Adverse selection in purchasing health insurance: Evidence from high-frequency logbook data

David I. Levine with Karen Zhang[1]

Theories of adverse selection suggest that those who purchase insurance will be in worse health than those that decline insurance. George Akerlof (1970) points out this adverse selection can be muted for those offered insurance at a steep discount.

We are interested in whether the logbook data can help us understand the dynamics of joining and leaving SKY micro-health insurance in rural Cambodia. We run three tests, one examining members who join SKY versus decline SKY, one comparing those who will versus will not soon join SKY, and one comparing those who will join SKY at regular versus deeply discounted prices.

Theory

As discussed the Polimeni and Levine[2], the economic theory of adverse selection predicts that those who purchase insurance are those who anticipate high future health care costs.

If health is a fixed or very slow-moving attribute of a family, we have the prediction:

> H1: Those who never join SKY are healthier than those who have joined.

In some cases households may observe persistent news about their health, and join after bad news. That scenario leads to:

> H2A: On average future SKY member are healthier many months prior to joining SKY than just a month or two prior to joining SKY.

In other cases, consumers delay receiving medical care until after an insurance policy covers them.[3] If households postpone care prior to joining SKY, we have:

[2] Rachel Polimeni and David I. Levine, "Adverse Selection based on Observable and Unobservable Factors in Health Insurance," working paper, UC Berkeley, 2011.
[3] Marika Cabral, "Claim Timing and Ex Post Insurance Selection: Evidence from Dental Insurance," working paper, January 4, 2011.

H2A': On average future SKY member are healthier many months prior to joining SKY than just a month or two prior to joining SKY.

Importantly, both scenarios lead to high usage of SKY when a member (see H1).

If SKY member are more likely to leave SKY if their health has improved, we have

H2B: Among households that have been SKY member: They are healthier after quitting than they were prior to joining

If SKY improves health we have a second reason for this hypothesis. We can only test this hypothesis on households that joined SKY during the data collection phase.

Finally, the economic theory of adverse selection suggests that it will occur more often when consumers are offered insurance at the regular price than when offered insurance at a a steep discount. This reasoning leads to:

H3: The worse health of SKY members (as in hypothesis H1) show up more strongly among consumers offered insurance at the regular price than when offered insurance at a steep discount.

Data

As part of a larger evaluation of SKY health insurance, in July and August 2008 we distributed logbooks to 2,456 households (14,145 individuals). These households were in 142 villages across rural areas in 3 Cambodian provinces: Kampot, Kandal and Takeo.

The logbooks had one page per month, and enumerators asked respondents to fill out the form every day that something happens to a household member's health or whenever there is a health-related expenditure. Enumerators asked respondents to record all health-related activities, problems and expenditures, ranging from immunizations to purchasing aspirin to fatal injuries. We asked them to list the cost of each treatment and the days lost from each accident or illness.

We hired a data monitor in each village (usually the village chief or his assistant) to collect the logbook pages each month. To protect anonymity, we gave each data monitor a locked box with a slot on top and asked that he carry it to each participating household each month. These forms were then brought back to Phnom Penh by SKY staff as they visited villages as part of their regular job.

We analyse data through August of 2010. However, with limited resources to follow up, compliance dropped steadily over time. For example, we have logbook pages from 1930 households in September 2008, but from only 639 a year later. Some logbooks were not collected regularly, as some data monitors delivered large bundles after we called to remind them.

Statistical Methods

We focus on two outcomes: total days disabled and total health care costs reported by each household each month. Because households can report a health problem on one line (with its corresponding days disabled) and a health care treatment on a different line (with its corresponding costs), we collapse the data to the household-month level, summing all days disabled and all costs reported by each household each month

Note that a household can report more than 31 days disabled in a month if more than one person is disabled or if the respondent report total days disabled after the fact, and recorded more than one month's disability for one illness or injury. For example, a respondent might list a medicine for an illness that has disabled someone for 2 months and write "60" as days disabled.

Both costs and days disabled have large positive outliers. Thus, we focus our analysis on the natural logs of (days disabled in a month + 1) and of (total health care costs + $1). This transformed variable has neither positive nor negative outliers.

The data have very strong trends in compliance and in the types of households that turned in a logbook sheet in a month. Thus, we control for time trends and analyse the first two months (with very high compliance) separately from later months.

We have multiple observations from almost all households, so we adjust our standard errors for any clustering by household.

Results

Descriptive statistics

Figure 1A shows the distribution of days lost from major activity ("days disabled") by all members of a household in a month.

Figure 1A: Days lost by a household per month, (%)
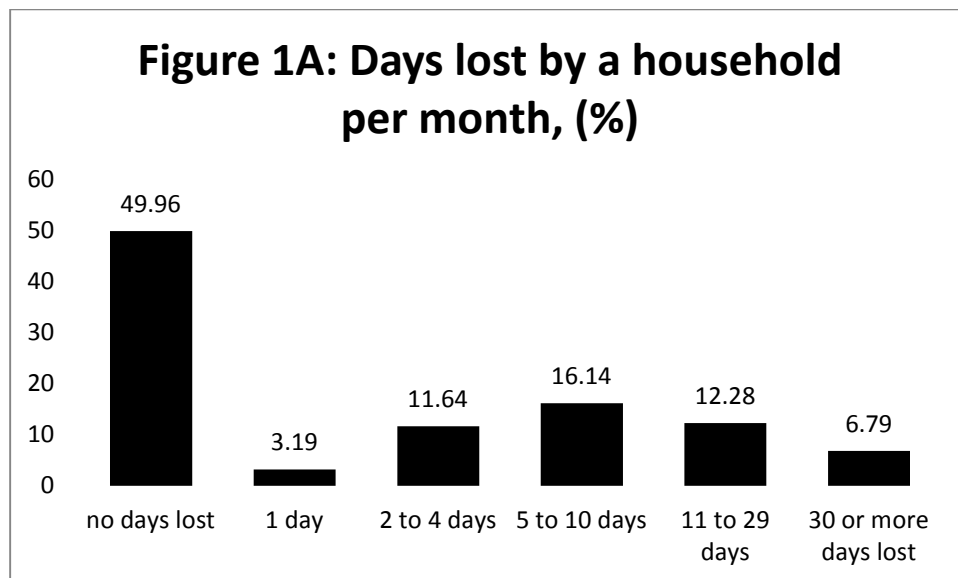
Note: N = 24,431 household-months.

Almost exactly half of households reported no days disabled in a given month (50.0%). The median household with any days disabled had 7 days disabled. About one out of 15 households (6.8%) reported more than 30 days disabled in a month, either due to multiple people with disabilities or due to respondents reporting the entire duration of a multi-month disability.

$Health care spending per household / month, (%)
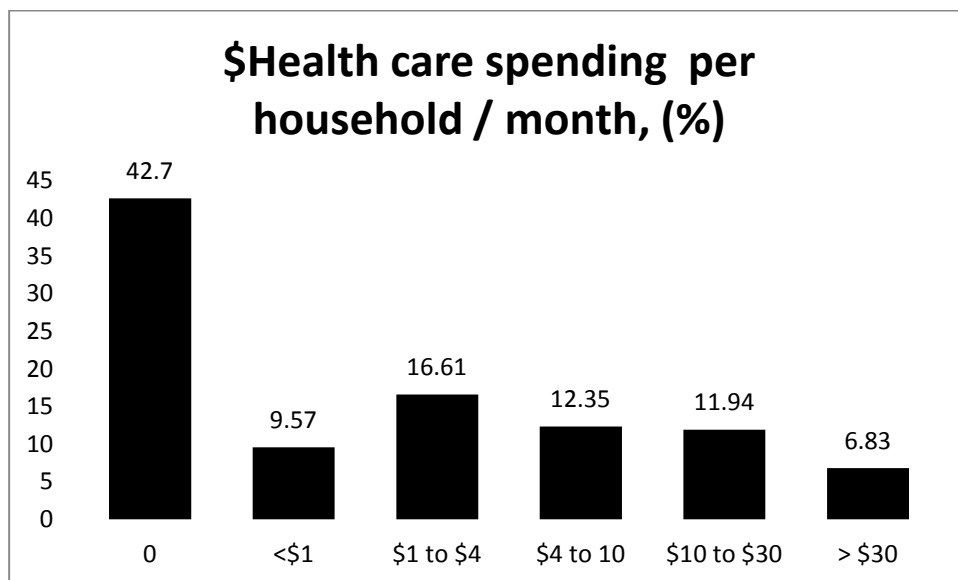

Notes: N= 24,237 household-months. Riel costs converted to US$ at an exchange rate of 4000 riel per 1 US$.

Almost as many households (42.7%) reported no healthcare spending in a month as reported no days disabled. Among those with any spending, the median spending was $5. About one out of 15 household-months (6.8%) reported spending over $30 on healthcare in a month.

In the first 2 months of the logbooks (August and September of 2008) only 22% reported no days disabled and only 9.4% of household-months reported zero spending (See figure 1C for the distribution of reported health care spending during the first 2 months of logbook data collection). Both figures were far below the rates of zero days disabled and zero spending for the whole sample (42.7% and 50.0%).

It is possible that health was poor those first two months. At the same time, it is likely the total sample under-reports true spending and disability as respondents lost interest in completing the form and as data monitors became less diligent in reminding participants.

Figure 1C: $Health care spending per household / month, only first 2 months of logbook collection, (%)
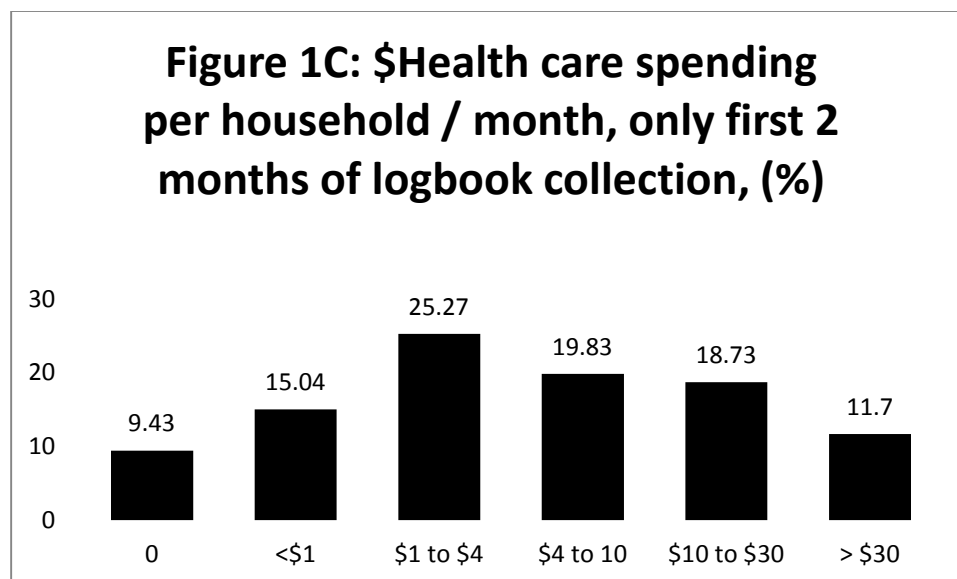
Note: Includes 3,530 household-months from August and September, 2008.

The distributions of those reporting any days disabled or any spending have a bit lower medians than for the full sample ($4.27 median spending among those reporting any spending and 5 days disabled among those reporting any days disabled in the first 2 months, vs. $5.00 and 7 days for the entire sample). This pattern is consistent with some (but not all) of any under-reporting being concentrated among relatively minor incidents.

Regardless of whether we examine the first 2 months or the entire sample, the basic story is as expected: The burden of ill health is very unevenly distributed across households and across time, and that burden is very high for a subset of households in any given month.

Are those who join SKY in poor health?

We find not support for our first hypothesis that those who never join SKY are healthier than those who have joined (Table 1). The point estimate on ever being a SKY member is a small negative effect (-0.11, SE = 0.10, not statistically significant) – the opposite sign (if not significantly so) of what is predicted by theory (column 1).

Results are similar if we exclude those who are or were SKY members, so are comparing future SKY members with those who decline to join (column 2).

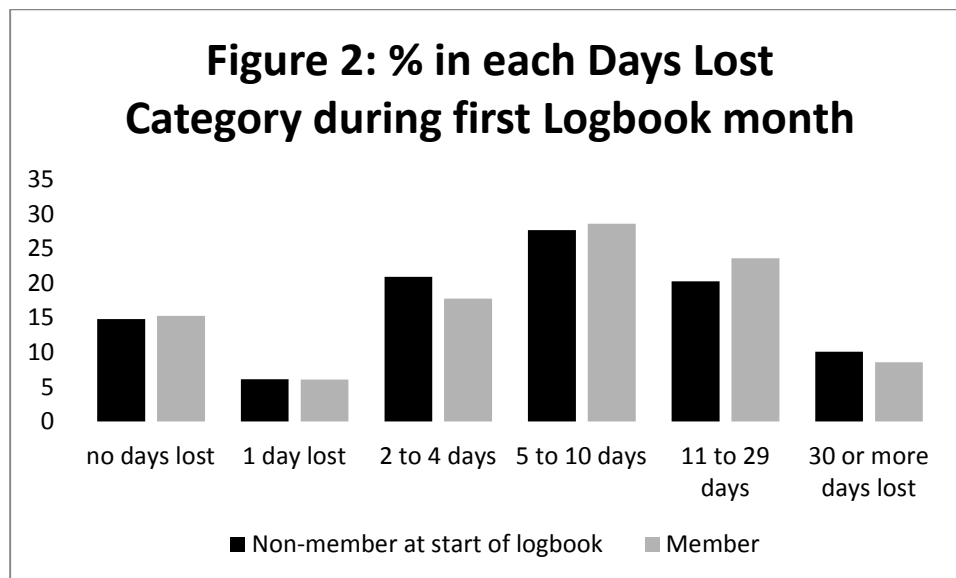## Table 1: Are SKY members disabled more days per month than non-members?

	ln(1+days)	ln(1+days)	ln(1+days)
Sample		Excludes current and former SKY members	
Column	1	2	3
Ever a SKY member	-0.113	-0.206	
	(-1.10)	(-1.81)	
Will join SKY within 2 months			0.373
			(-0.32)
Constant	2.026***		
	(-19.86)		
Month dummies		Yes	Yes
Observations	8587	1823	1823

Notes: *t statistics (in parentheses) adjusted for clustering at the household level.*
* p<0.05, ** p<0.01, *** p<0.001

If news about long-lasting bad health shocks arrives unexpectedly, but those who receive such news are then more likely to join SKY, then those who join SKY should be in poorer health than their peers just prior to joining. In fact, we find that those who will join SKY within the next month or two are less healthy than peers who remain non-members, but the effect is nowhere close to statistically significant (col. 3).

By excluding past members, columns 2 and 3 automatically drop those who are members when the logbook data collection began. If we focus solely at those who were SKY members when the logbook data collection began, there is no evidence they are disabled more days than their peers (see Figure 2). This test is biased against finding adverse selection to the extent SKY reduces disability.

Figure 2: % in each Days Lost Category during first Logbook month

35
30
25
20
15
10
5
0

no days lost | 1 day lost | 2 to 4 days | 5 to 10 days | 11 to 29 days | 30 or more days lost

■ Non-member at start of logbook　■ Member

Note: 1,237 non-member and 476 SKY members. Distributions are not statistically significantly different using the Wilcoxon rank-sum test (P = 0.52).

Do people join SKY when their health worsens and leave when health improves?

Columns 2 and 3 of Table 1 compare in the cross section the health of those who will versus will not join SKY. In this section we analyze within the time series of health for families that were members of Sky for some, but not all, of their logbook months. We compare days disabled during the months before, during, and after a family joins SKY. The basic regression of ln(days disabled + 1) with a fixed effect for each household is in Table 2. These regressions have as a baseline the health of those who are past members.

We find no support for hypothesis H2A that future SKY member are healthier many months prior to joining SKY than just a month or two prior to joining SKY. Among those who will join SKY, those who will not join for more than 2 or more months and those who will join within the next two months have almost identical coefficients (-0.14, SE = .2 and -0.12, SE = 0.17, difference n.s.; column 1). When we look at costs, we find slightly higher health care expenditures for those several months from joining SKY relative to those about to join SKY .14 (.22) and 0.04 (0.19), difference n.s., col. 2). Conversely, our estimates do not show that SKY members delay care just prior to joining (hypothesis H2A'). While we do not support either hypothesis, it could be both are true but the forces for high care prior to joining (due to recent bad health news) and for low care (due to delaying any possible care until insurance begins) roughly cancel out.

We also find no support for hypothesis H2B that households that have been SKY member are healthier after quitting than they were prior to joining. Specifically, the two coefficients on pre-sky outcomes listed in the previous paragraph are both of the wrong sign; that is, the point estimate is that families had slightly *better* health prior to joining SKY compared to after

leaving; although the coefficients are jointly not statistically significant). Specifically, there are small negative and statistically insignificant coefficients comparing ln(days disabled + 1) of those about to be members by 2 or more months (-.14, SE = .2, n.s.) and within the next two months (-.12, SE = .17, n.s.) compared to those who have been members. Results are similarly small and not statistically significant if we examine costs (column 2), although they are of the expected sign.

For those who were SKY members at the start of logbook data collection, the coefficient comparing ln(days disabled+1) when a member versus after leaving is of the predicted sign, but is tiny (-0.03, SE = 0.08) and not statistically significant (col. 3).

Table 2: Fixed effect analysis of SKY members' health before during and after membership

	ln (1+days disabled)	ln ($1+health care cost)	ln (1+days disabled)
Column	1	2	3
Sample	Member at some time, but not at start of logbook data collection		Member at start of logbook data collection
Will join, but not within 2 months	-0.19 (0.204)	0.09 (0.215)	
With join SKY within 2 months	-0.13 (0.095)	-0.02 (0.106)	
Currently a member	0.05 (0.117)	0.22 (0.129)	-0.03 (0.075)
Former member	--	--	--
Month	0.0078 0.0089	0.024* 0.009	-0.007 0.006
# of household fixed effects	336	333	505
Observations	2,798	2,791	4,760
R-squared	0.39	0.34	0.34
Adj. R-squared	0.31	0.25	0.27

* $p<0.01$, ** $p<0.05$
Robust standard errors in parentheses adjust for clustering by household. Baseline group is former members.

Is adverse selection more severe among those who bought insurance at the regular price?

We repeated the above analyses comparing those who were offered insurance at the regular price versus a steep discount. A typical result is in Figure 3. As with the entire sample, when we focus only on those who paid the full price, we do not find evidence of adverse selection.

Figure 3: Were members less healthy if they paid full price?
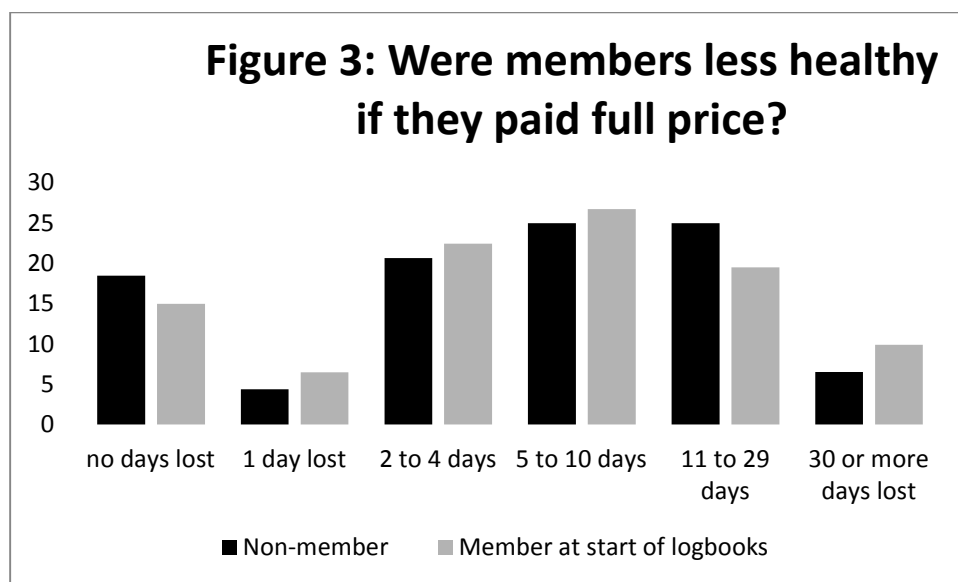
Notes: N = 789 decliners and 92 purchasers, all offered SKY insurance at the regular price. Distributions do not differ significantly on the Wilcoxon rank-sum test.

Thus, we do not support hypothesis 3 that the worse health of SKY members (as in hypothesis H1) show up more strongly among consumers offered insurance at the regular price than among those offered insurance at a steep discount.

Conclusions

Our analysis of the logbook data has three main conclusions:

1. The descriptive analyses finds that about one out of 15 rural households in Cambodia households have more than disabled in a month or spend more than $30 on health care in a month. Either of these situations puts a family's standard of living at risk.

2. The analysis of who buys SKY does not find evidence of adverse selection. For example, new members are not disproportionately those who had high days disabled prior to joining. Future research will explore why these results differ from those of Polimeni and Levine[4].

3. As a methodological point, the logbooks capture useful high-frequency data. At the same time, we find adherence declines substantially over time. Future data collection using logbooks should have more reminders and/or incentives to ensure high adherence.

---

[4] Rachel Polimeni and David I. Levine, "Adverse Selection based on Observable and Unobservable Factors in Health Insurance," working paper, UC Berkeley, 2011.