Research Institute for Environmental Economics and Management

Discussion Paper Series No.2003

Subjective risk belief function in the field:

Evidence from cooking fuel choices and health in India

Hide-Fumi Yokoo, Toshi H. Arimura,

Mriduchhanda Chattopadhyay & Hajime Katayama

August 2020

WASEDA

Subjective risk belief function in the field: Evidence from cooking fuel choices and health in India

Hide-Fumi Yokoo,[1*] Toshi H. Arimura,[2] Mriduchhanda Chattopadhyay,[2] and Hajime Katayama[2]

1. *Hitotsubashi University* 2. *Waseda University*

August 2020

Abstract

We investigate the accuracy of the perceptions of health risks in India. To examine systematic risk misperception, which is relevant to policy debates, we present the concept of the subjective risk belief function (SRBF). The context of our study is the risk of developing physical symptoms related to household air pollution caused by cooking. Using field data collected from 588 respondents in 17 villages in West Bengal, we regress the probability of symptoms conditional on fuel choices to estimate the respondent-specific health risk changes. Then, we elicit the subjective probabilistic beliefs using an interactive method with visual aids. Considering the estimated risks as objective risks, we estimate the linear SRBF. Our estimated coefficient of the average SRBF is in the range of 0.58 to 0.79, which implies a slight underestimation of the change in risk when switching from cooking with firewood to cooking with liquefied petroleum gas, although the respondents have a qualitatively accurate belief. We further find that risk misperception is correlated with religion but not with age or education.

Keywords: Belief, Cooking fuel choice, Health risk, India, Risk misperception, Subjective probabilistic expectation
JEL classification: D83, D84, I12, O13, Q53

We thank Takeshi Aida, Masayuki Yagasaki, and participants in seminars at GRIPS, Hitotsubashi University, Keio University, Kobe University, Osaka University, and the University of Tsukuba, as well as conference participants at EfD 2018, JEA (Spring) 2018 and WCERE 2018 for helpful comments. We thank Global Climate Change, Jadavpur University for their infrastructural assistance with our field survey. Special thanks to Mari Sakudo for comments in the early stages of this project. Toshi Arimura and Hajime Katayama appreciate the financial support of JSPS KAKENHI Grant Number 16K13364. This research is also supported by the Research Institute for Environmental Economics and Management, Waseda University. There are no conflicts of interest to declare.
* Corresponding author. E-mail: hidefumi.yokoo@r.hit-u.ac.jp

1. Introduction

Information campaigns are frequently employed in developing countries to promote health risk avoidance behavior (Dupas, 2011a; Pattanayak and Pfaf, 2009). One approach adopted in these campaigns is to provide information on the likelihood of developing a disease conditional on a certain behavior. For example, Dupas (2011b) examines the impact of information about the prevalence rates of HIV, disaggregated by age group, on the sexual behavior of teenage girls in Kenya. The result shows evidence of changing sexual partners from older (riskier) to younger (less risky) men.[1] While Dupas finds a positive impact of risk information on avoidance behavior, its effectiveness might depend on the accuracy of ex ante beliefs about the probabilities of possible states (Godlonton and Thornton, 2013).

Investigating whether systematic risk misperceptions exist is not only useful in formulating efficient public policies but also crucial for understanding the nature of risk attitudes. Preferences and subjective beliefs are considered two potential sources of variation in attitudes toward risk. For example, Savage (1954) extends the expected utility (EU) theory to allow decision-makers to maximize EU based on their preferences and the subjective probabilities of different states. More recently, several non-EU models of risk preferences have been proposed.[2] Among these, both rank-dependent EU theory and cumulative prospect theory use a two-step framework of preferences and beliefs to understand decision making (Barberis, 2013; Fox and Tversky, 1998). A number of empirical studies have estimated risk preferences from field data by using one or more of these models and assuming that subjective beliefs correspond to objective probabilities (for a review, Barseghyan et al., 2018). However, if this assumption does not hold, a basic identification problem will occur because many preference and belief combinations can lead to the same choice (Manski, 2004), meaning that a quantitative study on the accuracy of risk perception is required.[3]

---

[1] Another study that examines behavioral responses to information regarding relative risk is Godlonton et al. (2016), who examine asymmetric behavioral responses to information on the results of experimental studies reporting that male circumcision is partially protective against the risk of HIV transmission. Another strand of the literature conducts policy simulations to investigate how information campaigns change behaviors (for example, Delavande and Kohler, 2016; Viscusi, 1990).

[2] For example, rank-dependent EU theory developed by Kahneman and Tversky (1979) and Quiggin (1982), cumulative prospect theory developed by Tversky and Kahneman (1992), and the model of reference-dependent preferences developed by Köszegi and Rabin (2006).

[3] For example, Barseghyan et al. (2013b) find evidence of a probability distortion characterized by the substantial overweighting of small risks and only mild insensitivity to risk changes. Furthermore, they argue that neither Gul's (1991) model of disappointment aversion nor Kőszegi and Rabin's (2006) model alone can explain probability distortions. However, they cannot determine whether probability distortions occur because individuals engage in probability weighting or whether they misperceive risks at the beginning.

There is longstanding literature on risk misperception in many areas, such as smoking, food, terrorism, healthcare, and air pollution.[4] For example, Breyer (1993) finds that experts believe that hazardous waste sites pose *medium-to-low* risks to the public, while household air pollution poses a *high* risk, although public perceptions have driven policies to focus on hazardous waste sites rather than air quality within houses.[5] Note that earlier studies—including Breyer (1993)—elicit subjective *non-probabilistic* beliefs using ordered categories such as the Likert scale. Several recent studies differ from earlier contributions by eliciting subjective *probabilistic* beliefs.[6] A seminal work by Viscusi (1990), for example, examines whether smokers underestimate the risk of lung cancer by eliciting subjective *probabilistic* beliefs. Using a national telephone survey in the US, he finds that the average value of subjective beliefs on the risk to smokers is approximately 0.4, while the *true* value is estimated as ranging from 0.05 to 0.1, suggesting a high overestimation on average.[7]

Several other studies further compare elicited beliefs and estimated risks at the individual level. Oster et al. (2013) examine the beliefs of US citizens on their probability of having Huntington disease and compare them with an evaluation performed by a doctor based on the results of clinical tests for each individual enrolled in the study. Carman and Kooreman (2014) elicit the beliefs of Dutch citizens on their probability of having influenza, heart disease and breast cancer with and without preventive care. They also compare beliefs with individual-specific risk levels calculated by using epidemiological models. Khwaja et al. (2007) assess the accuracy of subjective probabilistic beliefs about the 10-year mortality hazard collected in the Health and Retirement Study in the US by comparing the survey results with econometrically estimated hazards for individuals in the same sample. Relatedly, Khwaja et al. (2009) compare smokers' subjective beliefs on future survival with corresponding individual-specific probabilities estimated from regression analyses.

This paper extends the above line of research to the developing world, where imperfect information regarding health risk is more pronounced. There are at least two challenges in quantifying individual-specific misperception in regions such as rural India: the elicitation of subjective beliefs of

---

[4] Wright and Ayton (1994) provide a comprehensive review of early studies on subjective risk belief.

[5] Another example that studies subjective non-probabilistic beliefs is Lange (2011), who explores the role of education in cancer screening behavior.

[6] To the best of our knowledge, Viscusi and O'Connor's (1984) study is the first that elicits a continuous risk belief measure and create a probabilistic variable.

[7] In Viscusi (1990), subjective probabilistic beliefs are elicited by using the question, "Among 100 cigarette smokers, how many of them do you think will get lung cancer because they smoke?" The individual's response to this question is divided by 100 to obtain the lung cancer belief. The average is 0.426 for a sample of 3,119 respondents. The estimate for *true* probability has been calculated using information from US surgeon general's reports. Lundborg and Lindgren (2002) and Schoenbaum (1997) also compare subjective probabilistic beliefs with epidemiological predictions. Note that these three studies use estimates that are obtained outside the study as the true probabilities.

individuals with low education and the estimation of objective probabilities without high-skilled doctors. To address the first problem, we adopt an interactive method developed by Delavande et al. (2011b) and Delavande (2014). This method utilizes visual aids to elicit subjective probability since simply asking the percentage chance of the occurrence of an event is too abstract and complex for some respondents, especially in the developing world.[8] To overcome the second problem, we use econometric analyses following Khwaja et al. (2007) and Khwaja et al. (2009).[9] We conduct regression analyses and calculate the respondent-specific risk changes using estimated coefficients. To address potential endogeneity concerns, we adopt an instrumental variable (IV) method. To create an instrument, we exploit the practice of arranged marriage in India. Since we rely on recall data, we further address possible mismeasurement (measurement error) by using a parametric method.

To compare subjective beliefs and objective probabilities, we adopt a concept that we call the "subjective risk belief function (SRBF)," originally proposed by Johansson-Stenman (2008). The SRBF represents subjective risk belief as a function of objective risk. Note that the SRBF becomes flatter (steeper) than a slope of one if the individual overestimates small (large) risks and underestimates large (small) risks. According to Johansson-Stenman (2008), the degree of bias in risk belief on the change, i.e., the coefficient of the SRBF, is crucial to designing efficient information provision policies. By considering our econometrically estimated risks as *true* objective risks, our research framework enables us to estimate the SRBF, which has been considered theoretically (Barseghyan et al., 2013a; Johansson-Stenman, 2008).[10]

The specific context of risk examined in this paper is the risk of physical symptoms potentially related to household air pollution caused by cooking with solid fuels (for a review of this topic, see Jeuland, Pattanayak, and Bluffstone, 2015). In various developing countries, household air pollution from primitive household cooking fires is considered the leading environmental cause of death (Hanna et al., 2016). Estimates of the burden in India alone show that approximately 1.04 million premature deaths and 31.4 million disability-adjusted life years are attributable to household air pollution (Balakrishnan et al., 2014). At our research site, it is possible to switch to cleaner cooking fuel, such as liquefied petroleum gas (LPG) by paying additional fixed and variable costs. However, a substantial

8 Okeke et al. (2013) elicit subjective probabilistic beliefs on cervical cancer risk in Nigeria but without using visual aids.

9 Similarly, Brown et al. (2017) study the effect of the beliefs regarding water safety on avoidance behavior in Cambodia and utilize the method of Delavande et al. (2011b). They elicit subjective beliefs and create a dummy variable for being "optimistic," using the sample mean as the reference point. In contrast to Brown et al. (2017), the present study estimates the objective risk for each respondent using econometric analyses and compares it with beliefs to incorporate possible variation in objective risks.

10 One exception is Khwaja et al. (2007), who present a local linear smooth plot of subjective and objective mortality hazard that corresponds to the SRBF.

proportion of households continue to use dirty fuel (for example, firewood). One possible reason for this choice is their underestimation of the change in health risk; hence, we quantitatively examine whether misperceptions of the change in risk exist to predict the effectiveness of policies such as information campaigns.

From our elicitation of subjective risk beliefs, we find that all of our 588 respondents believe that the risk of having symptoms when using solid fuel is higher than the risk when using LPG. This suggests that the entire sample qualitatively correctly believes that there are health risks of using solid fuel. However, our estimate of the coefficient of the SRBF is 0.7, which is statistically significantly smaller than one. The results also show that the SRBF estimates decrease when we econometrically incorporate the potential for mismeasurement of episodes of symptoms. We further add characteristic variables (age, religion, and years of education) and their interaction terms with estimated objective risk to the SRBF. The estimation results show that Muslim respondents are more likely to underestimate the risk levels and change, implying an association between religious faith and risk beliefs. In summary, this paper shows that it is possible to quantitatively examine the accuracy of belief on health risk, even in the less developed world. Furthermore, our research framework allows us to empirically examine the source of biased beliefs.

In Section 2, we present the conceptual framework of SRBF, which provides useful information for policymaking from the parameter of the coefficient. In Section 3, we describe our data. We conducted household surveys in 17 villages in West Bengal to create a dataset on cooking fuel choices, physical symptoms, and subjective beliefs related to fuel use and health status. In Section 4, we econometrically estimate the health function and calculate two respondent-specific probabilities of experiencing symptoms. In Section 5, we present the results of the elicitation and calculation of the two subjective probabilistic beliefs for each respondent. In Section 6, we compare the subjective beliefs and the objective probabilities and estimate the SRBF. Section 7 discusses the limitations of the study and Section 8 concludes the paper.

2. Conceptual framework

Consider a health risk $r_i \in [0,1]$ of individual i associated with a certain action taken by an individual. Health risk can be, for example, the probability that an individual will have a symptom of a respiratory infection. Let s_i be the subjective belief of individual i regarding this probability. Johansson-Stenman (2008) assumes this subjective belief to be a function of the objective risk, that is, $s_i = \psi(r_i)$.[11] We

[11] Barseghyan et al. (2013a) also propose a utility function that includes $\psi(r_i)$. They develop a strategy to distinguish the model of rank-dependent probability weighting from systematic risk misperceptions in field data

refer to this function, ψ, as the SRBF.

Johansson-Stenman (2008) extends EU theory and presents a model in which risk misperceptions are allowed. The model provides an important implication regarding the effectiveness of an information provision policy. Specifically, misperception of risk levels ($s_i - r_i > 0$ or $r_i - s_i > 0$) is not a necessary condition for information provision to be effective; what matters is whether there is misperception regarding the change from a risky choice to a relatively safe alternative.

Importantly, the steepness of the SRBF can succinctly express a signal of misperception in risk changes that relates to policymaking. Assume a linear SRBF to simplify the analyses:

$$s_i = \psi(r_i) = \rho_0 + \rho_1 r_i.$$

If an individual correctly perceives the change in risk, then $\frac{\partial\psi}{\partial r} = 1$. Instead, if the SRBF is flat (steep) such that $\frac{\partial\psi}{\partial r} < 1 (> 1)$, then the change in risk is underestimated (overestimated); therefore, information provision may improve the efficiency of choice. Johansson-Stenman (2008) regards $\frac{\partial\psi}{\partial r}$ as an important parameter when examining the optimal information provision in the second-best world where taxing risky goods is not allowed.[12]

3. Data

3.1. Background

Approximately 60% of the world's population currently uses either gas or electricity for cooking, while the remaining 40% uses solid fuels. Solid fuels include coal, charcoal, animal dung, agricultural residue, and firewood. Burning such solid fuels for cooking produces carbon monoxide, PM2.5, and other toxic chemicals. Many epidemiological studies provide evidence linking cooking-related household air pollution with various diseases (reviews include Smith and Pillarisetti, 2017). Such diseases include acute lower respiratory infections (ALRIs), lung cancer, cataracts, and chronic obstructive pulmonary disease (COPD). Smith et al. (2014) estimate that household air pollution caused 3.9 million premature deaths worldwide in 2010 and as much as a 4.8% reduction in disability-adjusted life years. Among those who cook using solid fuels, one-fourth live in India.

There are several ways to reduce health risks related to household air pollution. A traditional

without directly measuring subjective beliefs.

[12] Johansson-Stenman (2008) models information provision as a costly public policy that reduces the discrepancy between subjective beliefs and objective risks both in levels and changes. Another notable feature of the model is the fear (or mental suffering) associated with risk belief, which is directly included in the utility function. Despite its potential importance in modeling utility, fear is excluded from the present study since it is unrelated to the existence (or nonexistence) of a systematic bias in risk beliefs.

approach is to improve the cooking stoves used to burn solid fuels. For example, a chimney could be attached to stoves to let the polluted air out of the room. However, several studies suggest that improving stoves may not reduce health risks (Anenberg et al., 2013; Hanna et al., 2016). For this reason, gas and electricity, which are cleaner, are widely promoted (Smith and Pillarisetti, 2017). Since 2015, the Indian government (along with the world's three largest oil companies) has been phasing in several measures to promote LPG, such as the provision of subsidies and the free distribution of LPG gas stoves (Gould and Urpelainen, 2018). Nevertheless, the use of LPG remains limited.[13]

3.2. Sample construction

We use a dataset collected for concurrent work by the authors (Chattopadhyay et al., 2020).[14] In addition to the dataset used in the concurrent work, we collected subjective risk belief data for the present paper. We selected Dhapdhapi-II gram panchayat (GP)[15] in the state of West Bengal as our research site since the use of dirty cooking fuels is prevalent there (see Chattopadhyay et al., 2020). There are seventeen villages in this GP, and the majority of residents still use traditional solid fuels for cooking, although LPG distribution networks are already established. Switching from solid fuels to LPG incurs both fixed and variable costs. At the time of our field research, the average cost of switching, including the cost of purchasing an LPG stove, was approximately 5,000 Indian rupees (INR), which was approximately 75% of the average monthly income of our sample households. In addition, households have to purchase LPG cylinders distributed by traders.[16] These costs may make it less attractive to switch from solid fuels to LPG since households can collect their own firewood.[17]

Following a preliminary survey, we conducted two rounds of field surveys. The first round was conducted from December 2016 to January 2017. We used a stratified random sampling method to choose 600 household heads among the 13,024 adults listed on the voter list of Dhapdhapi-II GP,

[13] According to Gould and Urpelainen (2018), approximately 13% of the respondents to a survey administered to 8,568 households in 714 Indian villages between 2015 and 2017 reported that they used LPG as their primary cooking fuel. However, a majority of them used both solid fuels and LPG. Less than 4% of the respondents used LPG only.

[14] Chattopadhyay et al. (2020) examine the impact of subjective risk beliefs on fuel choice and health. They find that beliefs of becoming sick from dirty fuel usage reduce the fraction of days with dirty fuel usage that degrades the health of the respondent.

[15] A GP is village-level unit of self-government in India. A typical GP comprises several villages.

[16] Usually, one LPG cylinder (14.2 kg) is considered a unit of LPG consumed for domestic purposes. In 2016-2017 in this region, the average price of one LPG cylinder was 640 INR, and the subsidized cost was 420 INR. Kar et al. (2019) provide more detailed information on the LPG market in India. Gupta and Köhlin (2006) and Gould and Urpelainen (2018) present a more comprehensive explanation of cooking fuel markets in India.

[17] It is also possible to buy firewood at a market. In our research site, approximately 40 kg of the firewood costs 200-300 INR. From our preliminary survey, we found that 36% of the firewood users at our research site bought firewood at the market.

which was published online.[18] In the first round, our enumerators visited the selected 600 households, and 596 participated (four declined to participate). Of the variables collected in the first survey, this paper uses the individual and household characteristic variables. The second round was conducted from December 2017 to January 2018. Our enumerators visited the same 596 households that had participated in the first round and obtained responses from 588 (a further eight households declined to participate in the second round). This paper uses data on subjective beliefs and the self-reported experience of symptoms as well as fuel usage from these 588 households. We define our respondent as the primary cook in a household.

3.3. Definition of the variables

Smith and Pillarisetti (2017) discuss lung function, eye opacity, blood pressure, and electrocardiogram ST-segment as biomarkers of the effects of household air pollution. Thus, in this current study, we would ideally collect data on these biomarkers for each respondent to evaluate objective health risk. For example, Hanna et al. (2016) conducted spirometry tests with approximately 2,500 subjects to evaluate the impact of improved cooking stoves on lung function. However, it is costly to conduct such clinical tests at all the visits with the cooperation of, for example, doctors. Hanna et al. (2016) further conducted recall surveys on physical symptoms to complement the results of the spirometry tests. From the questionnaire used in their study, we selected ten physical symptoms and conducted a preliminary survey to examine prevalence at our research site. From the results of our survey, we defined the three most frequently observed symptoms as signals of diseases potentially caused by household air pollution, namely, *dry cough*, *sore or runny eyes*, and *difficulty breathing*. In the second round of our survey, we asked, "Did you experience Y in the last 30 days?" for each symptom "Y." We then created an indicator variable to denote the self-reported experience of symptoms ($Symp_i$), which took value 1 if the respondent had experienced at least one of the three symptoms in the past 30 days and 0 otherwise.

Next, we surveyed each household's cooking fuel usage patterns. As noted above, both firewood and LPG, as well as other cooking fuels, are used in this area. Furthermore, some households use different cooking fuels within a month or even within a day. Such fuel stacking is well known in the literature (Gould and Urpelainen, 2018; Kar et al., 2019). We asked respondents, "In the last 30 days before the last month, how many days did you use X for cooking?" For fuel "X," we asked about seven types of fuels: electricity, LPG, kerosene, coal/charcoal, solid fuels such as cow dung cakes/straw,

[18] A part (a stratification unit within each electoral constituency) was our stratification unit. As the size of the population in each part was not uniform, we sampled proportionally according to the population size of each part.

firewood, and others . While most of the sample households used either firewood or LPG, we considered two categories of fuels to include the minor options and to simplify the questions on risk belief. Following Gupta and Köhlin (2006) and Heltberg (2005), we define the sum of the days of LPG, kerosene, and electricity usage as the number of days of *clean* fuel usage and the sum of the days of coal/charcoal, solid fuels, firewood, and "others" usage as the number of days of *dirty* fuel usage. By dividing the number of days of dirty fuel usage by 30, we create a variable of the fraction of days of dirty fuel usage by household *i*, denoted as $Dirty_i \in [0,1]$, as our variable of interest. The English versions of the questionnaires used in our surveys are shown in the Online Appendix.

In Section 4, we estimate an objective probability that respondent *i* will have one of the three symptoms in the next month if he or she uses dirty fuel for all 30 days in this month:

$$r_i(Dirty_i = 1) = \Pr_i(Symp_i \ = 1|Dirty_i = 1).$$

We also elicit respondent i's subjective belief about this probability:

$$s_i(Dirty_i = 1) = \psi\big(\Pr_i(Symp_i \ = 1|Dirty_i = 1)\big).$$

In Section 5, we report the methods and results of elicitation of subjective risk beliefs. A comparison of the two enables us to identify misperceptions regarding the risk level of dirty fuel. To identify misperception in the risk level of clean fuels and the change in risk, we estimate an objective probability that respondent *i* will have one of the three symptoms in the next month if the respondent uses clean fuel for all 30 days in this month:

$$r_i(Dirty_i = 0) = \Pr_i(Symp_i \ = 1|Dirty_i = 0).$$

and elicit a subjective belief about it:

$$s_i(Dirty_i = 0) = \psi\big(\Pr_i(Symp_i \ = 1|Dirty_i = 0)\big).$$

3.4. Summary statistics

Table 1, Panel A reports the summary statistics of the variables. Seventy-six percent of the respondents reported that they had experienced at least one of the three symptoms in the last month. Online Appendix Figure A1 shows the distribution of the fraction of days of dirty fuel usage ($Dirty_i$), which shows two pileups at values 0 and 1. Approximately half of the respondents (45.2%) indicated that they used only dirty fuel for all 30 days before the last month, while 13.1% used clean fuel only.[19] Other respondents used both clean and dirty cooking fuel within the same month. As a result, the mean of $Dirty_i$ is 0.68. Almost all of our respondents are women; there is only one household in which a man is the primary cook.

[19] Among those who used dirty fuels for all 30 days, 82.1% used only firewood, while the remainder also used other solid fuels, such as cow dung cakes. Among those who used clean fuels for all 30 days, 96.1% used only LPG.

The average age of our respondents is 38.5 years, while the average number of years of education is 4.7. In our sample, 69.4% of households follow the Hindu religion, while the others follow the Muslim religion. To control for the long-term and/or cumulative impact of household air pollution, we create the variable "Cumulative years of clean fuel usage until the first round." For this variable, 78.6% of our sample has a zero value, since they used only dirty fuel until the first round. The remaining 21.4% of the sample has a value larger than zero and the average of those who have a value larger than zero is 6.8 years.

4. Estimation of objective risks

4.1. Probit model

To quantitatively identify risk misperception in the field, we first estimate the respondent-specific risk of symptoms using data collected in the survey. As a benchmark, we consider the probit model by assuming that $Dirty_i$ is exogenous:

$$E[Symp_i = 1 | Dirty_i, \boldsymbol{X}_i] = \Phi(\beta_0 + \beta_1 Dirty_i + \boldsymbol{X}_i{}'\delta_1), \tag{1}$$

where $\boldsymbol{X}_i$ is the vector of individual and household characteristics and $\Phi(\cdot)$ is the cumulative distribution function (CDF) for the standard normal distribution. Those characteristics include age of the respondent, household size, years of education of the respondent, an indicator of whether the respondent's household follows the Hindu religion, monthly household income, an indicator of whether the respondent is a housewife, the number of cooks in the household, an indicator of whether the kitchen is located outside the dwelling space, an indicator of whether the household owns a personal computer, and cumulative years of clean fuel usage until the first round.

Using the results of the regression, the objective probability that respondent i will have one of three symptoms in the next month if the respondent uses dirty fuel for all 30 days in this month is calculated as:

$$r_i(Dirty_i = 1) = \Phi\left(\hat{\beta}_0 + \hat{\beta}_1 + \boldsymbol{X}_i{}'\hat{\delta}_1\right),$$

where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\delta}$ are the estimators of the probit model. Similarly, the probability if the respondent uses clean fuel for all 30 days in the month is calculated as:

$$r_i(Dirty_i = 0) = \Phi\left(\hat{\beta}_0 + \boldsymbol{X}_i{}'\hat{\delta}_1\right).$$

We refer to the above two probabilities as *estimated risks*.

There are two primary concerns regarding the above model. First, there is a possibility of endogeneity problems. For example, those who are more likely to have symptoms may be less likely to use dirty fuel. Second, there may be a case in which a respondent misreports their experience of symptoms since we rely on recall data. We address these two concerns as detailed in the subsections below.

4.2. Two-stage residual inclusion (2SRI) model

To address the possible endogeneity of $Dirty_i$, we adopt IV methods. As our instrument, we create a variable based on the question: "Do you have some opportunity to obtain cooking fuel from neighbors, friends or relatives?" Since households can collect firewood by themselves (see Section 3.2), some households may give it to neighbors.[20] Using this question, we create a dummy variable, "Have an opportunity to obtain fuel from neighbors, etc." to use as an instrument. According to the arguments below, we consider this instrument to be valid.

First, we can imagine that a respondent who has an opportunity to obtain dirty fuel uses it more often than one who does not. Thus, we expect that our instrument sufficiently correlates with $Dirty_i$. Second, and more important, because of marital practice in this region, it is not unreasonable to assume that our instrument does not correlate with unobserved individual-specific factors affecting $Symp_i$. In India, arranged marriage remains dominant, especially in rural areas (Allendorf and Pandian, 2016).[21] Under this practice, married women have limited or no chance to choose their husband and the location of his house (which becomes her house) or its neighbors.[22] Note that almost all of our respondents are married women (Table 1). Thus, it is reasonable to consider that our respondents have a very limited chance to choose their neighbors and thus the opportunity to obtain fuel thereby, meaning that the respondent's health cannot be directly associated with the instrument. This implies the exogeneity of our instrument. On average, 23.8% of the respondents had a value of one for the IV (Table 1).

Due to the nonlinearity of equation (1), we use the two-stage residual inclusion (2SRI) approach. Note that the 2SRI is consistent but the two-stage least squares estimator is not if the model is nonlinear (Terza et al., 2008). Furthermore, we adopt the fractional response variable framework (Papke and Wooldridge, 1996) in the first stage since $Dirty_i$ is a proportion that ranges from zero to

[20] From our survey, we collected data on monthly expenditure on cooking fuel. We found that 27.0% of our sample took "0" for this variable, meaning that they collected firewood or other solid fuels themselves. The sample mean of monthly fuel expenditure was 442.4 INR, while the mean when excluding "0" observations was 606.3 INR. Note that the mean of monthly household income was 7,428 INR (see Table 1). Although not reported in the tables, 26.5% of our sample households possessed a ration card for household below the poverty line.

[21] According to Allendorf and Pandian (2016), among their female sample who married in the 2000s, only 6.4% chose their spouses; 62% of women chose their husbands jointly with their parents, while parents alone chose for 31.0%. In this sample, 64% met their husbands for the first time on their wedding day.

[22] Most of northern India is characterized by a patrilocal residence system where the married male brings his wife to live with his father's family (Dalmia and Lawrence, 2005).

one.

The first stage of our model is a fractional probit where the conditional mean function is specified as

$$E[Dirty_i|z_i, \boldsymbol{X}_i] = \Phi(\beta_2 + \beta_3 z_i + \boldsymbol{X}_i'\gamma),$$

where z_i is the instrument. We obtain the Bernoulli quasi-maximum likelihood estimators of β_2, β_3 and γ ($\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\gamma}$, respectively). For the second stage, we use the nonlinear least squares method to the following regression model:

$$Symp_i = \Phi(\beta_4 + \beta_5 Dirty_i + \boldsymbol{X}_i{}'\delta_2 + \beta_u \hat{u}_i) + e_i^{2SRI},$$

where e_i^{2SRI} is the regression error term and $\hat{u}_i$ is the residual defined as $\hat{u}_i = Dirty_i - \Phi\left(\hat{\beta}_2 + \hat{\beta}_3 z_i + \boldsymbol{X}_i'\hat{\gamma}\right)$. Two estimated risks of this model are calculated using the estimators $\hat{\beta}_4$, $\hat{\beta}_5$ and $\hat{\delta}_2$.

4.3. Hausman, Abrevaya and Scott-Morton (HAS) model

Another econometric concern is the misclassification of $Symp_i$. It is possible that, although an individual responded that she/he had one of the three symptoms in the last month, the response was incorrect. The opposite might also have happened, where there was a response of no symptom while the respondent actually had a symptom. Since we asked about the experience of symptoms that are minor and common for the past several weeks, these misclassifications can occur. Possible reasons for this are the limited ability to recall and/or a biased perception of one's own health.

Measurement errors in dependent variables can lead to estimators that are biased and inconsistent if a regression model is nonlinear (Hausman, 2001). Hausman et al. (1998) propose a parametric method for estimating a binary outcome model with misclassification (HAS model). We apply this method to address the possible misclassification of $Symp_i$. We define the probabilities of false positives and false negatives conditional on the true status of symptoms as

$$\Pr(Symp_i = 1 | Symp_i^T = 0) = \alpha_{0i},$$
$$\Pr(Symp_i = 0 | Symp_i^T = 1) = \alpha_{1i},$$

where $Symp_i^T$ is the true indicator for $Symp_i$. Hausman et al. (1998) assume that these probabilities are constants for all individuals, that is, $\alpha_{0i} = \alpha_0$ and $\alpha_{1i} = \alpha_1$ for all i. They therefore propose a regression model that allows for misclassification as

$$Symp_i = \alpha_0 + (1 - \alpha_0 - \alpha_1)\Phi(\beta_6 + \beta_7 Dirty_i + \boldsymbol{X}_i{}'\delta_3) + e_i^{HAS},$$

where e_i^{HAS} is the error term.[23] We, again, adopt the probit model. Hausman et al. (1998) show that parameters ($\alpha_0, \alpha_1, \beta_6, \beta_7,$ and δ_3) are identified due to the nonlinearity of the normal CDF as long as $\alpha_0 + \alpha_1 < 1$.[24] They further demonstrate that the maximum likelihood estimation of this equation provides straightforward and consistent estimates.[25] Note that the estimated coefficients of α_0 and α_1 provide a specification test for whether misclassification is a problem. The two estimated risks of this model are calculated using the estimators $\hat{\beta}_6$, $\hat{\beta}_7$, and $\hat{\delta}_3$ to obtain reasonable results even allowing for the possibility of misclassification.

4.4. Results of the estimation of objective risks

Table 2 reports the estimation results for the five models: the probit model, probit model with interaction terms, 2SRI model, 2SRI model with interaction terms, and HAS-probit model. Panel A reports the estimated coefficients. Appendix Table A1 reports the average marginal effects for all the variables included in the models and the first stage of the 2SRI model. The results from the five models consistently show that $Dirty_i$ is positively associated with the experience of symptoms in the subsequent month. Columns 3 and 4 report the results of the 2SRI models. The residual of the first stage ($\hat{u}_i$) is not statistically significant, suggesting that there is no endogeneity of $Dirty_i$.[26] Column 5 reports the result of the HAS-probit model. The estimated probability of a false positive (α_0) is 0.14, while that of a false negative (α_1) is 0.03. This suggests that some respondents who answered "yes" to the symptom question may not have suffered from it. Note that the coefficient of $Dirty_i$ in the HAS-probit (Column 5) is larger than that in the standard probit (Column 1), meaning that the health risk of dirty fuels can be underevaluated if the possibility of misclassification is ignored. This attenuation effect due to misclassification is consistent with a previous study (Meyer and Mittag, 2017). In summary, we conclude that a positive and significant impact of dirty fuels exists and that the result is robust to model choices.

To evaluate the health risk of dirty fuels quantitatively, we estimate the average adjusted predictions (AAPs) at specific values of $Dirty_i$. Panel B reports the results for the five models. Column 1 reports the results of the probit model. If all the respondents take $Dirty_i = 1$, then the average of the probabilities that each respondent will have the symptoms in the next month is 0.98. All

---

23 With no misclassification, $\alpha_0 = \alpha_1 = 0$, and this equation becomes Equation (1).

24 Note that this condition is relatively weak, since it states that the combined probability of misclassification is not so high that, on average, one cannot tell which result actually occurred (Hausman, 2001). Further, note that the knowledge of or an assumption on the error distribution is necessary to obtain consistent estimators in the HAS model.

25 Meyer and Mittag (2017) refer to these parameters as the HAS-probit.

26 Our instrument is positively and statistically significantly associated with $Dirty_i$ (see Appendix Table A1).

five models show quantitatively similar results, meaning that using dirty fuels for all 30 days results in experiencing the symptoms almost certainly. This average probability decreases to 0.90 if $Dirty_i$ is decreased to 0.75. The probability becomes 0.71 (0.42) if the fraction of dirty fuel usage becomes half (a quarter) of a month.

Note that the probability of the symptoms may not be zero, even if an individual uses clean fuels for all 30 days, since the symptoms examined in this study are quite common and can be caused by factors other than cooking. The estimated average probability at $Dirty_i = 0$ shows a larger variation than that of $Dirty_i = 1$. The AAP for the 2SRI model is 0.26, while the AAP for the HAS model is 0.02. Due to this sensitivity regarding the choice of health risk models, we estimate the SRBF for each of the five models in the next section. Using estimated coefficients, we calculated two probabilities, $r_i(Dirty_i = 0)$ and $r_i(Dirty_i = 1)$, for each model for each respondent. We consider these estimated risks as objective probabilities.

5. Elicitation and aggregation of subjective risk beliefs

5.1. Methods for eliciting subjective probabilistic beliefs

The elicitation of subjective probabilities began in developed countries. Manski (2004) and Hurd (2009) review the literature on elicitation method. More recently, economists have begun to elicit subjective probabilities in less developed countries. Delavande (2014) summarizes the challenges and methods to elicit subjective probabilities in developing countries. Delavande et al. (2011b) conclude that, even in developing countries, survey respondents can generally understand and answer probabilistic questions.

Several notable designs are proposed that differ from those for developed countries. First, using visual aids and physical objects is encouraged in developing countries since simply asking for a percent chance is too abstract.[27] Asking respondents to allocate stones, marbles, or beans helps them to express probabilistic concepts, even if they are less literate. The use of 10 or 20 physical objects is now quite standard in the literature.[28] Second, asking about a binary event is easier for respondents than asking about the distribution of a continuous outcome. Third, asking respondents to imagine an event that will be experienced by "people like you" is commonly adopted. This type of wording is

[27] In developing countries, the collection of subjective probabilistic expectations in the course of a one-on-one interview is common, unlike in developed countries where the use of mail, phone, or online surveys is common (Delavande et al., 2011b).

[28] Delavande et al. (2011a) examine the sensitivity of elicited subjective probabilities to variations in elicitation designs such as the number of beans. They conduct a methodological randomized experiment with boat owners in India and elicit expectations about future fish catching.

appealing since it helps respondents formulate expectations separately for idiosyncratic and aggregated risks. Delavande (2014) provides more detailed and other related discussions.

5.2. Subjective beliefs on the risk of the three symptoms

Prior to the first round, we conducted a pilot test targeting 70 households in August 2016. In this test, we attempted to elicit people's beliefs regarding several physical symptoms, including the three that are considered in this paper.[29] Unlike the second round, which focuses on the health risk that may appear in the next month, we also elicited beliefs on risk in the subsequent three-month, six-month, one-year, and two-year periods.

From the pilot test, we obtained two intriguing findings. First, many respondents believed that the probability that they would have symptoms in the future depended on whether they had any symptoms currently. They believed that future symptoms depended on their current health condition and the type of fuel they used. Thus, in the main surveys, we decided to elicit people's subjective beliefs conditional on their current symptom status.

Second, regarding the three symptoms (dry cough, sore or runny eyes, and difficulty breathing), most of the respondents believed that the probability of having these symptoms in the following month was less than one, even if they had the symptoms currently. In other words, they believed that they could be healed naturally.[30] This is different from beliefs on HIV/AIDS elicited in other work (for example, Delavande and Kohler, 2016). In the case of such a disease, once a person has HIV, it is not worth considering probabilistic beliefs that the person will not become infected with HIV.

Based on these observations, we assume that individuals at our research site separately form the following two beliefs, conditional on their fuel usage patterns:

$$\psi\left(\Pr_i\left(Symp_{i,t+1} = 1 \middle| Dirty_{i,t} = a, Symp_{i,t} = 0\right)\right) = \psi_{ia0},$$

$$\psi\left(\Pr_i\left(Symp_{i,t+1} = 1 \middle| Dirty_{i,t} = a, Symp_{i,t} = 1\right)\right) = \psi_{ia1},$$

where t denotes the time period, where a 30-day period constitutes one period, and $a \in [0,1]$.

Two remarks are worth noting. First, it is not an *a priori* hypothesis that individuals believe

[29] Furthermore, we elicited subjective beliefs on the risks of symptoms in the respondents' spouses and children, in addition to themselves.

[30] Many respondents expressed their belief that the probability of developing any symptoms in six months would be lower than the probability of having symptoms in three months and that the probability of having symptoms in one year would be lower than the probability of having symptoms in six months. We also elicited subjective risk beliefs by proposing hypothetical situations regarding treatment, such as a situation in which they would receive medications from a physician and a situation in which they would make their own remedies using homemade medicines. See the Online Appendix for details of the preliminary survey questionnaire.

that the experience of symptoms in the next period depends on the experience of symptoms in the current period. Rather, this model is based on the observations of our pilot test.[31] Second, as a result of our modeling, an individual's belief becomes a two-state Markov chain conditional on a given fuel usage pattern $a \in [0,1]$.

Following Delavande (2014), as well as other previous studies, we elicited the subjective risk beliefs of our respondents using ten candies, allowing them to express probabilities in units of 0.1.[32] Our enumerators explicitly asked the respondents to link the number of candies allocated to the perceived likelihood of experiencing the three symptoms.[33] In the survey, the term *sick* was used if a respondent had one of three symptoms and *healthy* if not. The following question was used to elicit subjective risk beliefs:

> *Consider a hypothetical individual who is identical to you. Imagine that there are options regarding the primary fuel for cooking. In each health status situation H, please answer how likely you think it is that she (or he) will become (remain) sick in the next 30 days if she (or he) used fuels X in all the previous 30 days.*

where X has two options "LPG, kerosene, or electricity" and "Firewood, cow dung cakes, or coal." Note that we did not use the term *dirty* or *clean*. For the health status situation, "H," the options were "she is *healthy*" and "she is *sick*." The instructions used in the elicitation are shown in Appendix Table A2.[34] The above questions enabled us to elicit the four subjective risk beliefs of individual i: ψ_{i00}, ψ_{i01}, ψ_{i10}, and ψ_{i11}.

Since we model individual subjective beliefs as a Markov process, it is possible to calculate a stationary distribution conditional on each fuel choice $a = \{0,1\}$ using the following equation:

---

31 We acknowledge that a different argument can be made about individuals' beliefs. Specifically, individuals may believe that the probability of experiencing symptoms in the next period depends not only on symptoms in the current period but also on symptoms in the previous n(>1) periods. However, if we assume that an individual's beliefs depend on symptoms in the past *n* periods (and fuel usage pattern), the beliefs that must be elicited would increase exponentially. That would impose a greater burden on the survey respondents, and the survey would then become unwieldy, making respondents more likely to skip questions, which leads to biases. We therefore sought a compromise by assuming that individuals believe that the probability of having symptoms in the next period only depends on their present symptoms.

32 While several previous studies elicit subjective probabilistic beliefs by simply asking people to rate the risk from zero to 10 (for example, Khwaja et al., 2007), Viscusi and Hakes (2003) raise the concern that this scale does not succeed as a probability metric and recommend the use of visual aids.

33 Before the question on risks of cooking, our enumerators elicited respondents' subjective beliefs on the likelihood of rainfall on that particular day to check whether the respondents had understood how to express their beliefs.

34 The complete questionnaires used in the first and the second rounds are included in the Online Appendix.

$$\psi\big(r_i(Dirty_i = a)\big) = \frac{\psi_{ia0}}{1 + \psi_{ia0} - \psi_{ia1}}.$$

To compare the estimated risks and the elicited beliefs, we use this concept of a stationary distribution of the subjective probabilities.

5.3. Results of the elicitation of subjective risk beliefs

Table 1, Panel B reports the mean and standard deviation of the four elicited risk beliefs, while Figure 1 shows the histogram of the four beliefs. The left panels of Figure 1 show the distributions of the subjective belief that an individual will experience one of the three symptoms in the next month ($Symp_i = 1$) when he or she uses clean cooking fuel and if he or she currently has symptoms or is healthy. The right panels of Figure 1 show the distributions of the subjective belief that an individual will become sick in the next month when he or she uses dirty cooking fuel and if he or she is currently sick or healthy. Although the subjective beliefs range from zero to one, the distribution of the subjective belief that a healthy individual will have a symptom in the next month if he or she uses clean fuel is concentrated at a very low value (approximately 0.1), while the distribution of the subjective belief that a sick individual will have symptoms in the next month if he or she uses clean fuel is concentrated at a moderately low value. The distribution of the subjective belief that a healthy individual will have symptoms in the next month if one uses dirty fuel is concentrated at moderately high values (approximately 0.6), while the distribution of the subjective belief that a sick individual will have symptoms in the next month if one uses dirty fuel is concentrated over high values (around 0.9). The comparison between Panels A and C (similarly, B and D) suggests that, on average, respondents believe that the probability of experiencing symptoms is high if one experiences symptoms currently compared to the case in which one is currently healthy.

Figure 2 shows the histograms of the two subjective beliefs, which are the stationary distributions of the two Markov chains. This figure shows that the histogram for the case in which one uses clean fuel (Panel A) is skewed to the left, while the case in which one uses dirty fuel (Panel B) is skewed to the right. This suggests that, on average, our respondents believe that using LPG (or kerosene or electricity) leads to a lower probability of symptoms than using firewood (or cow dung cakes or coal).

6. Estimation of the subjective risk belief function

Figure 3 plots two elicited subjective beliefs (see Figure 2) on the y-axis and two estimated risks on the x-axis for the 588 respondents. Each panel (A to E) corresponds to the estimated risks calculated using probit, probit with interaction terms, 2SRI, 2SRI with interaction terms, and HAS, respectively. The red X depicts $s_i = \psi(r_i(Dirty_i = 0))$, and the blue cross depicts $s_i = \psi(r_i(Dirty_i = 1))$. The green line shows $s_i = r_i$, implying that the observation indicates a correct belief in terms of risk levels if the plot is on the line.

Panels A, B, C, and D display similar graphs. Overall, the blue crosses are plotted to the upper-right side of the red Xs. This implies that dirty fuels are worse than clean fuels in terms of objective health risks and that the respondents' subjective beliefs are in line with these objective risks. Regarding the red Xs, several responses are concentrated at the bottom of the graphs. These come from those respondents who believe that there is no risk from clean fuels. However, along the x-axis, all observations are located above zero, with considerable variation, suggesting that respondents underestimate risk of the symptoms. Other red Xs are scattered both above and below the green line. Regarding the blue crosses, a certain number of respondents are below the green line, even though other respondents express a subjective belief of one. Panel E displays a slightly different graph. Red Xs are clustered on the left, and blue crosses are clustered on the right side of the graph. This reflects the results in Table 3, Panel B, where the AAP at $Dirty_i = 0$ is quite small and $Dirty_i = 1$ is fairly large in the HAS model compared to other models. What the five panels have in common is that the variation in subjective beliefs is larger than the variation in estimated risks. This implies that while our respondents qualitatively correctly perceive the risk of household air pollution, both under- and overestimation of risk levels exist.

To investigate misperception in risk changes on average, we examine the steepness of the SRBF. We model the linear SRBF and estimate $\frac{\partial \psi}{\partial r}$ using the regression equation:

$$s_i = \rho_0 + \rho_1 r_i + \sigma_i + u_i,$$

where σ_i represents the fixed effects, u_i is the error term, and ρ_1 gives $\frac{\partial \psi}{\partial r}$.

Table 3 reports the results of the estimation of the SRBF for each pair of estimated risks obtained by the five models. A null hypothesis of $\frac{\partial \psi}{\partial r} = 1$ is tested and reported. The estimated coefficient of the probit model is $\frac{\partial \psi}{\partial r} = 0.70$ and significantly different from both zero and one. The estimated intercept is 0.07. This means that, on average, our respondents almost accurately perceive the risk of clean fuel but underestimate the risk of dirty fuel, leading to a slight underestimation of the change in risk. The probit with interactions model and the 2SRI with interactions model result in surprisingly similar results. The 2SRI model without interaction yields that $\frac{\partial \psi}{\partial r} = 0.79$, which implies a slightly more accurate perception than in the previous models, while it is still significantly different from one. In contrast, the HAS model yields $\frac{\partial\psi}{\partial r} = 0.58$ with an intercept of 0.18, meaning additional misperception in the risk level of clean fuel, which results in a larger misperception of the change in risk.

Since we obtained two plots for each respondent, we consider the respondent-specific SRBF and examine its coefficient for each respondent. To visually depict the respondent-level heterogeneity in the SRBF, we calculate the two differences below:

$$\Delta s_i = \psi\big(r_i(Dirty_i = 1)\big) - \psi\big(r_i(Dirty_i = 0)\big),\ \Delta r_i = r_i(Dirty_i = 1) - r_i(Dirty_i = 0),$$

where $\frac{\Delta s_i}{\Delta r_i}$ yields the coefficient of individual i's linear SRBF. Again, Δr_i is calculated using five models of health risk.

Figure 4 plots $(\Delta s_i, \Delta r_i)$. The red line illustrates $\Delta s_i = \Delta r_i$, indicating $\frac{\partial\psi_i}{\partial r} = 1$. First, there is no observation with $\Delta s_i < 0$, meaning that all 588 respondents believe that firewood (and other solid fuels) entails a higher risk of the three symptoms than LPG (and kerosene). Second, the variation in Δs_i is greater than the differences in our estimated risks (Δr_i). This pattern is observed in all five panels. Identifying whether a specific individual misperceives risk is interesting; however, Figure 4 shows that the results can differ with the choice of health risk model. In that sense, our methodology is less robust in the estimation of objective risks.

To investigate characteristics that correlate with misperception in risk levels and changes, we estimate the following regression equation:

$$s_i = \rho_0 + \rho_1 r_i + \rho_2 Age_i + \rho_3 Hindu_i + \rho_4 Education_i$$

$$+\rho_5(r_i \times Age_i) + \rho_6(r_i \times Hindu_i) + \rho_7(r_i \times Education_i) + u_i,$$

where Age_i and $Education_i$ are the age and years of education of the respondent, respectively. $Hindu_i$ takes value one if the respondent's household follows the Hindu religion and zero otherwise, implying that the respondent's household follows the Muslim religion.

Table 4 reports the estimation results. $Hindu_i$ shows a negative and statistically significant association with subjective belief, while the constant term is larger than zero, meaning that the Hindu respondents have relatively accurate beliefs regarding the risk levels. Furthermore, the coefficient on $r_i \times Hindu_i$ is positive and significant, suggesting that, on average, our respondents underestimate the change in risk when switching from dirty to clean fuel; however, the Hindu respondents' belief is relatively accurate in the sense that $\frac{\partial \psi_i}{\partial r}$ is closer to one than that of Muslim respondents, which is approximately 0.1 points. This result implies a relationship between subjective risk belief and religious faith. Age_i is not associated with subjective belief, while there is a very weak negative association between $r_i \times Age_i$ and subjective belief. No significant and robust association is observed with regard to $Education_i$.

7. Discussion

This study has the following limitations that should be noted. First, this study has limited policy implications due to its scope. In general, a study on the risk of serious diseases, such as ALRI or COPD, would have greater policy implications. However, it is more difficult to evaluate the objective risk of such serious diseases since they occur less frequently. For this reason, we targeted three minor symptoms; hence, careful interpretation of the results is required. If respondents misperceive the risk of serious symptoms, their fuel choice becomes less efficient even if they quite accurately perceive the risk of minor symptoms. We admit that there is a tradeoff between the policy implications of the research subject and the feasibility of objective risk estimation. Furthermore, this study focused on short-term effects, that is, the impact that the use of fuel for one month might have on health in the following month. A study on the SRBF for long-term risks is another interesting topic. However, another limitation of this study is that it only examined the health impacts on individuals who are actually involved in cooking. Health risks for their children are also an important policy issue. Whether parents or other decision-makers in the household correctly perceive the health risks of their fuel choice for children is an extremely important question.

Second, this paper does not consider confidence in respondents' subjective beliefs. For example, several respondents expressed their belief on the risk of LPG as 0.5. It is possible to imagine that the respondents believed that the risk was half; however, it is also possible that the respondents were not confident in their risk estimates. We did not consider how confidently the respondents formulated certain beliefs.

Third, the issue, the sensitivity of the regression analysis used to estimate objective risks must be addressed. The SRBF estimates were sensitive to the selection of the econometric model, indicating the difficulty, and the limitations, of estimating objective risks. Nevertheless, we believe that the previous studies that use the average value obtained from epidemiological predictions share similar limitations. Improving the quality of data collection and econometric analysis is required to obtain better estimates of the SRBF.

8. Conclusion

This paper proposes a research framework to quantify individuals' misperceptions regarding changes in risk, presents the concept of the subjective risk belief function (SRBF), and estimates it in West Bengal. From our elicitation of subjective beliefs, we find evidence that people who are involved in cooking in West Bengal believe that firewood has higher health risks than LPG. This perception is qualitatively correct. The Indian government is already implementing several programs, such as providing subsidies for LPG stove purchases (see Kar et al., 2019 for further details). In fact, some households at our research site had already received subsidies to cover the costs of an LPG stove. This implies that the superiority of LPG with respect to health may now be widely known. On the other hand, when the estimated risks are quantitatively considered, we find slight misperceptions of health risk regarding switching from firewood to LPG. The coefficient of the average linear SRBF is estimated to be between 0.58 and 0.79, which is statistically significantly smaller than one (though it is also statistically significantly larger than zero). For this reason, it can be expected that the effects of additional information-provision policies on behavioral change may exist, but they may not be significant.

There is another finding on the association between misperception and a characteristic variable. The SRBF estimates, which include interaction terms, show no significant correlations in terms of age or educational levels, but they show a significant correlation between religion and risk misperception. Note that our estimation result from the first stage shows that Muslim respondents are more likely to use dirty fuels (see Appendix Table A1), which is consistent with previous studies (see a review by Lewis and Pattanayak, 2012); this preference for solid fuels can be interpreted as the result of cultural practices and taste associated with meals (Gupta and Köhlin, 2006). These findings suggest an association between preferences and beliefs.

The above evidence leads to at least two important directions for future research. First, investigating further correlations between observables and misperceptions would make it possible to target information provision by observables. Second, further empirical studies on belief formation using our framework are possible.

The potential impact of our results on the literature on risk attitudes is worth noting. Previous studies show that risk preference is biased toward overweighting (underweighting) to extremely small (large) probabilities or shows an inverted S-shape (Barseghyan et al., 2018). Our results suggest that risk belief is also biased in a similar way. Further studies on risk attitudes that incorporate both preferences and beliefs are required, possibly extending our research framework.

References

Allendorf K, Pandian RK. The Decline of Arranged Marriage? Marital Change and Continuity in India. *Population and Development Review* 2016;42; 435-464.

Anenberg S, Balakrishnan K, Jetter J, Masera O, Mehta S, Moss J, Ramanathan V. Cleaner cooking solutions to achieve health, climate, and economic cobenefits. *Environmental Science & Technology* 2013;47; 3944.

Balakrishnan K, Cohen A, Smith KR. Addressing the burden of disease attributable to air pollution in India: The need to integrate across household and ambient air pollution exposures. *Environmental Health Perspectives* 2014;122; A6-A7.

Barberis N. The psychology of tail events: Progress and challenges. *American Economic Review* 2013;103; 611-616.

Barseghyan L, Molinari F, O'Donoghue T, Teitelbaum JC. Distinguishing probability weighting from risk misperceptions in field data. *American Economic Review* 2013a;103; 580-585.

Barseghyan L, Molinari F, O'Donoghue T, Teitelbaum JC. The nature of risk preferences: Evidence from insurance choices. *American Economic Review* 2013b;103; 2499-2529.

Barseghyan L, Molinari F, O'Donoghue T, Teitelbaum JC. Estimating risk preferences in the field. *Journal of Economic Literature* 2018;56; 501-564.

Breyer S. Breaking and Vicious Circle: Toward Effective Risk Regulations. Harvard University Press: Cambridge, MA; 1993.

Brown J, Hamoudi A, Jeuland M, Turrini G. Seeing, believing, and behaving: Heterogeneous effects of an information intervention on household water treatment. *Journal of Environmental Economics and Management* 2017;86; 141-159.

Carman KG, Kooreman P. Probability perceptions and preventive health care. *Journal of Risk and Uncertainty* 2014;49; 43-71.

Chattopadhyay M, Arimura TH, Katayama H, Sakudo M, Yokoo H-F. Subjective Probabilistic Expectations, Household Air Pollution, and Health: Evidence from cooking fuel use patterns in India. Discussion Papers 2020-05, Graduate School of Economics, Hitotsubashi University 2020.

Dalmia S, Lawrence PG. The Institution of Dowry in India: Why It Continues to Prevail. *Journal of Developing Areas* 2005;38; 71-93.

Delavande A. Probabilistic expectations in developing countries. *Annual Review of Economics* 2014;6; 1-20.

Delavande A, Giné X, McKenzie D. Eliciting probabilistic expectations with visual aids in developing countries: How sensitive are answers to variations in elicitation design? *Journal of Applied*

Econometrics 2011a;3; 479-497.

Delavande A, Giné X, McKenzie D. Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics* 2011b;94; 151-163.

Delavande A, Kohler H-P. HIV/AIDS-related expectations and risky sexual behaviour in Malawi. *Review of Economic Studies* 2016;83; 118-164.

Dupas P. Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics* 2011a;3; 1-34.

Dupas P. Health behavior in developing countries. *Annual Review of Economics* 2011b;3; 425-449.

Fox CR, Tversky A. A Belief-Based Account of Decision Under Uncertainty. *Management Science* 1998;44; 879-895.

Godlonton S, Munthali A, Thornton R. Responding to risk: Circumcision, information, and HIV prevention. *Review of Economics and Statistics* 2016;98; 333-349.

Godlonton S, Thornton RL. Learning from Others' HIV Testing: Updating Beliefs and Responding to Risk. *American Economic Review* 2013;103; 439-444.

Gould CF, Urpelainen J. LPG as a clean cooking fuel: Adoption, use, and impact in rural India. *Energy Policy* 2018;122; 395-408.

Gul F. A theory of disappointment aversion. *Econometrica* 1991;59; 667-686.

Gupta G, Köhlin G. Preferences for domestic fuel: Analysis with socio-economic factors and rankings in Kolkata, India. *Ecological Economics* 2006;57; 107-121.

Hanna R, Duflo E, Greenstone M. Up in smoke: The influence of household behavior on the long-run impact of improved cooking stoves. *American Economic Journal: Economic Policy* 2016;8; 80-114.

Hausman J. Mismeasured variables in econometric analysis: problems from the right and problems from the left. *Journal of Economic Perspectives* 2001;15; 57-67.

Hausman JA, Abrevaya J, Scott-Morton FM. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* 1998;87; 239-269.

Heltberg R. Factors determining household fuel choice in Guatemala. *Environment and Development Economics* 2005;10; 337-361.

Hurd MD. Subjective Probabilities in Household Surveys. *Annual Review of Economics* 2009;1; 543-562.

Jeuland M, Pattanayak SK, Bluffstone R. The economics of household air pollution. *Annual Review of Resource Economics* 2015;7; 81-108.

Johansson-Stenman O. Mad cows, terrorism and junk food: Should public policy reflect perceived or objective risks? *Journal of Health Economics* 2008;27; 234-248.

Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 1979;47; 263-291.

Kar A, Pachauri S, Bailis R, Zerriffi H. Using sales data to assess cooking gas adoption and the impact of India's Ujjwala programme in rural Karnataka. *Nature Energy* 2019.

Khwaja A, Silverman D, Sloan F, Wang Y. Are mature smokers misinformed? *Journal of Health Economics* 2009;28; 385-397.

Khwaja A, Sloan F, Chung S. The relationship between individual expectations and behaviors: Mortality expectations and smoking decisions. *Journal of Risk and Uncertainty* 2007;35; 179-201.

Kőszegi B, Rabin M. A model of reference-dependent preferences. *Quarterly Journal of Economics* 2006;121; 1133-1165.

Lange F. The role of education in complex health decisions: Evidence from cancer screening. *Journal of Health Economics* 2011;30; 43-54.

Lewis JJ, Pattanayak SK. Who adopts improved fuels and cookstoves? A systematic review. *Environmental Health Perspectives* 2012;120; 637-645.

Lundborg P, Lindgren B. Risk perceptions and alcohol consumption among young people. *Journal of Risk and Uncertainty* 2002;25; 165-183.

Manski CF. Measuring expectations. *Econometrica* 2004;72; 1329-1376.

Meyer BD, Mittag N. Misclassification in binary choice models. *Journal of Econometrics* 2017;200; 295-311.

Okeke EN, Adepiti CA, Ajenifuja KO. What is the price of prevention? New evidence from a field experiment. *Journal of Health Economics* 2013;32; 207-218.

Oster E, Shoulson I, Dorsey ER. Optimal expectations and limited medical testing: Evidence from Huntington disease. *American Economic Review* 2013;103; 804-830.

Papke LE, Wooldridge JM. Econometric Methods for Fractional Response Variables With an Application to 401 (K) Plan Participation Rates. *Journal of Applied Econometrics* 1996;11; 619-632.

Pattanayak SK, Pfaff A. Behavior, environment, and health in developing countries: Evaluation and valuation. *Annual Review of Resource Economics* 2009;1; 183-217.

Quiggin J. A theory of anticipated utility. *Journal of Economic Behavior & Organization* 1982;3; 323-343.

Savage LJ. The Foundations of Statisticss. John Wiley: New York; 1954.

Schoenbaum M. Do smokers understand the mortality effects of smoking? Evidence from the Health and Retirement Survey. *American Journal of Public Health* 1997;87; 755-759.

Smith KR, Bruce N, Balakrishnan K, Adair-Rohani H, Balmes J, Chafe Z, Dherani M, Hosgood HD, Mehta S, Pope D. Millions dead: how do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution. *Annual Review of Public Health* 2014;35; 185-206.

Smith KR, Pillarisetti A. 2017. Household Air Pollution from Solid Cookfuels and Its Effects on Health. In: Mock CN, Nugent R, Kobusingye O, Smith KR (Eds), Injury Prevention and Environmental Health, vol. 7. World Bank: Washington, DC; 2017.

Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics* 2008;27; 531-543.

Tversky A, Kahneman D. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 1992;5; 297-323.

Viscusi WK. Do smokers underestimate risks? *Journal of Political Economy* 1990;98; 1253-1269.

Viscusi WK, Hakes J. Risk ratings that do not measure probabilities. *Journal of Risk Research* 2003;6; 23-43.

Viscusi WK, O'Connor CJ. Adaptive responses to chemical labeling: Are workers Bayesian decision makers? *American Economic Review* 1984;74; 942-956.

Wright G, Ayton P. Subjective probability. John Wiley & Sons: Oxford, England; 1994.

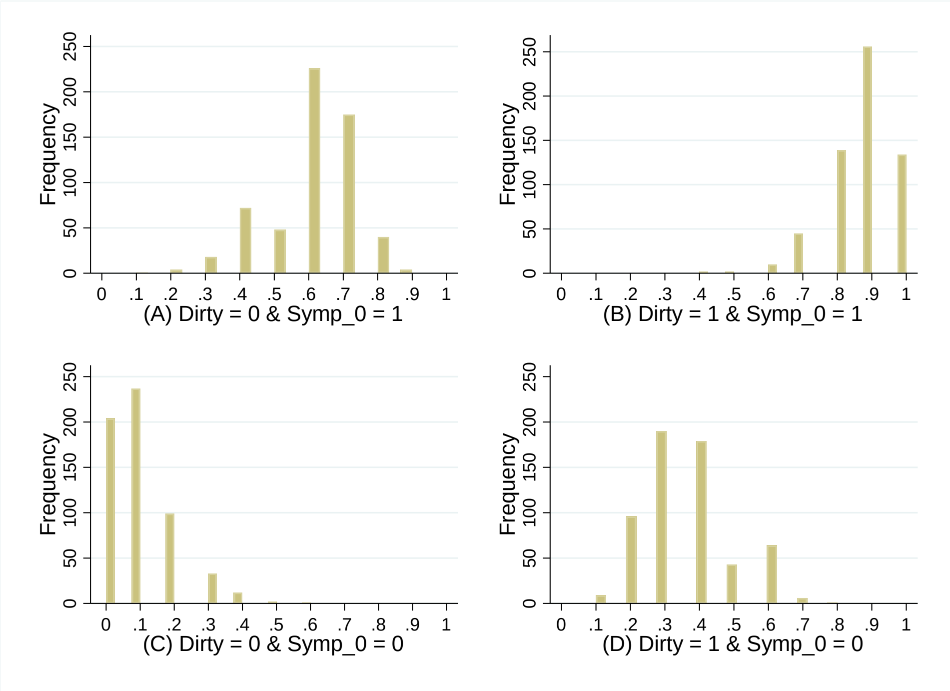


Figure 1. Distribution of four elicited subjective beliefs

Notes: This figure shows the distribution of the elicited subjective beliefs. Ten candies were used in our field survey, allowing respondents to express probabilities in units of 0.10. Panels A, B, C, and D show elicited ψ_{i01}, ψ_{i11}, ψ_{i00}, and ψ_{i10}, respectively.

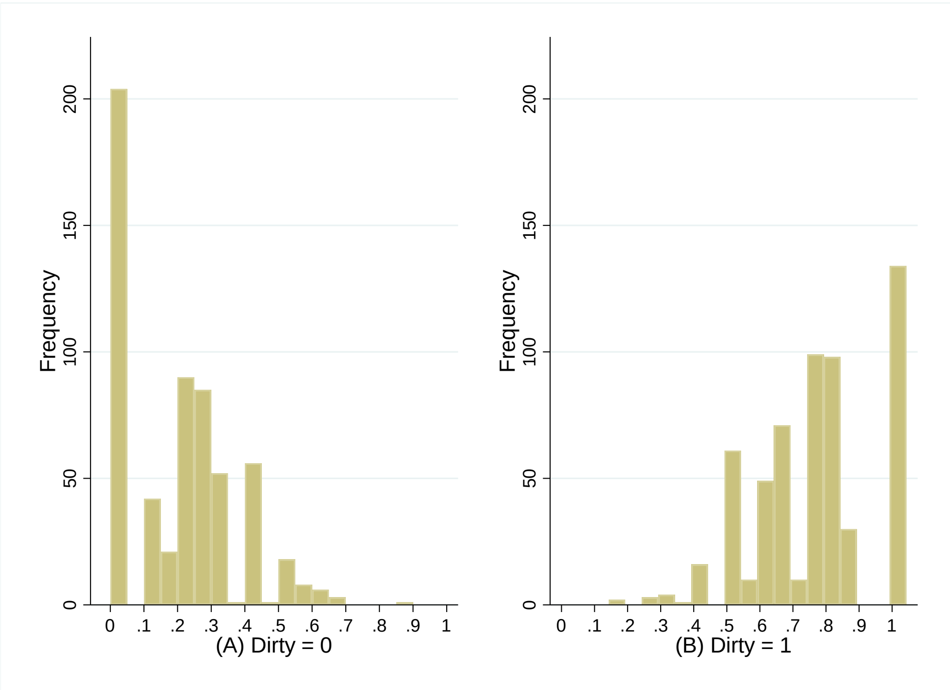
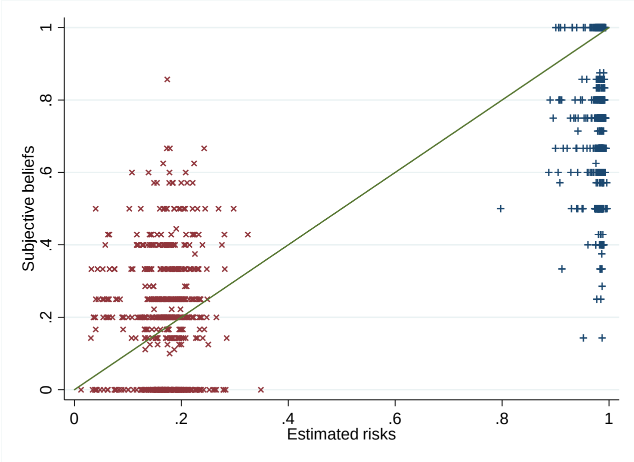
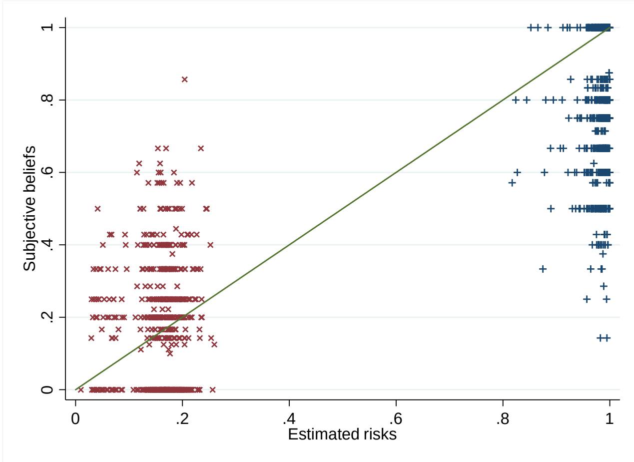


Figure 2. Distribution of the two subjective beliefs

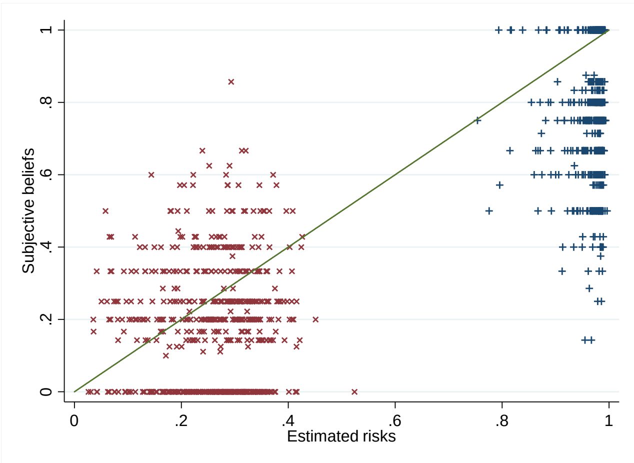
Notes: This figure shows the stationary distribution of subjective probabilities conditional on each fuel choice. Panel A shows $\psi\left(r_i(Dirty_i = 0)\right)$, and Panel B shows $\psi\left(r_i(Dirty_i = 1)\right)$.



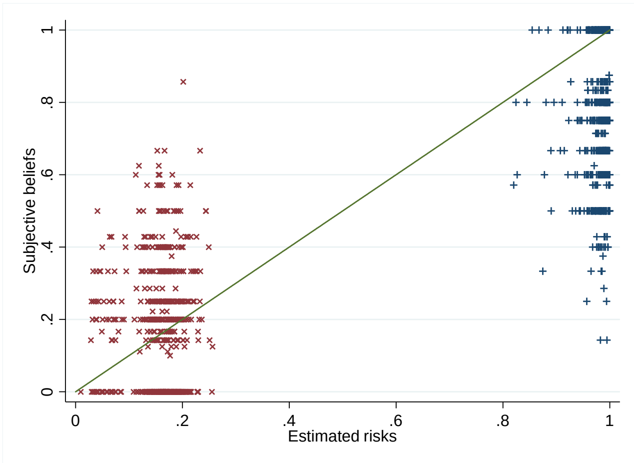
(A) Probit



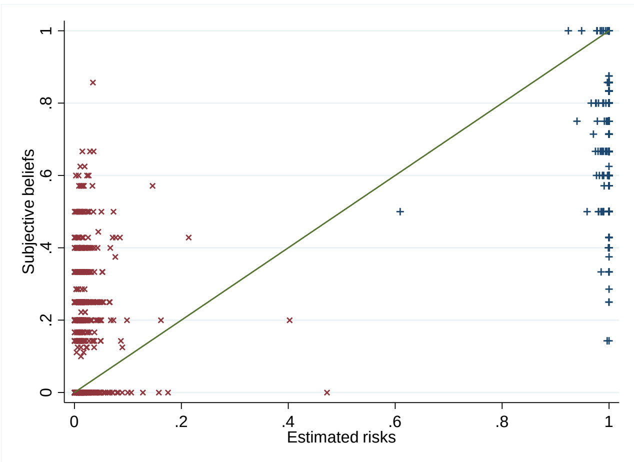
(B) Probit with interaction terms



(C) 2SRI



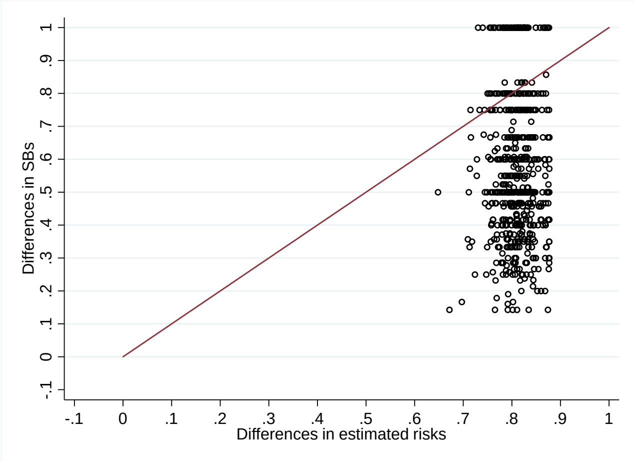
(D) 2SRI with interaction terms



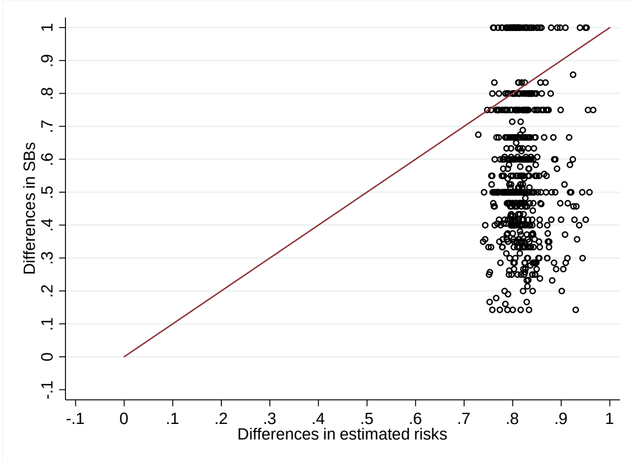
(E) HAS-probit

Figure 3. Subjective beliefs and estimated risks

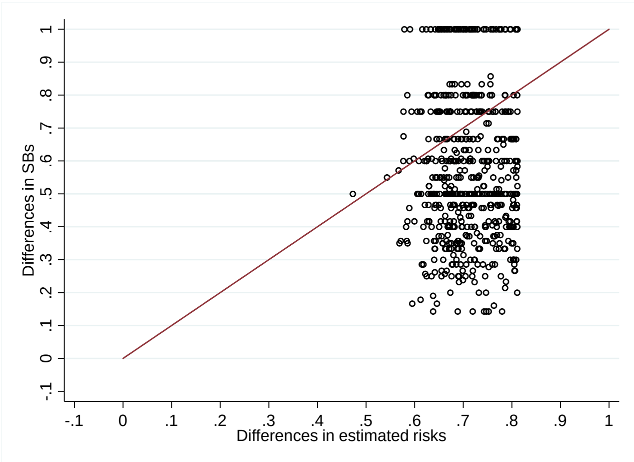
Notes: This figure shows the empirical results of the relationship between subjective belief and objective estimated risk for 588 respondents. Panels A, B, C, D, and E correspond to estimated risks calculated using probit, probit with interaction terms, 2SRI, 2SRI with interaction terms, and HAS, respectively. The red X depicts $s_i = \psi\big(r_i(Dirty_i = 0)\big)$, and the blue cross depicts $s_i = \psi\big(r_i(Dirty_i = 1)\big)$. The green line shows $s_i = r_i$.



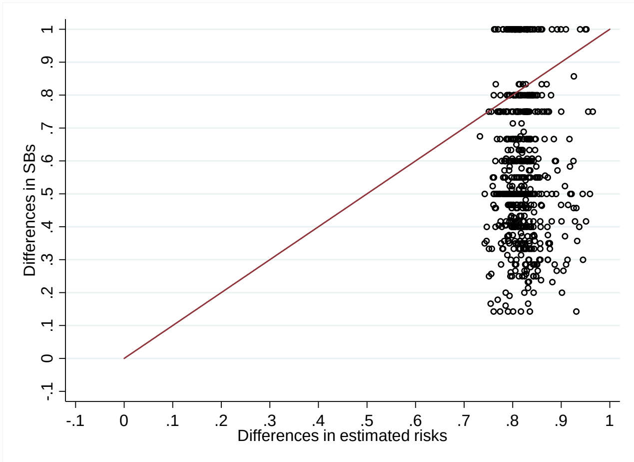
(A) Probit



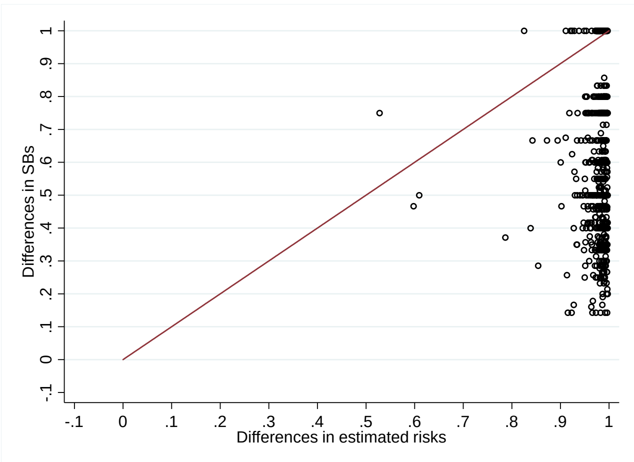
(B) Probit with interaction terms



(C) 2SRI



(D) 2SRI with interaction terms



(E) HAS-probit

Figure 4. Subjective beliefs and estimated risks in risk changes

Notes: This figure plots $(\Delta s_i, \Delta r_i)$. Panels A, B, C, D, and E correspond to the estimated risks calculated using probit, probit with interaction terms, 2SRI, 2SRI with interaction terms, and HAS, respectively. The red line illustrates $\Delta s_i = \Delta r_i$.

Table 1. Summary statistics

	Mean	Standard deviation
Panel A: Characteristic variables		
Symptoms in the past 30 days ($Symp_i$) (binary)	0.755	0.430
Fraction of days of dirty fuel usage ($Dirty_i$) before the last month	0.679	0.379
Female (binary)	0.995	0.071
Age of the respondent	38.548	11.221
Respondent household follows the Hindu religion (binary)	0.694	0.461
Years of education of the respondent	4.713	4.141
Monthly household income (*1000 INR)	7.428	3.690
Household size	4.612	2.054
Respondent is a housewife (binary)	0.952	0.213
Number of cooks in the household	1.128	0.403
Kitchen is located outside the dwelling space (binary)	0.158	0.365
Cumulative years of clean fuel usage until the first round	1.466	4.167
Household owns a personal computer (binary)	0.065	0.246
Have an opportunity to obtain fuel from neighbors, etc. (binary)	0.238	0.426
Panel B: Four elicited subjective beliefs		
$\psi(\Pr(Symp_{i,t=1} = 1 \mid Dirty_i = 1 \text{ and } Symp_{i,t=0} = 0))$	0.363	0.127
$\psi(\Pr(Symp_{i,t=1} = 1 \mid Dirty_i = 1 \text{ and } Symp_{i,t=0} = 1))$	0.876	0.100
$\psi(\Pr(Symp_{i,t=1} = 1 \mid Dirty_i = 0 \text{ and } Symp_{i,t=0} = 0))$	0.102	0.101
$\psi(\Pr(Symp_{i,t=1} = 1 \mid Dirty_i = 0 \text{ and } Symp_{i,t=0} = 1))$	0.600	0.128

Notes: The number of observations is 588.

Table 2. Risk of dirty fuel on physical symptoms (probit, 2SRI, and HAS)

Dependent variable: $Symp_i$	(1)	(2)	(3)	(4)	(5)
Probit models:	Standard	Standard	2SRI	2SRI	HAS
Panel A: Coefficients					
$Dirty_i$	3.084***	1.806**	2.629***	1.814	5.939***
	(0.246)	(0.883)	(0.818)	(1.270)	(1.300)
$Dirty_i$ × Age of the respondent		0.008		0.008	
		(0.020)		(0.023)	
$Dirty_i$ × Monthly household income		0.155*		0.155*	
		(0.080)		(0.085)	
First-stage residual ($\hat{u}_i$)			0.500	-0.012	
			(0.839)	(0.925)	
Misclassification α_0					0.138***
					(0.046)
Misclassification α_1					0.027***
					(0.010)
Other control variables	Yes	Yes	Yes	Yes	Yes
Panel B: Average Adjusted Predictions					
AAP at $Dirty_i = 0$	0.168	0.160	0.258	0.158	0.021
	(0.037)	(0.037)	(0.185)	(0.149)	(0.024)
AAP at $Dirty_i = 0.25$	0.418	0.422	0.493	0.420	0.246
	(0.040)	(0.040)	(0.139)	(0.150)	(0.081)
AAP at $Dirty_i = 0.5$	0.706	0.720	0.729	0.720	0.735
	(0.027)	(0.028)	(0.046)	(0.063)	(0.062)
AAP at $Dirty_i = 0.75$	0.901	0.908	0.891	0.908	0.965
	(0.016)	(0.015)	(0.030)	(0.019)	(0.023)
AAP at $Dirty_i = 1$	0.979	0.979	0.968	0.979	0.998
	(0.007)	(0.007)	(0.027)	(0.013)	(0.003)
Observations	588	588	588	588	588
Log Likelihood	-166.198	-163.988	-166.017	-163.988	-160.836
AIC	356.397	355.976	358.034	357.976	349.673
BIC	408.917	417.250	414.932	423.627	410.947

Notes: Panel A reports estimated coefficients for each model. The results for the constant term and control variables are not reported. Appendix Table A1 reports the results for all the control variables. Numbers in parentheses are standard errors for columns 1, 2, and 5 and the bootstrap estimate of the standard errors for columns 3 and 4. ***, **, and * indicate significance at the 1%, 5%, and 10% levels of statistical significance, respectively. Panel B reports the average adjusted predictions (AAPs) at each value of $Dirty_i$. Numbers in parentheses are delta-method standard errors.

Table 3. Estimation of the subjective risk belief function (fixed effects)

Dependent variable: s_i	(1)	(2)	(3)	(4)	(5)
Model of the health risk	Standard	Standard	2SRI	2SRI	HAS
Interaction terms	No	Yes	No	Yes	No
Estimated risk (r_i)	0.696***	0.688***	0.791***	0.687***	0.578***
	(0.011)	(0.011)	(0.013)	(0.011)	(0.009)
Constant	0.071***	0.078***	-0.015*	0.079***	0.175***
	(0.006)	(0.006)	(0.008)	(0.006)	(0.005)
p-value (H_0: $\frac{\partial\psi}{\partial r} = 1$)	0.000	0.000	0.000	0.000	0.000
Fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	1176	1176	1176	1176	1176
R squared	0.869	0.868	0.865	0.868	0.869

Notes: This table reports the results of the estimation of the subjective risk belief function. Numbers in parentheses are standard errors robust to heteroskedasticity. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4. Estimation of the subjective risk belief function with characteristics (OLS)

Dependent variable: s_i	(1)	(2)	(3)	(4)	(5)
Model of the health risk	Probit	Probit	2SRI	2SRI	HAS
Interaction terms	No	Yes	No	Yes	No
Estimated risk (r_i)	0.667***	0.700***	0.784***	0.698***	0.584***
	(0.048)	(0.048)	(0.056)	(0.048)	(0.040)
Age of the respondent	0.000	0.001	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hindu religion	-0.066***	-0.067***	-0.051**	-0.068***	-0.049***
	(0.017)	(0.017)	(0.020)	(0.017)	(0.015)
Years of education of the	-0.000	0.001	0.004*	0.001	0.000
respondent	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
$r_i \times$ Age	-0.001	-0.002*	-0.001	-0.002*	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$r_i \times$ Hindu religion	0.099***	0.101***	0.087***	0.101***	0.080***
	(0.023)	(0.023)	(0.027)	(0.023)	(0.020)
$r_i \times$ Years of education	-0.001	-0.003	-0.005	-0.003	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Constant	0.117***	0.094***	0.004	0.096***	0.183***
	(0.037)	(0.036)	(0.043)	(0.036)	(0.032)
Observations	1176	1176	1176	1176	1176
R squared	0.725	0.725	0.712	0.725	0.726

Notes: This table reports the results of the estimation of the subjective risk belief function. Numbers in parentheses are standard errors that are clustered at the respondent level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix

Table A1. Risk of dirty fuel for physical symptoms (average marginal effects)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	$Symp_i$	$Symp_i$	$Dirty_i$	$Symp_i$	$Symp_i$	$Symp_i$
Probit models:	Standard	Standard	Fractional	2SRI	2SRI	HAS
$Dirty_i$	0.485***	0.506***		0.413***	0.508***	0.459***
	(0.024)	(0.027)		(0.128)	(0.146)	(0.026)
Age of the respondent	0.002	0.002	-0.001	0.002	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hindu religion	0.013	0.014	-0.143***	0.001	0.014	0.010
	(0.031)	(0.031)	(0.027)	(0.037)	(0.038)	(0.033)
Years of education of the	0.002	0.002	-0.013***	0.000	0.002	0.000
respondent	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Monthly household income	0.002	0.008	-0.022***	0.001	0.009	0.001
(thousand INR)	(0.004)	(0.005)	(0.005)	(0.004)	(0.006)	(0.003)
Household size	0.007	0.006	0.025***	0.009	0.006	0.008
	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)	(0.009)
Respondent is a housewife	0.123**	0.148**	-0.040	0.121**	0.148**	0.116***
	(0.053)	(0.070)	(0.061)	(0.047)	(0.065)	(0.044)
Number of cooks in the	-0.024	-0.023	-0.011	-0.026	-0.023	-0.024
Household	(0.041)	(0.043)	(0.040)	(0.042)	(0.044)	(0.041)
Kitchen is located outside	0.007	0.013	-0.013	0.007	0.013	0.032
the dwelling space	(0.035)	(0.033)	(0.036)	(0.040)	(0.038)	(0.037)
Cumulative years of clean fuel	-0.001	-0.000	-0.027***	-0.003	-0.000	0.004*
Usage	(0.003)	(0.003)	(0.009)	(0.005)	(0.006)	(0.002)
Household owns a personal	-0.088*	-0.087	-0.117*	-0.098*	-0.087	-0.136**
Computer	(0.048)	(0.060)	(0.060)	(0.055)	(0.074)	(0.067)
Have an opportunity to get			0.173***			
fuel from neighbors, etc.			(0.033)			
First-stage residual ($\hat{u}_i$)				0.079	-0.002	
				(0.131)	(0.143)	
$Dirty_i$ × Age	No	Yes	No	No	Yes	No
$Dirty_i$ × Monthly income	No	Yes	No	No	Yes	No
Misclassification α_0	No	No	No	No	No	Yes
Misclassification α_1	No	No	No	No	No	Yes
Observations	588	588	588	588	588	588
Log Likelihood	-166.198	-163.988	-290.1	-166.017	-163.988	-160.836
AIC	356.397	355.976	604.2	358.034	357.976	349.673
BIC	408.917	417.250	656.7	414.932	423.627	410.947

Notes: This table reports the average marginal effects. The results for a constant term are not reported. Numbers in parentheses are delta-method standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A2. The instructions used in the elicitation of subjective beliefs

<table>
<tr><th colspan="6">Subjective Probability-related Information</th></tr>
<tr><td colspan="6">I will now ask you a few questions regarding the likelihood of the occurrence of the following events. There is no right or wrong answer. I just want to know what you think. There are 10 candies in front of you. One candy denotes one chance of the occurrence of any event out of 10. To express how likely you think it is that a specific event will occur, please choose and put aside some candies from the lot. If you put ZERO candies on the plate, this means that you are SURE that the event will NOT happen. As you ADD candies, this means you think that the LIKELIHOOD that the event will happen INCREASES. If you put one or two candies, it means that you think the event is unlikely to happen but is still possible. If you pick five candies, this means that it is just as likely to happen as it is likely not to happen. If you pick eight candies, this means that the event is more likely to happen than not to happen. If you put TEN candies on the plate, this means that you are SURE the event WILL HAPPEN.</td></tr>
<tr><td colspan="6">To the enumerator: If SCORE calculated from Q3a is > 0, go to 10. If the SCORE is 0, skip 10 and go to 11</td></tr>
<tr><td>10</td><td colspan="4">How likely do you think it is that exposure to smoke from burning cooking fuel caused your disease symptoms?</td><td></td></tr>
<tr><td colspan="6">To the enumerator: Please explain the health status definitions in section VA of <u>Note to the Enumerators.</u></td></tr>
<tr><td>11</td><td colspan="5">Consider a hypothetical individual who is identical to you. Imagine that there are options regarding the primary fuel for cooking. In each health status situation, please answer how likely you think it is that she will become/remain sick in the next 30 days if she used [fuels] in all the previous 30 days?</td></tr>
<tr><td colspan="6">To the enumerator: Please ask only regarding the likelihood of falling <u>Sick</u>. Please calculate 10 minus [candies for the likelihood of falling Sick] and confirm the likelihood of staying Healthy.</td></tr>
<tr><td colspan="2">Description of health status</td><td colspan="2">Case-I: She is <u>Healthy</u></td><td colspan="2">Case-II: She is <u>Sick</u></td></tr>
<tr><td colspan="2">Fuel used for cooking on all 30 days in the last month</td><td>LPG/Kerosene/ Electricity</td><td>Firewood/ Cow dung cakes/Coal</td><td>LPG/Kerosene/ Electricity</td><td>Firewood/Cow dung cakes/Coal</td></tr>
<tr><td>a</td><td>Sick</td><td></td><td></td><td></td><td></td></tr>
<tr><td>b = 10–a</td><td>Healthy</td><td></td><td></td><td></td><td></td></tr>
</table>

Notes: This is an English version of the subjective risk section in the second-round survey. See the Online Appendix for the full version of the questionnaire.