a measure of solid fuel use and a measure of clean fuel use - and their components. Table [3.2](#) shows the results for the four primary outcomes together with the sharpened q values.

Table 3.2: Effects of Information Intervention

	(1) Belief Index(PC)	(2) Belief Index(HH)	(3) Solid Fuel Index	(4) Clean Fuel Index
HH informed (PC & children)	0.062 (0.098) [0.525] {1.000}	0.089 (0.094) [0.342] {1.000}	0.024 (0.076) [0.752] {1.000}	-0.075 (0.068) [0.264] {1.000}
HH informed (PC only)	0.009 (0.084) [0.913] {1.000}	0.069 (0.086) [0.425] {1.000}	0.081 (0.060) [0.180] {1.000}	-0.085 (0.079) [0.283] {1.000}
PC informed (PC & children)	0.217 (0.078) [0.006]*** {0.047}**	0.081 (0.089) [0.363] {1.000}	-0.017 (0.093) [0.852] {1.000}	0.068 (0.074) [0.358] {1.000}
N	1994	1995	1996	1996
T1=T2	0.55	0.80	0.42	0.89
T1=T3	0.07	0.92	0.68	0.03
T2=T3	0.00	0.87	0.26	0.05

Standard errors clustered at the village level in parentheses. Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. p-values in square brackets, q-values in curly brackets. Ti=Tj is the p-value from a t-test of equality of the Ti and Tj coefficients. *** p < 0.01, ** p < 0.05, * p < 0.1.

3.4.1 Beliefs about health risks

Primary Cooks' Beliefs

We first look at the effects on the primary cooks' beliefs in the treatment group in which they were the targeted recipient of the information intervention (T3). We see in Column (1) of Table 3.2 that their index of beliefs about health risks to women and children increases by 0.22 standard deviations (an increase in the value of the index indicates a stronger belief that solid fuel use has health risks). We see in row 3 (PC informed) of Table 3.3 that the increase is driven by changes in beliefs about long-term outcomes. Appendix table C.7 shows that the treated primary cooks correctly identified 7.5% more diseases (out of 7) as being associated with solid fuel use than the control mean of 41%, 5% more risk factors (out of 4 for women and out of 3 for children) as being less deadly than cooking with solid fuels over a control mean of 22% and 26% for women and

children respectively. They are also 8.3 percentage points more likely to believe that children's lifespan is adversely affected due to solid fuel usage, an increase of 23% over the control mean of 36%, and are more likely to believe that the smoke from solid fuels significantly affects the health of women and children. Therefore, primary cooks' beliefs about long-term risks to both primary cooks and young children change as a result of the intervention (row 3, Table 3.4).

Table 3.3: Effect of Information Intervention on PC's Beliefs

	(1) Belief Index	(2) Belief Index (LT)	(3) Belief Index (ST)
HH informed (PC & children)	0.062 (0.098)	0.125 (0.112)	-0.044 (0.075)
HH informed (PC only)	0.009 (0.084)	0.114 (0.096)	-0.104 (0.064)
PC informed (PC & children)	0.217*** (0.078)	0.320*** (0.092)	0.067 (0.069)
N	1994	1994	1994
T1=T2	0.55	0.91	0.38
T1=T3	0.07	0.04	0.13
T2=T3	0.00	0.01	0.01

Standard errors clustered at the village level in parentheses. Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. Ti=Tj is the p-value from a t-test of equality of the Ti and Tj coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

However, we see small and statistically insignificant effects on beliefs about short-term health risks (column (3), Table 3.3), i.e. the incidence of the five aforementioned symptoms among women and children (appendix tables C.8 and C.9). This may in part be due to a recall effect and uncertainty about the information given. Exact numbers may have been difficult for the respondents to recall four months after the intervention.

Household Heads' Beliefs

We now discuss the effects on household heads' beliefs for the treatment groups in which they received the information intervention (T1 and T2). For household heads who were informed about

Table 3.4: Effect of Information Intervention on PC's Beliefs

	(1) Belief Index	(2) Belief Index (About PC)	(3) Belief Index (About Child)
HH informed (PC & children)	0.062 (0.098)	0.069 (0.094)	0.087 (0.113)
HH informed (PC only)	0.009 (0.084)	0.050 (0.084)	0.016 (0.092)
PC informed (PC & children)	0.217*** (0.078)	0.251*** (0.080)	0.216** (0.090)
N	1994	1994	1994
T1=T2	0.55	0.83	0.48
T1=T3	0.07	0.03	0.20
T2=T3	0.00	0.01	0.01

Standard errors clustered at the village level in parentheses. Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. Ti=Tj is the p-value from a t-test of equality of the Ti and Tj coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

risks to both women and children, we see in Table 3.2 that their index of beliefs about health risks to women and children increases by 0.09 standard deviations, though this is not statistically significant at 10%. Additionally, as seen in table 3.5, while their beliefs about long-term health consequences increase by 0.13 standard deviations, beliefs about the incidence of short-term symptoms are slightly negative (albeit not significant).

We see similar results for household heads who were informed only about risks to women. Their belief index about risks to women increases by 0.10 standard deviation (table 3.6). Results from t-tests of equality of coefficients of T1 and T2 for all the indices and their components show that equality cannot be rejected, suggesting that the effect of providing additional information about risks to children is similar to informing household heads about risks to primary cooks only.

Table 3.5: Effect of Information Intervention on HH's Beliefs

	(1) Belief Index	(2) Belief Index (LT)	(3) Belief Index (ST)
HH informed (PC & children)	0.089 (0.094)	0.134 (0.101)	-0.034 (0.069)
HH informed (PC only)	0.069 (0.086)	0.141 (0.094)	-0.084 (0.082)
PC informed (PC & children)	0.081 (0.089)	0.204* (0.109)	-0.085 (0.061)
N	1995	1995	1995
T1=T2	0.80	0.93	0.55
T1=T3	0.92	0.46	0.41
T2=T3	0.87	0.47	0.99

Standard errors clustered at the village level in parentheses. Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. Ti=Tj is the p-value from a t-test of equality of the Ti and Tj coefficients. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 3.6: Effect of Information Intervention on HH's Beliefs

	(1) Belief Index	(2) Belief Index (About PC)	(3) Belief Index (About Child)
HH informed (PC & children)	0.089 (0.094)	0.144 (0.096)	0.075 (0.099)
HH informed (PC only)	0.069 (0.086)	0.100 (0.092)	0.075 (0.086)
PC informed (PC & children)	0.081 (0.089)	0.134 (0.086)	0.079 (0.098)
N	1995	1995	1995
T1=T2	0.80	0.63	1.00
T1=T3	0.92	0.91	0.97
T2=T3	0.87	0.66	0.96

Standard errors clustered at the village level in parentheses. Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. Ti=Tj is the p-value from a t-test of equality of the Ti and Tj coefficients. *** p < 0.01, ** p < 0.05, * p < 0.1.

Intrahousehold Information Transmission

Since in the treated households, only one of the household head or the primary cook received the intervention, we can compare the beliefs of the targeted and non-targeted recipients to examine the existence and direction of intrahousehold transmission of information. Table 3.5 shows that when primary cooks are informed, household heads' overall beliefs increase by 0.081 (very similar to giving information directly in T1 and T2). There is also an increase in their long-term beliefs by 0.20 standard deviations. This suggests that information is being shared by the primary cook and is also believed by the household head (particularly about risks to primary cooks as can be seen in table 3.6).

When household heads are given the intervention, we find that the effects on primary cooks' beliefs are similar to those for household heads' beliefs, suggesting transmission of information from the household head to the primary cook as well (Tables 3.3 and 3.5). However, the magnitude of change in primary cooks' beliefs is significantly lower than when they are directly informed. These results suggest that while information is being shared within the household by both the household members, informing primary cooks is more effective than informing household heads at changing both of their beliefs because of the larger direct effect of the intervention on primary cooks.

3.4.2 Solid and Clean Fuel Use

We do not find evidence of overall change in either solid fuel or clean fuel use as a result of the information intervention, despite the aforementioned changes in beliefs (columns 3 and 4 of table 3.2). Neither the indices nor any of their components see a significant change in any of the treatment arms (tables C.13 and C.14). We explore possible reasons for this in our discussions on secondary outcomes and heterogeneity analyses below.

3.4.3 Secondary Outcomes and Mechanisms

Interestingly, from table 3.7, we see that targeted primary cooks' beliefs about risks to household heads also increases by 0.21 standard deviations, despite the information being about risks to women and children only. This suggests that the information led primary cooks to believe more strongly in the overall negative health consequences of using solid fuels for cooking. On the other hand, treatment effects on household heads' beliefs about household heads are very small (0.04 s.d.)⁷ and not statistically significant. This result, along with the observed heterogeneity along the primary cooks' index of empowerment discussed in the next section, helps rationalize the lack of change in fuel usage despite changes in primary cooks' beliefs and changes in household heads' beliefs about risks to primary cooks.

Additionally, we find that the probability of the primary cook raising concerns about the smoke from solid fuels to anyone in the household is also higher in all treatment groups compared to the control (column 1 of table 3.8). When directly informed, primary cooks are 9.2 percentage points more likely to report raising a concern, an increase of 31% over the control mean of 29%. When household heads were informed, the treatment effects are smaller at 6.6 percentage points (though the t-test does not reject the null of equality of these coefficients). This is likely stemming from the intrahousehold communication of the information discussed above.

Interestingly, there is also evidence of improvement in self-reported mental health outcomes of primary cooks in the group in which they were directly informed as measured by the GAD-7 score, with higher improvements for primary cooks (columns 3 and 4 of table 3.8).⁸ Household heads see an improvement in mental health in both the groups where they were the information recipients

⁷The treatment effects on beliefs about primary cooks are 0.14 s.d. and 0.10 s.d. in T1 and T2 respectively and about children are 0.08 s.d. in both T1 and T2 as in table 3.6.

⁸The GAD-7 score is calculated by assigning values 0, 1, 2, 3 to responses "not at all," "several days," "more than half the days," and "nearly every day" respectively, for 7 questions. Scores of 0-4 are associated with minimal anxiety, 5-9 with mild anxiety, 10-14 with moderate anxiety, and 15-21 with severe anxiety.

and the group where the primary cook was the information recipient. The household head might be less anxious if he views the information as indicative of the fact that he is not susceptible to risks from solid fuels. The cook would be less anxious in the group in which she was directly informed because she is able to effect some (albeit minor changes) in cooking behavior.

Table 3.7: Effect of Information Intervention on Beliefs about Household Head

	(1) PC's beliefs about HH	(2) HH's beliefs about HH
HH informed	-0.003 (0.072)	0.035 (0.077)
PC informed	0.207*** (0.068)	0.054 (0.085)
N	1994	1995
Control mean	-0.00	-0.00
Baseline value as control	Yes	Yes
HH informed=PC informed	0.00	0.77

Standard errors clustered at the village level in parentheses. Indices are standardized with mean 0 in the control group. Strata fixed effects, and PDS-LASSO controls included. HH informed=PC informed is the p-value from a t-test of equality of the pooled T1, T2 and T3 coefficients. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 3.8: Effect of Information Intervention (Secondary Outcomes)

	(1) PC raised concern	(2) PC's say in fuel choice	(3) PC's GAD-7 score	(4) HH's GAD-7 score
HH informed	0.066* (0.035)	-0.077 (0.080)	-0.328** (0.154)	-0.256** (0.129)
PC informed	0.092** (0.037)	-0.096 (0.087)	-0.332** (0.160)	-0.007 (0.141)
N	1994	1994	1994	1995
Control mean	0.29	-0.00	3.72	3.75
Baseline value as control	Yes	Yes	No	No
HH informed=PC informed	0.39	0.81	0.98	0.04

Standard errors clustered at the village level in parentheses. Indices are standardized with mean 0 in the control group. Strata fixed effects and PDS-LASSO controls included. HH informed=PC informed is the p-value from a t-test of equality of the pooled T1, T2 and T3 coefficients. *** p < 0.01, ** p < 0.05, * p < 0.1.

3.4.4 Heterogeneity Analysis

Heterogeneity analysis by the primary cook's empowerment index offers an insight into the potential reason behind the null effects on fuel use. We find that in the group where primary cook was directly

informed, the treatment effect on the primary cook's beliefs about themselves and young children, as well as about household heads are higher for those whose empowerment index was below median at baseline (Tables 3.9 and 3.10). On the other hand, while the treatment effect on household heads' beliefs about primary cooks and children are also higher by 0.2 standard deviations when the primary cook is less empowered (column 2, table 3.9) , their beliefs about risks to household heads is not statistically significantly different from the control group (column 2, table 3.10). This suggests that the treatment effects on beliefs about risks to primary cooks and children are higher for women who do not have a say in household purchases and choice of stove and fuel, and they are therefore unable to effect a change in the fuel usage. While household heads' (who by definition have a say in these decisions) beliefs about the risks associated with the primary cook also change when the primary cooks are informed, their beliefs about themselves do not change. Therefore, despite them having a say, the households' fuel usage does not change. This uncovers one key reason for why information interventions might fail to bring about behavioral change. When the benefits accrue mainly to the individuals who lack the bargaining power to influence collective decisions, such interventions are unlikely to produce desired effects.

Other interesting heterogeneity in treatment effects are reported in appendix tables C.17 and C.18 . For primary cooks with education level beyond grade 5, when directly informed, the treatment effects are close to 0.35 s.d., higher than 0.16 s.d. for those less educated. In this sub-group, treatment effect on solid fuel use is -0.19 standard deviations , and for clean fuel use is 0.06 standard deviations consistent with the change in beliefs .

Table 3.9: Heterogeneity in beliefs about risks to PCs and Children by PC's baseline empowerment index

	(1) Belief Index (PC)	(2) Belief Index (HH)
HH informed	0.158 (0.114)	0.084 (0.088)
PC informed	0.298*** (0.114)	0.201** (0.098)
Above-median PC empowerment index	0.003 (0.074)	-0.085 (0.083)
HH informed x Above-median	-0.261** (0.107)	-0.029 (0.111)
PC informed x Above-median	-0.160 (0.119)	-0.268** (0.118)
N	1994.000	1995.000
HH informed (Below-median Empowerment)	0.158	0.084
HH informed (Above-median Empowerment)	-0.102	0.055
PC informed (Below-median Empowerment)	0.298***	0.201**
PC informed (Above-median Empowerment)	0.139	-0.068

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Table 3.10: Heterogeneity in beliefs about risks to Household Head by PC's baseline empowerment index

	(1) Belief about HH (PC)	(2) Belief about HH (HH)
HH informed	0.075 (0.100)	0.009 (0.083)
PC informed	0.247*** (0.089)	0.052 (0.090)
Above-median PC empowerment index	-0.102 (0.089)	-0.221** (0.107)
HH informed x Above-median	-0.189 (0.118)	0.044 (0.128)
PC informed x Above-median	-0.092 (0.127)	-0.006 (0.135)
N	1994.000	1995.000
HH informed (Below-median Empowerment)	0.075	0.009
HH informed (Above-median Empowerment)	-0.114	0.053
PC informed (Below-median Empowerment)	0.247***	0.052
PC informed (Above-median Empowerment)	0.155	0.046

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Table 3.11: Heterogeneity by PC's baseline empowerment index

	(1) Solid Fuel Index	(2) Clean Fuel Index
HH informed	0.159* (0.087)	-0.097 (0.102)
PC informed	-0.008 (0.101)	0.014 (0.113)
Above-median PC empowerment index	0.059 (0.093)	-0.019 (0.106)
HH informed x Above-median	-0.228* (0.117)	0.131 (0.123)
PC informed x Above-median	0.012 (0.137)	-0.008 (0.136)
N	1996.000	1996.000
HH informed (Below-median Empowerment)	0.159*	-0.097
HH informed (Above-median Empowerment)	-0.069	0.034
PC informed (Below-median Empowerment)	-0.008	0.014
PC informed (Above-median Empowerment)	0.004	0.006

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

3.5 Conclusion

We study whether limited awareness about health risks explains the continued use of solid fuels for cooking, by conducting a cluster randomized controlled trial in which either the primary cook or the household head was informed about the actual health risks. While the intervention increased beliefs about long-term health risks, most notably when primary cooks were the informed party, it had no effect on household fuel usage. Treated primary cooks correctly identified more diseases and risk factors as being associated with solid fuels and were more likely to believe that children's lifespan is adversely affected. Informing household heads had weaker effects on beliefs. Evidence of intrahousehold information sharing emerges, but the strength of belief change is greater when primary cooks are directly informed.

Despite changes in beliefs, there was no detectable shift in solid or clean fuel use. Secondary

outcomes and heterogeneity analyses point to mechanisms: primary cooks' beliefs about risks to household heads also increased, but household heads did not update beliefs about themselves. Heterogeneity analysis shows that such belief updating was stronger among primary cooks who have less of a say in household decisions, which might explain why even these groups did not significantly alter behavior. We also find some evidence of unintended consequences of the information intervention if it is directed at members that are not direct beneficiaries.

Taken together, the findings suggest that while information can shift beliefs, particularly for those who the information is perceived to benefit, intrahousehold dynamics and the locus of decision-making limit behavioral change. For policy makers, this suggests that information interventions need to be designed carefully, especially in contexts where benefits accrue to one member whereas decision making authority is with another.

BIBLIOGRAPHY

- Afridi, F., S. Debnath, and E. Somanathan (2021). A breath of fresh air: Raising awareness for clean fuel adoption. *Journal of Development Economics* 151, 102674.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association* 103(484), 1481–1495.
- Ashraf, N., E. Field, A. Voena, and R. Ziparo (2022). Gendered spheres of learning and household decision making over fertility. Technical report, Working Paper.
- Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3), 491–507.
- Berkouwer, S. B. and J. T. Dean (2022). Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households. *American Economic Review* 112(10), 3291–3330.
- Chattopadhyay, M., T. H. Arimura, H. Katayama, M. Sakudo, and H.-F. Yokoo (2021). Subjective probabilistic expectations, household air pollution, and health: Evidence from cooking fuel use patterns in west bengal, india. *Resource and Energy Economics* 66, 101262.
- Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of political Economy* 100(3), 437–467.
- Conlon, J. J., M. Mani, G. Rao, M. W. Ridley, and F. Schilbach (2021). Learning in the household. Technical report, National Bureau of Economic Research.

- Gupta, G. and G. Köhlin (2006). Preferences for domestic fuel: analysis with socio-economic factors and rankings in kolkata, india. *Ecological Economics* 57(1), 107–121.
- Haaland, I., C. Roth, and J. Wohlfart (2023). Designing information provision experiments. *Journal of economic literature* 61(1), 3–40.
- Hanna, R., E. Duflo, and M. Greenstone (2016). Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves. *American Economic Journal: Economic Policy* 8(1), 80–114.
- International Institute for Population Sciences (IIPS) (2021). National family health survey (nfhs-5), 2019-21: India. Mumbai: IIPS.
- Jeuland, M., V. Bhojvaid, A. Kar, J. Lewis, O. Patange, S. Pattanayak, N. Ramanathan, I. Rehman, J. T. Soo, and V. Ramanathan (2015). Preferences for improved cook stoves: Evidence from rural villages in north india. *Energy Economics* 52, 287–298.
- Jeuland, M., M. A. Desai, E. F. Bair, N. M. A. Cader, D. Natesan, W. J. Isaac, S. Sambandam, K. Balakrishnan, G. Thangavel, and H. Thirumurthy (2023). A randomized trial of price subsidies for liquefied petroleum cooking gas among low-income households in rural india. *World Development Perspectives* 30, 100490.
- Krishnapriya, P. (2022). Gendered information and adoption of improved energy technologies in rural india.
- Lundberg, S. and R. A. Pollak (1996). Bargaining and distribution in marriage. *Journal of economic perspectives* 10(4), 139–158.
- Miller, G. and A. M. Mobarak (2013). Gender differences in preferences, intra-household externalities, and low demand for improved cookstoves. Technical report, National Bureau of Economic Research.
- Murray, C. J. (2024). Findings from the global burden of disease study 2021. *The Lancet* 403(10440), 2259–2262.

- National Statistical Office, Ministry of Statistics and Programme Implementation (2020). Time use in india–2019 (january–december 2019). Available at: https://mospi.gov.in/documents/213904/301563/Report_TUS_2019_0.pdf.
- Pattanayak, S. K., M. Jeuland, J. J. Lewis, F. Usmani, N. Brooks, V. Bhojvaid, A. Kar, L. Lipinski, L. Morrison, O. Patange, et al. (2019). Experimental evidence on promotion of electric and improved biomass cookstoves. *Proceedings of the national Academy of Sciences* 116(27), 13282–13287.
- Somanathan, E., T. Ninan, M. H. Bergin, M. Jeuland, E. Gupta, U. Kumar, R. Kamdar, V. Chowdhury, S. Chandna, K. Barkjohn, et al. (2022). *Electric stoves as a solution for household air pollution: Evidence from rural India*. Indian Statistical Institute, Delhi, Economics and Planning Unit.
- Yokoo, H.-F., T. H. Arimura, M. Chattopadhyay, and H. Katayama (2023). Subjective risk belief function in the field: Evidence from cooking fuel choices and health in india. *Journal of Development Economics* 161, 103000.
- Ziparo, R. (2016). Why do spouses communicate: love or interest? a model and some evidence from cameroon. *Unpublished mimeo*.

Appendix A

APPENDIX FOR CHAPTER 1

Parallel Trends

Table A.1: Parallel Pre Trends test using data from 2007-2014

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Year (0-7)	-0.026*** (0.003)	-0.026*** (0.003)	-0.022*** (0.003)	-0.007*** (0.002)	0.001 (0.002)
High intensity x Year	-0.024*** (0.004)	-0.023*** (0.004)	-0.024*** (0.004)	-0.016*** (0.003)	-0.016*** (0.003)
N	2,092,204	2,074,982	2,085,418	231,726	232,499

High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Year is a linear variable with year=0 for year 2007 and year=7 for year 2014 All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Table A.2: Parallel Pre Trends test using data from 2007-2014

	(1)	(2)	(3)	(4)	(5)	(6)
	Recognise 1-9	Recognise 10-99	Division	Recognise letter	Grade 1 text	Grade 2 text
Year (0-7)	-0.005*** (0.001)	-0.008*** (0.002)	-0.006*** (0.001)	-0.011*** (0.001)	-0.007*** (0.001)	-0.001 (0.001)
High intensity x Year	-0.008*** (0.001)	-0.014*** (0.002)	-0.002* (0.001)	-0.010*** (0.001)	-0.007*** (0.002)	-0.003*** (0.001)
N	2074982	2074982	2074982	2085418	2085418	2085418

High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Year is a linear variable with year=0 for year 2007 and year=7 for year 2014 All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * $p < 0.001$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Parallel Pre Trends test using data from 2011-2014

	(1)	(2)	(3)	(4)	(5)	(6)
	Recognise 1-9	Recognise 10-99	Division	Recognise letter	Grade 1 text	Grade 2 text
Year (0-3)	-0.013*** (0.002)	-0.006** (0.003)	-0.001 (0.002)	-0.023*** (0.002)	-0.007*** (0.003)	0.003 (0.002)
High intensity x Year	0.003 (0.002)	0.002 (0.003)	0.003* (0.002)	0.001 (0.003)	0.003 (0.003)	0.001 (0.002)
N	902482	902482	902482	905549	905549	905549
R2	0.06	0.09	0.04	0.07	0.05	0.04

This data only uses ASER data for years 2011- 2014. High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Year is a linear variable with year=0 for year 2007 and year=7 for year 2014. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Table A.4: Impact of PMUY connections on learning for years between 2011-2016 only

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Post	-0.011 (0.017)	0.001 (0.016)	-0.021 (0.019)	0.005 (0.014)	0.010 (0.018)
Post X PMUY connections	0.005 (0.005)	0.002 (0.005)	0.007 (0.005)	0.003 (0.004)	0.003 (0.004)
Mean Dependent	-0.00	0.00	-0.00	0.31	0.50
Mean Dependent:Pre	-0.09	-0.10	-0.08	0.26	0.46
N	419796	418489	419107	43136	43193

Treatment variable is the number of PMUY connections released in each district per 10 households until August 2021. Post=1 for Year =2016 and 0 for Year=2013. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * $p < 0.001$, ** $p < 0.05$, *** $p < 0.01$

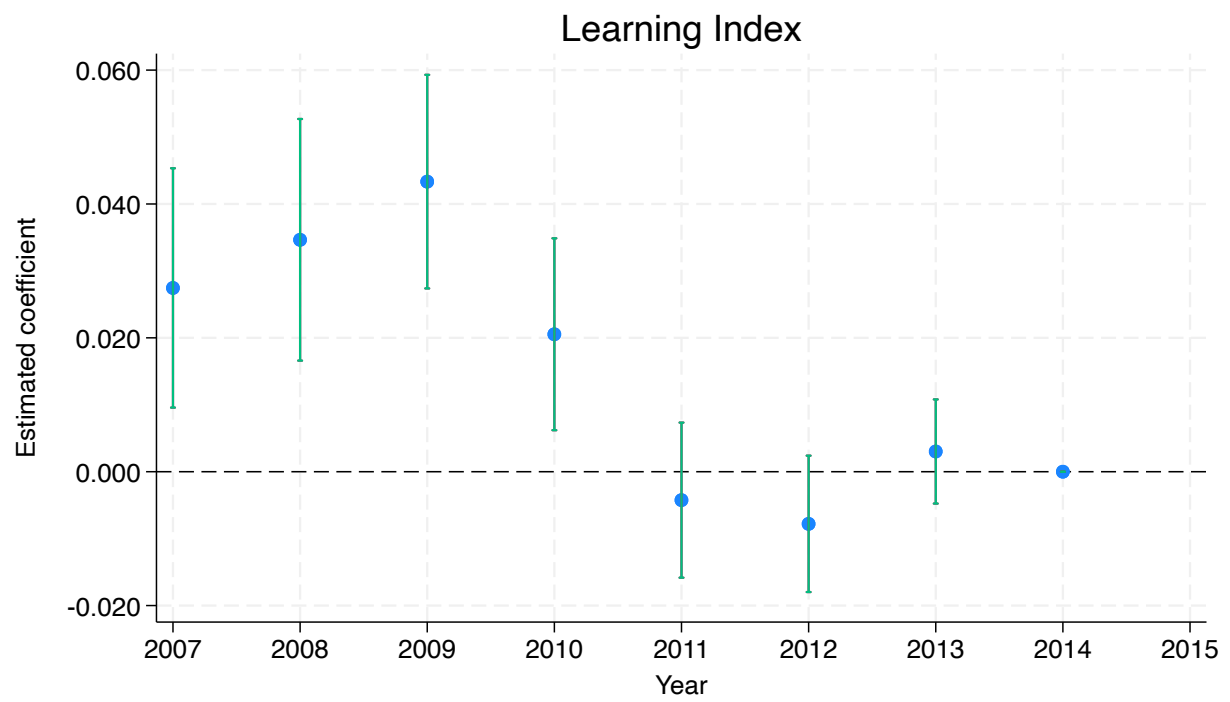


Figure A.1: Pre-treatment coefficient for overall learning

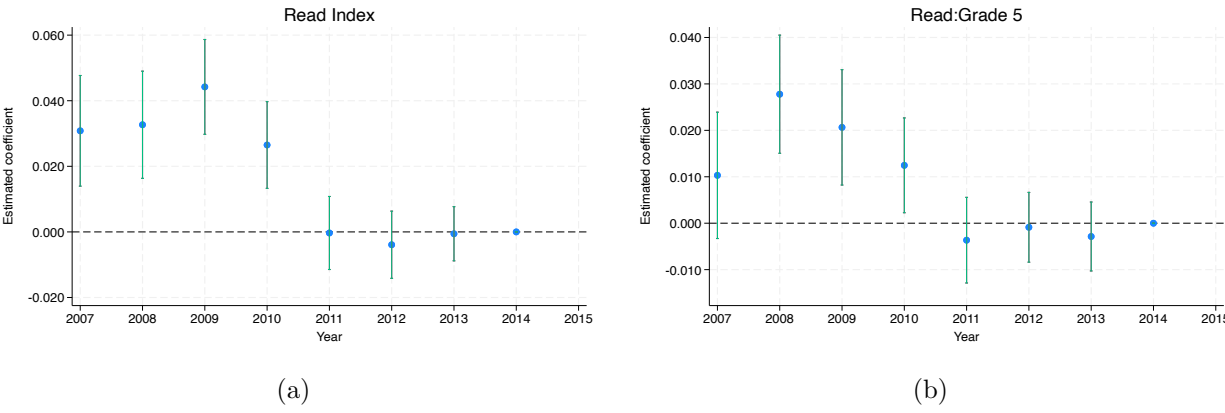


Figure A.2: Pre-treatment coefficient for reading ability

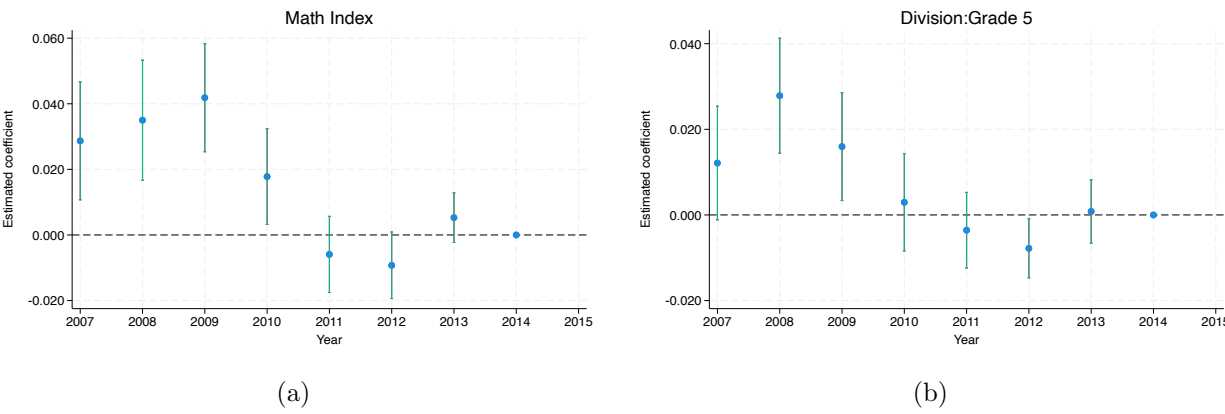


Figure A.3: Pre-treatment coefficient for Math ability

Additional Results

Table A.5: Impact of PMUY connections on learning

	(1)	(2)	(3)	(4)	(5)	(6)
	Recognise 1-9	Recognise 10-99	Division	Recognise letter	Grade 1 text	Grade 2 text
Post	-0.007 (0.005)	0.008 (0.007)	-0.017*** (0.004)	-0.023*** (0.006)	-0.056*** (0.007)	-0.043*** (0.005)
Post X PMUY connections	0.007*** (0.001)	0.009*** (0.002)	0.008*** (0.001)	0.007*** (0.002)	0.010*** (0.002)	0.011*** (0.002)
Mean Dependent	0.83	0.56	0.11	0.81	0.36	0.20
Mean Dependent:Pre	0.81	0.52	0.09	0.78	0.33	0.19
N	1170387	1170387	1170387	1173418	1173418	1173418

Treatment variable is the number of PMUY connections released in each district per household until August 2021 Post=1 for Year =2022 and 0 otherwise All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Table A.6: Impact of PMUY connections on learning: Binary Treatment

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Post	-0.042*** (0.013)	-0.012 (0.013)	-0.080*** (0.014)	-0.027*** (0.008)	-0.086*** (0.011)
Post X High Intensity	0.107*** (0.016)	0.106*** (0.015)	0.092*** (0.017)	0.084*** (0.010)	0.090*** (0.013)
Mean Dependent	-0.00	0.00	-0.00	0.31	0.50
Mean Dependent:Pre	-0.14	-0.14	-0.12	0.26	0.47
N	1175618	1170387	1173418	127086	127313

High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * $p < 0.001$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Impact of PMUY connections on learning: Binary Treatment

	(1)	(2)	(3)	(4)	(5)	(6)
	Recognise 1-9	Recognise 10-99	Division	Recognise letter	Grade 1 text	Grade 2 text
Post	0.000 (0.004)	0.019*** (0.006)	-0.006* (0.003)	-0.015*** (0.005)	-0.044*** (0.006)	-0.030*** (0.005)
Post X High Intensity	0.030*** (0.006)	0.029*** (0.007)	0.024*** (0.003)	0.027*** (0.007)	0.033*** (0.007)	0.034*** (0.005)
Mean Dependent	0.83	0.56	0.11	0.81	0.36	0.20
Mean Dependent:Pre	0.79	0.51	0.08	0.74	0.32	0.20
N	1170387	1170387	1170387	1173418	1173418	1173418

High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * $p < 0.001$, ** $p < 0.05$, *** $p < 0.01$

Binary treatment variable

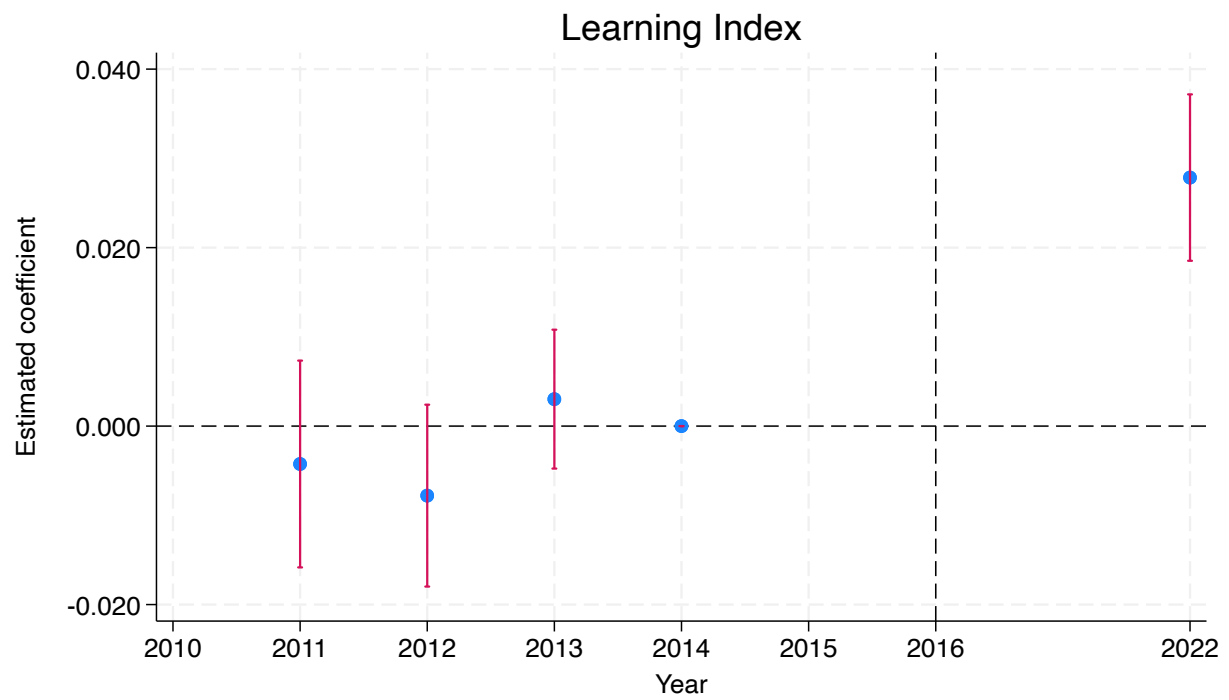
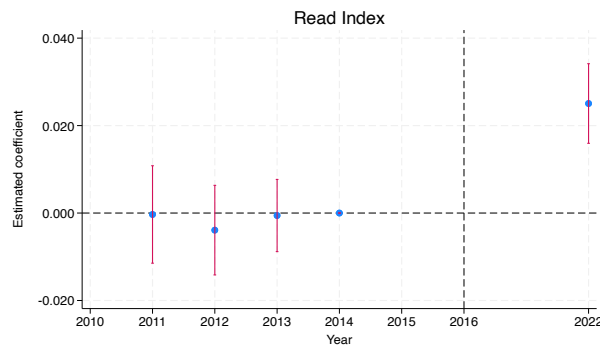
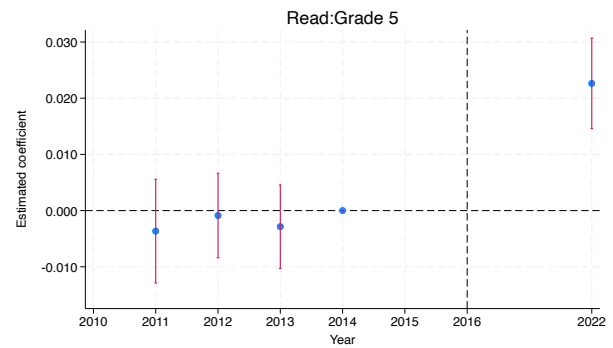


Figure A.4: Event Study style graph for overall learning ability: Binary treatment

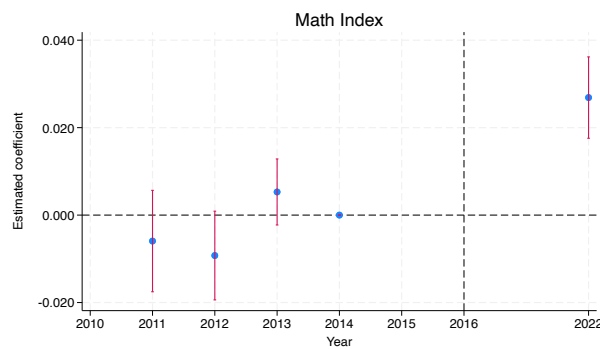


(a)

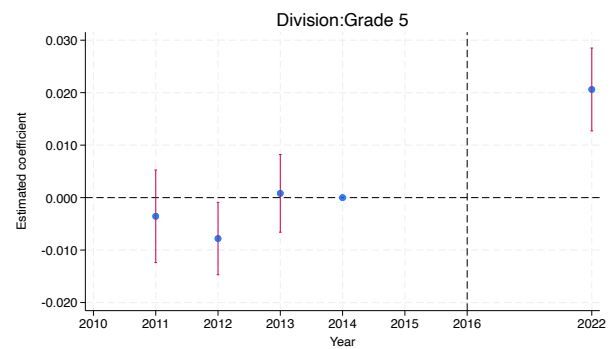


(b)

Figure A.5: Event study style graph for reading ability: Binary treatment



(a)



(b)

Figure A.6: Event Study style graphs for Math ability: Binary treatment

With additional controls

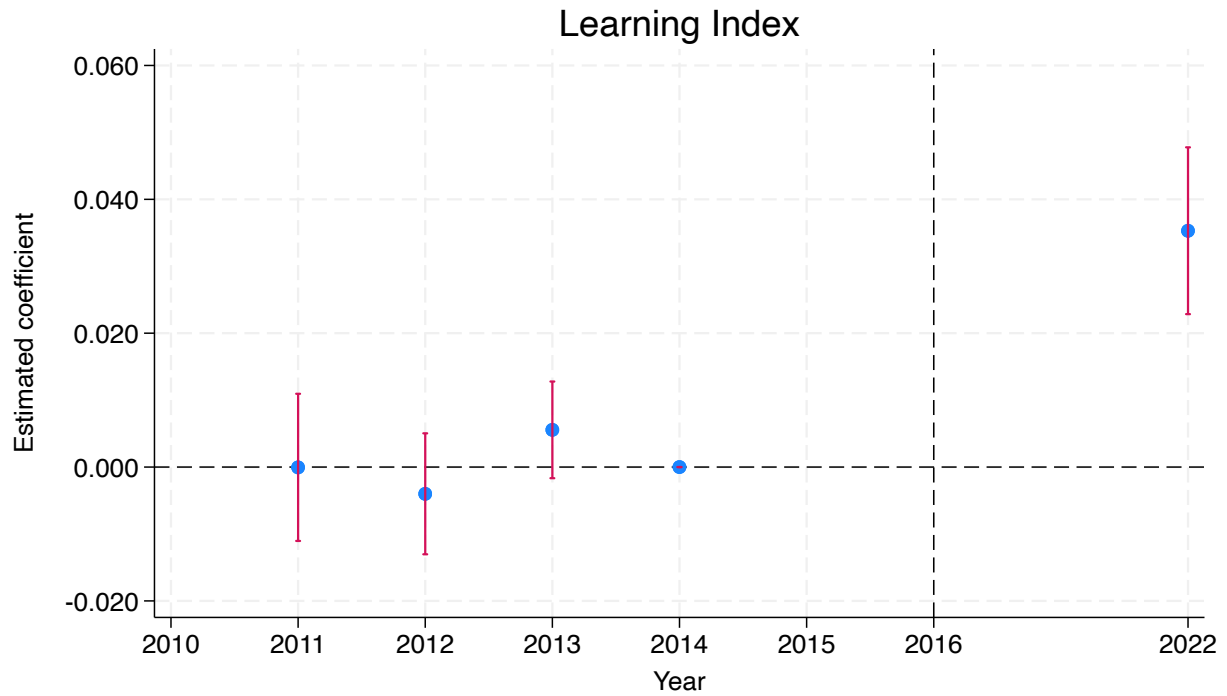


Figure A.7: Event study style plot for overall learning with additional time varying controls

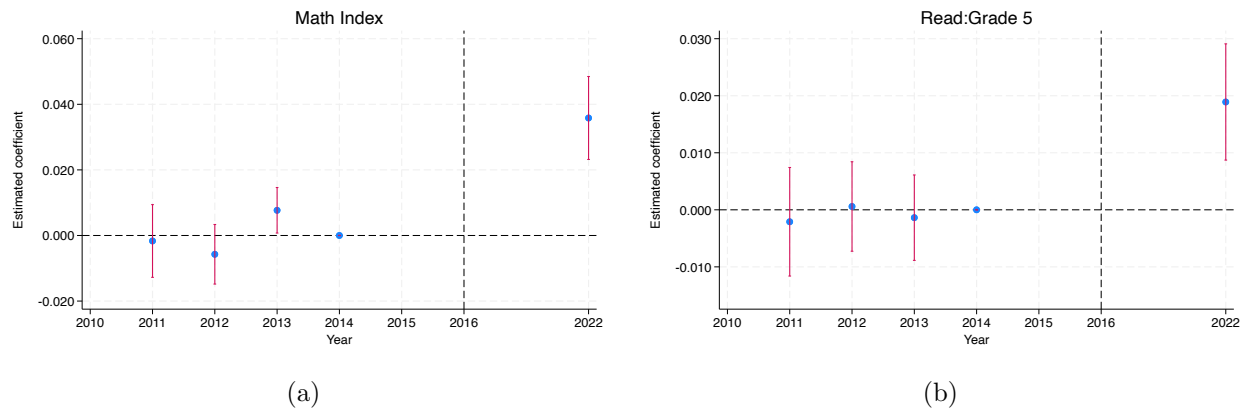


Figure A.8: Event study style plot for reading ability with additional time varying controls

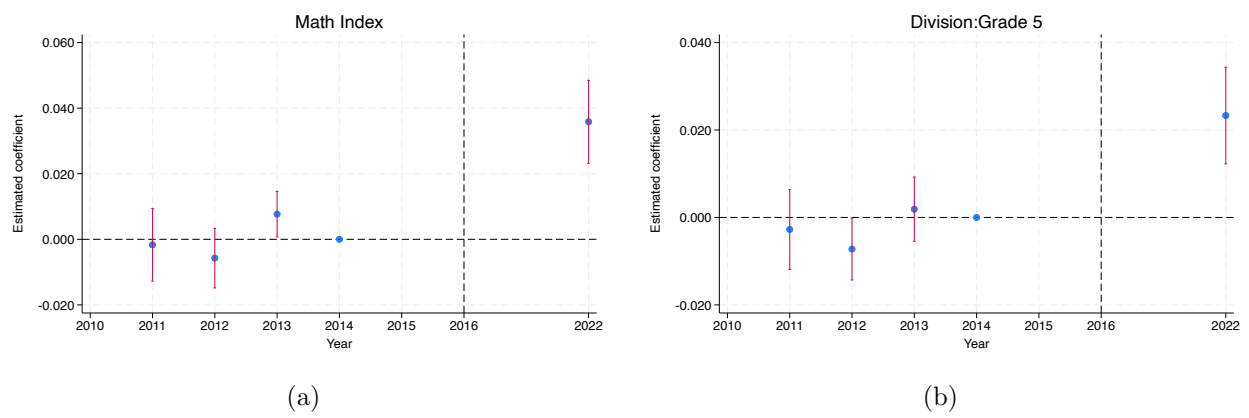


Figure A.9: Event study style plot for Math ability with additional time varying controls

Synthetic Difference-in-Differences

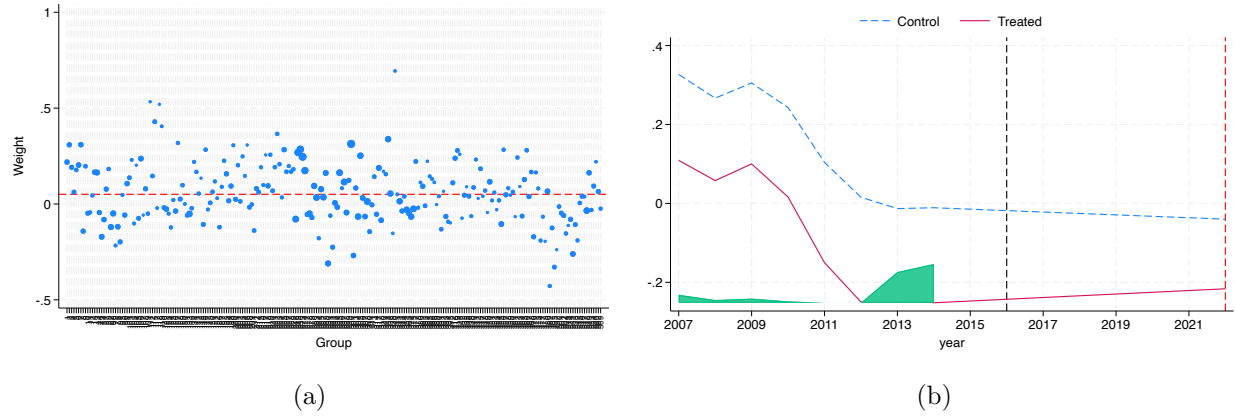


Figure A.10: Synthetic difference in differences unit weights, time weights and parallel trends for the learning index

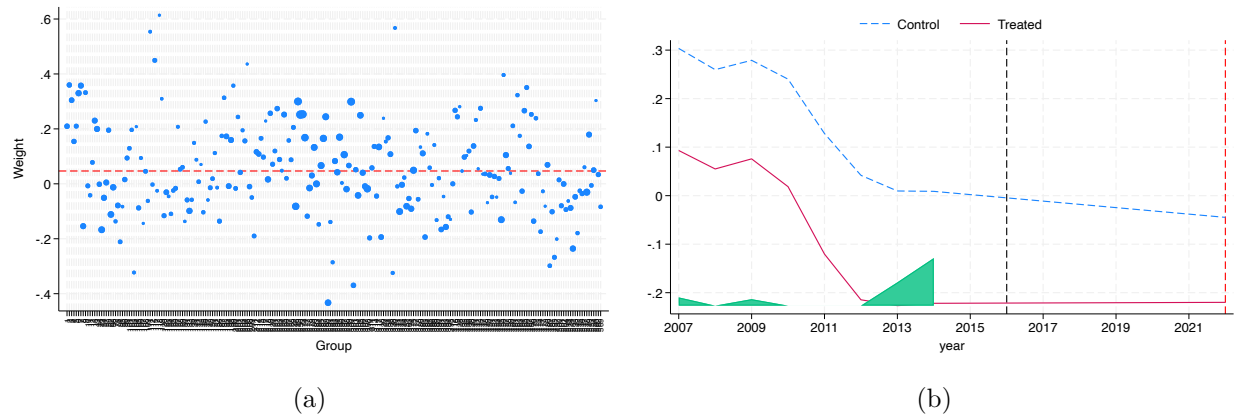


Figure A.11: Unit weights, time weights and parallel trends for the reading index

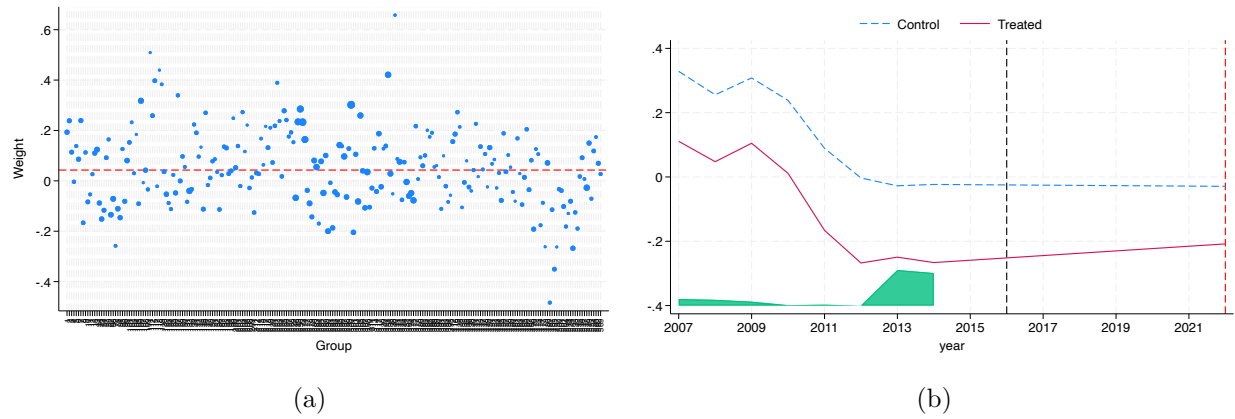


Figure A.12: Synthetic difference in differences unit weights, time weights and parallel trends for math index

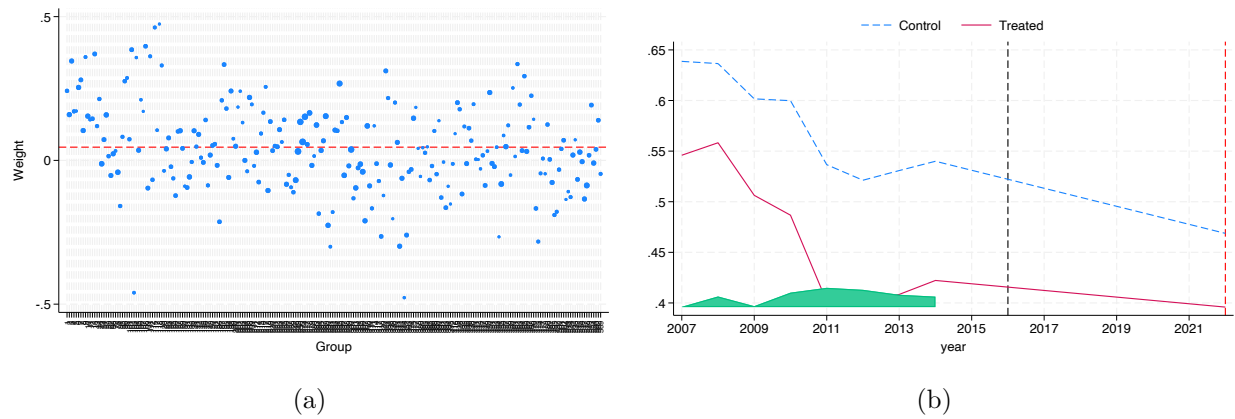


Figure A.13: Synthetic difference in differences unit weights, time weights and parallel trends for grade 5 reading

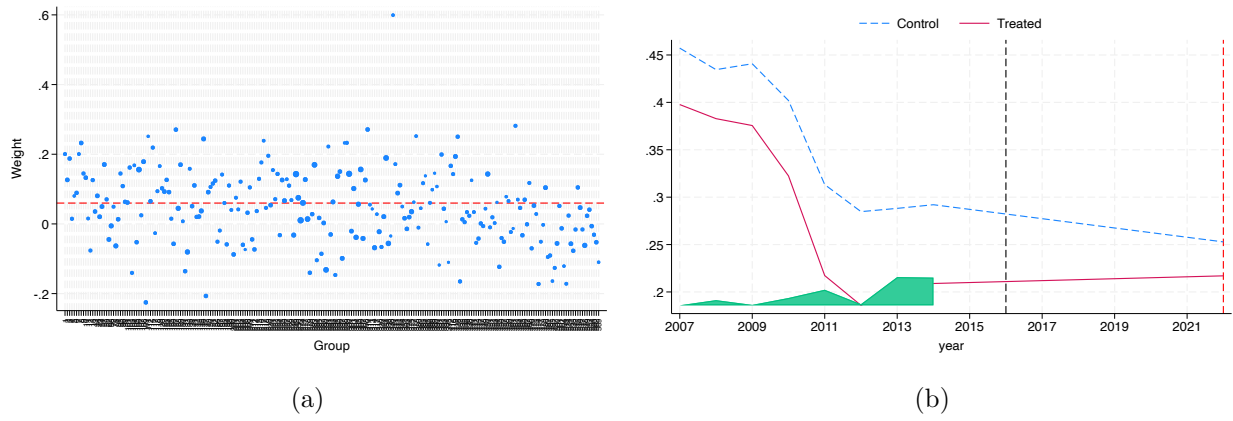


Figure A.14: Synthetic difference in differences unit weights, time weights and parallel trends for grade 5 math

Heterogeneity

Table A.8: Heterogenous impact of PMUY connections on learning by gender: Binary treatment

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
post=1	-0.060*** (0.014)	-0.030** (0.013)	-0.094*** (0.014)	-0.041*** (0.008)	-0.102*** (0.012)
Post X High intensity district	0.091*** (0.017)	0.097*** (0.016)	0.069*** (0.018)	0.079*** (0.011)	0.081*** (0.014)
Female	0.016*** (0.005)	-0.010** (0.005)	0.051*** (0.005)	-0.013** (0.005)	0.034*** (0.005)
Female X High intensity district	-0.084*** (0.007)	-0.073*** (0.007)	-0.091*** (0.007)	-0.038*** (0.007)	-0.059*** (0.008)
Post x Female	0.035*** (0.007)	0.038*** (0.007)	0.028*** (0.008)	0.028*** (0.008)	0.031*** (0.010)
Post X High intensity districtx Female	0.035*** (0.010)	0.022** (0.009)	0.049*** (0.011)	0.012 (0.011)	0.019 (0.014)
N	1170034	1164843	1167856	126622	126847

Treatment variable is the number of PMUY connections released in each district per household until August 2021. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses.

Table A.9: Heterogenous impact of PMUY connections on learning by wealth: Binary treatment

	(1)	(2)	(3)	(4)	(5)
	Learning Index	Math Index	Read Index	Division:Grade 5	Read:Grade 5
Post	-0.330*** (0.025)	-0.311*** (0.024)	-0.326*** (0.026)	-0.193*** (0.020)	-0.180*** (0.025)
Post X High intensity district	-0.076** (0.030)	-0.049* (0.029)	-0.112*** (0.032)	0.012 (0.027)	-0.081*** (0.030)
Middle Wealth	-0.245*** (0.010)	-0.254*** (0.010)	-0.207*** (0.010)	-0.140*** (0.010)	-0.087*** (0.009)
Post X High intensity district X Mid	0.121*** (0.032)	0.088*** (0.031)	0.153*** (0.033)	0.028 (0.028)	0.152*** (0.031)
Poorest	-0.534*** (0.016)	-0.520*** (0.015)	-0.508*** (0.016)	-0.242*** (0.014)	-0.251*** (0.014)
Post X High intensity district X Poor	0.075* (0.042)	0.046 (0.041)	0.118*** (0.043)	-0.006 (0.032)	0.095** (0.040)
Richest	-0.047*** (0.012)	-0.054*** (0.012)	-0.031*** (0.011)	-0.010 (0.012)	0.013 (0.010)
N	1186174	1180904	1183962	128302	128531

Treatment variable is the number of PMUY connections released in each district per household until August 2021. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses.

Mechanisms

Impact on teacher's attendance

Table A.10: Impact of PMUY connections on attendance: continuous treatment

	(1) Teacher Attendance
Post	0.005 (0.005)
PMUY connections X Post	-0.000 (0.002)
Mean Dependent	0.72
Mean Dependent:Pre	0.71
N	106559

Treatment variable is the number of PMUY connections released in each district per household until August 2021. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses.

Impact on out of school children

Table A.11: Parallel pre-trends test for out of school children using data from 2007-2014

	(1) Learning Index	(2) Math Index	(3) Read Index
Year (0-7)	-0.009 (0.011)	-0.008 (0.011)	-0.009 (0.010)
High intensity x Year	-0.013 (0.012)	-0.008 (0.012)	-0.017 (0.011)
N	23600	23252	23452

High intensity=1 if number of PMUY connections per household in that district is greater than the median district. Year is a linear variable with year=0 for year 2007 and year=7 for year 2014. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * $p < 0.001$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Impact of PMUY connections on learning on out-of-school children

	(1) Learning Index	(2) Math Index	(3) Read Index
Post	-0.177** (0.083)	-0.172* (0.091)	-0.186*** (0.066)
Post X PMUY connections	-0.010 (0.019)	-0.007 (0.021)	-0.008 (0.015)
Mean Dependent	-0.00	0.00	-0.00
Mean Dependent:Pre	-0.09	-0.10	-0.08
N	25879	25527	25726

Treatment variable is the number of PMUY connections released in each district per 10 households until August 2021. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Table A.13: Impact of PMUY connections on Enrollment (=1 if child enrolled in school)

	(1) Enrolled in School
Post	-0.021*** (0.003)
Post X PMUY connections	-0.004*** (0.001)
Mean Dependent	0.98
Mean Dependent:Pre	0.99
N	1640469

Treatment variable is the number of PMUY connections released in each district per 10 households until August 2021. Post=1 for Year =2022 and 0 otherwise. All regression include district fixed effects. Heteroskedasticity-robust standard errors clustered at the district level are in parentheses. * p<0.001 , ** p<0.05, *** p<0.01

Appendix B

APPENDIX FOR CHAPTER 2

Tagging FIL/MIL and DIL pairs In this section I describe how I identified the households that had at least 1 FIL-DIL pair using the IHDS household roster for the first wave. I identified four groups of households using the identity of the household head and each member's relation to the household head ¹. The union of these four groups are all the FILs (9056) and DILs (11,312) in our sample. The unique households containing these FIL-DIL pairs is the baseline sample of 8962 households. The groups are as follows:

Group 1: The household head is male. Here the household head is the FIL and the daughter-in-law is the DIL. There are 9,824 DILs and 7,578 FILs in this group in 7,578 households.

Group 2: The household head is male. Here the father of the household head is the FIL and the wife of the household head is the DIL. There are 1,441 DILs and 1,432 FILs in 1,432 households.

Group 3: The household head is female. Here the husband of the household head is the FIL and the daughter-in-law is the DIL. There 17 DILs and 15 FILs in this group in 15 households.

Group 4: The household head is female. Here the father in law of the household head is the FIL and the household head is the DIL. There are 30 DILs and 30 FILs in 30 households.

¹These groups are not exhaustive to identify all FIL-DIL pairs within a household. For example, if the household head is male, then his male sibling and the sibling's daughter in law form an FIL-DIL pair. However, we cannot identify them because the DIL's relationship to the head is noted in a catchall 'Other relation'. These are only 0.5% of the household roster though

Similarly, we tagged MIL-DIL pairs from four groups and identified unique households where they co-resided in during the first wave. The groups are as follows:

Group 1: The household head is male. Here the wife of the head is the MIL and the daughter in law is the DIL. There are 6,802 MILs and 8,801 DILs living in 6,767 households.

Group 2: The household head is male. Here the mother of the head is the MIL and the wife of the head is the DIL. There are 4,630 MILs and 4,651 DILs living in 4,628 households.

Group 3: The household head is female. Here the head is the MIL and the daughter in law of the head is the DIL. There are 1,169 MILs and 1,456 DILs living in 1,169 households.

Group 4: The household head is female. Here the head is the DIL and the mother in law of the head is the MIL. There are 111 MILs and 112 DILs living in 111 households.

The union of these four groups are all the MILs (12,713) and DILs (15,019) in our sample. They live in 12,309 unique households in the baseline ².

²Note that there are 369 females who are tagged as both MIL and DIL in this exercise

Additional Tables and Figures

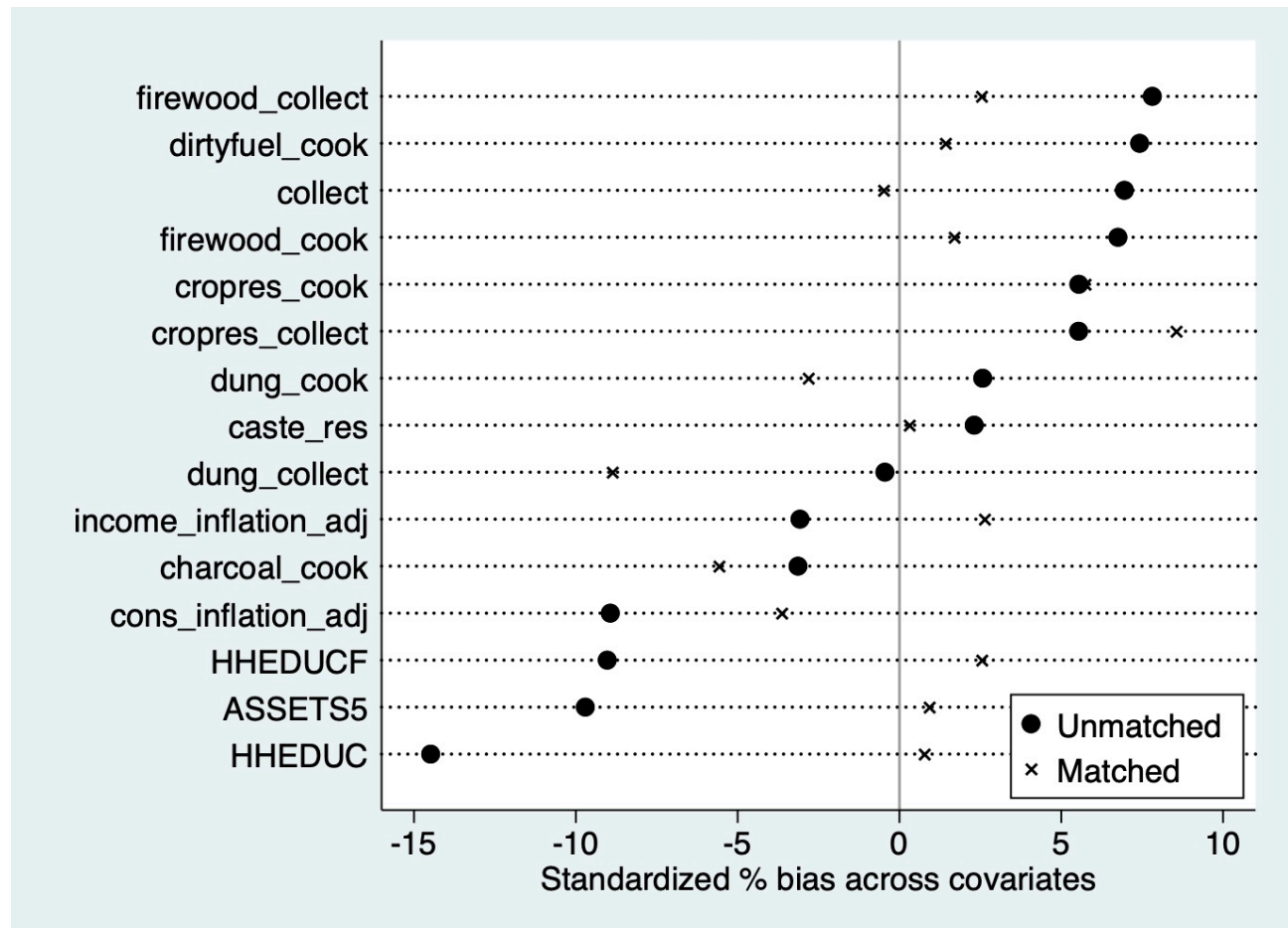


Figure B.1: Bias in covariates in the original sample versus the matched sample

Table B.1 : Impact of the death of the FIL on solid fuel use with additional time varying controls

	(1) Any Dirty Fuel	(2) Firewood	(3) Dung	(4) Crop Residue	(5) Charcoal
Post	-0.001 (0.030)	0.055* (0.031)	-0.001 (0.028)	0.197*** (0.030)	-0.001 (0.007)
Post x FIL dead	-0.027* (0.015)	-0.027* (0.016)	-0.021 (0.013)	-0.031** (0.013)	0.002 (0.004)
Constant	0.666*** (0.030)	0.567*** (0.030)	0.378*** (0.026)	0.087*** (0.024)	0.024*** (0.008)
Baseline Mean	0.54	0.49	0.34	0.10	0.03
N	20288	20265	20160	19888	19629

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with either MIL or FIL at baseline. The regression includes controls for baseline level income, wealth and adult members multiplied by Post

Table B.2 : Impact of the death of the FIL on clean fuel use with additional time varying controls

	(1) Any Clean Fuel	(2) Kerosene	(3) LPG
Post	-0.194*** (0.024)	-0.121*** (0.020)	-0.072*** (0.017)
Post x FIL dead	0.011 (0.014)	0.010 (0.012)	0.002 (0.011)
Constant	0.184*** (0.028)	0.179*** (0.022)	-0.025 (0.022)
Baseline Mean	0.42	0.14	0.32
N	20289	20379	20379

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with either MIL or FIL at baseline. The regression includes controls for baseline level income, wealth and adult members multiplied by Post

Table B.3 : Impact of the death of the FIL on solid fuel collection with additional time varying controls

	(1) Collects any Fuel	(2) Collects Firewood	(3) Collects Dung	(4) Collects Crop Residue
Post	0.019 (0.017)	0.071*** (0.022)	-0.078*** (0.023)	0.183*** (0.032)
Post x FIL dead	-0.004 (0.011)	-0.021 (0.013)	-0.016 (0.012)	-0.029** (0.014)
Constant	0.696*** (0.018)	0.609*** (0.023)	0.482*** (0.024)	0.138*** (0.027)
Baseline Mean	0.68	0.59	0.44	0.17
N	20284	20264	20153	19885

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes all those households where DIL coresides with either MIL or FIL at baseline. The regression includes controls for baseline income, wealth and number of adult members multiplied by Post

Table B.4 : Impact of the death of only the FIL on solid fuel use

	(1) Any Dirty Fuel	(2) Firewood	(3) Dung	(4) Crop Residue	(5) Charcoal
Post	0.056*** (0.012)	0.052*** (0.012)	0.048*** (0.010)	0.105*** (0.010)	-0.004* (0.002)
Post x Only FIL dead	-0.031 (0.033)	-0.045 (0.035)	-0.052* (0.028)	-0.025 (0.031)	-0.004 (0.010)
Constant	0.571*** (0.007)	0.513*** (0.007)	0.364*** (0.006)	0.111*** (0.006)	0.029*** (0.001)
Baseline Mean	0.55	0.49	0.34	0.10	0.03
N	17504	17485	17403	17171	16944

Post is an indicator for the second wave of the IHDS survey. Only FIL dead=1 if the household is one where only the FIL had died by the second wave. All regressions include household fixed effects. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.5 : Impact of the death of only the FIL on clean fuel use

	(1) Any Clean Fuel	(2) Kerosene	(3) LPG
Post	0.038*** (0.009)	-0.054*** (0.008)	0.095*** (0.007)
Post x Only FIL dead	0.048 (0.033)	0.014 (0.028)	0.028 (0.027)
Constant	0.398*** (0.005)	0.146*** (0.004)	0.288*** (0.004)
Baseline Mean	0.42	0.15	0.31
N	17503	17587	17587

Post is an indicator for the second wave of the IHDS survey. Only FIL dead=1 if the household is one where only the FIL had died by the second wave. All regressions include household fixed effects. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.6 : Impact of the death of only the FIL on solid fuel collection

	(1) Collects Firewood	(2) Collects Dung	(3) Collects Crop Residue
Post	-0.028*** (0.008)	-0.049*** (0.008)	0.091*** (0.011)
Post x Only FIL dead	-0.021 (0.032)	-0.017 (0.029)	-0.031 (0.034)
Constant	0.615*** (0.004)	0.473*** (0.004)	0.183*** (0.006)
Baseline Mean	0.59	0.44	0.17
N	17484	17396	17168

Post is an indicator for the second wave of the IHDS survey. Only FIL dead=1 if the household is one where only the FIL had died by the second wave. All regressions include household fixed effects. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.7 : Summary Statistics: Households with FIL dead vs MIL dead

	(1) FIL alive	(2) FIL dead	(3) Difference
Income	134,423.15	135,112.09	688.94
Household Consumption	128,347.30	125,752.63	-2,594.66
Household Assets- quintile (1-5)	3.33	3.35	0.02
Below poverty line	0.25	0.26	0.01
Urban	0.21	0.24	0.03
Highest adult education	8.98	8.66	-0.32
Highest female education	5.45	5.51	0.07
Household size	8.03	7.72	-0.31**
Number of adult males	2.45	2.44	-0.02
Number of adult females	2.49	2.22	-0.26***
Firewood used for cooking	0.52	0.52	0.00
Dung used for cooking	0.38	0.33	-0.05*
Biomass used for cooking	0.56	0.57	0.01
Uses dirty stove	0.80	0.80	0.01
Collects firewood	0.63	0.63	0.00
Collects dung	0.51	0.44	-0.07***
Collects any fuel	0.72	0.71	-0.01
HH uses LPG	0.38	0.38	-0.00

This tables compares households where a coresiding FIL had died between IHDS 1 and IHDS 2 and households where they had not. It is obtained by regression of an indicator variable equal to 1 if the FIL is dead and 0 if the MIL dies on various variable obtained from the 2005 survey

All values are in 2005 Rupees

Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.8 : Impact of the death of the FIL on income and household size

	(1) Annual Income	(2) Income Per Capita	(3) HH size	(4) Number of adults
Post	20327.969*** (3737.583)	-3539.612*** (370.083)	-1.683*** (0.052)	-0.843*** (0.029)
Post x FIL dead	-22343.961*** (8364.637)	-58.229 (929.840)	-0.621*** (0.092)	-0.649*** (0.051)
Constant	125673.843*** (1895.843)	18065.249*** (197.431)	7.465*** (0.025)	4.369*** (0.014)
Baseline Mean	136102.27	19409.52	7.49	4.45
N	20382	20382	20382	20382

Post is an indicator for the second wave of the IHDS survey. FIL dead=1 if the household is one where the FIL had died by the second wave. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.9 : Impact of the death of the FIL on household splitting

	(1) Split HH
FIL dead	-0.059*** (0.011)
Constant	0.282*** (0.005)
Baseline Mean	0.28
N	8962

Split HH=1 if the undivided household in the first wave had split into two or more by the second wave of the survey. FIL dead=1 if the household is one where the FIL had died by the second wave. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.10 : Impact of household splitting on solid fuel use

	(1) Any Dirty Fuel	(2) Firewood	(3) Dung	(4) Crop Residue	(5) Charcoal
Post	0.045*** (0.012)	0.042*** (0.012)	0.043*** (0.010)	0.091*** (0.010)	-0.004 (0.003)
Split HH * Post	0.005 (0.014)	0.005 (0.015)	0.005 (0.013)	0.030** (0.012)	-0.002 (0.004)
Baseline Mean	0.54	0.49	0.34	0.10	0.03
N	20289	20266	20161	19889	19630

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.11 : Impact of the death of the FIL on solid fuel collection

	(1) Collects Firewood	(2) Collects Dung	(3) Collects Crop Residue
Post	-0.024*** (0.008)	-0.030*** (0.007)	0.083*** (0.011)
Split HH * Post	-0.015 (0.012)	-0.052*** (0.012)	0.006 (0.014)
Baseline Mean	0.59	0.44	0.17
N	20265	20154	19886

Post is an indicator for the second wave of the IHDS survey Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.12 : Impact of the death of FIL on DIL's autonomy

	(1) Autonomy Index	(2) Financial Autonomy Index
Post	0.303*** (0.028)	0.545*** (0.023)
Post x FIL dead	0.053 (0.041)	0.010 (0.036)
Constant	0.003 (0.013)	-0.016 (0.011)
Baseline Mean	0.00	0.00
N	15202	15181

Post is an indicator for the second wave of the IHDS survey. Heteroskedasticity-robust standard errors clustered at PSU level are in parentheses. Sample includes only those households where DIL coresides with both MIL and FIL at baseline.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C

APPENDIX FOR CHAPTER 3

Underestimation of health risks

Table C.1 : Underestimation of Primary Cook's Symptoms

	(1)		(2)	
	Primary	Cook	Household	Head
	mean	sd	mean	sd
Proportion underestimating PCs dry cough	0.93	0.25	0.96	0.20
Proportion underestimating PCs sore eyes	1.00	0.05	1.00	0.05
Proportion underestimating PCs difficulty breathing	0.61	0.49	0.63	0.48
Proportion underestimating PCs skin itching	0.49	0.50	0.52	0.50
Proportion underestimating PCs headaches	0.96	0.21	0.97	0.17
Observations	2000		2000	

Table C.2 : Underestimation of Children's Symptoms

	(1)		(2)	
	Primary	Cook	Household	Head
	mean	sd	mean	sd
Proportion underestimating children's dry cough	0.71	0.45	0.75	0.43
Proportion underestimating children's sore eyes	0.25	0.43	0.26	0.44
Proportion underestimating children's difficulty breathing	0.28	0.45	0.26	0.44
Proportion underestimating children's skin itching	0.36	0.48	0.35	0.48
Proportion underestimating children's headaches	0.31	0.46	0.29	0.45
Observations	2812		2812	

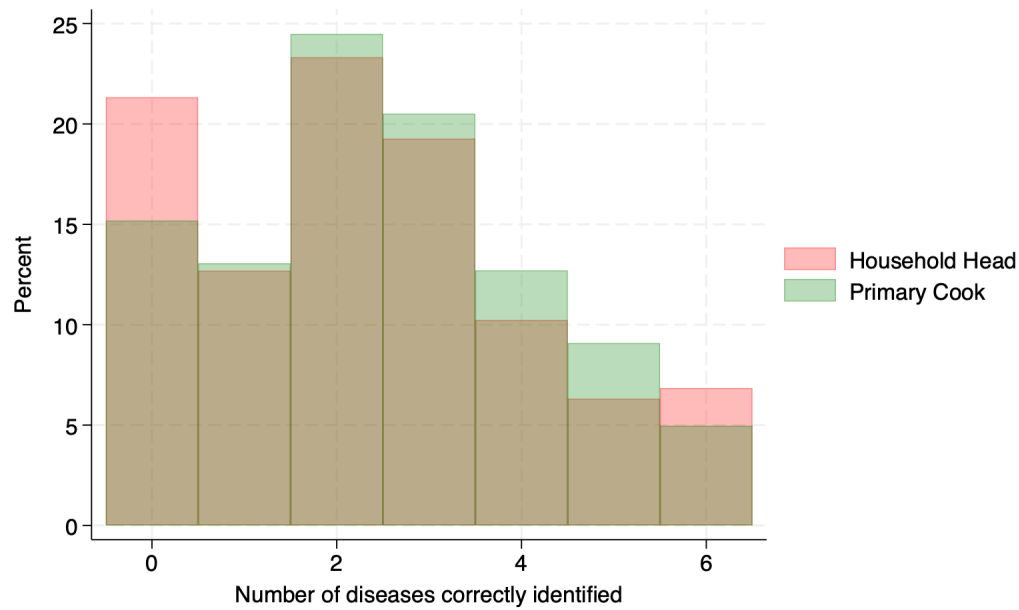


Figure C.1: Number of diseases out of 6 correctly identified by PC and HH

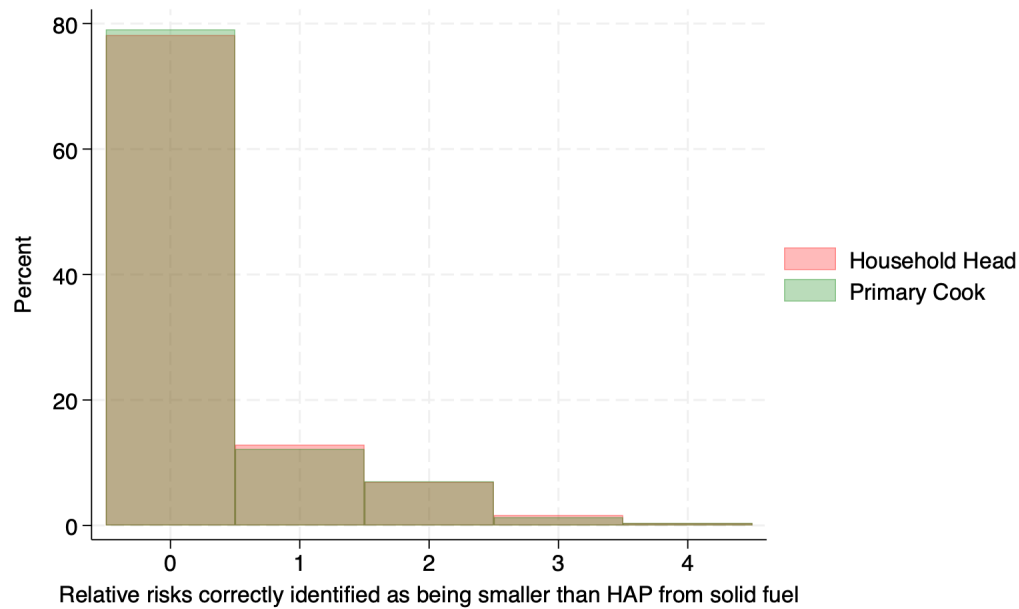


Figure C.2: Number of factors correctly identified as being less risky than solid fuels for mortality and morbidity

C.1 Information Script:

C.1.1 Treatment Group 1: Target Individual is the Household Head (This should be the same person who was interviewed in the baseline survey)

(1.) Hello! My name is ... and I am a researcher with Datakosh India, a development advisory firm based in New Delhi. We are talking to you today regarding a research project being conducted in collaboration with researchers from the University of Washington situated in the United States of America. We had talked to you about a month ago where you had agreed to be a part of this study. We had asked you about your beliefs about the use of different kinds of fuels for cooking. Today we want to give you some information about the health risks of using solid fuels like firewood, animal dung and crop residue for cooking. This will take about 15-20 minutes. Are you willing to hear this information?

(1) Yes

(0) No

Long term health risks for the primary cook

First, I want to give you some information about the long-term health risks of cooking with solid fuels like firewood for the primary cook of the household. By primary cook we mean individual in the household who spends the most time cooking.

According to the Global Burden of Disease report, 2021 prepared by the Institute for Health Metrics and Evaluation, which is one of the most reputed scientific studies about the causes of death and disability around the world, out of 100 deaths of females in Odisha in 2021, about 11 deaths could be attributed to household air pollution generated from the use of solid fuels.

According to this report household air pollution from solid fuels is a risk factor (risk factor is any behavior that increases the likelihood that a person will experience a negative health event) for females above age 15 in Odisha for the following diseases:

1. Stroke- which is a disease where the blood flow to the brain stops
2. Ischemic heart disease which commonly leads to a heart attack is a disease where the blood flow to heart is reduced
3. COPD or Chronic inflammatory lung diseases- this a major disease of the lung where there is obstruction of airflow in the lungs. Symptoms include breathing difficulty, cough, mucus (sputum) production and wheezing. Bronchitis which is inflammation of the breathing tubes in the lungs is often a part of COPD
4. Diabetes commonly known as ‘sugar’
5. Blindness and Vision Loss including cataract
6. Lung Cancer

Stroke, Heart Disease, COPD, and Lung Cancer are some of the leading causes of death and disability for women in India and using solid fuels for cooking is one of the biggest factors that contributes to these diseases.

For women in India, the use of solid fuels for cooking is a bigger risk factor for diseases and death than unsafe water and sanitation (use of unimproved toilet facilities like pit latrines without a slab, hanging or bucket latrines, and open defecation), chewing or smoking (including secondhand smoking) tobacco, high alcohol use, low physical activity, and very high or low temperatures (for example, a heat wave or a cold wave.)

All the above information shows that the use of solid fuels for cooking could reduce the lifespan of the primary cook as well as lead to major diseases which could even result in death or disability. This means that by using more clean fuels like LPG instead of firewood, cooking in well-ventilated areas instead of closed areas as well as adopting other methods to reduce exposure to smoke from burning solid fuels, it is possible that the primary cook will suffer from fewer serious and chronic diseases, will be healthier and will need to spend less money on medical expenses.

Short term health risks for Primary Cook

Now, I want to give you some information about the short-term impacts and the prevalence of different symptoms from using solid fuels for cooking. By short term impacts I mean that you will observe these symptoms during or a short time after using solid fuels for cooking or being in close proximity with the burning solid fuels. These symptoms are not only uncomfortable and bothersome for the person suffering them but could also be a precursor to or indicate a more serious disease. To collect this information, we conducted a survey in June and July with more than 2000 primary cooks in Koraput district in rural households that were similar to yours. We found that

1. Out of 100 primary cooks in households that primarily use solid fuels for cooking, around 73 primary cooks reported to have suffered from cough/wheezing in the one month before we conducted the survey. In households that primarily use LPG for cooking, out of 100 primary cooks, around 23 reported to have suffered from cough/wheezing in the one month before we conducted the survey
2. Out of 100 primary cooks in households that primarily use solid fuels for cooking, around 91 primary cooks reported to have suffered from runny eyes/sore eyes/redness in the one month before we conducted the survey. In households that primarily use LPG for cooking, out of 100 primary cooks, around 8 reported to have suffered from runny eyes/sore eyes/redness in the one month before we conducted the survey.
3. Out of 100 primary cooks in households that primarily use solid fuels for cooking, around 33

primary cooks reported to have suffered from shortness of breath/tightness in chest in the one month before we conducted the survey. In households that primarily use LPG for cooking, out of 100 primary cooks, around 4 reported to have suffered from shortness of breath/tightness in chest in the one month before we conducted the survey.

4. Out of 100 primary cooks in households that primarily use solid fuels for cooking, around 16 primary cooks reported to have suffered from skin rashes/itching/redness in the one month before we conducted the survey. In households that primarily use LPG for cooking, out of 100 primary cooks, around 9 reported to have suffered from skin rashes/itching/redness in the one month before we conducted the survey.

5. Out of 100 primary cooks in households that primarily use solid fuels for cooking, around 71 primary cooks reported to have suffered from headaches/dizziness in the one month before we conducted the survey. In households that primarily use LPG for cooking, out of 100 primary cooks, around 30 reported to have suffered from headaches/dizziness in the one month before we conducted the survey.

Long term health risks for children

Now, I want to give you some information about the long-term health risks of cooking with solid fuels like firewood, animal dung and crop residue for cooking for little children in the household.

The World Health Organization (WHO) which is the main organization responsible for promoting health and safety globally, says that children are particularly harmed due to the use of solid fuels for cooking. A report from them says that in 2020, the use of solid fuels for cooking had caused deaths of more than 2 lakh 30 thousand children under the age of 5 across the world. The Global Burden of Disease (GBD) report, 2021 also talks about the diseases and disabilities that specifically affect children and could be associated with the use of solid fuels. Out of 100 deaths of children aged under 5 in Odisha in 2021, around 18 deaths could be attributed to household air pollution

generated from the use of solid fuels.

According to the GBD report, household air pollution from solid fuels is a risk factor for children under age 5 in Odisha for the following diseases:

1. Lower respiratory Infections: These are infections that are caused in the lower respiratory tract (windpipe, lungs, and the tubes that carry air into the lungs). These could be often accompanied by pneumonia, high grade fever and cough.
2. Neonatal Disorders: These are diseases that are associated with newly born babies. These include preterm birth, low birth weight, brain dysfunction, jaundice, a life-threatening response to a blood infection known as sepsis.
3. Diarrhea.

Taken together, neonatal disorders, lower respiratory infection and diarrhea contributed to about 69% of all deaths of children under the age of 5 in Odisha in 2021.

For children under age 5 in India, the use of solid fuels for cooking is a bigger risk factor for diseases and death than unsafe water and sanitation as well as very high or low temperature.

The above information implies that the use of solid fuels for cooking could reduce the lifespan of a child as well as could lead to major diseases which could prevent the healthy growth and development of the child or even lead to death.

Short term health risks for Children under the age of 6

Now, I want to give you some information about the short-term impacts and the prevalence of different symptoms from using solid fuels for cooking. To collect this information, we conducted a

survey in June and July collecting information about more than 2000 children under the age of 6 in Koraput district in rural households that were similar to yours. We found that:

1. Out of 100 children under age 6 in households primarily using solid fuel, 40 children reported to have suffered from cough/wheezing in the one month before we conducted the survey.
2. Out of 100 children under age 6 in households primarily using solid fuel, 18 children reported to have suffered from runny eyes/sore eyes/redness in the one month before we conducted the survey.
3. Out of 100 children under age 6 in households primarily using solid fuel, 8 children reported to have suffered from shortness of breath/tightness in chest in the one month before we conducted the survey.
4. Out of 100 children under age 6 in households primarily using solid fuel, 10 children reported to have suffered from skin rashes/itching/redness in the one month before we conducted the survey.
5. Out of 100 children under age 6 in households primarily using solid fuel, 6 children reported to have suffered from headaches/dizziness in the one month before we conducted the survey.

Even though these are smaller than the symptoms for primary cooks, it does not mean that children are not at risk or at lower risk. Children under the age of 5 are often not able to communicate their discomfort and the true impact may be much higher than this. Therefore, effort should be made to make sure that children's exposure to smoke from cooking is as low as possible.

C.1.2 Treatment Group 2: Target Individual is the Household Head (This should be the same person who was interviewed in the baseline survey)

Same as sections [C.1.1](#) through [C.1.1](#) in treatment group 1.

C.1.3 Treatment Group 3: Target Individual is the Primary Cook (This should be the same person who was interviewed in the baseline survey)

Same as sections C.1.1 through C.1.1 in treatment group 1.

C.1.4 Answers to commonly asked questions

1. Why are you giving the information about Primary cooks or children to us (Household Heads) and not to them directly?

We are giving the information to you so that as the household head, you are aware about the risks that are posed to the primary cook by cooking with solid fuels. Since you are the household head and final decision maker in the household, we think that making you aware about the situation might lead to some changes in your cooking fuel choice.

2. Why are you not giving information about household heads/males to us (Household Heads)?

As stated above, primary cooks are particularly vulnerable to the health impacts from pollution that is generated from burning solid fuels since they will have the most exposure. However, anybody who is exposed to the smoke is at risk. At this time however, we do not have specific information for the household heads that we can give you.

3. How is diabetes(sugar) caused by using solid fuels?

Diabetes is a disease which occurs when the body is not able to regulate blood glucose levels. Insulin, a hormone released by the pancreas, is responsible for this regulation. When we breathe in air pollutant particles, it could lead to inflammation and this inflammation could lead to insulin not doing its job properly leading to risk of diabetes.

4. How is stroke and heart disease caused by using solid fuels?

Stroke occurs when blood flow to the brain reduces and a common cause for heart attacks is restricted blood flow from and to the heart. Breathing in air pollutant particles can lead to inflammation, damaging blood vessels and hardening arteries. This can increase blood pressure and make it difficult for blood to flow freely and put a strain on the heart and other parts of the cardiovascular system.

5. How is vision loss and blindness caused by using solid fuels?

Constant exposure of the eye to air pollutant particles could lead to cataract and also damage to other parts of the eye like the cornea and retina. They can also reduce blood flow to the eyes causing irreversible damage to vision.

6. Why are children more vulnerable than adults to smoke from solid fuels?

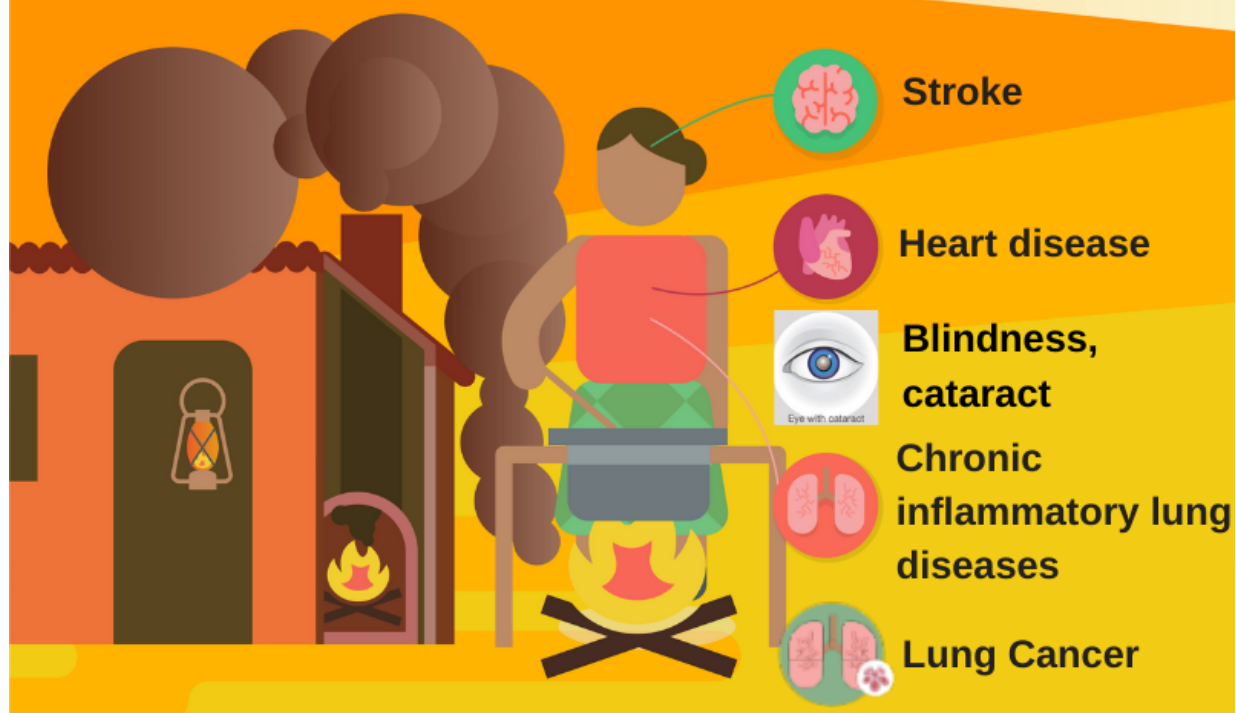
Children breathe faster and therefore take in more air and pollutants than adults. Children also have lower immunity levels than adults. Children do not have fully developed organs and therefore the stress from air pollution has an impact on the organs' development and healthy functioning of the body.

7. How can we reduce our exposure to air pollution to solid fuels?

The best way is to use clean sources of cooking like LPG or electric induction stoves. However, other methods of reducing exposure to solid fuels like using improved cookstoves that burn wood more efficiently and have a chimney in addition to cooking in very well ventilated rooms with a fan or exhaust fan could also be effective.

Household air pollution

Out of 100 deaths of females in Odisha in 2021, about 11 deaths could be attributed to household air pollution generated from the use of solid fuels.



Household air pollution

In 2020, the use of solid fuels for cooking had caused deaths of more than 2 lakh 30 thousand children under the age of 5 across the world



Table C.3 : Summary Statistics and Balance Test

Variable	Control	T1	T2	T3	Control-T1	Control-T2	Control-T3
HH size	5.462 (1.597)	5.190 (1.555)	5.416 (1.532)	5.350 (1.654)	-0.286** (0.119)	-0.043 (0.123)	-0.111 (0.143)
Head edu grade 8	0.378 (0.485)	0.410 (0.492)	0.388 (0.488)	0.352 (0.478)	0.031 (0.042)	0.009 (0.046)	-0.026 (0.038)
HH has toilet facility	0.120 (0.325)	0.142 (0.349)	0.146 (0.353)	0.130 (0.337)	0.017 (0.035)	0.027 (0.035)	0.008 (0.031)
HH has electricity	0.978 (0.147)	0.978 (0.147)	0.990 (0.100)	0.986 (0.118)	-0.000 (0.009)	0.012 (0.008)	0.007 (0.009)
LPG refill time	45.064 (41.738)	47.834 (42.705)	51.648 (48.249)	46.540 (42.498)	2.320 (7.001)	5.342 (6.298)	0.384 (6.694)
No. of LPG refills (12months)	1.354 (1.456)	1.598 (1.852)	1.280 (1.624)	1.188 (1.436)	0.214 (0.198)	-0.068 (0.175)	-0.145 (0.156)
PCs belief index (solid fuel)	0.027 (1.010)	-0.024 (1.026)	-0.033 (0.938)	0.030 (1.025)	-0.038 (0.069)	-0.056 (0.060)	0.020 (0.070)
HHs belief index (solid fuel)	0.050 (1.001)	-0.024 (0.966)	-0.022 (0.990)	-0.005 (1.042)	-0.057 (0.064)	-0.068 (0.063)	-0.036 (0.068)
Solid fuel use index	0.005 (0.991)	-0.088 (1.044)	0.063 (1.000)	0.019 (0.960)	-0.075 (0.089)	0.062 (0.095)	0.022 (0.087)
Clean fuel use index	0.039 (1.114)	0.047 (0.965)	-0.003 (1.023)	-0.083 (0.881)	-0.004 (0.093)	-0.037 (0.073)	-0.113 (0.074)
Avg time children in kitchen	0.351 (0.605)	0.518 (1.084)	0.392 (0.633)	0.391 (0.606)	0.166** (0.078)	0.038 (0.046)	0.036 (0.045)
No. of kids 6 or below	1.442 (0.596)	1.372 (0.542)	1.404 (0.549)	1.400 (0.545)	-0.071 (0.045)	-0.040 (0.041)	-0.041 (0.037)
Activities PC has a say	6.136 (2.319)	6.156 (2.393)	5.970 (2.460)	5.966 (2.250)	0.031 (0.288)	-0.162 (0.278)	-0.133 (0.218)
Annual per capita consumption	31,487.963 (16,783.361)	33,067.352 (16,165.121)	32,342.957 (19,864.928)	32,515.418 (19,873.986)	1,630.073 (1,937.234)	821.402 (1,910.587)	1,005.539 (1,862.506)
Observations	500	500	500	500	1,000	1,000	1,000

Standard errors are clustered at the village level and are shown in parentheses. Strata fixed effects are included.

Table C.4 : Belief Index Components

Belief Index
• Proportion (out of 100) of primary cooks and children under 6 who would have experienced each of the five symptoms in households similar to their household in their district that primarily use solid fuels in the last 30 days – sore/runny eyes or redness – difficulty breathing – dry cough/wheezing – headaches/dizziness – skin rashes/itching • Proportion of diseases (out of 7) correctly identified as being associated with solid fuel usage • Proportion of risk factors (out of 4) correctly identified as being smaller than solid fuel use for deaths of women in India • Proportion of risk factors (out of 3) correctly identified as being smaller than solid fuel use for deaths of children 5 and under in India • Indicator variable for whether lifespan of children 5 and under is adversely affected (=1 if they think it is) • Sum of indicator variables for whether they think that the smoke from cooking with solid fuels significantly affects primary cook's and children under 6's health

Table C.5 : Solid Fuel Use Index Components

Solid Fuel Use Index
• Self-reported solid fuel used (kg) from intervention to start of winter • Self-reported solid fuel used (kg) from start of winter to endline survey • Indicator variable for whether any cookstoves that use solid fuels (except improved' cookstoves (ICS)) have been used once per day or more since intervention to start of winter • Indicator variable for whether any cookstoves that use solid fuels (except ICS) have been used once per day or more since start of winter until just before the endline survey • Sum of the time (in minutes) that each solid fuel cookstove, except ICS, was used on a typical day since after intervention to start of winter • Sum of the time (in minutes) that each solid fuel cookstove, except ICS, was used on a typical day from the start of winter until just before the endline survey • Sum of time spent by all household members (in hours) in collecting/preparing the solid fuel since intervention to before endline • Self-reported amount spent by household (in Rs.) on purchasing solid fuel since intervention to before endline

Table C.6 : Clean Fuel Use Index Components

Clean Fuel Use Index
• Amount of LPG used (in kilograms per day) • Self-reported amount spent by household (in Rs.) on purchasing LPG since intervention to before endline for own consumption • Indicator variable for whether LPG, electric or improved cookstove was used since intervention • Indicator variable for whether LPG, electric or improved cookstove was used once per week or more since intervention to start of winter • Indicator variable for whether LPG, electric or improved cookstove was used once per week or more since start of winter to before endline • Sum of the time (in minutes) that LPG, electric and improved stove cookstoves was used on a typical day since intervention to start of winter • Sum of the time (in minutes) that LPG, electric and improved stove cookstoves was used on a typical day since since start of winter to before endline

Additional Tables

Table C.7 : Effect of Information Intervention on PC's Beliefs

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief Index (LT)	Disease Corr	Risk Corr(PC)	Risk Corr(child)	Lifespan(child)	Affect Health
PC informed (PC & children)	0.320*** (0.092)	0.075*** (0.021)	0.049** (0.023)	0.054** (0.024)	0.083* (0.044)	0.105** (0.050)
N	1994	1994	1994	1994	1994	1994
Control mean	0.00	0.41	0.22	0.26	0.36	0.32

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(6). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table C.8 : Effect of Information Intervention on PC's Beliefs (PC's Symptoms)

	(1) Belief Index (ST)	(2) Cough	(3) Sore eyes	(4) Breathlessness	(5) Skin Rashes	(6) Headaches
PC informed (PC & children)	0.091 (0.067)	0.013 (0.014)	0.003 (0.013)	0.007 (0.016)	0.021* (0.012)	0.012 (0.012)
N	1994	1994	1994	1994	1994	1994
Control mean	0.00	0.45	0.51	0.38	0.30	0.41

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(6). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table C.9 : Effect of Information Intervention on PC's Beliefs(Children's Symptoms)

	(1) Belief Index (ST)	(2) Cough	(3) Sore eyes	(4) Breathlessness	(5) Skin Rashes	(6) Headaches
PC informed (PC & children)	0.024 (0.071)	-0.007 (0.015)	-0.003 (0.015)	0.002 (0.016)	0.018 (0.012)	0.000 (0.013)
N	1994	1994	1994	1994	1994	1994
Control mean	0.00	0.40	0.43	0.35	0.28	0.31

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(6). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table C.10 : Effect of Information Intervention on HH's Beliefs

	(1) Belief Index (LT)	(2) Disease Corr	(3) Risk Corr(PC)	(4) Risk Corr(child)	(5) Lifespan(child)	(6) Affect Health
HH informed (PC & children)	0.134 (0.101)	0.050** (0.023)	0.043* (0.026)	0.026 (0.028)	0.013 (0.046)	0.041 (0.043)
HH informed (PC only)	0.141 (0.094)	0.049** (0.020)	0.019 (0.024)	-0.008 (0.028)	0.036 (0.046)	0.080** (0.040)
N	1995	1995	1995	1995	1995	1995
Control mean	-0.00	0.40	0.21	0.25	0.37	0.27
T1=T2	0.93	0.96	0.25	0.13	0.55	0.36

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(6). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. T1=T2 is the p-value from a t-test of equality of the T1 and T2 coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.11 : Effect of Information Intervention on HH's Beliefs (PC's Symptoms)

	(1) Belief Index (ST)	(2) Cough	(3) Sore eyes	(4) Breathlessness	(5) Skin Rashes	(6) Headaches
HH informed (PC & children)	-0.038 (0.064)	-0.022 (0.015)	0.000 (0.012)	-0.018 (0.015)	-0.002 (0.013)	0.004 (0.014)
HH informed (PC only)	-0.067 (0.081)	-0.019 (0.016)	-0.013 (0.014)	-0.025 (0.015)	-0.005 (0.016)	-0.007 (0.017)
N	1995	1995	1995	1995	1995	1995
Control mean	0.00	0.41	0.46	0.36	0.28	0.38
T1=T2	0.71	0.85	0.31	0.64	0.83	0.57

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(6). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. T1=T2 is the p-value from a t-test of equality of the T1 and T2 coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.12 : Effect of Information Intervention on HH's Beliefs(Children's Symptoms)

	(1) Belief Index (ST)	(2) Cough	(3) Sore eyes	(4) Breathlessness	(5) Skin Rashes	(6) Headaches
HH informed (PC & children)	-0.036 (0.077)	-0.019 (0.016)	-0.007 (0.013)	-0.013 (0.015)	-0.001 (0.013)	-0.006 (0.018)
HH informed (PC only)	-0.081 (0.078)	-0.022 (0.014)	-0.013 (0.013)	-0.022 (0.014)	-0.007 (0.015)	-0.017 (0.018)
N	1995	1995	1995	1995	1995	1995
Control mean	0.00	0.37	0.40	0.32	0.27	0.30
T1=T2	0.58	0.84	0.68	0.55	0.66	0.58

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(6). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. T1=T2 is the p-value from a t-test of equality of the T1 and T2 coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.13 : Effect of Information Intervention on Solid Fuel Use

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Solid Fuel Index	Collection time (hrs)	Expenditure (Rs.)	Qty (kg) (Au)	Used daily (Au)	Mins used daily (Au)	Qty (kg) (Wi)	Used daily (Wi)	Mins used daily (Wi)
HH informed (PC & children)	0.024 (0.076)	2.579 (5.001)	-5.869 (19.713)	13.044 (21.254)	-0.007 (0.023)	3.780 (4.322)	1.976 (3.485)	-0.016 (0.019)	1.423 (4.716)
HH informed (PC only)	0.081 (0.060)	4.142 (4.374)	6.583 (22.117)	22.379 (19.923)	0.000 (0.023)	4.878 (4.151)	-0.682 (3.174)	-0.020 (0.020)	0.724 (4.598)
PC informed (PC & children)	-0.017 (0.093)	-2.037 (4.856)	31.998 (37.601)	-18.396 (22.730)	-0.033 (0.030)	-5.807 (5.008)	-5.361 (3.799)	-0.032 (0.023)	-7.431 (5.874)
N	1996	1996	1996	1996	1996	1996	1996	1996	1996
Control mean	-0.00	74.85	60.62	750.35	0.92	156.97	127.57	0.95	185.48
T1=T2	0.42	0.75	0.53	0.69	0.77	0.78	0.52	0.84	0.88
T1=T3	0.68	0.37	0.28	0.21	0.38	0.04	0.10	0.51	0.13
T2=T3	0.26	0.17	0.49	0.09	0.26	0.02	0.28	0.64	0.15

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(9). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. Ti=Tj is the p-value from a t-test of equality of the Ti and Tj coefficients. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table C.14 : Effect of Information Intervention on Clean Fuel Use

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Clean Fuel Index	Used since intervention	LPG per day (kg)	Expenditure (Rs.)	Mins used daily (Au)	Used weekly (Au)	Mins used daily (Wi)	Used weekly (Wi)
HH informed (PC & children)	-0.075 (0.068)	-0.005 (0.035)	-0.010 (0.007)	-65.668 (64.277)	0.455 (3.732)	-0.006 (0.033)	-2.892 (3.541)	-0.030 (0.033)
HH informed (PC only)	-0.085 (0.079)	-0.017 (0.037)	-0.011 (0.008)	-84.256 (70.896)	-0.754 (4.662)	-0.015 (0.035)	0.179 (4.675)	-0.018 (0.035)
PC informed (PC & children)	0.068 (0.074)	0.020 (0.035)	0.005 (0.008)	21.326 (69.511)	3.932 (4.273)	0.007 (0.033)	7.109 (4.380)	0.037 (0.033)
N	1996	1996	1996	1996	1996	1996	1996	1996
Control mean	0.00	0.56	0.09	769.74	52.35	0.54	41.31	0.39
T1=T2	0.89	0.71	0.85	0.78	0.78	0.77	0.52	0.73
T1=T3	0.03	0.40	0.05	0.17	0.38	0.64	0.02	0.04
T2=T3	0.05	0.25	0.06	0.13	0.34	0.46	0.19	0.11

Standard errors clustered at the village level in parentheses. Column (1) is an index of the components of columns (2)-(8). Indices are standardized with mean 0 in the control group. Strata fixed effects, baseline value of outcome and PDS-LASSO controls included. Ti=Tj is the p-value from a t-test of equality of the Ti and Tj coefficients. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table C.15 : Heterogeneity by HH's baseline belief index

	(1) Belief Index (HH)
HH informed	0.113 (0.113)
PC informed	0.070 (0.124)
Below-median HH belief (baseline)	-0.372*** (0.099)
HH informed x Below-median	-0.067 (0.124)
PC informed x Below-median	0.036 (0.143)
<i>N</i>	1995

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Table C.16 : Heterogeneity by PC's baseline belief index

	(1) Belief Index (PC)
HH informed	0.072 (0.103)
PC informed	0.221* (0.117)
Below-median PC belief (baseline)	-0.355*** (0.118)
HH informed x Below-median	-0.071 (0.134)
PC informed x Below-median	0.007 (0.156)
<i>N</i>	1994

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Table C.17 : Heterogeneity by PC's education

	(1) Belief Index (HH)	(2) Belief Index (PC)
HH informed	0.080 (0.098)	0.029 (0.120)
PC informed	0.105 (0.112)	0.158 (0.116)
PC Grade 6 or above	0.021 (0.097)	0.114 (0.116)
HH informed x Grade 6 or above	-0.025 (0.121)	0.012 (0.135)
PC informed x Grade 6 or above	-0.072 (0.126)	0.190 (0.139)
<i>N</i>	1995	1994

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Table C.18 : Heterogeneity by PC's education

	(1)	(2)
	Solid Fuel Index	Clean Fuel Index
HH informed	0.101* (0.059)	0.004 (0.098)
PC informed	0.113 (0.076)	0.008 (0.106)
PC Grade 6 or above	0.081 (0.080)	0.455*** (0.081)
HH informed x Grade 6 or above	-0.132 (0.111)	-0.071 (0.103)
PC informed x Grade 6 or above	-0.301** (0.135)	0.053 (0.115)
<i>N</i>	1996	1996

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Table C.19 : Heterogeneity by Child's Gender

	(1)	(2)
	Belief Index (HH)	Belief Index (PC)
HH informed	0.068 (0.119)	-0.111 (0.106)
PC informed	0.041 (0.127)	0.015 (0.107)
Atleast one child is male	0.022 (0.101)	-0.191** (0.084)
HH informed x Male child	0.001 (0.123)	0.222** (0.105)
PC informed x Male child	0.054 (0.129)	0.330*** (0.121)
<i>N</i>	1995	1994

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Table C.20 : Heterogeneity by Child's Gender

	(1) Solid Fuel Index	(2) Clean Fuel Index
HH informed	0.008 (0.097)	-0.050 (0.107)
PC informed	-0.034 (0.116)	-0.012 (0.140)
Atleast one child is male	0.029 (0.111)	-0.099 (0.078)
HH informed x Male child	0.062 (0.122)	0.026 (0.095)
PC informed x Male child	0.047 (0.133)	0.038 (0.125)
<i>N</i>	1996	1996

Standard errors clustered at the village level in parentheses. Strata fixed effects included.

Administrative information

Funding: This work was supported by the Weiss Fund for Research in Development Economics at the University of Chicago (grant number SUB000292).

Institutional Review Board (ethics approval): IRB approval was obtained from University of Washington with ID STUDY00019649.

Declaration of interest: We have no competing interests to declare.

Acknowledgments: We thank Rachel Heath, Alan Griffith, Emma Riley, Isabelle Cohen, Chris Anderson, Victor Menaldo and participants at the Labor-Development Brownbag at University of Washington for helpful comments. We are grateful to Sanjay Prasad, Manas Kumar and their team at DATAKOSH for support in the field.