

Yilma, Zelalem; O'Donnell, Owen; Mebratie, Anagaw; Alemu, Getnet; Bedi, Arjun S.

Working Paper

Subjective Expectations of Medical Expenditures and Insurance in Rural Ethiopia

Tinbergen Institute Discussion Paper, No. 15-120/V

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Yilma, Zelalem; O'Donnell, Owen; Mebratie, Anagaw; Alemu, Getnet; Bedi, Arjun S. (2015) : Subjective Expectations of Medical Expenditures and Insurance in Rural Ethiopia, Tinbergen Institute Discussion Paper, No. 15-120/V, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/125126>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2015-120/V
Tinbergen Institute Discussion Paper



Subjective Expectations of Medical Expenditures and Insurance in Rural Ethiopia

*Zelalem Yilma*¹

Owen O'Donnell^{1,2}

*Anagaw Mebratie*¹

*Getnet Alemu*³

*Arjun S. Bedi*¹

¹ Erasmus University Rotterdam, the Netherlands;

² Tinbergen Institute, the Netherlands;

³ Addis Ababa University, Ethiopia.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Subjective Expectations of Medical Expenditures and Insurance in Rural Ethiopia[†]

Zelalem Yilma^a, Owen O'Donnell^b, Anagaw Mebratie^c, Getnet Alemu^d, Arjun S. Bedi^e

October, 2015

Abstract

Little is known about perceptions of medical expenditure risks despite their presumed relevance to health insurance demand. This paper reports on a unique elicitation of subjective probabilities of medical expenditures from rural Ethiopians who are offered the opportunity to purchase health insurance. We find that expectations are positively correlated with past expenses to a degree that exceeds the serial correlation in realized expenditures, suggesting overestimation of persistence and underestimation of the potential gains from insurance. Despite the fact that forecast expenditures do predict realized expenditures to some extent, there is no evidence that expectations influence the decision to take out health insurance, although plans to insure are positively related to the perceived dispersion of medical expenses.

Key words: Subjective probability, medical expenditure, out-of-pocket payments, adverse selection, health insurance, Ethiopia

JEL Classifications: D82, D84, I13, O12

[†] The authors acknowledge the financial support of the Netherlands Organisation for Scientific Research (NWO-WOTRO), grant number W07.45.103.00, the University of Goettingen and the Ethiopian Economics Association. ^a Erasmus University Rotterdam, debebe@iss.nl; ^b Erasmus University Rotterdam, Tinbergen Institute, and University of Macedonia, odonnell@ese.eur.nl; ^c Erasmus University Rotterdam, mebratie@iss.nl; ^d Addis Ababa University, galemu2001@yahoo.com; ^e Erasmus University Rotterdam.

1. Introduction

Reduction in exposure to medical expenditure risk underpins the case for social health insurance and is presumed to motivate enrolment in voluntary health insurance. Yet little is known of the incidence and magnitude of such risk in developing countries, and even less is known about the formation of risk perceptions. Cross-sectional measurement of inter-household variation in medical expenditures confounds uncertainty faced by each household with differences in spending across households that are predictable from characteristics not fully observable in the data but known to the household. Panel data of sufficient length to identify the stochastic properties of medical expenditures, which would feed into a measure of risk at the household level, are rare (Feenberg and Skinner 1994; French and Jones 2004); possibly even non-existent in a developing country context. Even if such data were available, using them to infer perceived medical expenditure risk would impose the assumption that expectations are formed rationally on the basis of all information available. The validity of this assumption has not been tested. In fact, very little is known about how individuals forecast their health expenditures and their ability to do so in any context (Breyer et al., 2012).

Improved knowledge of the formation of expectations regarding health care expenditure, perceptions of the associated financial risks and the degree to which the demand for insurance is related to such risks is essential for a better understanding of the functioning of insurance markets and for evaluation of social insurance programmes. Tests for adverse selection in health insurance based on the association between coverage and realized expenditures are plagued by the difficulty of disentangling selection from moral hazard (Chiappori, 2000). The relationship between forecast expenditure and enrolment potentially offers a more direct test (Hendren, 2013).

This paper reports on the elicitation of subjective probabilities of medical expenditures in rural Ethiopia. It examines the extent to which expectations are based on past realized expenditures and whether uptake of health insurance is related to both the mean and dispersion of the distribution of forecast medical expenses. If individuals base their expectations and insurance purchases on determinants of medical expenditures that the insurer is either not able to observe or, as in the Ethiopian programme, is prevented from utilising in setting the premium, then adverse selection will arise. While rational agents will behave in this way, it is by no means obvious that unsophisticated consumers, with little or no experience of health insurance, perhaps little appreciation of factors that raise medical spending and possibly limited ability to predict their expenses, will do so. Particularly, but not only, in low-income contexts, the questions of how expectations of medical expenses are formed and utilized are very much open. We address them directly.

Besides the relevance of our analysis to the operation of voluntary health insurance in low-income settings -- particularly in Ethiopia, where nationwide scale-up of the pilot scheme examined here is planned -- the paper makes a modest methodological contribution. The high degree of risk exposure in low-income, rural settings has prompted experimentation with the elicitation of subjective probabilities of economic outcomes (e.g. Delavande 2008; Bellemare 2009; Giné et al 2009; Santos and Barrett 2011; Delavande and Kohler 2012; McKenzie et al. 2013). The emerging consensus is that it is feasible to elicit informative expectations data from low-income, poorly educated populations (Attanasio 2009; Delavande et al. 2011; Delavande, 2014). To the best of our knowledge, this is the first study to elicit beliefs about future spending on health care in any setting -- high- or low-income. Respondents were asked to report probabilities that their spending on medical care would exceed certain thresholds. Clearly, the information content of such data depends on the ability of individuals with limited

education to understand relatively complex survey questions and to express beliefs consistent with the basic laws of probability. A central aim of the study is to establish whether it is possible to collect data on expectations that are valid in the sense of being logically consistent and plausibly informative of beliefs about future spending on health care.

We find that the bulk of the sample is able to respond sensibly to a seemingly abstract exercise of assessing the likelihood of future health scenarios and their associated medical expenses. Having verified the basic validity of the data, we make a distributional assumption that allows us to derive the first two moments of the distribution of future medical expenditure perceived by each household. Subsequently, we examine the information respondents utilise in forming expectations of medical expenditure, before evaluating the predictive value of expectations and, finally, assessing the extent to which health insurance enrolment is predicated on the expectation and perceived risk of medical expenses.

We find that the sample average of the second moment of each household's forecast medical expenditure is substantially smaller than the cross-sectional variance in realized medical expenditure, confirming that cross-sectional measures confound risk with predictable heterogeneity across households. Expected spending is positively related to past levels to an extent that exceeds the serial correlation in realized expenditure. This suggests that households underestimate volatility, which would reduce the perceived gains from insurance. Expected spending is positively correlated with realized spending in the period over which the forecast is made. After conditioning on observable covariates, expected expenditure still predicts realized spending (albeit weakly) suggesting that respondents hold some additional relevant information that is incorporated in their reported subjective probabilities. However, there is little or no evidence that expectations influence the decision

to take out health insurance, although plans to insure are positively related to the perceived dispersion of medical expenses.

The paper is organized as follows. Section 2 describes the sampling and the elicitation of subjective probabilities of health payments. Section 3 assesses the validity of the expectations data. In section 4, we compare moments of forecast medical expenditure with past and realized spending. Section 5 examines factors associated with the formation of beliefs concerning the mean and dispersion of future health spending. Section 6 assesses the extent to which expectations add value in predicting future spending. Section 7 examines whether the perceived distribution of expenditure is utilised in the decision to enrol in insurance. The final section offers concluding remarks.

2. Survey data and belief elicitation instrument

2.1 Sampling design

Three rounds of household panel data were collected in 16 rural districts located in four of the nine regions of Ethiopia (Tigray, Amhara, Oromiya, and SNNPR). These regions account for about 86 percent of the country's population (Population Census Commission, 2008). Within each district, the first survey round was fielded in six randomly chosen Kebeles (lowest administrative unit) in March-April 2011. In each of the 96 Kebeles, 17 households were randomly selected yielding a total of 1,632 households comprising 9,455 individuals. After the introduction of a government-sponsored voluntary health insurance scheme in 12 of the 16 districts in June-July 2011, two follow-up surveys of the same households, on which the bulk of the paper is based, were fielded in March-April 2012 (N=1599 households) and March-April 2013 (N=1583).

The sample households are predominantly engaged in agriculture (84% farmers), have very low levels of education (45% uneducated) and are very poor (23% qualify for a safety net food programme).¹

Within the 12 districts in which the insurance scheme operates, enrolment is voluntary at the household level. Premia vary across regions but not across households within regions.² The benefit package is comprehensive including both outpatient and inpatient care with very few exclusions. There are no co-payments provided higher levels of care are accessed through referral from a health center. Without a referral, households are liable for 50 percent of the cost of hospital treatment. By April 2012, the scheme had achieved an enrolment rate of around two-fifths of households residing in the targeted districts, and by April 2013 coverage had reached around one half.

2.2 Elicitation of expectations

A respondent, usually the head of the household or the spouse of the head, was asked about anticipated payments for health care and medicines over the next 12 months. This proceeds by first asking preliminary questions to fix the range of the distribution of future expenditure and then asking the respondent to report the probability of exceeding thresholds within this range (Dominitz and Manski 1997; Manski 2004; Attanasio and Augsburg 2012). The specific questions used to determine the minimum and maximum amounts were:

*Imagine that no member of your household contracts a **NEW serious illness or injury** in the next 12 months. In such a case what would be the **MINIMUM***

¹ The statistics quoted refer to the 2012 data and are percentages of heads of households. The safety net programme is the Productive Safety Net Programme (PSNP) that operates in food insecure districts and targets food insecure households in such districts. For further descriptive statistics, see Appendix Table A3.

² The premium ranges from 10 Birr (US\$ 0.6) to 15 Birr per household per month. In one region (Amhara), the premium varies with household size (3 Birr per individual per month but all individuals in the household must still be enrolled). The central government subsidizes a quarter of the premium while district and regional governments are expected to cover the costs of providing a fee waiver to the poorest 10 percent of the population.

amount of money your household would have to pay for health care and medicines (including transport costs) over the next 12 months?

*Imagine that at least one member of your household contracts a **NEW serious illness or injury** that requires treatment in a hospital in the next 12 months. In such a case what would be the **MAXIMUM** amount of money your household would have to pay for health care and medicines (including transport costs) over the next 12 months?*

The reference to a new illness is intended to ensure that long-standing health conditions that will continue to require treatment and medication are not overlooked in reporting the minimum and that the maximum refers to expenses that would be incurred were health to deteriorate markedly. The reference to hospital treatment in the maximum question is intended to prompt thought of the most expensive scenario.

After establishing the range, enumerators were instructed to compute three thresholds that divide the range into four equal intervals.³ Respondents were then asked:

*How likely is it that the amount your household will spend on health care and medicines (including transport costs) in the next 12 months will be greater than: [amount defined by each of three thresholds]
(see scale below which goes from 0 to 10 where 0 indicates no chance of happening and 10 indicates will definitely happen) [ruler shown].*

If the respondent was unsure but thought it was more likely that the household would spend more than the given threshold than not, then s/he was instructed to point somewhere between 0 and 10 but closer to 10 than 0 (and vice versa). Enumerators were instructed not to prompt for revision even if the responses did not seem sensible, e.g., because the reported likelihood of exceeding a higher threshold was greater than that reported for a lower threshold.

³ Enumerators were instructed to compute the three cut offs as: $A = \min + X$, $B = A + X$, $C = B + X$, where $X = (\max - \min) / 4$.

The expectations questions were asked immediately after questions about the affordability of insurance and after a series of questions about the incidence and economic consequences of illness and death in the household in the last 12 months. Respondents had therefore been primed to contemplate health-related expenses.

2.3 Medical expenditure data

Prior to the expectations module, another recorded health care utilization and expenditures incurred. For each episode of illness experienced by any household member in the previous two months, the respondent was asked to report the type and quantity of ambulatory health care received and the payments made for consultation and diagnostics, medicine, transport and other associated health care costs. Additionally, all hospitalizations of household members in the previous 12 months were reported along with the costs incurred. The information on actual medical expenditure was recorded in all rounds of the survey, while the expectations were elicited in the last two rounds.

3. Validity of the expectations data

Our first objective is to establish whether it is feasible to elicit informative data on expectations of medical expenditures in the context of a poor rural economy.

3.1 Response rates

Only 19 of 1,599 households surveyed in 2012 did not respond to all of the expectations questions, and there were only two such households in 2013 (Table 1). Most of these incomplete responses did not report the minimum and/or maximum (12), or gave zero values for both (7), and were not asked to report subjective probabilities.

Table 1: Expectations data: non-response, enumerator errors & illogical responses

	2012	2013
Total number of observations	1599	1583
Non response	19	2
Enumerator error in calculation of thresholds	229	88
Enumerator error resulted in wrong ordering of thresholds	38	30
Total usable observations	1542	1551
Illogical responses		
One violation of monotonicity	66	127
Two violations of monotonicity	65	100
All reported probabilities zero	46	28
<i>Observations with illogical responses</i>	<i>177</i>	<i>255</i>
Total observations with logical responses	1365	1303
	(85.4%)	(82.3%)

Notes: Observations are household respondents. Enumerator error refers to mistakes in computation of $k=(\max-\min)/4$, $A=\min+k$, $B=A+k$ or $C=B+k$. 27/32 (2012) and 33/57 (2013) enumerators made such errors. ‘...wrong ordering of thresholds’ refers to violation of $\max>C>B>A>\min$. Total usable observations = 1599-19-38 in 2012 and 1583-2-30 in 2013. Violation of monotonicity corresponds to e.g. $P(X>A)<P(X>B)$ with $A<B$.

A non-response rate of only 0.7% is reassuring, suggesting that the exercise is not too abstract. But mere participation is no guarantee that responses contain information on beliefs held with respect to future spending on health care. A basic requirement for the data to be informative is that the expenditure reported in the negative scenario -- onset of new serious illness/injury that required hospitalization -- should not be lower than the expenditure reported for the most positive scenario -- no onset of any new illness/injury. Enumerators were instructed to prompt respondents to revise the amounts reported if this occurred, which did happen relatively frequently. 364 households (23%) in 2012 and 314 households (20%) in 2013 were prompted to make a revision, suggesting a degree of difficulty in recognizing the brief scenarios as indicative of the least and most expensive outcomes with respect to medical

expenditures.⁴ Those prompted to revise their max/min values were more likely to report logically inconsistent probabilities (p value <0.001).

Enumerators made mistakes in calculating the intended evenly-spaced thresholds for 229 households (15%) in 2012 and 88 households (6%) in 2013 (Table 1). These errors do not emanate only from a few enumerators – in 2012 (2013) more than four-fifths (three-fifths) of the enumerators made at least one mistake. They are not fatal for the belief elicitation exercise provided that the thresholds calculated are in increasing order, even if they are not evenly spaced. This condition is violated only for 38 households in 2012 and 30 households in 2013.

In total, we lose 57 out of 1599 observations in 2012 and 32/1583 in 2013 due to non-response or calculation errors. Households dropped are smaller, poorer and more likely to be headed by a female, to forgo health care when sick and to have a member with some sort of disability (Appendix Table A1).

3.2 Illogical responses

Consistency of responses with the axioms of probability requires that the reported probabilities of incurring health expenditure (X) in excess of a series of increasing thresholds ($A < B < C$) satisfy monotonicity: $P(X > A) \geq P(X > B) \geq P(X > C)$. Responses that do not satisfy this condition are labelled illogical. Of the usable observations, 131 (8.5%) in 2012 and 227 (14.6%) in 2013 violated monotonicity at least once (Table 1). A further 46 and 28 in 2012 and 2013 respectively reported zero probability of spending more than all three thresholds, which we also consider an error since the same respondents report a maximum possible expenditure in excess of all three thresholds. In total, 177 respondents in 2012

⁴ For 37 (in 2012) and 24 (2013) households, information on whether they were prompted to revise their responses is missing.

(11.5%) and 255 in 2013 (16.4%) reported probabilities that are logically inconsistent. This is higher than the 4 percent rate of logical response errors found by Attanasio and Augsburg (2012) in their study of subjective income expectations in rural India using a similar instrument but is comparable to Dominitz and Manski's (1997) finding of a 10 percent error incidence for subjective expectations of income in Wisconsin.⁵

Similar to those excluded because of non-response or enumerator errors, households providing illogical responses are more likely to be poorer, smaller, less healthy and more likely to forgo health care for economic reasons (Appendix Table A1). They are also less likely to be enrolled in the health insurance scheme and to be engaged in agricultural activities as their main occupation.

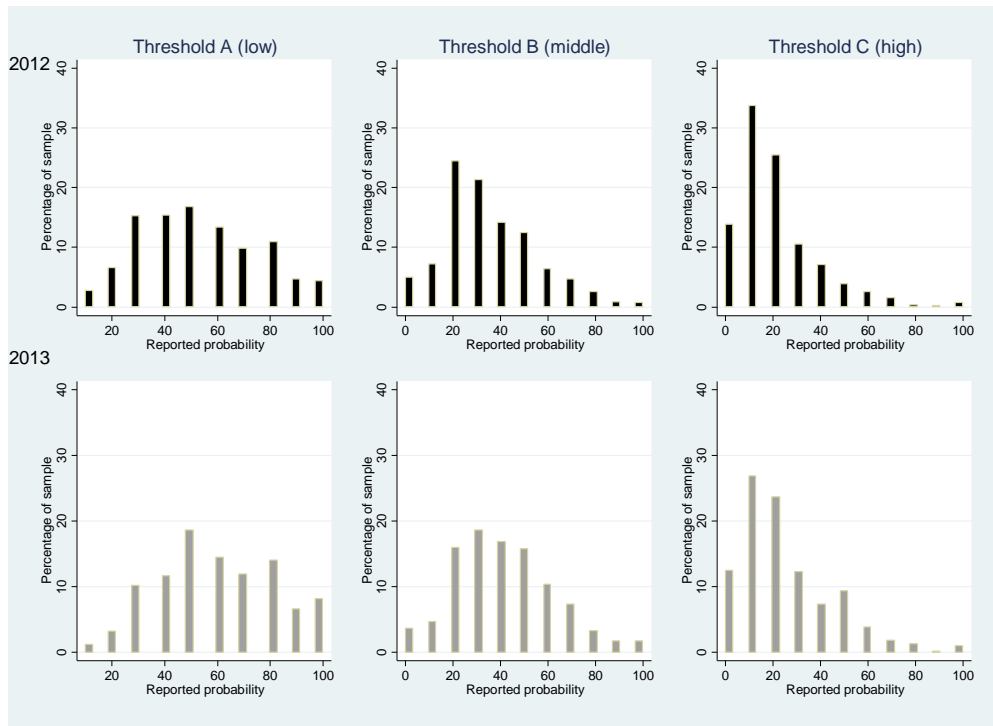
3.3 Distribution of responses

Distributions of the reported probabilities for each of the thresholds are presented in Figure 1. There is no bunching at focal responses of 0%, 50% or 100%, which is often a feature of subjective probability data (Kleinjans and van Soest 2014). The most likely explanation is our use of a 0-10 reporting scale (subsequently transformed), rather than 0-100. While the narrower scale will result in less variation, the lack of bunching suggests that the loss of information may not be so severe.

There is a clear shift in the mass of the distribution from right to left as the threshold is raised. While this is inevitable since observations that violate monotonicity have been dropped, the degree of the shift is indicative of respondents understanding the question and reporting substantially lower probabilities as the threshold is raised.

⁵ After prompting for revisions, the incidence of illogical responses dropped to 5 percent in Dominitz and Manski (1997). Since a principal aim of the current study is to examine the ability of low educated individuals to report subjective probabilities, we decided not to allow revision, except with respect to the reported minimum and maximum values.

Figure 1: Distributions of reported probabilities of medical spending exceeding various thresholds



Overall, the high response rate, the majority of usable, logically consistent responses (85 percent in 2012 and 82 percent in 2013), the lack of bunching and the anticipated shift in the mass of the distribution suggest that the bulk of the sample is able to respond sensibly to a seemingly abstract exercise of contemplating and assessing the likelihood of future health scenarios and their associated medical expenses. This lends support to interpretation of the responses of each household as points on its perceived cumulative distribution function of future medical expenditure.

4. Distributions and correlations of forecast and realized medical expenditure

4.1. Moments of forecast medical expenditure distributions

We assume that a piecewise uniform distribution provides a reasonable approximation to the underlying probability distribution for each household and use this, along with the elicited subjective probabilities, to compute the first and second moments of the distribution of health expenditure in the next 12 months perceived by each household.⁶ We do the same for the distribution of the logarithm of expenditures.

Table 2: Sample statistics of forecast and realized medical expenditure, 2012 (ETB)

	Cross-section statistics					
	Mean	Std. dev.	Min.	Median	Max.	N
<i>Forecast Expenditure (t+1)</i>						
Mean	461	562	2.40	305	6375	1365
Standard deviation	173	256	1.08	101	3320	1365
Coeff. of variation	0.35	0.18	0.02	0.34	1.08	1365
<i>Realized Expenditure (t)</i>						
Simple extrapolation	393	1441	0	0	19830	1365
Regression extrapolation	390	1047	~0	120	16476	1355
If realized expenditure >0						
<i>Mean forecast expenditure (t+1)</i>	520	641	21	326	6375	398
<i>Realized expenditure (t)</i>	1349	2418	6	480	19830	398

Notes: Simple and regression extrapolation of realized expenditures refer to method of estimating annual spending on outpatient care from reported expenditure in past two months, as explained in text. Sample size for regression extrapolation is slightly smaller due to missing values on covariates. ETB = Ethiopian Birr, US\$1=ETB 17.42 (April 2012)

The top panel of Table 2 provides sample summary statistics of parameters of the household-specific distribution of forecast medical expenditure derived from the probabilities reported in the 2012 survey (see Appendix Table A2 for 2013). The household-specific mean of forecast annual expenditure (i.e. the first moment of the subjective distribution) ranges from ETB 2.40 (US\$ 0.14) to ETB 6375 (US\$ 365.97), with a sample mean of ETB 461 (US\$ 26.41). Comparison with the distribution of realized expenditures in the *prior year* is complicated by the fact that spending on outpatient care is reported for a period of two

⁶ From only three subjective probabilities it is not possible to establish which type of distribution fits the data best (Attanasio and Augsburg, 2012).

months preceding the survey. We approximate annual expenditure by *simple extrapolation*, i.e. multiplying outpatient expenditure by six and adding the result to inpatient expenditure reported for the past twelve months. While this provides a credible estimate of mean annual expenditure across the sample, it will overestimate the inter-household variance. Therefore, we also predict outpatient expenditure in the last two months for each household from a Poisson regression of reported expenditure on covariates (Jones, 2011) and multiply this prediction by six before adding it to reported inpatient expenditure (referred to in Table 2 as *regression extrapolation*).⁷ This gives the same sample mean as the simple extrapolation method but a smaller variance, which in this case is an underestimate of the cross-sectional variance in annual expenses.⁸

The sample mean of the first moment of forecast expenditure is about one-sixth greater than the mean of realized expenditures. The discrepancy appears to be attributable to the large difference in the propensity of zero expenditures. Despite the large proportion of households with no reported spending on health care (71% by the simple extrapolation method), respondents seldom contemplate making no payments even when told to think of a scenario in which no new serious illness or injury is contracted and health spending is at a minimum.⁹ Restricting attention to households that incurred medical expenditures, the sample mean of these positive expenditures is 2.7 times the mean of the expected

⁷ Covariates (see Appendix Table A3) include household size and demographics, the head of household's education and occupation, measures of economic status, dummies for household level chronic health problems, paralysis, self-assessed health status, difficulty in hearing or speaking and a death in the household in the previous year and 96 village dummies. We deliberately do not include indicators of acute illness since the aim is to predict recurring expenditures based on stable characteristics.

⁸ The small difference in the two sample means is due to missing values of some of the covariates included in the regression model.

⁹ In 2012, only 3 respondents (of the 1599) stated a minimum health expenditure of zero. Admittedly, this may reflect a deficiency in the question, which did not emphasise that zero expenditure was a permitted response.

expenditures. This difference partly arises from the simple extrapolation of two-monthly expenditures on outpatient care, which will greatly overestimate annual expenditures for some households. The median of realized (positive) expenditure is closer to that of forecast expenditure, although the difference is still substantial.

It is important to emphasise that the sample distribution of expected expenditures should not resemble that of realized expenditures. The former is a distribution of expectations, while the latter is a distribution of stochastic outcomes. The range and standard deviation of realized expenditures should be, and are, much greater than the corresponding sample statistics for the expected expenditure.

The sample mean of the household-specific standard deviation of forecast expenditure is only one-sixth of the cross-section standard deviation of realized expenditure based on the regression extrapolation approach, and an even lower proportion of the estimate using simple extrapolation of outpatient expenditures. On average across the sample, the household-specific standard deviation is less than the respective mean (the coefficient of variation is a little more than a third, on average), while, as is common with cross-section data on health care spending (e.g. Van Doorslaer et al, 2007), the standard deviation across households is substantially larger than the sample mean. These comparisons suggest there is substantial overestimation of risk exposure using measures based on cross-sectional variance. Such measures confound risk with predictable heterogeneity across households. Measures of dispersion in the distribution of forecast expenditure do not. They capture perceptions of risk. The discrepancy between the two approaches would be substantially reduced by examining the cross-sectional variation in realized expenditures conditional on observable determinants (Flores and O'Donnell, 2013). But not all predictors that are known to the household are likely to be documented in the data, such that cross-

sectional variance (of residuals) is still likely to overestimate risk perceived by the household, and consequently, according to theory, its demand for single-period insurance.

4.2 Correlation between expected and past medical expenditure

The degree to which medical expenditures display persistence over time and the extent to which this is taken into account in the formation of expectations about future medical spending are of considerable importance to the operation of health insurance markets (Breyer et al. 2012).

The serial correlation between actual health expenditure in ETB money values incurred in consecutive years is small and not significantly different from zero for either pair of years (Table 3, top panel). The correlation is even lower for expenditure on inpatient care, but it is not significant for spending on outpatient care either. Serial correlation is more evident in the logarithm of expenditure (bottom panel), particularly for spending on outpatient care, which suggests that extremely high expenditures that are not repeated from year to year reduce the linear correlation in amounts.¹⁰ Still, the correlations are considerably smaller than those obtained for health care costs in Europe and the US.¹¹ A likely reason is that the higher burden of infectious disease and acute illness in Ethiopia reduces the serial correlation in medical expenditures relative to that observed in high-income, more elderly

¹⁰ Further evidence of this is that censoring expenses at the 99th percentile of the distribution of positive expenditures increases the magnitudes of the correlations of monetary amounts and results in those for total expenditure and outpatient expenditure in 2011-2012 becoming significant.

¹¹ For example, Di Nardi et al (2015) find a correlation coefficient of 0.57 for total health care costs in consecutive years for older persons in the US. Bakx et al (2015) report a correlation of 0.66 for total health expenditures calculated for the whole population of the Netherlands.

populations experiencing a greater relative burden of chronic illness.¹² The apparently high degree of volatility in medical expenses indicates substantial potential gains from risk pooling.

Table 4 shows correlations between expected and realized medical spending. Using actual expenditures reported in all three survey rounds and subjective probabilities reported in the latter two rounds, we can examine the correlation between expected expenditure for the year ahead and actual spending up to three years previous.¹³ All correlations are positive. The one between expectations and three year lagged actual spending is the smallest and is not significant. For both survey rounds in which expectations are reported, they are more closely correlated with spending over the last year than with spending in the year prior to the last (but this decline is statistically significant only for the 2013 survey). This suggests that expectations are based on more recent information. The positive correlations are not simply driven by the large proportion of households incurring no health expenditure expecting to spend little in the future. In fact, when attention is restricted to households with positive past levels of spending, the correlations increase (Table 4, bottom panel).

The significant positive correlation of expectations with past expenditure is in contrast to the absence of serial correlation in actual spending (at least when measured in money values, not logs). While we must acknowledge that our ability to estimate the latter is limited by the two-month recall period for outpatient expenses (see footnote 12), this discrepancy is at least indicative of adaptive expectations, rather than rational expectations, with the degree

¹² Restricting attention to the sample of households who reported chronic illness in both periods raises the correlation coefficient of total medical expenditure substantially to 0.062 for 2011-2012 and 0.097 for 2012-2013, although neither reaches significance. The low serial correlation will be partly due to spending on outpatient care being reported only for the last two months. Month-to-month fluctuations in spending will lower the correlation over two periods of two months separated by one year relative to the correlation of annual expenditures over consecutive years.

¹³ In the table, we show correlations with realized total expenditure computed by the simple extrapolation method (i.e. outpatient*6+inpatient). If instead with simply add two-monthly outpatient expenditure to inpatient expenditure, the correlations generally decline somewhat in magnitude but maintain significance.

of persistence being overestimated. However, expectations are not entirely inaccurate. There is a significant positive correlation between the expectation for the year ahead derived from the subjective probabilities reported in 2012 and the expenditure that materialised over the following year (Table 4, 1st column, 4th row). This suggests that households are, at least to some extent, able to predict their future spending on medical care.

Table 3. Correlation between realized medical expenditure in consecutive years

	2011-2012	2012-2013
ETB Amounts		
Total	0.033	0.020
Outpatient	0.030	0.015
Inpatient	0.001	0.007
Logs		
Total	0.122***	0.130***
Outpatient	0.129***	0.105***
Inpatient	0.058**	0.006
Number of observations	1365	1122

Notes: Total expenditure is computed by simple extrapolation method. Data in 2012 and 2013 are restricted to the sample with logical responses to expectations questions as detailed in Table 1. When taking logs, zero values are replaced by one half of the smallest positive value. ***, ** and * indicate significantly different from zero at 1%, 5% and 10% respectively.

Table 4. Correlations between expected and realized medical expenditure

Realized expenditure	Expenditure expected in t+1	
	t=2012	t=2013
t-2	N/A	0.043
t-1	0.128***	0.069**
t	0.136***	0.221***
t+1	0.138***	N/A
If realized exp. >0		
t-2	N/A	0.045
t-1	0.143***	0.089*
t	0.192***	0.334***
t+1	0.186***	N/A

Notes: Realized expenditure calculated by multiplying outpatient expenditure reported for last two months by six and adding result to inpatient expenditure reported for the last year. N/A - data on realized expenditure are not available. *** p<0.01, ** p<0.05, * p<0.1

5. Formation of medical expenditure expectations

5.1 Predictors of the household-specific mean

In Table 5, we present least squares regressions of the household-specific mean of log expected expenditure on past realized expenditures and other household characteristics potentially relevant to the formation of expectations.¹⁴ The regressions are estimated from the pooled 2012 and 2013 observations. The estimates may be interpreted as best linear predictions, revealing how expectations correlate with observable characteristics (Dominitz, 2001). Past outpatient and inpatient expenditures incurred in the year in which expectations were elicited as well as expenditures incurred in the previous year are entered separately because it is plausible that differences in the nature of health conditions that result in outpatient and inpatient expenditures lead to their differential weighting in the formation of expectations. The specification is consistent with adaptive expectations, which implies that expected spending is a weighted average of realized spending in previous periods (Nerlove, 1958), although we do not claim to be testing this hypothesis.¹⁵ Inclusion of the lagged value of outpatient expenditure is particularly appropriate because of the two-month reference period that results in a partial measure of the information on such spending the household may utilise to predict spending in the coming year. Besides past expenditures, the regressions include indicators of: i) health (illness, sensory impairment, disability, health self-assessed as poor/very poor and a recent death in the household); ii) socioeconomic status (wealth

¹⁴ Estimates from a least absolute deviation regression are generally consistent with those of least squares with respect to identification of characteristics that significantly predict the household-specific mean. See Appendix Table A4.

¹⁵ The analysis is intended to be descriptive with the purpose of assessing whether the subjective probabilities reported are sufficiently correlated with factors one would anticipate may be utilised in the formation of expectations such that there are grounds to interpret the data as expectations.

represented by quintile groups of a principal components score derived from assets and housing conditions, participation in a safety net programme, possession of a savings account, occupation and education); iii) health care coverage/access (health insurance enrolment, reported forgone care when sick and time to nearest health center); iv) reported experience of an economic/crime/conflict/natural shock in past year; and, v) household size and demographics. Definitions and means of the covariates are provided in the Appendix, Table A3.

The predictions account for around 30 percent of the variability in the household-specific means across the sample indicating a good deal of systematic variation. The reported expected medical expenditures are very far from being pure noise.

Inpatient expenditure in the last year and outpatient expenditure both in the last two months and in the two months preceding the previous survey are all statistically significant predictors of the mean of (log) forecast medical expenditure in the year to come. The latter is not significantly associated with expenses incurred on inpatient care in the year preceding the last. This may be because inpatient treatment is received for more acute conditions while outpatient expenditures include medication taken for more chronic conditions. Using the usual approximation, a 1% increase in inpatient expenditure over the last year is associated with a 0.14% increase in the expectation of total medical expenditure in the coming year. The elasticity with respect to outpatient expenditure in the last two months is smaller, which is to be expected given the discrepancy in the reference periods.¹⁶ But expectations are also positively related to spending on outpatient care in the two months prior to the last survey,

¹⁶ A one percent increase in outpatient expenses in the last two months is likely to imply a less than one percent increase in annual expenditure assuming that the increase is not repeated in every other two monthly period throughout the year.

which is consistent with respondents forming expectations on the basis of average spending over a number of periods.

Given the magnitude of expenditures on inpatient care, one expects that households can remember and report them accurately. Even if errors are made, provided the reported expenditures are those utilised in the formation of expectations of future spending, there will be no bias in the estimate of the association between expected and (perceived) past expenditure. However, expenditures on ambulatory care and medicines reported for the last two months are a noisy indicator of actual expenditures on these items in the last year. We deal with this by instrumenting household reported outpatient expenditure in the two months before each survey with respective mean outpatient expenditure in the Kebele in which the household is located. Two stage least squares coefficients reported in the second column of Table 5 show substantial increases in the coefficients on each of the two instrumented expenditure variables, which become comparable to the elasticity estimated for inpatient expenditure in the last year. Most of the coefficients of the other covariates remain similar in magnitude and significance to the OLS estimates.

Expected spending on medical care appears to be very closely correlated with ability to pay. On the basis of the OLS estimate, the wealthiest quintile group identified by the composite assets index spends two-thirds more on medical care in the coming year than the poorest fifth of households. Households in which the head has some form of education expect to spend 8.4% more than those without any education.

Enrolment in the health insurance scheme is positively but not significantly correlated with expected medical spending. The point estimate suggests that households that require more time to travel to a health center expect to spend more. Although this is (marginally)

significant only for OLS, it is consistent with the inclusion of transport costs in the expenses respondents are instructed to report.

Table 5. Pooled regression of household-specific mean of log forecast medical expenditure

	OLS		2SLS	
	Coeff.	Robust SE	Coeff.	Robust SE
Ln(outpatient expense) t	0.0225***	(0.00762)	0.101**	(0.0503)
Ln(outpatient expense) t-1	0.0418***	(0.00876)	0.130***	(0.0419)
Ln(inpatient expense) t	0.141***	(0.0219)	0.129***	(0.0251)
Ln(inpatient expense) t-1	0.00439	(0.0222)	-0.0126	(0.0244)
Someone in household has:				
Illness > 30 days	0.173**	(0.0754)	0.0585	(0.0948)
Sensory impairment	0.00896	(0.0613)	0.00626	(0.0617)
Paralysis/mobility issues	-0.0163	(0.0586)	-0.0463	(0.0628)
Poor/very poor health	0.0826	(0.0729)	-0.0264	(0.0828)
Died in last year	0.149	(0.129)	0.0967	(0.140)
Assets quintile groups (ref. is bottom)				
2 nd bottom	0.208***	(0.0592)	0.199***	(0.0621)
Middle	0.298***	(0.0594)	0.291***	(0.0606)
2 nd top	0.404***	(0.0635)	0.375***	(0.0690)
Top	0.669***	(0.0713)	0.621***	(0.0760)
Covered by safety net programme				
Has bank account	0.0916	(0.0677)	0.0799	(0.0680)
Non-agricultural employment	0.162	(0.129)	0.137	(0.126)
Educated head	0.0842**	(0.0389)	0.0725*	(0.0416)
Health insurance enrolled	0.0423	(0.0431)	0.0607	(0.0419)
Forgone care when sick	-0.116	(0.119)	-0.0911	(0.125)
Minutes to nearest health center	0.000778*	(0.00047)	0.000702	(0.00049)
Shock	0.106**	(0.0500)	0.103**	(0.0513)
Ln(household size)	0.000928	(0.0606)	-0.0635	(0.0697)
Year 2013	0.298***	(0.0690)	0.331***	(0.0647)
Constant	3.588***	(0.200)	3.670***	(0.208)
Observations	2,631		2,631	
R ²	0.298		0.241	
F-test of joint significance (p-values):				
All variables	0.000		0.000	
Health variables	0.055		0.833	
Age-sex composition	0.010		0.003	

Notes: Regressions also include the share of household member in ten gender specific age groups and district dummies. Standard errors in parentheses are corrected for clustering at Kebele level. *** p<0.01, ** p<0.05, * p<0.1

Of the health variables, only illness with symptoms that have lasted for at least one month is individually significantly correlated with the mean of log forecast medical expenditure. This variable loses significance when outpatient expenditure in the last two months is instrumented, which is consistent with both variables providing information on health that is utilised in the formation of expectations of medical expenses.¹⁷ The health variables are jointly significant when estimating by OLS but not by 2SLS. Households that reported having experienced a negative shock due to the local economy, weather, crime or conflict in the last year expect to incur greater expenditure on medical care in the coming year. Notwithstanding the fact that there has been some control for health, the most likely explanation would appear to be illness or injury as a result of such an event. The demographic composition of the household appears to matter. The coefficients are not presented but expected expenditure is lower in households that have a larger share of elderly (60+) females relative to all other demographic groups.

Regressions analogous to those presented in Table 5 reveal much less systematic variation in the standard deviation of log forecast expenditure (see Appendix Table A6). In particular, this measure does not appear to be associated with past realizations of expenditure or health suggesting that while the location of the subjective distribution shifts with these factors, its dispersion does not.

5.2 Revisions to expectations

¹⁷ If the health variables are excluded from the OLS regression, then the coefficients on the past expenditure variables increase (results available on request). This is further indication that both sets of variables are capturing the extent to which expectations are formed on the basis of health-related information.

The fact that the means of the distributions derived from reported subjective probabilities correlate with past realized expenditures, ability to pay proxies and health provides grounds to interpret these means as expectations of future spending on medical care. However, the associations could arise from confounding factors correlated with both the reporting of subjective probabilities and the regressors. We deal with time invariant confounders by estimating in first differences. If the associations are preserved, this would provide stronger evidence that the data contain information on expectations that are revised in response to changes in realized expenditures, economic circumstances and health.

The simple correlation between the change in expected expenditure and in realized outpatient expenditure is 0.061. The respective correlation with change in realized inpatient expenditure is 0.075. Both are statistically significant at 5% level. After controlling for covariates, the change in (log) expected medical expenditure continues to be positively and significantly associated with the change in realized outpatient and inpatient expenditures (Table 6). Expected spending does appear to be revised upward when actual spending in previous periods increases. The coefficient on both past outpatient expenditures increase somewhat relative to the levels estimates while that on inpatient expenditure in the last year falls. Inpatient expenditure in the previous to last year remains insignificant. Instrumenting the changes in outpatient spending with the Kebele specific mean changes results in increases in the outpatient coefficients. While most of the significant predictors in the models estimated in levels are no longer statistically significant, we continue to find that better-off households (in terms of asset quintiles and those with bank saving accounts) expect higher expenditure.¹⁸

¹⁸ The mean of log forecast expenditure is derived under the assumption that the distribution is piecewise uniform. This assumption can be avoided by examining variation in the log of the mid point of the reported minimum and maximum health expenditure.¹⁸ Regression estimates obtained using both levels and differences

Table 6: First difference regressions of household-specific means of log forecast medical expenditure

	OLS		2SLS	
	Coeff.	Robust SE	Coeff.	Robust SE
Ln(outpatient expense) t	0.0386**	(0.0148)	0.254***	(0.0778)
Ln(outpatient expense) t-1	0.0507***	(0.0155)	0.283***	(0.0891)
Ln(inpatient expense) t	0.115**	(0.0466)	0.120**	(0.0516)
Ln(inpatient expense) t-1	-0.0350	(0.0475)	-0.0586	(0.0570)
Someone in household has:				
Illness > 30 days	0.0490	(0.136)	0.0203	(0.139)
Sensory impairment	-0.0277	(0.0962)	-0.0445	(0.110)
Paralysis/ mobility problem	0.0674	(0.0863)	0.0477	(0.0872)
Poor/very poor health	0.141	(0.114)	0.00600	(0.122)
Died in last year	0.146	(0.238)	0.174	(0.273)
Assets quintile groups (ref. is bottom)				
2 nd bottom	-0.00513	(0.0955)	-0.0708	(0.0975)
Middle	0.0893	(0.115)	-0.00940	(0.129)
2 nd top	0.174	(0.120)	0.123	(0.132)
Top	0.562***	(0.148)	0.521***	(0.157)
Covered by safety net programme	0.176	(0.165)	0.252	(0.171)
Has bank account	0.186*	(0.111)	0.265**	(0.110)
Non-agricultural employment	0.182	(0.232)	0.0940	(0.244)
Educated head	0.137	(0.0849)	0.102	(0.0933)
Health insurance enrolled	0.0161	(0.0787)	0.0445	(0.0746)
Forgone care when sick	-0.114	(0.142)	-0.155	(0.174)
Minutes to nearest health center	0.000817	(0.00102)	0.000813	(0.00113)
Shock	0.0943	(0.0760)	0.107	(0.0789)
Ln(household size)	0.365	(0.248)	0.446	(0.273)
Constant	0.328***	(0.0687)	0.387***	(0.0616)
Observations	1,097		1,097	
R ²	0.079			
F-test of joint significance (p-values):				
All variables	0.000		0.000	
Health variables	0.636		0.977	
Age-sex composition	0.512		0.418	

Notes: Regressions also include the share of household member in gender specific age groups. Standard errors in parentheses are corrected for clustering at Kebele level. *** p<0.01, ** p<0.05, * p<0.1

of this measure as the dependent variable are similar to those presented in Tables 5 and 6 (see Appendix Table A5).

6. Predictive value of expectations

The regressions reported in the previous section reveal a good deal of systematic variation in expected expenditure with factors presumed relevant to future spending. But seventy percent of the variability in expectations remains unexplained. We now consider whether this unexplained variation is useful in predicting realized spending. If it is, then this would suggest that respondents hold information over and above that contained in covariates commonly available in surveys that enable them to form expectations that are somewhat accurate in predicting future medical expenses. And furthermore, that this private information is incorporated in reported subjective probabilities.

We employ a Generalised Linear Model (GLM) in which realized total medical expenditure is specified as an exponential function of covariates and estimate this by Gamma Pseudo Maximum Likelihood (GPML) (Gourieroux et al 1984; Manning and Mullahy 2001; Santos Silva and Tenreyro 2006). This is a commonly used estimator for medical expenditures (Jones 2011) and offers two main advantages in the present context. First, the estimator gives less weight to observations with a large conditional mean, which increases robustness to outliers that could arise from the potential measurement error in spending on outpatient care. Second, with a very large number of observations reporting no medical expenditure, substantial inconsistency could arise from adding an arbitrary constant to these zero values before taking logs and estimating in least squares (Santos Silva and Tenreyro 2006). GPML avoids this by taking the log of the conditional expectation and performs well even with a very large proportion of zero values in the sample (Santos Silva and Tenreyro 2011).

Modeling total medical expenditure in 2013 as a function of the mean of log forecast expenditure for that year without controls gives a significant GPML coefficient of 0.228 (SE=0.115).¹⁹ This is consistent with the significant correlation coefficient between expected and realized expenditure given in Table 4 (column 1, row 4) and implies that a one percent increase in expected expenditure is associated with an approximate increase of 0.23 percent in the spending that does actually materialize. So, expectations do have some predictive accuracy. Do they also have predictive value in addition to the forecast that can be made on the basis of observable covariates?

The first column of Table 7 gives GPML estimates from a model that includes the 2012 values (and the 2011 value of total expenditure) of the same covariates used to predict the mean of forecast medical expenditure in Table 5. Lagged values of the covariates are used since the objective is to test whether expectations have predictive power after conditioning on observable information available to be incorporated into those expectations. Both lags of past expenditure are positively and significantly associated with actual expenditure in 2013. So, the serial correlation in log expenditures shown in Table 3 remains after controlling for covariates. Medical expenditure is also significantly associated with past health and it is substantially greater for households that were poorest according to the assets index in 2012. This may reflect a higher (unmeasured) disease burden among the poorest. Strong dependence on ability to pay is reflected in the estimate of 86 percent lower spending by those who report forgoing health care in the past due to economic reasons.²⁰ A standard

¹⁹ Total realized expenditure is computed as outpatient expenditure in the last two months multiplied by six, plus inpatient expenditure in the last year. If we instead use total realized expenditure computed by simply adding outpatient expenditure in the last two months to inpatient expenditure in the last year, we find a significant GPML coefficient of 0.261 (SE=0.133). Coefficients in Table 7 generally increase in magnitude and are robust in terms of significance when this alternative calculation of total realized expenditure is used.

²⁰ $\exp(-1.995)-1=-0.864$.

deviation increase in travel time (=45 minutes) to the nearest health center is associated with a one-fifth increase in expenditure.²¹ This is presumably due to the cost of transport that is included in the measure of medical spending.

In the second column of Table 7, we add the mean of log forecast expenditure. Realized expenditure is positively related to the expectation. The point estimate suggests that a one percent increase in the expectation is associated with a 0.08 percent increase in actual expenditure. This coefficient is not significant using a z-test but a likelihood ratio test rejects the model that excludes the expectation (LR=4.2, p-value=0.0404). Further, in an alternative specification that uses the mean of forecast expenditure in money units, rather than the mean of the log, the expectation variable is significant (10%) even using a z-test.²²

The coefficients on the other covariates are generally robust to adding the expectations variable, while the coefficient on the latter falls by around two-thirds when covariates are added (from 0.228 to 0.0799). Hence, most of the predictive value of expected expenditure comes from processing information available in observable covariates. But respondents appear to have some residual information relevant to their future medical expenditure that they draw on in answering the subjective probability questions.

²¹ $(\exp(0.00454)-1)\times 45=0.1957$.

²² In the alternative specification, we also enter the lagged realized expenditures in levels rather than logs. As a further robustness check, we estimated by Poisson Pseudo Maximum Likelihood, which weights all observations equally and so may be more sensitive to outliers, rather than GPML. In this case, the mean of expected expenditure is significant at 1%. Results from these alternative specifications and estimators are available on request.

Table 7. Estimates of Generalised Linear Model of realized total medical expenditure in 2013 (Y_{t+1})

	w/o expectations		with expectations	
	Coeff.	Robust SE	Coeff.	Robust SE
$E_t[\ln Y_{t+1}]$			0.0799	(0.113)
$\ln Y_t$	0.119***	(0.0390)	0.117***	(0.0393)
$\ln Y_{t-1}$	0.0622*	(0.0333)	0.0604*	(0.0325)
Someone in the household has:				
Illness > 30 days	1.029***	(0.343)	1.033***	(0.343)
Sensory impairment	0.0293	(0.273)	0.0501	(0.273)
Paralysis/ mobility problem	0.144	(0.280)	0.127	(0.279)
Poor/very poor health	-0.342	(0.349)	-0.354	(0.348)
Died in last year	-0.867**	(0.365)	-0.875**	(0.362)
Assets quintile groups (ref. is bottom)				
2 nd bottom	-0.584**	(0.297)	-0.615**	(0.293)
Middle	-0.536*	(0.302)	-0.583**	(0.296)
2 nd top	-0.653**	(0.312)	-0.699**	(0.318)
Top	0.481	(0.352)	0.398	(0.357)
Covered by safety net programme	-0.0461	(0.380)	-0.0365	(0.382)
Has bank account	-0.00318	(0.261)	0.0416	(0.272)
Non-agricultural employment	-0.0949	(0.627)	-0.112	(0.621)
Educated head	-0.141	(0.232)	-0.144	(0.231)
Health insurance enrolled	0.109	(0.252)	0.0884	(0.255)
Forgone care when sick	-1.995***	(0.586)	-2.005***	(0.589)
Minutes to nearest health center	0.00434**	(0.00211)	0.00438**	(0.00210)
Shock	0.205	(0.186)	0.186	(0.187)
$\ln(\text{household size})$	0.685*	(0.359)	0.702*	(0.359)
Constant	2.612***	(0.947)	2.235**	(1.120)
Wald test of joint significance (p-values):				
Health variables	0.004		0.004	
District dummies	0.000		0.000	
Age-sex composition	0.000		0.000	
Log-likelihood	-8497.9		-8495.8	
Observations	1,330		1,330	

Notes: Table gives GPML estimates of total realized medical expenditure in 2013 ($t+1$) as an exponential function of covariates. $E_t[\ln Y_{t+1}]$ is the expectation at time t of log expenditure in $t+1$. All covariates are 2012 values, except that the 2011 value (in addition to the 2012 value) of total realized medical expenditure is included. Models also include the share of household member in gender specific age groups and district dummies. Standard errors in parentheses are corrected for clustering at Kebele level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

7. Do expectations influence the decision to insure?

If indeed respondents are able to use available information to anticipate, to some extent, their future medical expenses, then one might suppose that the information incorporated in the reported expectations will be utilised in deciding whether to take out health insurance. The regressions presented in Tables 5 and 6 show that expected medical expenditure is positively, but not significantly, associated with insurance enrolment. Even if one overlooks the lack of significance, it is not clear what to make of the fact that sample households with insurance report higher expected medical expenses.²³ Coverage would be expected to reduce OOP payments, although the scheme has been found to have no significant impact on households' medical spending while raising utilization of health services (Mebratie et al 2013). The positive correlation may therefore reflect selection – households with greater expected payments are more likely to be enrolled.²⁴ But testing this hypothesis requires examination of whether expectations subsequently influence the decision to enrol.

To do this, we restrict attention to districts in which the insurance scheme was offered and to households within those districts that had not yet taken the opportunity to enrol by the time of the 2012 survey when the expectations questions were first asked. For these households, we examine the relationship between the distribution of forecast medical expenditures for the year ahead and the propensity to take out coverage during the course of that year. Given the insurance premium does not depend on individual risk, those who

²³ The insurance scheme operates by issuing a card that entitles the holder to free care at contracted facilities. There is no payment and subsequent reimbursement. Hence, if a household takes its coverage into account in reporting expected expenditure, it is likely to be reporting OOP payments and not the gross value of the health care anticipated to be accessed through insurance.

²⁴ Since there are no copayments under the scheme, moral hazard (with a high price elasticity) would not result in increased payments unless increased access to low levels of treatment resulted in referral to higher levels of treatment with associated expenditures. But even if there were such an effect, anticipation of it would require very sophisticated agents, which villagers experiencing the first two years of operation of a health insurance scheme are unlikely to be.

report a higher expectation in 2012 have a greater incentive to enrol by 2013. Besides adverse selection on expected expenses, insurance is presumed to be motivated by its potential to reduce the variability of expenses. If households hold information on the degree of dispersion in their future medical expenses that they are able to express in the subjective probability questions and they do indeed seek to reduce this risk exposure, then, for a given degree of risk aversion, those who report a higher standard deviation of future medical expenditure in 2012 will be more likely to have enrolled by 2013.

In Table 8 we split households that were not enrolled in the insurance scheme in 2012 according to enrolment status in 2013. One quarter of households enrolled between the two surveys. For each group, we show in the top panel the sample means of the first two moments of the one-year-ahead medical expenditure distribution derived from the probabilities reported in 2012. Contrary to the hypotheses that enrolment is motivated by adverse selection and risk exposure, there is no difference in either the means or standard deviations by insurance status.²⁵ The bottom panel gives means of the parameters of the distributions reported in 2013. Those who enrolled between 2012 and 2013 reported subjective probabilities that imply a 38 percent higher mean and a 31 percent higher standard deviation of forecast expenditure than those who chose to remain without insurance. This is puzzling since coverage would be expected to reduce both the expectation and the dispersion of the distribution. It may be that changed circumstances, such as the onset of illness, caused households to revise their expectations upward and to insure.

²⁵ The expectations of medical expenditure reported in 2012 could possibly take into account plans to enrol in the insurance scheme in the coming year and the consequences this would have for OOP payments. To take account of this possibility, we have restricted the sample further to households not enrolled in 2012 and who declare at that point that they have no plans to enrol. This does not change the conclusion (from both bivariate and multivariate analyses) of there being no significant difference in expectations between those who do and do not subsequently enrol.

Table 8: Sample averages of mean and standard deviation of future medical expenditure by entry to insurance -- households not enrolled in 2012

	Insurance status in 2013		p-value of test equality of means
	Enrolled	Not enrolled	
<i>Expectations reported in 2012</i>			
Mean	404	376	0.488
Standard deviation	154	153	0.958
<i>Expectations reported in 2013</i>			
Mean	632	462	0.013
Standard deviation	247	169	0.020
No. of observations	127	353	

Notes: Sample restricted to districts in which insurance is offered and to households who had not enrolled by 2012. Only households who gave logical responses in both years are reported.

As is clear from the probit marginal effects estimates given in the first column of Table 9, conditioning on covariates does not change the conclusion that the probability of enrolling in the insurance scheme between 2013 and 2012 does not vary with either the mean or the standard deviation of (log) forecast medical expenditures reported in 2012.²⁶ Past medical expenditure and health variables are excluded from the model since these may provide information that is incorporated into expectations of future expenses that subsequently influence insurance enrolment.²⁷ Households in higher wealth quintile groups are more likely to insure, as are those living close to health facilities offering better quality care.²⁸

Table 9: Probit estimates of probability of actual and planned insurance enrolment for households not initially enrolled (marginal effects)

	Actually enrolled by 2013	Plan to enroll
--	---------------------------	----------------

²⁶ The mean and standard deviation of expected expenditure are highly correlated ($\rho=0.788$), which might be considered an explanation for neither being individually significant. When the model is re-estimated first including only the mean (and covariates) and then only the standard deviation, neither is significant.

²⁷ In fact, including lagged realized medical expenditure and health indicators has little or no influence on the magnitude and significance of the coefficients on the expectations variables.

²⁸ The perceived quality of care is reported by the head of the nearest health facility. The specific question is: "Do you think this health center is providing the expected standard of health care services?" This was asked in a survey of 48 health centers (3 from each of the 16 districts) conducted in April-May 2011 (Mebratie et al. 2015).

	Marginal effect	Robust SE	Marginal effect	Robust SE
Mean log forecast medical exp.	0.0254	(0.0235)	-0.00761	(0.0193)
Standard deviation log forecast medical exp.	-0.127	(0.0777)	0.323***	(0.0854)
Assets quintile groups (ref. is bottom)				
2 nd bottom	0.0789	(0.0713)	0.135***	(0.0470)
Middle	0.165**	(0.0715)	0.168***	(0.0531)
2 nd top	0.153**	(0.0770)	0.162**	(0.0638)
Top	0.159*	(0.0814)	0.0860	(0.0701)
Covered by safety net programme	0.0637	(0.0582)	0.0286	(0.0542)
Has bank account	0.0892	(0.0738)	-0.0847	(0.0625)
Educated head	0.0386	(0.0347)	-0.00759	(0.0351)
Head of hhold holds official position	0.00144	(0.0486)	0.0657*	(0.0392)
Forgone care when sick	-0.0358	(0.109)	0.0738	(0.0858)
Minutes to nearest health center	0.000290	(0.000443)	-0.000269	(0.000452)
Quality of care at nearest health center	0.142***	(0.0542)	-0.100	(0.0663)
Ln(household size)	0.00252	(0.0519)	0.220***	(0.0598)
Muslim	0.0335	(0.0761)	-0.159**	(0.0626)
Tigray region	0.297***	(0.113)	0.223**	(0.0943)
Amhara region	0.220**	(0.104)	0.0288	(0.102)
Oromiya region	0.0619	(0.0985)	0.288***	(0.0908)
Year dummy			-0.0304	(0.0393)
Observations	592		1,068	
Pseudo R ²	0.113		0.108	

Notes: The sample for the left-hand column includes all households that were not enrolled in 2012. The dependent variable is an indicator of whether the household had enrolled by 2013. The sample for the right-hand column includes all households that were not enrolled in 2012 and/or 2013. The respective dependent variable is an indicator of whether the respondent reported that the household planned to enrol. The models also include the share of household members in gender specific age groups. Standard errors in parentheses are corrected for clustering at Kebele level. *** p<0.01, ** p<0.05, * p<0.1

Ability to pay appears to be a strong constraint on enrolment. Given this, it could be that households form plans to enrol based on expectations of medical expenses that they are not able to realise. To examine this hypothesis, we restrict the sample to households that were not enrolled in 2012 and/or 2013 and model their declared intention to enrol as a function of expectations and covariates. The probit marginal effects presented in the right-hand column of Table 9 reveal that while the likelihood of planning to enrol does not vary with the mean of (log) forecast expenditure, it rises positively and significantly with the

standard deviation.²⁹ This is consistent with perceived high volatility of medical expenses motivating plans to insure, even if those plans are not ultimately implemented. Failure to fulfill plans may be due to time-inconsistency, although it is not necessarily indicative of this (Giné et al. 2014, Halevy 2015).³⁰

All in all, despite the fact that expectations do appear to be formed on the basis of relevant information, including past medical expenses, and are to a limited extent predictive of future expenses, there is no strong evidence that these expectations are used in the decision to take out health insurance. Greater perceived risk of medical expenses may raise the willingness to insure, if not actual enrolment.

8. Conclusion

Our unique elicitation of subjective probabilities of medical expenditures reveals that the majority of the rural Ethiopians sampled are able to provide logically consistent responses, in the sense of satisfying monotonicity, that correlate with past expenses and predict future spending. This positive finding echoes that of other exercises in the elicitation of probabilistic expectations of various outcomes conducted in developing countries (Attanasio 2009; Delavande et al. 2011; Delavande, 2014). It suggests that measurement of household-specific distributions of future expenditure offers a feasible alternative to reliance on the cross-sectional variance of medical expenses that, even after conditioning on covariates, is likely to substantially overstate the risk faced by any one household. But households do make

²⁹ A simple test of differences in sample means between those who plan to enrol and those who do not shows no significant difference in the expectation of future spending (ETB 437 vs ETB 412) but a significantly higher standard deviation for those planning to enrol (ETB 177 vs ETB 141).

³⁰ Of the households not enrolled in 2012, 59% (354/604) report that they plan to enrol in the insurance scheme. Of those households, only 26% (91/354) had actually enrolled by 2013. Both the mean and standard deviation of forecast expenditure reported in 2012 are slightly higher for those who do realise their plan to enrol (mean: 414 vs 381, standard deviation: 164 vs 153) but neither difference is significant.

mistakes in assessing their risk exposure. We find that the strength of the relationship between expected and past medical spending exceeds the degree of persistence in actual expenditure. Previously high spending households may therefore overestimate the extent to which they will gain from insurance. If these households were to respond to this misperception, then the selection would not be as adverse as would be presumed on the basis of the correlation between enrolment and past expenses. This is not to say that there is no scope for adverse selection. To a limited extent, expected expenditure predicts realized spending even after conditioning on covariates. But we find no evidence that households act on this private information. Expectations do not appear to influence the decision to take out health insurance. This is somewhat at odds with the literature, although evidence of behaviour responding to health expectations captured by subjective probabilities mostly comes from one country (Malawi) and one health condition (HIV/AIDS) (Delavande and Kohler 2012; Shapira 2013; de Paula et al. 2014). The finding that plans to insure, if not actual enrolment, are positively related to the perceived dispersion of medical expenses suggests that the desire for insurance is (partly) driven by risk exposure. We can only speculate on why this motivation does not become effective demand. It could be lack of ability to pay, which does appear to be a strong determinant of insurance enrolment. Households living close to subsistence may be particularly prone to time inconsistency (Giné et al. 2014). From a distance, enrolment in health insurance to relieve stress arising from volatile medical expenses can seem attractive. But when the time comes to pay the premium, the urgency of other needs may take precedence over a payment to cover medical care that may not even be needed.

To the best of our knowledge, this is the first study to elicit beliefs about future spending on health care. As such, it inevitably suffers from limitations that future research

should take care not to repeat. First, a feature of medical expenditure data is that many households spend nothing. Our instrument did not explicitly offer zero expenditure as a possible outcome. It may be preferable to first ask about the probability of spending anything at all on health care, and then the probabilities of spending within categories over the range of positive amounts. Second, the instrument asked about expectations of medical expenditure over one year, while actual expenditure on ambulatory care was recorded for the past two months. This inconsistency, which hampers comparison of the levels of expected and realized expenditure but is less of an impediment to correlation analysis, is not easy to resolve since extension of the recall period for ambulatory care is likely to increase measurement error. Third, the finding that expectations do not appear to influence the decision to insure may partly be due to the fact that two-fifths of households offered insurance had already enrolled by the time the first expectations data were collected. It would be preferable to collect expectations data from the time at which insurance is first offered. Notwithstanding these limitations, the findings from this study are sufficiently encouraging to warrant further research into how expectations of medical expenditures are formed and utilized in a number of contexts.

References

- Attanasio, O. (2009). "Expectations and Perceptions in Developing Countries: Their Measurement and Their Use" *American Economic Review: papers and proceedings* 99(2): 87-92.
- Attanasio, O. and Augsburg, B. (2012). "Subjective Expectations and Income Processes in Rural India." Universtiy College London. Mimeo.
- Bakx, P., O'Donnell, O. and van Doorslaer (2015). "Spending on healthcare in the Netherlands: Not going so Dutch?" Erasmus University Rotterdam, mimeo.
- Bellemare M. (2009). "When perception is reality: subjective expectations and contracting". *American Journal of Agricultural Economics* 91(5): 1377–81.
- Breyer, F., Bundorf, M.K. and Pauly, M.V. (2012) Health Care Spending Risk, Health Insurance, and Payment to Health Plans in Pauly, M.V., McGuire, T.G. and Barros, P.P. (Editors) *Handbook of Health Economics, Volume 2, Chapter 11: 691-762, Elsevier, Amsterdam*
- Chiappori, P. A. (2000). "Econometric models of insurance under asymmetric information". In Dionne, G. (Ed.), *Handbook of insurance*, pp. 363-392. Kluwer Academic Publishers, Norwell, MA.
- Delavande A. (2008). "Pill, Patch, or Shot? Subjective Expecations and Birth Control Choice". *International Economic Review* 49 (3): 999-1042.
- Delavande, A. (2014). "Probabilistic expectations in developing countries". *Annual Review of Economics*, 6(1): 1-20.
- Delavande, A., Gine, X. and McKenzie, D. (2011). "Measuring subjective expectations in developing countries: A critical review and new evidence". *Journal of Development Economics* 94(2): 151-163.
- Delavande, A. and Kohler, H-P. (2012). "The Impact of HIV Testing on Subjective Expectations and Risky Behavior in Malawi". *Demography* 49(3): 1011-1036.
- de Paula A, Shapira G, Todd P. (2014). "How Beliefs About HIV Affect Risky Behaviors: Evidence From Malawi". *Journal of Applied Econometrics* 29(6): 944-964
- Di Nardi, M.C., French, E., Bailey Jones, J. and McCauley, J. (2015). "Medical spending of the US elderly" NBER Working Paper 21270, Cambridge, MA: NBER.
- Dominitz, J. (2001). "Estimation of Income Expectations Models Using Expectations and Realization Data" *Journal of Econometrics* 102(2): 165-195
- Dominitz, J. and Manski, C.F. (1997). "Using Expectations Data to Study Subjective Income Expectations". *Journal of the American Statistical Association* 92(439):855-867
- Feenberg, D. and Skinner, J. (1994). "The Risk and Duration of Catastrophic Health Care Expenditures." *The Review of Economics and Statistics* 76(4): 633-647.

- Flores, G. and O'Donnell, O. (2013). "Catastrophic Medical Expenditure Risk" Tinbergen Institute Discussion Paper TI 2012-078/3, Amsterdam: Tinbergen Institute
- French, E. and Jones, J. B. (2004). "On the Distribution and Dynamics of Health Care Costs." *Journal of Applied Econometrics* 19(6): 705-721.
- Giné X, Townsend R, Vickery J. (2009). "Forecasting when it matters: evidence from semi-arid India". World Bank Working Paper Washington, DC: World Bank.
- Giné X., Goldberg, J., Silverman, D., Yang, D. (2014). "Revising Commitments: Field Evidence on the Adjustment of Prior Choices". Washington D.C.: World Bank.
- Gourieroux, C., Monfort, A., Trognon, A. (1984). "Pseudo maximum likelihood methods: applications to Poisson models" *Econometrica* 52: 701-720.
- Halevy, Y. (2015). "Time consistency: Stationarity and time invariance". *Econometrica* 83(1): 335-352.
- Hendren, N. (2013). "Private information and insurance rejection". *Econometrica* 81(5): 1713-1762.
- Jones, A.M. (2011) "Models for health care", in Hendy, D. and Clements, M. (eds.) Oxford Handbook of Economic Forecasting, Oxford: Oxford University Press,
- Kleinjans, K.J., van Soest A. (2014). "Rounding, focal point answers and nonresponse to subjective probability questions". *Journal of Applied Econometrics* 29(4): 567-585
- Kowalski, A.E. (forthcoming). "Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care" *Journal of Business and Economic Statistics*
- Kowalski, A.E. (2015). "What Do Longitudinal Data On Millions of Hospital Visits Tell Us About The Value of Public Health Insurance As A Safety Net For The Young And Privately Insured?" NBER WP20887, Cambridge, MA: NBER
- Manning, W.G., Mullahy, J. (2001). "Estimating log models: to transform or not to transform?" *Journal of Health Economics* 20 (4), 461-494.
- Manski, C.F. (2004). "Measuring Expectations". *Econometrica* 72(5): 1329-1376
- Mckenzie, D., Gibson, J. and Stillman, S. (2013). "A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad" *Journal of Development Economics*, 102: 116-127
- Mebratie, A.D., Sparrow, R., Yilma, Z., Abebaw, D., Alemu, G. and Bedi, A. (2013). "Impact of Ethiopian Pilot Community Based Health Insurance Scheme on Health Care Utilization: A household Panel Data Analysis" *The Lancet*, 381: S92
- Mebratie, A.D., Sparrow, R., Yilma, Z., Alemu, G. and Bedi, A. (2015). "Enrolment in Ethiopia's Community Based Health Insurance Scheme" *World Development*, 74: 58-76
- Nerlove, M. (1958). "Adaptive expectations and cobweb phenomena" *Quarterly Journal of Economics*, 72(2): 227-240.

- Parmar, D., Soares, A., de Allegri, M., Savadogo, G., and Sauerborn, R. (2012). "Adverse Selection in a community-based health insurance in rural Africa: Implications for introducing targeted subsidies" *BMC Health Services Research* 12 (1):181
- Population Census Commission, Federal Democratic Republic of Ethiopia (2008) "Summary and Statistical Report of the 2007 Population and Housing Census" Addis Ababa, Ethiopia.
- Preker, A.S. and Carrin, G. (2004). Health Financing for Poor People: Resource Mobilization and Risk Sharing. World Bank, Washington DC.
- Santos, P. and Barrett, C. (2011). "Persistent poverty and informal credit" *Journal of Development Economics*, 96(2): 337–47
- Santos Silva, J.M.C., Tenreyro, S. (2006). "The log of gravity", *Review of Economics and Statistics*, 88 (4), 641-658.
- Santos Silva, J.M.C., Tenreyro, S. (2011). "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator", *Economics Letters* 112 (2), 220-222.
- Shapira G. (2013). "How subjective beliefs about HIV infection affect life-cycle fertility: evidence from rural Malawi". Policy Research Working Paper 6343, World Bank, Washington DC.
- Van Doorslaer, E., O'Donnell et al (2007) "Catastrophic payments for health care in Asia", *Health Economics*, 16 (11): 1159-1184.
- Wang, H., Zhang, L., Yip, W. And Hsiao, W. (2006). "Adverse Selection in a Voluntary Rural Mutual Health Care Health Insurance Scheme in China." *Social Science and Medicine* 63 (5): 1236-1245.
- Yilma, Z., Mebratie, A., Sparrow, R., Dekker, M., Alemu, G., and Bedi, A.S. (2015) "Impact of Ethiopia's Community Based Health Insurance on household economic welfare". *World Bank Economic Review*. Doi:10.1093/wber/lhv009

Appendix

Table A1: Means of medical expenditures and selected covariates by subjective probability non-response, enumeration error and violation of logical consistency

	Excluded due to non-response / enumeration error			Subjective probabilities exhibit logical inconsistencies		
	Yes (N=89)	No (N=3093)	No difference p-value	Yes (N=432)	No (N=2668)	No difference p-value
Health insurance enrolled (%)	27.0	33.6	0.189	19.9	35.8	0.000
Outpatient expense (ETB)	42.1	53.7	0.648	50.8	54.1	0.788
Inpatient expense (ETB)	10.7	54.1	0.416	46.7	55.2	0.745
<i>Health measures</i>						
Illness > 30 days	12.4%	11.9%	0.895	13.9%	11.6%	0.177
Sensory impairment	18.0%	15.5%	0.523	18.1%	15.2%	0.126
Paralysis/mobility problem	24.7%	17.7%	0.088	21.8%	17.1%	0.019
Death in last year	2.2%	1.9%	0.818	1.2%	2.0%	0.222
Poor/very poor health	19.1%	14.8%	0.267	13.4%	15.1%	0.373
<i>Measures of socio-economic status</i>						
Poorest assets quintile group	36.0%	19.6%	0.000	30.6%	17.8%	0.000
Assets quintile 2	13.5%	20.2%	0.118	15.7%	20.9%	0.014
Assets quintile 3	21.4%	20.0%	0.745	21.1%	19.8%	0.527
Assets quintile 4	12.4%	20.2%	0.067	18.3%	20.6%	0.265
Richest assets quintile group	16.9%	20.1%	0.458	14.4%	20.9%	0.002
Has bank account	5.6%	12.1%	0.062	9.7%	12.5%	0.099
Covered by safety net programme	25.3%	20.5%	0.282	24.1%	20.0%	0.052
Forgone care when sick	6.7%	2.5%	0.013	3.7%	2.3%	0.089
Minutes to nearest health center	57.4	59.9	0.560	52.4	61.1	0.000
Educated head	46.6%	54.1%	0.162	54.4%	54.1%	0.884
Non-agricultural employment	2.3%	3.8%	0.461	5.6%	3.5%	0.036
Shock	39.3%	47.6%	0.125	48.6%	47.3%	0.623
<i>Demographic variables</i>						
Ln(household size)	1.458	1.666	0.000	1.616	1.674	0.013
Muslim	15.9%	26.9%	0.021	20.1%	28.0%	0.001

Note: For definitions of enumeration error and logical inconsistencies in subjective probability responses, see notes to Table 1. Definitions of the variables are provided in Table A3.

Table A2: Sample statistics of forecast and realized medical expenditure, 2013 (ETB)

	Cross-section statistics					Obs.
	Mean	Std. dev.	Min.	Median	Max.	
<i>Forecast Expenditure (t+1)</i>						
Mean	674	830	16.00	390	10325	1303
Standard deviation	255	427	3.14	123	8443	1303
Coeff. of variation	0.36	0.18	0.03	0.34	0.96	1303
<i>Realized Expenditure (t)</i>						
Simple extrapolation	365	1496	0	0	34800	1303
Regression extrapolation	367	1095	~0	120	25945	1287
If realized expenditure >0						
<i>Mean forecast expenditure (t+1)</i>						
	814	923	28	478	7038	362
<i>Realized expenditure (t)</i>						
	1315	2611	12	510	34800	362

Notes: Simple and regression extrapolation of realized expenditures refer to method of estimating annual spending on outpatient care from reported expenditure in past two months, as explained in text. Sample size for regression extrapolation is slightly smaller due to missing values on covariates. ETB = Ethiopian Birr, US\$1=ETB 18.45 (April 2013)

Table A3: Descriptions and summary statistics of variables

Variable	Description	2012 (N=1365)		2013 (N=1303)	
		Mean	Std. Dev.	Mean	Std. Dev.
Health insurance enrolled	Enrolled in community based health insurance	33.0%		38.7%	
Outpatient expense	Out-of-pocket outpatient payment (past 2 months)	56.5	217.3	51.5	233.1
Inpatient expense	Out-of-pocket inpatient payment (past 12 months)	54.3	541.1	56.1	503.7
Ln (average of Min & Max forecast OOP)	Log of simple average of minimum and maximum forecast health expenditure	5.8	1.0	6.1	1.1
$E_t[\ln Y_{t+1}]$	Expectation of log forecast health expenditure	5.6	0.9	5.8	1.1
Illness > 30 days	At least one member with chronic illness (symptoms stayed more than 30 days)	10.6%		12.7%	
Sensory impairment	At least one member has difficulty to hear/speak/ or see	12.6%		17.9%	
Paralysis/mobility problem	At least one member has some sort of paralysis or difficulty to stand up after sitting down	13.6%		20.8%	
Death in last year	Household member died in the past 12 months	2.6%		1.5%	
Poor/very poor health	At least one member has poor/very poor self-assessed health	12.5%		17.8%	
Bottom assets quintile group	Bottom 20% of full sample on principal components score based on possession of assets and housing conditions	18.6%		17.0%	
2 nd bottom assets quintile group	Assets index in 20-40% of full sample	20.5%		21.3%	
Middle assets quintile group	Assets index in 40-60% of full sample	19.4%		20.1%	
2 nd top assets quintile group	Assets index in 60-80% of full sample	20.5%		20.7%	
Top assets quintile group	Assets index in top 20% of full sample	21.0%		20.9%	
Has bank account	Household has saving bank account	11.6%		13.4%	
Covered by safety net programme	Household member of productive safety net program (PSNP), yes=1, no=0 (a targeted program in food insecure Woredas)	20.5%		19.5%	

Continued on next page

Table A3 (continued): Descriptions and summary statistics of variables

Variable	Description	2012 (N=1365)		2013 (N=1303)	
		Mean	Std. Dev.	Mean	Std. Dev.
Forgone care when sick	Someone was ill in last two months but did not receive treatment because of one of the following: a) health care/medicines are too expensive b) health facilities are too far c) could not take time off work / lose income	2.5%		2.1%	
Minutes to nearest health center	Travel time to nearest health center (minutes)	62.8	41.3	59.3	39.9
Educated head	Househd head has at least informal education	56.4%		51.6%	
Non-agri employment	Household head's main occupation: non-agricultural employment	2.9%		4.1%	
Ln(househod size)	Log household size	1.684	0.446	1.664	0.455
Muslim	Religion of the household head is Muslim	27.1%		29.0%	
Shock	Household reported experiencing shock in last 12 months arising from crime/conflict (divorce, land / water conflict, theft of crops, theft of livestock), economy (decline in price of output, unemployment, loss of equipment, death of livestock) or nature (flood, storm, fire, drought, untimely rain, insect damage).	44.2%		50.7%	

Notes: Standard deviations not given for binary variables. Regression models also include the share of household members in ten age-sex groups.

Table A4. Least absolute deviation regression of the Mean of Log Forecast Medical Expenditure

	LAD	
	Coeff.	Robust SE
Ln(outpatient expense) t	0.0170*	(0.00986)
Ln(outpatient expense) t-1	0.0259**	(0.0110)
Ln(inpatient expense) t	0.142***	(0.0305)
Ln(inpatient expense) t-1	0.0159	(0.0299)
Someone in household has:		
Illness > 30 days	0.228**	(0.0886)
Sensory impairment	0.0279	(0.0775)
Paralysis/mobility issues	-0.0547	(0.0741)
Poor/very poor health	0.118	(0.0731)
Died in last year	0.0583	(0.193)
Assets quintile groups (ref. is bottom)		
2 nd bottom	0.153*	(0.0893)
Middle	0.269***	(0.0965)
2 nd top	0.287***	(0.0909)
Top	0.531***	(0.102)
Covered by safety net programme	-0.154**	(0.0635)
Has bank account	0.103	(0.0664)
Non-agricultural employment	0.212	(0.149)
Educated head	0.115**	(0.0511)
Health insurance enrolled	0.0738	(0.0514)
Forgone care when sick	-0.133	(0.153)
Minutes to nearest health center	0.000766	(0.00067)
Shock	0.151***	(0.0474)
Ln(household size)	0.0870	(0.0935)
Year 2013	0.234***	(0.0442)
Constant	3.776***	(0.267)
Observations	2,631	
R ²	0.159	
F-test of joint significance (p-values):		
All variables	0.000	
Health variables	0.010	
Age-sex composition	0.036	

Notes: Regressions also include the share of household member in ten gender specific age groups and district dummies. Standard errors in parentheses are corrected for clustering at Kebele level. *** p<0.01, ** p<0.05, * p<0.1

Table A5. Regressions presented in Tables 5 and 6 with dependent variable replaced by log of simple average of minimum and maximum forecast medical expenditure

VARIABLES	Levels		First difference	
	OLS	2SLS	OLS	2SLS
Ln(outpatient expense) t	0.0251*** (0.00833)	0.104* (0.0537)	0.0424*** (0.0149)	0.251*** (0.0861)
Ln(outpatient expense) t-1	0.0392*** (0.00903)	0.121*** (0.0411)	0.0522*** (0.0158)	0.262*** (0.0878)
Ln(inpatient expense) t	0.140*** (0.0228)	0.128*** (0.0251)	0.112** (0.0504)	0.117** (0.0542)
Ln(inpatient expense) t-1	-0.00426 (0.0230)	-0.0203 (0.0245)	-0.0429 (0.0480)	-0.0643 (0.0554)
Someone in household has:				
Illness > 30 days	0.172** (0.0756)	0.0594 (0.0952)	0.0164 (0.146)	-0.0238 (0.147)
Sensory impairment	0.0112 (0.0599)	0.00769 (0.0611)	-0.0478 (0.0952)	-0.0658 (0.108)
Paralysis/ mobility problem	0.0241 (0.0561)	-0.00416 (0.0600)	0.177** (0.0825)	0.160* (0.0834)
Poor/very poor health	0.0619 (0.0790)	-0.0444 (0.0859)	0.0830 (0.121)	-0.0504 (0.134)
Death in last year	0.0969 (0.133)	0.0500 (0.142)	0.168 (0.248)	0.205 (0.273)
Assets quintile groups (ref. is bottom)				
2 nd bottom	0.222*** (0.0620)	0.213*** (0.0635)	0.0615 (0.102)	-0.00627 (0.102)
Middle	0.314*** (0.0655)	0.306*** (0.0658)	0.204 (0.123)	0.105 (0.128)
2 nd top	0.422*** (0.0683)	0.395*** (0.0727)	0.254** (0.122)	0.198 (0.134)
Top	0.660*** (0.0759)	0.614*** (0.0798)	0.630*** (0.152)	0.589*** (0.155)
Enrolled in PSNP	-0.0259 (0.0569)	-0.0154 (0.0620)	0.176 (0.168)	0.244 (0.171)
Has bank account	0.0706 (0.0672)	0.0591 (0.0691)	0.195* (0.118)	0.267** (0.118)
Non-agricultural employment	0.142 (0.127)	0.117 (0.124)	0.196 (0.235)	0.108 (0.245)
Educated head	0.0746* (0.0400)	0.0642 (0.0418)	0.123 (0.0897)	0.0927 (0.0959)

Continued on next page

Table A5 (continued). Regressions presented in Tables 5 and 6 with dependent variable replaced by log of simple average of minimum and maximum forecast medical expenditure

VARIABLES	Levels		First difference	
	OLS	2SLS	OLS	2SLS
Health insurance enrolled	0.0348 (0.0453)	0.0537 (0.0441)	0.0234 (0.0843)	0.0541 (0.0797)
Forgone care when sick	-0.0551 (0.121)	-0.0295 (0.128)	-0.0293 (0.167)	-0.0660 (0.192)
Minutes to nearest health center	0.000781* (0.000453)	0.000710 (0.000483)	0.00105 (0.00111)	0.00108 (0.00123)
Shock	0.147*** (0.0526)	0.144*** (0.0535)	0.135* (0.0794)	0.145* (0.0803)
Ln(household size)	-0.0522 (0.0708)	-0.114 (0.0797)	0.464* (0.239)	0.531** (0.256)
Year 2013	0.254*** (0.0729)	0.285*** (0.0697)		
Constant	3.840*** (0.192)	3.921*** (0.203)	0.281*** (0.0740)	0.336*** (0.0677)
Observations	2,631	2,631	1,097	1,097
R-squared	0.269	0.220	0.078	
F-test (P-value) for joint significance:				
All variables	0.000	0.000	0.000	0.000
Health variables	0.068	0.969	0.290	0.493
Age-sex composition	0.003	0.001	0.475	0.273

Notes: Regressions also include the share of household member in gender specific age groups. In columns 1 and 2, district dummies are also included. Standard errors in parentheses are corrected for clustering at Kebele level. *** p<0.01, ** p<0.05, * p<0.1

Table A6. Regressions of the standard deviation of log forecast medical expenditure

	OLS		2SLS	
	Coeff.	Robust SE	Coeff.	Robust SE
Ln(outpatient expense) t	-0.00149	(0.00317)	-0.00723	(0.0166)
Ln(outpatient expense) t-1	-0.000115	(0.00245)	0.000435	(0.0107)
Ln(inpatient expense) t	0.00780	(0.00600)	0.00870	(0.00620)
Ln(inpatient expense) t-1	-1.97e-05	(0.00743)	0.000283	(0.00764)
Someone in household has:				
Illness > 30 days	-0.0153	(0.0206)	-0.00920	(0.0262)
Sensory impairment	-0.00909	(0.0148)	-0.00811	(0.0147)
Paralysis/ mobility problem	0.0198	(0.0168)	0.0203	(0.0168)
Poor/very poor health	0.00275	(0.0202)	0.00811	(0.0245)
Death in last year	-0.0501*	(0.0291)	-0.0517*	(0.0294)
Assets quintile groups (ref. is bottom)				
2 nd bottom	0.0139	(0.0159)	0.0146	(0.0161)
Middle	0.0359*	(0.0186)	0.0365*	(0.0187)
2 nd top	0.0617***	(0.0194)	0.0626***	(0.0195)
Top	0.0688***	(0.0215)	0.0700***	(0.0223)
Covered by safety net programme	0.0276**	(0.0134)	0.0277**	(0.0132)
Has bank account	0.00268	(0.0160)	0.00345	(0.0156)
Non-agricultural employment	0.0132	(0.0306)	0.0154	(0.0308)
Educated head	0.00902	(0.0102)	0.00856	(0.0104)
Health insurance enrolled	-0.00756	(0.0129)	-0.00940	(0.0141)
Forgone care when sick	0.0151	(0.0323)	0.0123	(0.0321)
Minutes to nearest health center	0.000108	(0.00016)	0.000109	(0.000156)
Shock	0.00530	(0.0124)	0.00550	(0.0125)
Ln(household size)	-0.0274	(0.0187)	-0.0255	(0.0189)
Year 2013	0.00685	(0.0154)	0.00632	(0.0158)
Constant	0.289***	(0.0405)	0.284***	(0.0422)
Observations	2,631		2,631	
R-squared	0.099		0.097	
F-test (P-value) for joint significance:				
All variables	0.000		0.000	
Health variables	0.336		0.297	
Age-sex composition	0.790		0.765	

Notes: Regressions also include the share of household member in gender specific age groups and district dummies. Standard errors i(n parentheses) are corrected for clustering at Kebele level. *** p<0.01, ** p<0.05, * p<0.1