Adverse Selection in Low-Income Health Insurance Markets: Evidence from an RCT in Pakistan

Torben Fischer Markus Frölich Andreas Landmann

ONLINE APPENDIX

A. Supplementary Tables and Figures

Table A1 analyzes trends and non-linearity in insurance demand.

Table A1 - Insurance Uptake and Demand Elasticities

	P1	P1	P3	P3	P4	P4
Premium	-0.0066***	0.0320*	-0.0164***	-0.0110	-0.0164***	-0.0701**
	(0.0013)	(0.0173)	(0.0017)	(0.0337)	(0.0020)	(0.0276)
Premium^2		-0.0002**		-0.0000		0.0003**
		(0.0001)		(0.0002)		(0.0002)
Constant	0.8636***	-0.0.7413	1.8408***	1.6162	1.7726***	4.0090***
	(0.1133)	(0.7422)	(0.1613)	(1.4046)	(0.1825)	(1.887)
N	2981	2981	2937	2937	3156	3156

Notes: Results are from OLS regression. Standard errors based on bootstrapping the complete empirical process and clustered at the village level.

Table A2 – Determinants of Insurance Demand by Policy

	Household Level Uptake			Individual Level Uptake		
	Individual (P1)	Household (P3)	Group (P4)	Individual (P1)	Household (P3)	Group (P4)
Household Level						
Discount	0.011***	0.015***	0.014***	0.007***	0.017***	0.016***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
HH Size	-0.014*	-0.031***	-0.045***	-0.047***	-0.026***	-0.039***
	(0.008)	(0.008)	(0.008)	(0.005)	(0.009)	(0.007)
Income	0.000	-0.001	0.001	0.001	-0.000	0.001
(in ths PKR)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Saving	-0.000	0.000	0.000	-0.000	0.001	0.000
(in ths PKR)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Asset Index	-0.005	0.014	-0.001	0.004	0.009	0.001
	(0.008)	(0.010)	(0.008)	(0.005)	(0.011)	(0.007)
Head Female	0.023	-0.104**	-0.067	-0.034	-0.099*	-0.064
	(0.041)	(0.049)	(0.062)	(0.037)	(0.053)	(0.058)
No Education	-0.006	-0.109**	-0.105**	-0.004	-0.078	-0.095**
	(0.037)	(0.042)	(0.040)	(0.029)	(0.048)	(0.037)
Any Inpatient	0.096**	-0.019	-0.082	0.001	-0.038	-0.078
	(0.045)	(0.053)	(0.051)	(0.029)	(0.056)	(0.052)
Dependent Level						
Female				-0.115***	-0.024	-0.014
				(0.020)	(0.017)	(0.018)
Age (0-4)				0.158***	0.087*	0.092*
				(0.035)	(0.046)	(0.047)
Age (5-9)				0.094**	0.060	0.063
				(0.037)	(0.043)	(0.043)
Age (10-14)				0.083**	0.010	0.086**
				(0.033)	(0.040)	(0.040)
Age (15-19)				0.065**	0.009	0.010
				(0.029)	(0.035)	(0.031)
Age (20-29)						
Age (30-49)				0.025	-0.031	0.025
				(0.041)	(0.049)	(0.037)
Age (50-59)				0.045	0.111	0.053
				(0.072)	(0.068)	(0.054)
Age (60-69)				0.011	-0.009	0.017
				(0.051)	(0.059)	(0.060)
Age (70+)				0.105	0.046	0.112
				(0.083)	(0.073)	(0.092)
Low Health				0.169**	0.001	0.015
				(0.081)	(0.096)	(0.092)
Medium Health				0.089**	-0.003	0.001
				(0.039)	(0.038)	(0.044)
Inpatient Treatment				0.138**	-0.042	-0.083
				(0.056)	(0.090)	(0.051)
Outpatient Treatment				0.075**	0.047	0.006
				(0.031)	(0.034)	(0.033)
First Son				0.053**	0.023	0.018
				(0.027)	(0.020)	(0.019)
First Daughter				0.028	-0.021	0.036*
				(0.029)	(0.021)	(0.022)
Working				-0.062*	-0.022	0.005
				(0.033)	(0.028)	(0.028)
Constant	0.538***	0.581***	0.570***	0.476***	0.445***	0.419***
	(0.065)	(0.069)	(0.069)	(0.058)	(0.080)	(0.076)
N	856	830	877	2981	2937	3156
R^2	0.07	0.17	0.16	0.13	0.19	0.20

Notes: Point estimates result from OLS regression with standard errors clustered at the village level.

Table A3 – Correlation between Insurance Demand and Expected Costs Index

	(1)	(2)	(3)	(4)
Controls	none	non-health covariates^	observable by insurer~	all
P1 (N=2981)	30.230***	20.017***	19.609***	1.633
	(6.795)	(5.513)	(5.703)	(3.069)
P3 (N=2937)	8.679**	-0.130	0.620	-0.673
	(3.817)	(3.249)	(3.345)	(1.479)
P4 (N=3156)	7.351	-2.445	-2.860	1.007
	(4.859)	(3.583)	(3.505)	(1.708)

Notes: Result from OLS regression of the expected costs index on individual insurance uptake. Covariates are HH size, client gender, client education level dummy, age category dummies, HH income, HH savings, HH asset index, individual work status, individual health status, inpatient and outpatient treatment experience and related costs.
^ All variables except individual health status, inpatient and outpatient treatment experience and related costs.
~ HH size, client gender, client education level dummy, age category dummies.
Standard errors based on bootstrapping the complete empirical process and clustered at the village level.

Table A3 shows the result of regressing the expected costs index on individual insurance uptake under the different insurance policies. The first specification implements a simple positive correlation test. It reveals that the difference between insured and non-insured individuals is substantially larger in the individual (P1) than in the household (P3) and group (P4) insurance schemes. Specification (2) tests whether the positive correlation can be explained by selection based on non-health factors. The idea is that the purchase decision might be influenced by non-health factors which also correlate with health risk, thus creating a positive correlation without the intention of adverse selection. Controlling for such confounding factors would therefore lead to a change in the estimated coefficient compared to the first specification. The results from specification (2) show that some of the differences between insured and non-insured individuals can indeed be explained by non-health factors. Nonetheless, most of the correlation remains in policy P1, for which the coefficient is still highly significant.

As a next step, we control for characteristics that are easy to observe and verify. The idea of this exercise is to test whether an insurance company could in principle separate risk types when using information that is available and reliable in a low-income setting under realistic conditions. Specification (3) controls for such (mainly demographic) variables. Similar to the specification before, the coefficient remains positive and significant for the individual policy (P1), suggesting that classifying individuals based on observable baseline characteristics might not solve the adverse selection problem. For illustrative purposes, specification (4) uses all control variables – essentially the ones used to create the index. As expected, the correlation disappears.

Table A4 - Positive correlation test when using only (un)observable characteristics to predict expected cost index

	Observables^			Unobservables~		
	P1	P3	P4	P1	P3	P4
Insured	11.371***	9.583***	12.184***	19.979***	3.209	-0.933
	(4.776)	(3.501)	(4.076)	(5.099)	(4.064)	(5.280)
Constant	75.668***	75.650***	72.878***	70.689***	75.772***	75.721***
	(3.512)	(3.592)	(3.470)	(3.214)	(4.084)	(3.960)
N	2981	2937	3156	2981	2937	3156

Notes: Results from OLS regression of the expected costs index on individual insurance uptake with cost index predicted using only:

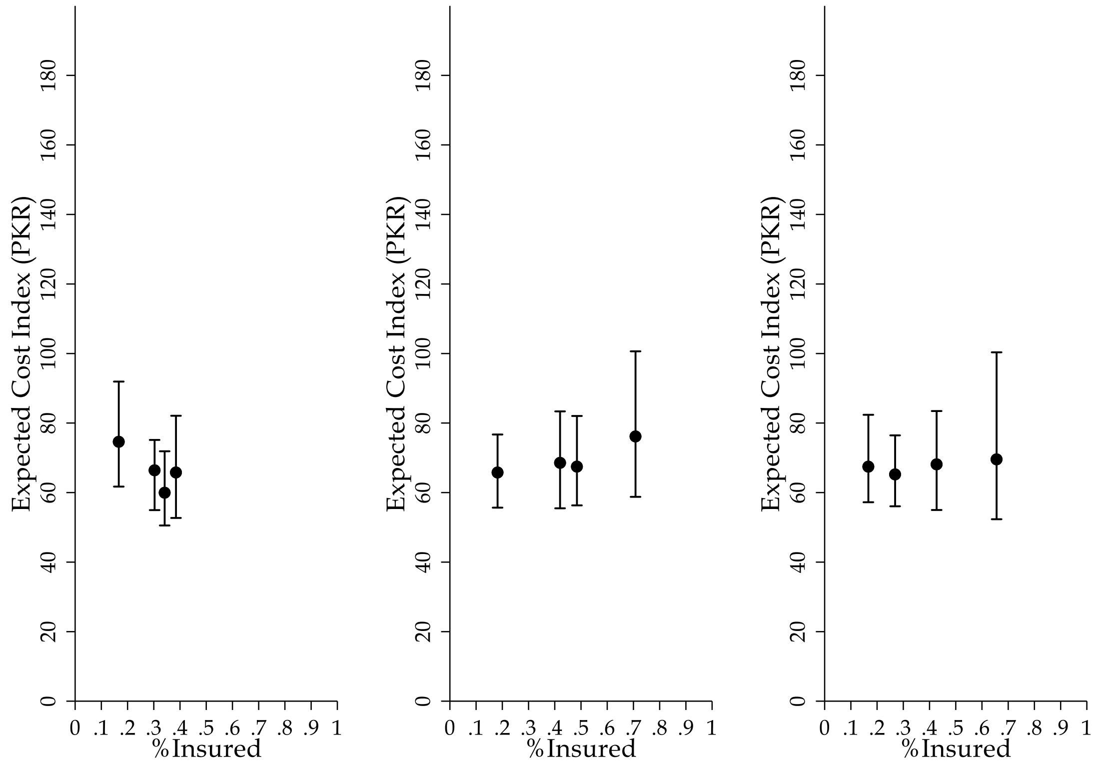
^ HH size, client gender, client education level dummy, age category dummies

~ HH income, HH savings, HH asset index, individual work status, individual health status, inpatient and outpatient treatment experience and related costs.

Standard errors based on bootstrapping the complete empirical process and clustered at the village level.

Figure A1 shows the distribution of costs across demand levels amongst the non-insured. For the individual policy, there appears to be a downward shift in the cost distribution when the share of insured becomes larger. Marginal individuals switching the insurance status in response to a change in price hence seem to be high risk relative to the non-insured but low risk relative to the insured. This is fully aligned with the economic theory on adverse selection discussed in Section II. In contrast, such a pattern for non-insured is not observed under household (P3) and group (P4) policies.

Figure A1 - Change in risk distribution across discounts, non-insured



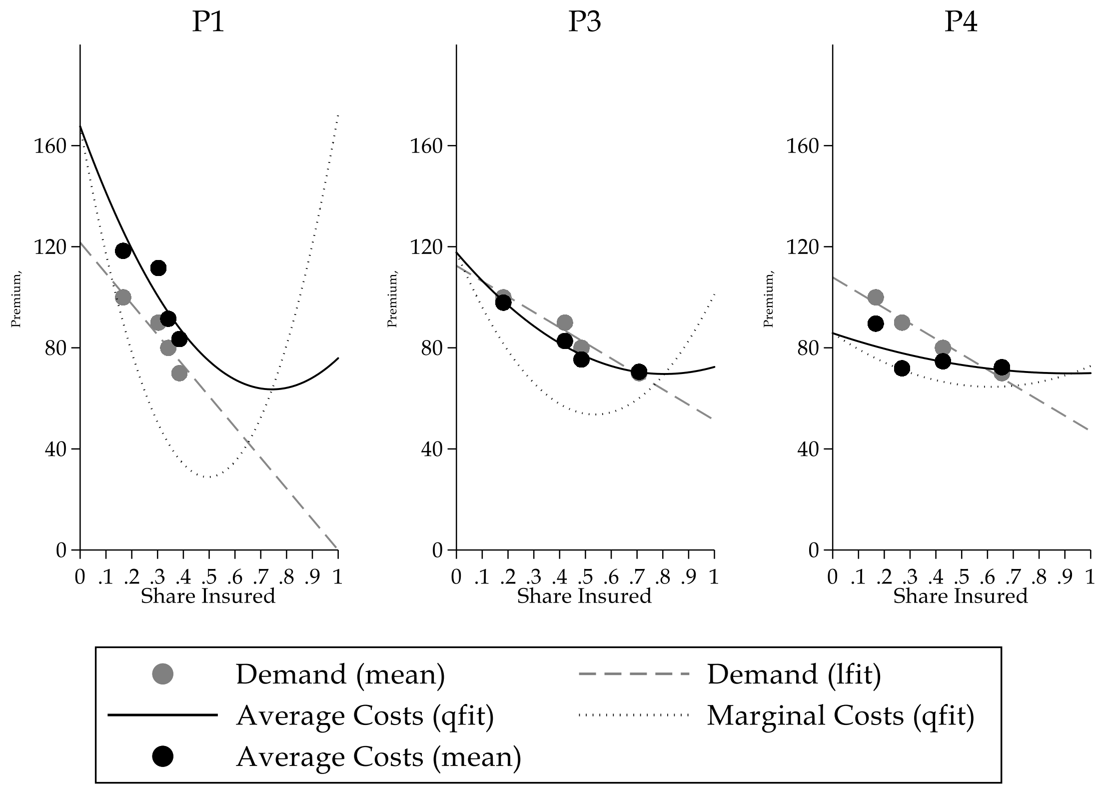
Notes: This figure illustrates shifts in the expected cost distribution by discount level and policy regime. The upper (lower) adjacent line depict the 2.5 and 97.5% quantiles of the bootstrap distribution of the expected cost index for a given policy and discount level. Note that this figure abstracts from the variation in uptake induced by bootstrap resampling by depicting demand levels observed in the original sample.

Table A5 – Slope of the Demand Curve, restricted

	Individual (P1)	Household (P3)	Household (P4)
Premium	-0.008***	-0.016***	-0.016***
	(0.000)	(0.002)	(0.002)
Constant	1.000	1.841***	1.773***
	(.)	(0.161)	(0.182)
N	2981	2937	3156

Notes: The slope of the demand curve is estimated from a linear regression of an individual take-up indicator on the premium, and a restriction of a constant larger or equal than 1 is imposed. Standard errors are not reported if the restriction is binding (only the case for P1). Standard errors are clustered at the village level.

Figure A2 - Market equilibrium and efficient allocation (quadratic cost curve), by policy



Notes: The figure plots the demand, average and marginal cost curves for the respective policies. Average demand for the corresponding premium is given by the dots in light grey. The slope of the demand curve is estimated from a linear regression of an individual take-up indicator on the premium for which a restriction of a constant larger or equal than 1 is imposed. Average costs of the insured for the corresponding demand are given by the dots in black. The slope of the average cost curve is estimated from a quadratic regression of the individual level expected cost index on average take-up at the corresponding premium level. The estimation is restricted to pass through the average cost index for the policy at a demand level of 1.

B. Randomization Procedure

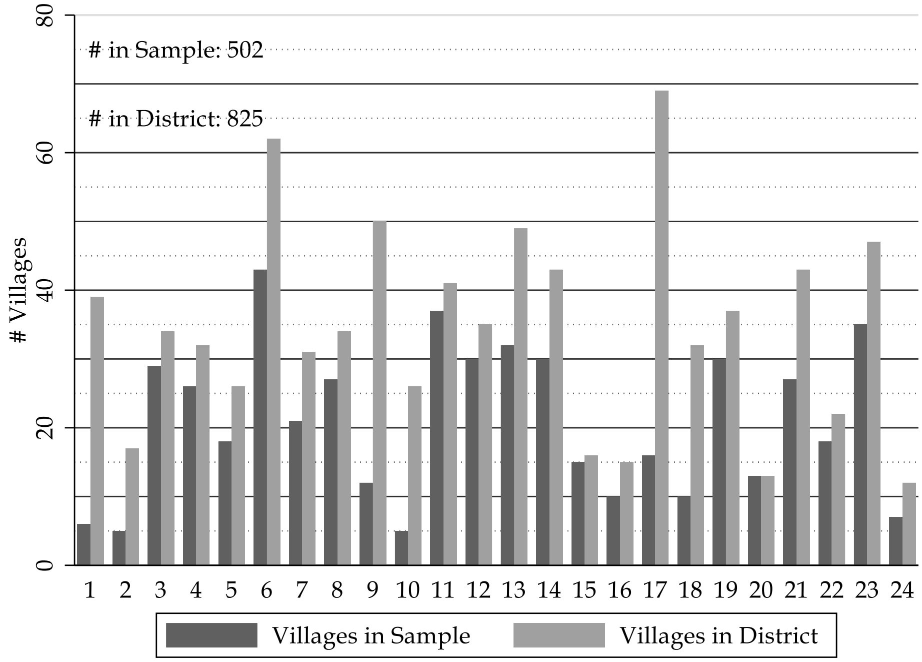
Sampling from incoming credit applications implies that we do not know the set of villages with incoming credit applications ex-ante. Instead, we start with a census of all villages in which our implementation partner operates. To achieve a balanced treatment allocation, we use a permuted block randomization procedure for dynamic treatment assignment. This procedure is used frequently in medical studies facing similar problems of patients stochastically entering the trial (McEntegart 2003). In addition, we stratify the treatment assignment across a set of ex-ante village characteristics to improve balance among treatments along a set of important characteristics.

We condition the randomization on the rural/urban status (4 categories), the historical origin of the village (2 categories) as well as the distance to the next hospital under NRSP's panel (3 categories). This leaves us with a categorization of villages into 24 strata. The treatment assignment proceeds as follows. In a first step, we generate a set of randomly permuted blocks of the six main treatment indicators for each of the 24 strata. In a second step, we produce a unique order in which the villages have entered the experiment. For this purpose, we rely on the timing of loan applications entered in the management information system (MIS). Using the list from step 2, we create strata specific lists of villages that are ordered according to the date and time they entered the MIS. In a final step, each village on this strata-specific list is matched with the corresponding treatment from the strata-specific permuted block of treatments.

This procedure guarantees a balanced distribution of treatments in each cluster, especially when there are sufficient villages per strata entering the experiment to cover full blocks. The reason is that within a full block, one village is assigned to each treatment and no imbalance can occur. Hence, the more full blocks are covered, the fewer imbalances can remain. Figure B1 shows the total number of villages in the district where the RCT takes place by strata and by the number of villages finally entering the experiment. Only three of 24 strata have fewer than six villages to create at least one full block.

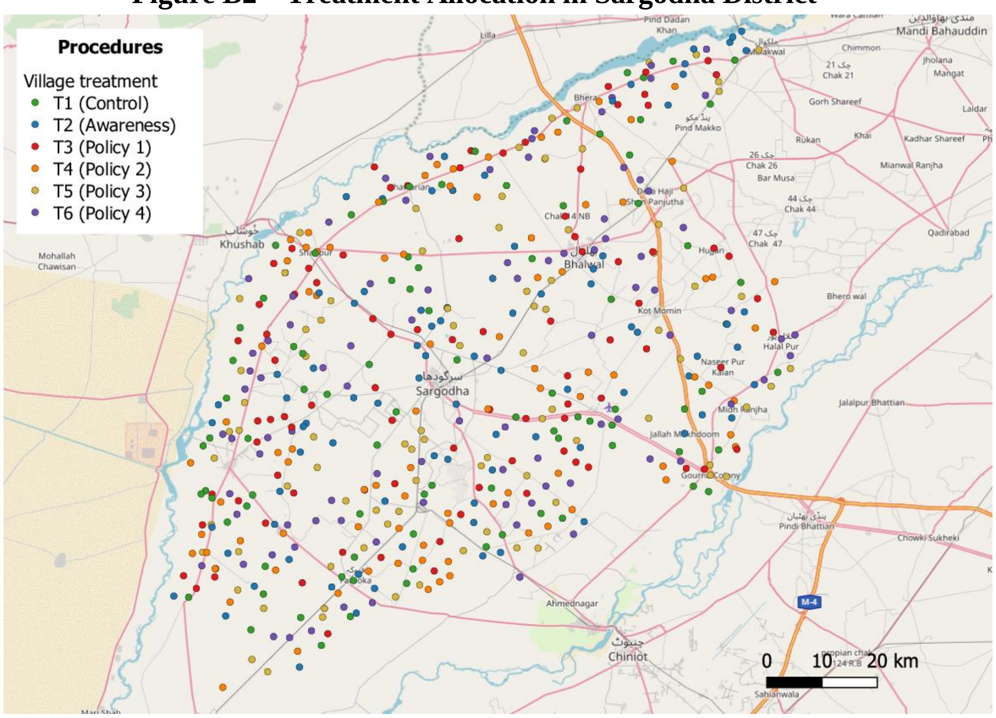
Figure B2 shows the geographical distribution of treatments. The treatment arms appear to be balanced across the whole district, suggesting that the randomization procedure worked as expected.

Figure B1 – Distribution of clusters across strata



Notes: The figures illustrates the distribution of treatment clusters across strata. The 24 strata are generated from ex-ante village level information on location (distance to closest panel hospital, 3 categories), historical origin (chak vs. no chak, 2 categories) and rural/urban status (percentage of agricultural loans, 4 categories).

Figure B2 – Treatment Allocation in Sargodha District



Notes: The figure illustrates the distribution of treatments across Sargodha district. The dots capture the center of the village as per administrative GPS readings. The legend gives the corresponding treatment. Note, however, that these readings are sometimes subject to error, such that some villages appear closer than they are in reality. The average distance to the closest villages is about 2 km, though travel distances are often larger due to indirect road connections.

C. Balancing Tests

We present balancing tests that assess whether our randomization indeed results in a similar distribution of covariates across treatment arms. The balancing tables have the following structure. The first column shows the overall means (standard deviations are in brackets). Subsequent columns provide means and standard deviations for each treatment arm separately. The final column contains the p-value from a joint test for model significance from the following estimation equation:

$$X_{iv} = \alpha + \beta_2 I_{\{T_{iv}=P2\}} + \beta_3 I_{\{T_{iv}=P3\}} + \beta_4 I_{\{T_{iv}=P4\}} + d_s S_v + \varepsilon_{iv} \ , \quad \text{(C1)}$$

where X_{iv} is the covariate, $I_{\{T_{iv}=Pk\}}$, k=2,3,4 are indicators for the treatments P2, P3 and P4 (P1 is the omitted category) and S_v with $v \in \{1, \dots, 24\}$ represents strata dummies.[33] The error term ε_{iv} is clustered at the village level. The test for joint significance of β_2, β_3 and β_4 is thus equivalent to a test for equal means in the treatment arms P1 to P4.

Table C1(a) provides summary statistics and balance tests for sociodemographic, economic and health indicators on the household and individual levels from the baseline survey. Comparing the means of sociodemographic indicators across treatment groups (first panel), we observe no significant differences. This is confirmed by the relatively high p-values of the joint test for model significance. The economic indicators (second panel), household health indicators (third panel) and individual health indicators (fourth panel) show no statistically significant differences across treatment groups. Table C1(b) provides summary statistics and balance tests for the bi-monthly phone survey data. Consent to the phone survey is above 90% and balanced across treatments.[34] About 2% of dependents report an inpatient event, leading to 14% of households having had some dependent admitted. These numbers are similar to the health-seeking behavior at baseline. Again, all variables appear to be balanced across treatment arms. Balancing also holds when the two

[33] Note that strata fixed effects are included only in the balance tests for the main treatments P1 to P4. Discounts are randomized on the level of the household and thus not stratified.

[34] We conduct a separate attrition analysis, but do not find any systematic differences in drop-out across the treatments.

control groups of villages are included where no additional insurance was available in the comparison.

Table C1-Balance Tests across Insurance Policy Treatments
(a) Baseline Survey

	Overall	P1	P2	P3	P4	P-val
Socio-Demographics - HH						
HH Size (Survey)	5.99	5.95	5.95	6.03	6.03	0.57
	(2.117)	(2.093)	(2.072)	(2.054)	(2.237)	
HH Size (Matched)	5.37	5.26	5.43	5.37	5.42	0.37
	(1.912)	(1.872)	(1.956)	(1.822)	(1.986)	
Dependents (Matched)	3.59	3.48	3.62	3.59	3.65	0.44
	(1.869)	(1.834)	(1.876)	(1.791)	(1.961)	
Age of Client	38.62	38.85	38.57	38.24	38.82	0.69
	(10.887)	(10.918)	(10.934)	(10.741)	(10.955)	
Client Female	0.53	0.57	0.51	0.50	0.54	0.33
	(0.499)	(0.495)	(0.500)	(0.500)	(0.499)	
Client No Education (D)	0.55	0.56	0.52	0.55	0.56	0.37
	5.99	5.95	5.95	6.03	6.03	
Economic - HH						
Income (month)	22691	21634	24515	22627	21953	0.28
	(24695)	(20018)	(34658)	(20225)	(20379)	
Asset Index	0.06	0.06	0.20	0.07	-0.09	0.37
	(2.422)	(2.433)	(2.539)	(2.319)	(2.387)	
Savings	12085	13548	13092	10147	11607	0.64
	(67986)	(71670)	(85948)	(31357)	(70158)	
Credit	30439	27603	33057	30112	30803	0.35
	(71910)	(54074)	(79531)	(78197)	(72204)	
Health & Insurance - HH						
Any Inpatient (D)	0.12	0.11	0.13	0.12	0.12	0.51
	(0.327)	(0.316)	(0.338)	(0.325)	(0.328)	
Knows Health Insurance (D)	0.18	0.20	0.19	0.18	0.16	0.62
	(0.385)	(0.397)	(0.390)	(0.383)	(0.369)	
Health - Dependents						
Health Step (1-5)	4.76	4.75	4.76	4.75	4.77	0.97
	(0.631)	(0.631)	(0.644)	(0.648)	(0.602)	
Outpatient Experience (D)	0.14	0.14	0.15	0.15	0.14	0.96
	(0.351)	(0.349)	(0.355)	(0.353)	(0.346)	
Inpatient Experience (D)	0.02	0.02	0.02	0.01	0.02	0.60
	(0.126)	(0.124)	(0.135)	(0.121)	(0.124)	
Outpatient Cost	212.05	190.04	230.09	206.45	219.11	0.51
	(878.726)	(817.807)	(923.245)	(855.271)	(907.079)	
Inpatient Cost	383.65	339.78	474.72	329.56	384.75	0.49
	(4181.490)	(3842.132)	(4689.559)	(3794.116)	(4290.000)	
N (Dependents)	15361	3560	3920	3796	4085	
N (HHs)	4283	1022	1083	1058	1120	

(b) Phone Survey

	Overall	P1	P2	P3	P4	P-val
Consent to participate (D)	0.93	0.92	0.93	0.93	0.93	0.82
	(0.259)	(0.269)	(0.254)	(0.261)	(0.253)	
Health - HH						
Any Inpatient (D)	0.14	0.15	0.13	0.15	0.15	0.46
	(0.351)	(0.353)	(0.334)	(0.360)	(0.355)	
Any Outpatient (D)	0.65	0.66	0.66	0.64	0.65	0.85
	(0.476)	(0.475)	(0.473)	(0.480)	(0.478)	
Health - Dependents						
Inpatient Experience (D)	0.02	0.02	0.01	0.01	0.02	0.96
	(0.124)	(0.130)	(0.121)	(0.120)	(0.124)	
Outpatient Experience (D)	0.14	0.14	0.14	0.14	0.14	0.88
	(0.348)	(0.348)	(0.349)	(0.350)	(0.344)	
Inpatient Cost	371.59	438.46	452.54	371.85	237.36	0.12
	(5537.914)	(5116.372)	(8022.091)	(4937.016)	(2872.399)	
Outpatient Cost	702.79	569.42	769.31	638.28	812.38	0.07
	(5415.117)	(3154.431)	(5475.952)	(5350.682)	(6772.168)	
N (Dependents)	14246	3275	3641	3496	3834	
N (HHs)	4283	1022	1083	1058	1120	

Notes: The table provides means and standard deviations (in parentheses) of the variables. Column 1 provides overall measures, while other columns indicate the policy. The last column contains the p-value from a joint test for model significance of equation (C1) including strata fixed effects. Standard errors are clustered at the village level. Binary variables are indicated with (D). Positive baseline health costs (outpatient and inpatient) are winsorized at the 90th percentile.

In a next step, we provide evidence for a balanced distribution of discount vouchers. Random assignment through household-level lotteries with replacement implies an expected uniform probability distribution of discounts. Table C2 illustrates the frequencies of the four discount levels across insurance policy as well as in general. In addition, we test the null-hypothesis of the expected uniform distribution by Pearson's Chi-square test, the p-value of which is reported in the second to last row. Our test does not reject the hypothesis of a uniform distribution, even though the share of zero discounts is lower than 25%. This holds true also for policy P1 for which we observe a stronger deviation from the expected uniform distribution.

Table C2 - Balance Check: Discount Allocation

	P1	P2	P3	P4	Overall
0	0.19	0.23	0.22	0.22	0.22
10	0.27	0.27	0.26	0.28	0.27
20	0.27	0.28	0.25	0.27	0.27
30	0.27	0.23	0.27	0.23	0.25
Pearson Chi2 P	0.2268	0.4632	0.5998	0.2290	0.2144
HHs	856	870	830	877	3432

Notes: Relative frequencies of discounts given the respective policy. Pearson Chi2 p provides the p-value from a chi-square test with H0 of a uniform distribution. The difference in number of observations to the main balance checks is explained by the fact that only households attending the community meeting received a discount.

To investigate potential systematic imbalances, we provide additional tests in Table C3. The idea is to investigate whether specific household characteristics, potentially related to health indicators and thus insurance demand, cause a jump in the probability of receiving a specific discount voucher. We replace the main treatment indicators in equation (C1) with discount-level indicators, where the zero discount group serves as the reference group. We test for discontinuous jumps in the probability of receiving a specific discount by conducting a joint test for model significance. The final column provides the corresponding p-value. We observe no statistically significant difference across discount levels for any of the health indicators. Similarly, there are no systematic differences in economic indicators. In terms of socio-demographic variables, it seems that there are statistically significant differences in the age and sex composition across discount levels. A clear, systematic pattern such as older individuals or females receiving higher discounts, however, is not visible. For this reason, we are confident that the randomization of discounts through household lotteries in the field is not subject to systematic imbalances.

Table C3 - Balance Checks (Discounts)

	Overall	D=0	D=10	D=20	D=30	P-val
Socio-Demographics – HH						
HH Size	5.99	5.98	5.96	6.01	6.01	0.94
	(2.10)	(2.03)	(2.05)	(2.24)	(2.08)	
Age of Client	38.72	38.33	39.52	39.03	37.86	0.01
	(10.96)	(10.92)	(11.22)	(11.19)	(10.40)	
Client Female (D)	0.53	0.50	0.52	0.57	0.54	0.03
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	
Client No Education (D)	0.54	0.53	0.55	0.57	0.52	0.17
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	
Economic – HH						
Avg. Inc. (month)	22723.69	22963.60	21587.96	24109.96	22264.71	0.12
	(25549.78)	(30839.78)	(16445.10)	(28174.47)	(25640.41)	
Land (acres)	1.40	1.29	1.48	1.41	1.42	0.65
	(3.26)	(2.91)	(3.29)	(3.12)	(3.64)	
Savings	12340.33	9757.70	14193.15	12826.61	12043.18	0.40
	(73120.94)	(33068.48)	(85167.08)	(90250.76)	(62995.56)	
Credit	30861.49	30574.92	32890.37	30272.55	29535.92	0.72
	(70137.66)	(80249.15)	(65293.04)	(73614.96)	(61564.62)	
Health & Insurance – HH						
Any Inpatient (D)	0.12	0.13	0.13	0.11	0.12	0.54
	(0.33)	(0.34)	(0.33)	(0.31)	(0.32)	
Total Inpatient Cost	2087.59	2223.27	2373.89	1510.28	2277.49	0.14
	(9852.24)	(10078.60)	(10921.39)	(7943.69)	(10259.55)	
Knows Health Ins. (D)	0.18	0.18	0.21	0.17	0.16	0.07
	(0.39)	(0.38)	(0.41)	(0.38)	(0.37)	
N (Dependents)	12283	2643	3283	3236	3121	
N (HHs)	3433	739	927	914	853	

Notes: The table provides means and standard deviations (in parentheses) of the respective variables. Column provides overall measures, while other columns indicate the respective policy. The last column contains the p-value from a joint test for model significance of equation (C1). Standard errors are clustered at the village level. Binary variables are indicated with (D). Positive baseline health costs (outpatient and inpatient) are winsorized at the 90th percentile.

Table C4 provides analogous balance tests for the group meeting attendance. We observe no statistically significant differences in observables between meeting attendants and non-attendants. The observed similarity supports external validity of our results for the population of credit clients in Sargodha district.

Table C4 - Balance Checks (Meeting Attendance)

	Overall	Not Attending	Attending	P-val
Health - Dependent				
Expected Reimbursement Cost (PKR)^	82.34	82.74	82.24	0.87
	(134.221)	(145.267)	(131.314)	
Subjective Health Status (1-5)	4.76	4.77	4.76	0.41
	(0.631)	(0.625)	(0.633)	
Outpatient Treatment (D)	0.14	0.14	0.14	0.69
	(0.351)	(0.348)	(0.351)	
Inpatient Treatment (D)	0.02	0.02	0.02	0.88
	(0.126)	(0.125)	(0.127)	
Outpatient Cost (PKR)	212.05	200.44	214.96	0.39
	(878.726)	(837.647)	(888.732)	
Inpatient Cost (PKR)	383.65	406.76	377.86	0.73
	(4181.490)	(4423.866)	(4118.686)	
Socio-Demographics - HH				
HH Size (Survey)	5.99	5.99	5.99	0.97
	(2.117)	(2.170)	(2.104)	
Age of Client	38.62	38.23	38.72	0.23
	(10.887)	(10.596)	(10.958)	
Client Female (D)	0.53	0.52	0.53	0.74
	(0.499)	(0.500)	(0.499)	
Client Has No Education (D)	0.55	0.56	0.54	0.34
	(0.498)	(0.497)	(0.498)	
Economic - HH				
Avg. Monthly Earning (PKR)	22691.34	22560.66	22723.69	0.86
	(24694.608)	(20900.437)	(25549.780)	
Asset Index	0.06	-0.05	0.09	0.13
	(2.422)	(2.419)	(2.423)	
Savings (PKR)	12085.10	11054.26	12340.33	0.48
	(67986.387)	(41200.112)	(73120.945)	
Total Credit (PKR)	30438.72	28731.23	30861.49	0.41
	(71910.002)	(78684.167)	(70137.665)	
Health & Insurance - HH				
Any Inpatient (D)	0.12	0.12	0.12	0.87
	(0.327)	(0.325)	(0.327)	
Inpatient Cost (HH)	2167.34	2489.41	2087.59	0.32
	(10155.145)	(11296.788)	(9852.237)	
Knows Insurance (D)	0.18	0.18	0.18	0.73
	(0.385)	(0.381)	(0.386)	
N (Dependents)	15361	3078	12283	
N (HHs)	4283	850	3433	

Notes: The table provides means and standard deviations (in parentheses) of the respective variables. Column 1 provides overall measures, while other columns indicate the attendance of the respective household in the group meeting. The last column contains the p-value from a joint test for model significance similar to equation (C1), excluding strata fixed effects. Standard errors are clustered at the village level. Binary variables are indicated with (D). Monetary variables are in Pakistani Rupees (PKR). Positive baseline health costs (outpatient and inpatient) are winsorized at the 90th percentile.

^ In line with the other balancing tables, we include all treatment arms in the test – including the individual high insurance (P2), which features higher expected costs. The mean of the costs index is therefore somewhat higher than in the standard coverage treatments only (P1, P3, P4).

D. Construction of the Expected Cost Index

The insurer's average cost curve constitutes the central element for testing adverse selection in this study. A straightforward estimate of the average cost curve would aggregate the insurer's reimbursed claims for a given insurance product and price level.[35] Since hospitalization is a rare event, we cannot – despite the large sample size – directly estimate the average cost curve based on these reimbursed claims. Instead, we use detailed baseline health and demographic information (X_{i0}) to predict the insurance provider's reimbursement costs for each individual i (C_{i1}). Time is indicated with 0 at baseline and with 1 at the end of the insurance period. We are interested in obtaining a good estimate of the conditional expectation of the provider's reimbursement cost at the end of the insurance period: $\hat{E}[C_{i1}|X_{i0}]$.

Again, a direct approach would use the observed reimbursement cost to estimate their relation to baseline characteristics. However, claims are too rare in our data to obtain a good estimate (only 39 claim cases are reported). This is partly because claims can only be made by people who are insured, which excludes the non-insured part of our sample from such an analysis. Furthermore, not all hospitalization cases lead to a claim.[36] We therefore use detailed information on inpatient health events and costs, gathered in our bi-monthly phone survey during the one-year product cycle. Hospitalization events in the phone survey are reported for 334 of the 21,470 dependents in the phone survey sample. Based on the aggregated inpatient expenditures during the insurance period, we calculate the maximum amount for each individual that could be reimbursed under the insurance policy ($\bar{C}_{i1}$). Subsequently, $\hat{E}[\bar{C}_{i1}|X_{i0}]$ can be predicted using an adequate regression. We account for the fact that not all of these costs are claimed by adjusting the final expected cost index (ECI_{i1}) as follows:

$$ECI_{i1} = \hat{E}[\bar{C}_{i1}|X_{i0}] \times \frac{\sum_P \sum_{i \in Insured} C_{iP1}}{\sum_P \sum_{i \in Insured} \hat{E}[\bar{C}_{i1}|X_{i0}]} \qquad \text{(D1)}$$

[35] As described in section III, there are four insurance products and four price levels.

[36] To gain insights into this phenomenon we conducted in-depth interviews with some households that were insured, reported a hospitalization event and yet did not claim reimbursement of their expenses. These interviews were conducted after the end of the insurance period to avoid interfering with the research study. The reported reasons for this behavior are manifold. While some incidences can be explained with unawareness about the claim procedure or frustration about the process, other cases are related to missing trust, preference for alternative (more expensive) coping strategies and recall problems about having bought the insurance product.

This means the prediction is made based on all potentially claimable costs, which maximizes statistical power. At the same time, the index is scaled by the ratio between actual claim amounts relative to the maximal claimable amount according to the policy (PKR 15,000 for P1, P3, and P4).

We estimate $\hat{E}[\bar{C}_{i1}|X_{i0}]$ using a Tobit model, controlling for a broad range of baseline household- and individual-level characteristics.[37] The household-level variables account for the economic situation, the household size and client characteristics. The individual-level characteristics include demographic information such as age, gender and whether the individual is contributing to the household income as well as detailed health information. The latter includes individuals' subjective health status, inpatient and outpatient health history, associated costs, type of health events experienced and subjective magnitude of the shocks experienced. Table D1 reports the estimated coefficients in the Tobit regression for eligible dependents. The estimated coefficients suggest that dependents in younger age groups cause lower reimbursable claims than the reference group of 30- to 49-year-olds. Further, better subjective health and better self-reported health history result in lower reimbursable costs.

Column 2 of Table D1 reports the results of an identical estimation that considers only the eligible dependents in the control groups. The purpose of this additional regression is to assess the robustness of our results to the existence of moral hazard. As described in Section III, the control groups are not offered any additional insurance and hospitalization expenditure for dependents in this group and hence should not be affected. Thus, comparing the coefficient estimates in columns 1 and 2 shows whether moral hazard changes the mapping from baseline characteristics to hospitalization expenditures. The resulting coefficient estimates are mostly similar to the ones reported in column 1 in sign and magnitude. Based on a Hausman specification test, we cannot reject that both models are equivalent (p-value: 0.62). This is consistent with the fact that there is no significant joint treatment effect of the insurance treatments on inpatient expenditures (see Table D3). The choice between including all observations and using the control groups only therefore does not make a large difference. To maximize the precision of our estimates, we include all observations (i.e. specification 1).

[37] A Tobit model is a natural choice, as maximum claimable amounts cannot be lower than zero and are restricted to PKR 15,000 in policies P1, P3 and P4.

Table D1 - Predicting Inpatient Expenditure using Baseline Characteristics

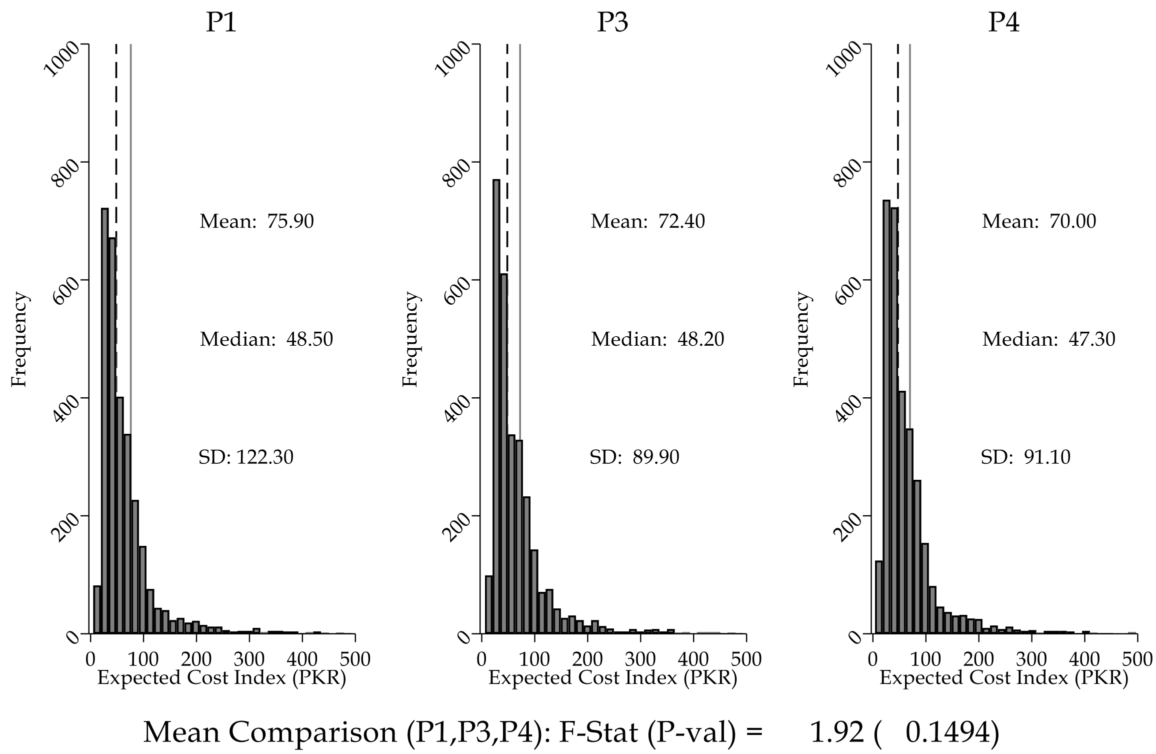
	1	2
	All T	Controls only
Household Level Info		
HH Size	-2101.73***	-2158.89*
	(614.76)	(1152.94)
Income (in 1000 Rs.)	52.63	-6.14
	(43.93)	(53.90)
Saving (in 1000 Rs.)	15.46	19.90
	(10.53)	(23.59)
Asset Index	166.29	-613.68
	(518.96)	(748.14)
Client Female	-2443.19	-3585.02
	(2480.65)	(3439.78)
Client has no education	-313.96	-2107.62
	(2455.84)	(3785.38)
Individual Level Info		
Age (0-4)	-11013.44**	-2201.52
	(5136.17)	(7848.51)
Age (5-9)	-23164.43***	-16746.32**
	(5532.99)	(8167.88)
Age (10-14)	-25380.22***	-17503.40**
	(5845.65)	(8581.06)
Age (15-19)	-12640.54***	-8662.90
	(4811.04)	(7051.92)
Age (20-29)	-8964.95*	-9270.62
	(5125.54)	(7693.66)
Age (50-59)	1807.44	-1494.23
	(6532.04)	(10008.00)
Age (60-69)	-5133.75	-6295.64
	(6621.86)	(9848.43)
Age (70+)	-3420.91	1022.77
	(6847.05)	(9406.45)
Working	-14354.92***	-15784.16**
	(4159.49)	(6591.64)
Female	398.16	-1904.10
	(2445.71)	(3409.65)
Subjective Health Status (1-5)	-6044.82***	-6917.62***
	(1711.41)	(2558.89)
Outpatient Treatment	6590.10	-3834.26
	(4676.39)	(7435.71)
Inpatient Cost (PKR)	0.32**	0.29
	(0.16)	(0.20)
Outpatient Cost (PKR)	1.02	2.50*
	(1.02)	(1.41)
Chronic Inpatient Disease	28285.87***	9173.34
	(9204.14)	(13300.04)
# Inpatient Cases	1677.53**	5201.63
	(852.42)	(3805.72)
# Neglected Inpatient Care	6081.98	-3615.92
	(9410.28)	(14967.62)
Drop in Subj. Health (Inpatient)	-4084.47	-2811.45
	(2615.04)	(4310.80)
Drop in Subj. Health (Outpatient)	84.00	-1664.36
	(1615.31)	(2381.89)
Constant	-51793.27***	-32862.80**
	(11026.38)	(16515.33)
sigma	49136.77***	42527.07***
	(3476.99)	(4715.93)
N	21473	7227
F value	6.04	3.12
P-value of F statistic	0.0000	0.0000

Notes: The table provides results from a Tobit model that explains the maximal claimable costs as a function of household- and individual-level variables. Standard error in parentheses are clustered at the village level. Monetary amounts are in Pakistani rupees (PKR), where 101 PKR ≈ USD 1. Positive baseline health costs (outpatient and inpatient) are winsorized at the 90th percentile.

We predict expected claimable inpatient expenditures $\hat{E}[\bar{C}_{i1}|X_{i0}]$ for each individual using specification 1 of Table D1. Consistent with Equation D1, we then apply a scaling factor of 0.4552 to predict the expected cost index ECI_{i1} for each individual under the policy.[38]

Figure D1 illustrates the distribution of the expected insurer costs across policies P1, P3, and P4. The mean of the distribution is shown as a grey solid and the median as a black dashed line. The figure reveals that the cost distribution is right-skewed in a similar way for all policies. A test for equality of their means cannot be rejected (p-value: 0.1494).

Figure D1 – Histogram of the expected cost index, by policy



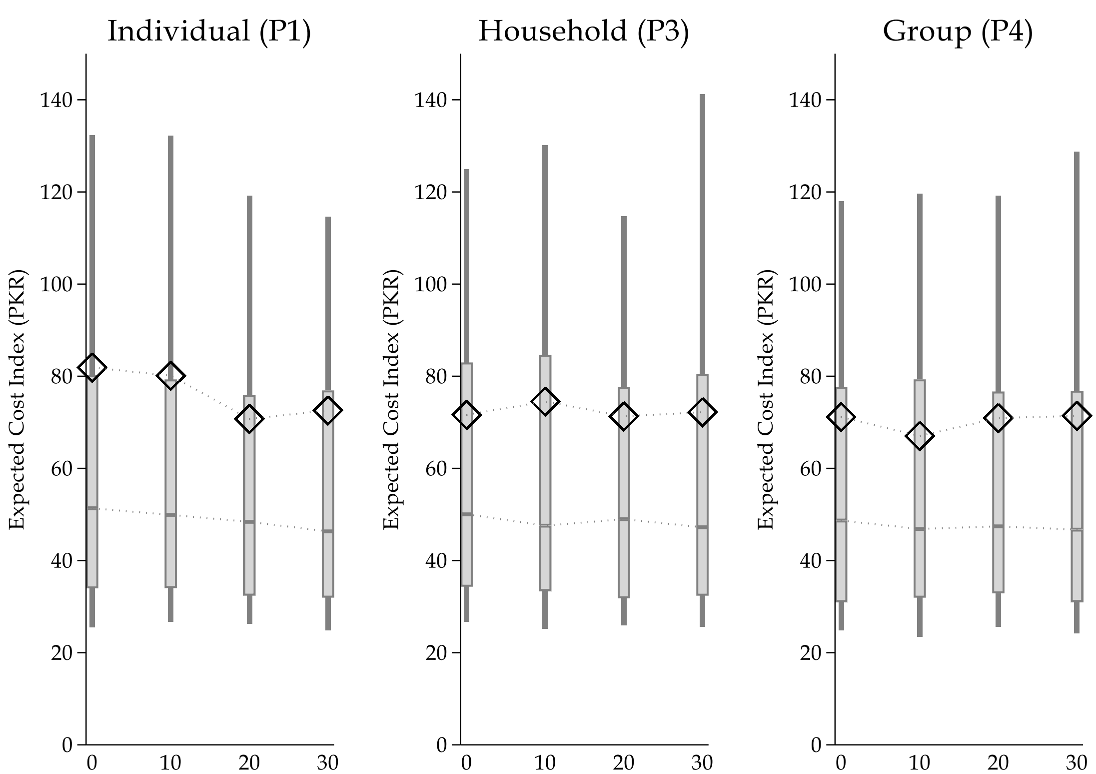
Notes: The figures shows histograms of the provider's expected reimbursement costs across the four policies. The mean and median are illustrated through the solid and dashed line respectively. The predicted reimbursement costs are measured in Pakistani rupees (PKR), where 101 PKR ≈ USD 1.

Figure D2 shows the balancing of the cost index across policies and prices. The box plots illustrate the interquartile rage (IQR), in addition to the 10th and the 90th percentile of the distribution. The distributions appear balanced across prices in all policies.

[38] The scaling factor is based on hospitalization expenditure and claim data during the insurance period as summarized in Table D2.

Table D2 summarizes and compares hospitalization costs up to the theoretical coverage limit ("Claimable Inpatient Costs"), number of claims reimbursed and average payouts under the different insurance policies. Reimbursed claims are based on all observations in the insurance data set. Claimable costs are based on the self-reported information from the bi-monthly phone survey and restricted to the observations that can be matched with insurance data (the dataset used in the paper). Matched and non-matched observations from the survey data are not significantly different, though. Besides illustrating the ratio between insurance payouts and potentially claimable amounts (0.3885), the table reveals that there are indeed strong differences in paid claims between products. The payout frequency tends to be higher in individual policies (P1, P2) than in households or group policies (P3, P4) and despite the limited number of cases, several comparisons via two-sample proportion tests are significant: P1 vs. P4 (p-value: 0.0904), P2 vs. P3 (p-value: 0.0173) , P2 vs. P4 (p-value: 0.0110) and P1+P2 vs. P3+P4 (p-value: 0.0057). Comparisons of P1 vs. P2 and P3 vs. P4 are all insignificant.

Figure D2 – Distribution of risk across discounts and policy regimes



Notes: This figure illustrates the distribution of the expected cost index by discount level and policy regime. The box plot illustrates the interquartile range (IQR), with the median indicated by the line separating the box. The lower (upper) adjacent line shows the 90th (10th) percentile, respectively. The diamond indicates the value of the mean.

Table D2 - Summary Statistics of Inpatient Expenditure and Claim Behavior

	N Insured	N Insured (Matched)	Mean Claimable Inpatient Costs^	Mean Predicted Claimable Inpatient Costs^	N Claims (Total)~	Mean Amount Claimed~
P1	1054	922	350.00	212.63	12	114.18
P2	663	615	450.90	325.75	11	202.36
P3	1505	1350	166.80	169.44	9	59.21
P4	1344	1211	122.79	163.70	7	55.04
Total	4566	4098	235.55	200.92	39	91.46
Test type			t		proportion	t
P1&2 vs. P3&4			0.0001***		0.0057***	0.0059***
P1 vs. P3			0.0188**		0.1358	0.1649
P1 vs. P4			0.0024***		0.0904*	0.1461

Notes: Monetary amounts are in Pakistani rupees (PKR), where 101 PKR ≈ USD 1. "Insured" are all individuals appearing the insurance management information system, "Insured (Matched)" are those Insured that can be matched with our survey data. Differences are explained by having to match on names when collating these two data bases for dependents without a unique national ID number (mostly for minor). ^ Based on "Insured (Matched)", ~ based on "Insured".

Table D3 – Treatment Effect of Insurance Policies on Reported Inpatient Cost

	Inpatient Cost (PKR)
P1	158.3321*
	(92.7487)
P3	109.9905
	(106.1380)
P4	-44.5723
	(62.1140)
Strata FE	yes
N	17832
R^2	0.0014
Wald	1.6900
p(Wald)	0.1685

Notes: Reported inpatient costs are in Pakistani rupees (PKR), where 101 PKR ≈ USD 1. The pure control group (excluding awareness) serves as the reference group. The OLS regression includes strata fixed effects and standard errors are clustered at the village level. The Wald test statistic is from a joint test of significance of the main treatment indicators. The estimation sample contains eligible dependents of all policies, excluding policy P2, for which there is information from the follow-up phone survey.

E. Simulation of Selection

In this appendix, we simulate how costs curves should look like under (i) perfect selection and (ii) fuzzy selection. The goal of the exercise is to understand the scope for selection remaining in the different policies. This exercise allows us to better gauge to which extend reduced selection in the bundled policies is a 'mechanical' design effect versus an additional behavioral effect.

To **simulate perfect selection**, we first order individuals (or households or groups) by (average) expected costs first, and following this order gradually allocate them to the insurance pool with increasing demand. We implement this logic by first creating percentiles of expected costs. For simplicity (and to reduce noise in the "fuzzy" simulations described below), we replace expected costs of all individuals (or households or groups) in each percentile bin with the average cost in the bin.[39] Note that this does not change the simulation of perfect selection, as even with a finer classification, the same average cost would be predicted for each complete percentile. The average cost curves resulting from perfect selection are shown as the blue lines in the graphs of Panel A in Figure E1. As expected, these lines get more and more flat when moving from the individual to the household and group policy, as more and more variation in costs has averaged out.

In a second step, we **simulate fuzzy selection** by allowing individuals (or households or groups) to make (random) mistakes when allocating themselves to "their" demand percentile. We implement this feature by adding a normally distributed error term to their percentile *number*, based on which we re-rank everyone and generate new "fuzzy" percentile bins.[40] The higher the standard deviation of the percentile error, the lower the correlation between the perfect selection percentiles and the fuzzy percentiles. This correlation is given in the graphs (e.g. "pctile corr = 1.0" for perfect selection). To further reduce noise in the simulation, we create 100 duplicates of each observation (before adding the error term to the perfect selection percentiles).

The advantage of modeling fuzzy selection in this manner is that it allows us to decompose deviations from the most extreme case – perfect selection in the individual scheme – into two parts:

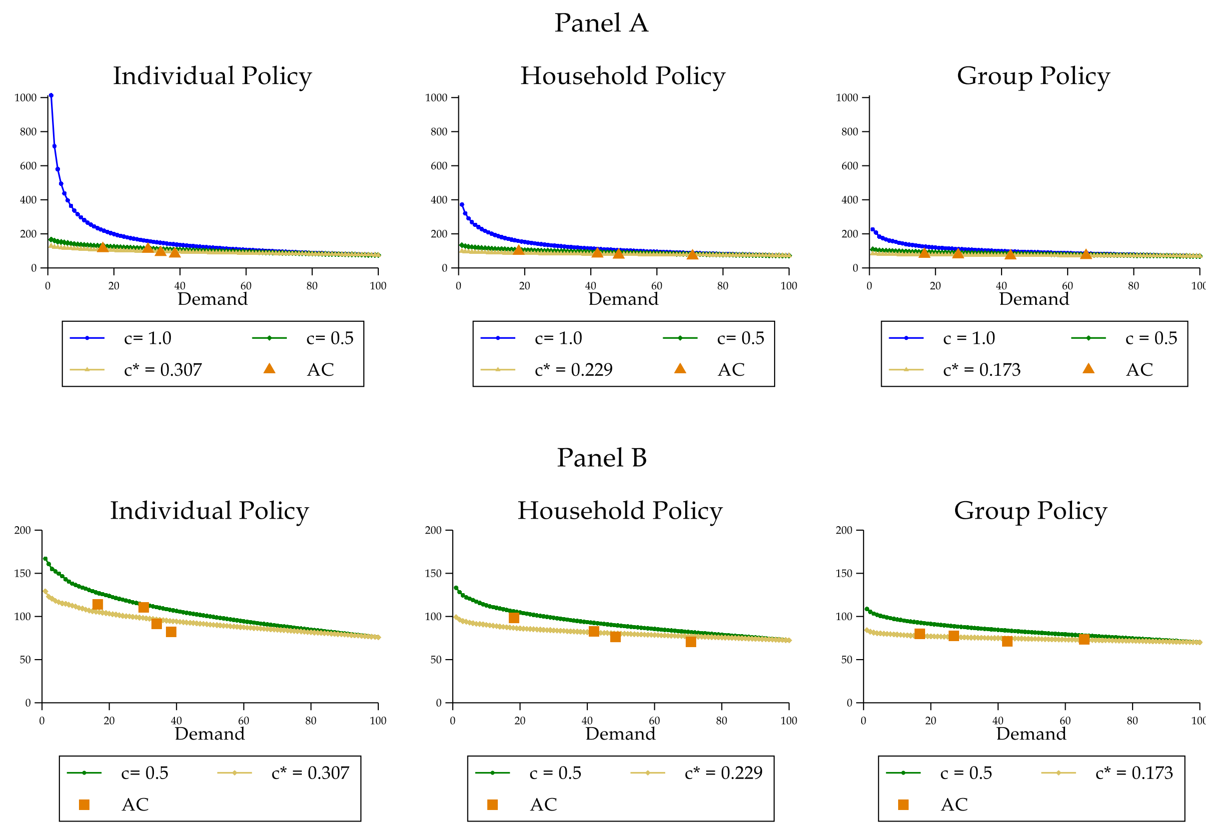
[39] In other words, we order the population of individuals (or households or groups) into 100 equally sized risk classes, within which we assume individuals (or households or groups) to be identical in terms if expected costs.
[40] Individuals (or households or groups) retain their expected cost, but we randomly permute their percentile rank, such that they would self-select into the insurance scheme earlier or later than predicted by their risk type. This feature implies that different observations of the same risk type might select into the scheme at different 'times'.

(i) selection governed by the standard deviation of the percentile error term (i.e. the extent to which individual (or households or groups) make errors in their insurance decision) and (ii) lower selection possibilities “by design” of the respective policy. As a result, holding constant the level of the percentile error, we would still predict cost curves to become less and less steep when moving from the individual to household and group scheme. This is illustrated by the green lines in the figures of Panel A and B, which show the result of a 0.5 correlation between the perfect selection and the fuzzy percentiles.

In a third step, we **fit the simulated curves to the empirically observed average cost points** in each policy adjusting the correlation accordingly. For an optimal fit, we use a weighted least squares criterion by selecting the correlation coefficient that minimizes the squared difference between the simulated and empirically observed average cost at the 4 observed demand points. Additionally, we weight the squared differences with the demand at each point, as this is exactly the factor by which the inverse of the standard error (of average costs) should differ. (Lower demand means that average costs are based on fewer observations and should be taken ‘less seriously’.) Basically, we run a WLS regression, where the regression line is not linear, but the result of a simulation (governed by a single free parameter). The average cost curves predicted by this exercise are shown as the yellow-beige lines in the figures below.

As described before, maximum possible selection (blue line in Panel A) decreases with higher levels of pooling. Not surprisingly, this holds also for fuzzy selection at a given “percentile error” in each policy (green line in Panel A and B). More interestingly, calibrating the simulation of fuzzy selection to our data, we find that selection under the household and group insurance is lower not only by design of the policies, but also because of a lower tendency to select within the scope possible. Specifically, the correlation between the perfect selection and the fuzzy percentiles decreases from 0.307 to 0.229 in the household and 0.173 in the group policy. Also note, that 0.173 is an upper bound for the correlation estimate, as full group uptake was not even mandated, such that our simulation (fully pooling groups) should overstate the policy design effect.

Figure E1 – Simulation of Selection



c = percentile correlation, AC = empirical average cost