7-9-2013

Can Consumers Make Affordable Care Affordable? The Value of Choice Architecture

Eric J. Johnson
Columbia University

Ran Hassin
Hebrew University of Jerusalem

Tom Baker
University of Pennsylvania Carey Law School

Allison T. Bajger
Columbia University

Galen Treuer
University of Miami

Can Consumers Make Affordable Care Affordable?

The Value of Choice Architecture

Eric J. Johnson[1*], Ran Hassin[2], Tom Baker[3], Allison T. Bajger[4], and Galen Treuer[5]

[1] Center for Decision Sciences and Graduate School of Business, Columbia University. New York, NY USA.

[2] Center for Rationality, The Hebrew University of Jerusalem. Jerusalem, Israel.

[3] Law School, University of Pennsylvania. Philadelphia, PA USA.

[4] Department of Psychology, Columbia University. New York, NY USA

[5] Abess Center for Ecosystem Science & Policy, University of Miami. Coral Gables, FL USA.

*To whom correspondence may be addressed: Eric J. Johnson (ejj3@columbia.edu)

Abstract

Starting this October, tens of millions will be choosing health coverage on a state or federal health insurance exchange as part of the Patient Protection and Affordable Care Act. We examine how well people make these choices, how well they think they do, and what can be done to improve these choices. We conducted 6 experiments asking people to choose the most cost-effective policy using websites modeled on current exchanges. Our results suggest there is significant room for improvement. Without interventions, respondents perform at near chance levels and show a significant bias, overweighting out-of-pocket expenses and deductibles. Financial incentives do not improve performance, and decision-makers do not realize that they are performing poorly. However, performance can be improved quite markedly by providing calculation aids, and by choosing a “smart” default. Implementing these psychologically-based principles could save purchasers of policies and taxpayers approximately 10 billion dollars every year.

Introduction

Starting this October, tens of millions of Americans, along with members of Congress, will participate in a grand experiment in consumer choice: They will select health insurance using a marketplace or *health insurance exchange* operated by states and federal governments as part of the Patient Protection and Affordable Care and Act. The success of these exchanges depends upon two related premises: First that consumers will be able to select the best policy for their needs, and second that price competition, driven by effective consumer choice, will lower prices. This hope is shared by divergent participants: Kathleen Sibelius, Secretary of Health and Human Services, and a Democrat, characterizes an exchange as "… a transparent, level playing field, driving down costs; … giv[ing] individuals and small businesses the same purchasing power as big businesses and a choice of plans to fit their needs." [1] Bill Frist, a physician and former Republican Senate Majority Leader, argues "State exchanges are good from a conservative standpoint because they involve consumer choice and markets." [2]

These premises are critical not only to the new exchanges, but also for all government administered health insurance markets and for the efficiency of privately provided benefit choices.[1] Yet, a large literature in psychology suggests that this may not be the case, since, as we shall see, these exchanges may not provide a helpful *choice architecture* to support decision-making. In this paper, we examine three related questions: Can people select the best policies? Do they know how well they are doing? Does the design of the sites change their performance?

Our results suggest there is significant room to improve these decisions. Without any intervention, respondents perform at near chance levels and show a significant bias, overweighting out-of-pocket costs and deductibles. Financial incentives do not improve performance, and decision-makers do not realize that they are performing badly. Without aids, only one population examined here, Columbia MBA students, perform reasonably well at this task. However, performance can be improved quite markedly by providing calculation aids, and by choosing a "smart" default, raising the performance of ordinary respondents to that of the MBA students.

Prior Research

The quality of choices on prior health insurance exchanges has been, at best, mixed. For example, when examining the exchanges implementing Medicare Part D, a prescription drug plan for seniors, Heiss, McFadden, and Winter [3] conclude "consumers are likely to have difficulty choosing among plans to fine tune their prescription drug coverage." Abaluck and Gruber [4] find that only 12.2% of seniors pick the most cost effective plan.

While the economics analysis of choice suggests that issues surrounding incentives and information may determine success, a more psychological analysis suggests that good decisions depend, critically, on subtle elements of how the choices are presented to the consumer, as described in an evolving literature on choice architecture [5-7]. Designing an exchange involves many design decisions including specifying the number and kind of options and attributes offered, determining the arrangement of options and the format and order of attributes, and selecting default options and computational aids.

The Massachusetts "Connector," an exchange operating since 2006, illustrates the impact of choice architecture: Before late 2009, the Connector simultaneously presented 25 plans from 6 insurance providers. In 2009, plans were reorganized into 3 tiers of coverage, categorized by premiums and out-of-pocket costs. Consumers first chose one of these levels and then viewed a smaller set of 6 standardized plans within a level. Work by Ericson and Starc [8] shows that this simple change markedly altered behavior: Consumers were increasingly sensitive to premium costs and out-of-pocket costs, changing market shares for some carriers by a factor of 2.

Thus, the advent of health exchanges presents a challenge: The choice could be daunting for consumers, resulting in suboptimal choices of policies that provide the wrong features or are too expensive. We are interested in how a prudent design of health exchanges based on psychological research could improve choice. We are also interested in a parallel question: Do people know if they are making good decisions? This is important because if people know that they are not doing well, they could seek assistance, potentially remedying their poor performance. If people are unaware of their inadequate performance, simply providing access to assistance will not improve their decision-making.

Choosing health insurance

When choosing insurance, consumers face two tasks. The first, which we do not examine, is to estimate their expected usage and out-of-pocket expenses for the upcoming year, and to consider the uncertainty around these quantities as a risky choice. The second is to select the right plan given their expected usage.

The following studies concentrated on people's ability to select cost effective policies and remove risk and usage prediction considerations. While economists analyze insurance choice by examining uncertainty, risk, and asymmetric information, we investigated the impact of psychological variables such as calculation costs as a major barrier to optimal choices. We examined a simplified version of the health insurance choice that allow us to assess the performance of choice architecture interventions, much like a wind tunnel might be used to evaluate candidate airplane designs.[2]

The reader might consider selecting the most cost effective plan in Figure 1, assuming, as did respondents in one of our experimental conditions, that he or she will make 9 doctor visits and incur $900 in out-of-pocket costs in the upcoming year.. This calculation might seem difficult, but some would argue that there might be heuristic strategies that perform well [9]. Yet we feel that there are two reasons for concern: First, users of these exchanges will be largely unfamiliar with selecting health insurance and will not be highly educated (seventy-seven percent will have a High School diploma or less) [10][3]. Second, this is an economically significant decision for these households: Even with subsidies, premiums will represent between 4 and 9.5% of the modest income of $48,529 for a family of 4 [10]. Consequently, mistakes may have large economic consequences.

Can Consumers Choose The Right Plan?

We examined consumers' decision-making abilities and conditions that might facilitate better decisions in a series of six framed field experiments [11], all but one using participants with demographics similar to those projected to use the exchanges. In addition to specifying the number of doctor visits one would make and the out-of-pocket costs one would incur in a given year, we also limited the number of plans available to either 4 or 8, a figure markedly lower than the number to be used in future exchanges (e.g., the Massachusetts Connector currently presents 47 plans, a discussion of choice set size) [12].

In all six experiments, subjects were asked to imagine they were choosing health insurance for a family of three—themselves, a partner and one child—with an anticipated number of doctor visits and out-of-pocket health care costs over the next year. Each subject was required to choose one plan from a set of 4 plans and one from a separate set of 8 plans. Plan set order was counter-balanced so half of the subjects chose from the 4-plan set first and half chose from the 8-

plan set first. Within each set of 4 and 8 plans, the display order of plans was also varied. In some experiments the number of visits or anticipated costs were varied (described below).

All studies shared certain features: All responses were collected online (see Table 1 in the Supporting Information) for demographics and other details). To isolate the effect of making a choice from a misunderstanding of the basic mechanics of health insurance, each session included explanations about insurance terms, such as premium, co-pay, and deductible, and required respondents to pass a comprehension test before proceeding (see methodological details for the content of these instructions and tests). Only those participants who passed this test were included in our analyses. Respondents viewed a table modeled after prototypes of exchanges (Figure 1) and chose an insurance plan. In Experiment 1 and 2, all components of prices resembled current prices and relationships among prices seen in existing and prototype exchanges.[4]

Experiment 1 provided a baseline measure of the proportion of people who choose the most cost-effective policy from 4 or 8 options. Figure 2 shows the outcomes from all experiments. The top half of each bar, in blue, represents the proportion of correct choices, and the bottom half, in red, plots the average dollar error, across respondents.[5] The dashed line represents expected choice quality by a random chooser. Panel A of Figure 2 shows a rather dramatic outcome: With 4 choice options, respondents selected the best option only 42 percent of the time, and made an average mistake of over $200 dollars. With eight options, they selected the correct option 21 percent of the time, a figure not different than chance ($p < .05$).

Experiment 2 added monetary incentives: Selecting the most cost-effective policy increased payment by $1 and generated an entry to a lottery that paid $200 to one correct chooser.[6] As can be seen in the next two bars of Figure 2 (A), incentives did not improve outcomes, and performance was close to chance.

This failure might be due to individuals' inability to perform the daunting calculations. One obvious intervention, used in Experiments 3 and 4, involves the use of a cost-calculator stating the annual total cost. In fact, several existing web sites, including Medicare.gov, provide such a tool. The present studies emphasized another important change designed to help diagnose the cause of poor performance: Plan attributes were drawn from an orthogonal experimental design, allowing us to estimate the weight participants give to the three cost components, premiums, co-payments and deductibles. According to economic theory, these costs should be approximately equally weighted since they all occur over the course of a year, and all contribute to the annual cost of the policy. However, past research has indicated that some costs (usually deductibles) are overweighted while others, like premiums are underweighted [4,13,14]. In addition, Experiment 4 also simplified the choice by removing quality information for half of the respondents.[7]

The results, shown in the third and fourth columns of Figure 2 (A), are not markedly different. Again respondents chose the most cost effective option less than half the time, and made large financial errors. The unaided decisions makers averaged errors of $611 in Experiment 3 and chose the correct option 32% of the time. Providing the calculators marginally helped but only in Experiment 4: Respondents provided with calculators chose the correct option 10.1% more often, and reduced the size of errors by $216, but still were only correct 47% of the time and made mean errors of $364.

Why was performance so poor? Answering this question may suggest interventions. While the math alone is challenging, the failure of the calculator to improve choice suggests that something else may be going on. Recall that past research shows that deductibles may be overweighted [13-16]. If this is the case, consumers may, arguably, have an incorrect notion of how deductibles contribute to overall cost. Figure 3 shows the weight given to each price component in Experiment 4. The results show a strong and consistent bias, compared to the ideal of equal weighting: Participants overweight the out-of-pocket costs and deductibles. Their improved performance with calculators is due, in part, to reducing this bias, as illustrated by the red bar. In other words, the presence of a calculator suggests that respondents came closer to treating all dollars as having the same cost.

Is this task simply impossible? Experiment 5 used a very different population to see how highly trained, financially literate individuals might do. We presented MBA students enrolled in a class on consumer finance with the same task as in Experiment 4. The average GMAT of students at this school was 716, and 59% of students came from consulting or financial services and related fields. As seen in the first column of Figure 2(B), they performed appreciably better, choosing the right option 73% of the time, and making an average mistake of $126. Their self-reports of how they accomplished the task are interesting: Forty percent reported using excel (this group performed quite well, selecting the correct option 85% of the time, and making an average error of only $47). This suggests that having *both* the right mental model *and* the ability to execute these calculations may be a basic requirement to make good choices.

In Experiment 6, we explored the possibility that mental models in conjunction with different possible interventions would produce good performance by individuals who will be using the exchange. To ensure understanding, and encourage the use of the correct mental model, all conditions received a tutorial about computing the annual cost and completed a quiz requiring one correct choice. We believe that this kind of *just-in-time* education might help both aided and unaided choice, and further eliminate a lack of knowledge (as opposed to computational complexity) as a barrier to better performance. We then compared this control condition to four different manipulations. An incentive group received a more extreme and sophisticated incentive regime that penalized respondents 10 cents for every $100 extra that was spent on insurance. We contrasted this to three choice architecture interventions. The first provided a *calculator,* explained what the calculator did, and tested that understanding. The second provided a *smart default* that preselected the most cost effective options given individuals' usage. Individuals could, and did, change that selection if desired. Finally, we combined *defaults and calculators*. The presence of incentives and our choice architecture manipulations allowed us to compare the cost effectiveness of these interventions.

The last four bars in Figure 2(B), which average data over the number of options, show that the treatments vary widely in effectiveness. The controls, despite having received instruction and tests of understanding, chose about as well as respondents in prior experiments. The second bar indicates that incentives did not have a significant effect on outcomes, even though individuals in the incentive condition took 38% longer to make their decisions, a significant increase relative to controls. Calculators (with education), in contrast, produced better decisions, having resulted in a significant decrease in the size of the loss and an increase in the proportion correct. The smart default option had a similar effect, as it reduced losses and increased the percentage correct.[8] Finally, when combined, the defaults and calculators seemed to complement each other, leading to performance levels that are comparable to those of the highly trained MBA students. This last result suggests, perhaps, that the use of calculators justifies the selection of the default, increasing the transparency of their selection. It also suggests that providing just-in-time education along with calculation and choice aids produces better performance.

While these interventions are effective, are they appreciated? This is an important question about meta-cognition that has important policy implications: If deciders are doing badly and need help, do they realize it? When they get help, do they appreciate it? We asked respondents how confident they were of making the correct choice in Experiments 3, 4 and 6, using a 1-9 point scale: While participants performed poorly, this was not reflected in their confidence ratings (mean rating 6.6, 6.75, and 7.6, respectively, in Experiment 3, 4, and the control condition in Experiment 6) and there was no correlation between these ratings and selecting the most cost effective plan (.09 averaged across these three studies). It appears that individuals did not realize the need for these interventions. They also did not appreciate the effect of the interventions consistently: Calculators created a marginal increase in confidence (+.23 relative to control, p <.06); defaults did not (+.14, p >.2). Finally, incentives did not increase performance, but they did increase effort and produced an unwarranted increase in confidence (+. 34, p < .03). All told, the picture that emerges is that of overconfident decision-makers who do poorly and do not realize it, and who do not realize that decision-architecture helped.

Conclusion

Our results present a bad news/good news story of particular importance. The bad news: Consumers left to their own devices seem to make large errors when choosing health insurance, suggesting that they will select options that are not cost efficient and they seem to be unaware of their failure. If consumers cannot identify cost efficient plans, then the Exchanges will not produce competitive pressures on health plan costs, one of the main advantages of relying upon choice and markets. It is possible that other factors, such as advertising and brokers may make the market more or less competitive. The impact of such institutions is a question for further research.

The good news is that we have demonstrated that exchange designers can improve consumers' performance markedly through the use of just-in-time education, smart defaults, and cost calculators. This list of potential design improvements is not exhaustive, and there are many other interventions that may improve choices. These include sorting by cost, the presence of quality cues, or limiting the number of options to those that meet criteria of cost effectiveness. These suggestions are not without precedent: In evaluations of Medicare Part D, Abaluck and Gruber [4] suggest that "restricting the choice set to the 3 lowest average cost options would have likely raised welfare for the elders". However, this limits consumer choice and we note that some design features, such as calculators, improve outcomes by make choice easier, without impinging upon consumer sovereignty.

The results of these studies allow us to approximately estimate the benefits of these kinds of choice architecture interventions. These estimates should be treated with appropriate caution because they are based on the particular set of policies used in our studies. However, our control group in Experiment 6 made an average error of $533, roughly 10% of the cost of the cheapest policy, compared to an error of $77 when both the default option and calculator were available, producing an estimated value to these features of $456 dollars per decision. At the individual level, unaided choice is expensive: It represents about 1% of the income of the proposed median buyers' household income. But, in the aggregate, an error of $456 represents staggering sums: If 20 million individuals make choices using the exchanges, a figure suggested by Congressional Budget Office estimates, unaided choice represents a cost to consumers of $9.12 billion dollars each year. Since almost all of these policies are subsidized through tax credits, good choice architecture would produce substantial savings to the federal budget and taxpayers.

This sizable impact is more significant since the improvement is largely a function of psychological factors that can be implemented inexpensively by being built into the choice engines powering the exchanges. Clearly, further research identifying the best mix of choice architecture tools in exchanges is both scientifically interesting and economically justified. While the success of the health exchanges will depend, in part, on the provision of cost-efficient products, it also will depend, to some extent, on the design of exchanges that will allow consumers to identify them and to choose plans that are good fit to their needs. Ignoring the impact of choice architecture and the psychological factors we examine could be an expensive mistake.

Acknowledgement: We thank Margaret Lee and Jon Westfall for their assistance. We also thank participants at the Sloan and Sage Consumer Finance Working Group and the Preference as Memories group for helpful comments at presentations. The authors have no conflicts of interest, and data from all experiments will be archived at www.plosone.org

References

1. Department of Health and Human Resources (2011) HHS and states move to establish affordable insurance exchanges, give Americans the same insurance choices as members of congress. Available: http://www.hhs.gov/news/press/2011pres/07/20110711a.html. Accessed 1 August 2012.
2. Alonso-Zaldivar R (2011) Health care overhaul debate now shifts to state. Available: http://www.newsmax.com/Newsfront/HealthCareStates/2011/01/22/id/383588. Accessed 11 July 2012.
3. Heiss F, McFadden D, Winter J (2010) Mind the gap! Consumer perceptions and choices of Medicare Part D prescription drug plans. In Wise DA, editors. Research findings in the economics of aging. London: Chicago University Press. pp. 413-381
4. Abaluck J, Gruber J (2011) Heterogeneity in choice inconsistencies among the elderly: Evidence from prescription drug plan choice. The American Economic Review 101: 377-381.
5. Goldstein DG, Johnson EJ, Herrmann A, Heitmann M (2008). Nudge your customers toward better choices. Harvard Business Review 86: 99-105.
6. Thaler RH, Sunstein CR (2008) Nudge: Improving decisions about health, wealth, and happiness. New Haven: Yale University Press.
7. Johnson EJ, Shu SB, Dellaert BGC, Fox C, Goldstein DG, et el. (2012) Beyond nudges: Tools of a choice architecture. Marketing Letters 23: 487-504.
8. Ericson KM, Starc A (2012) Heuristics and heterogeneity in health insurance exchanges: Evidence from the Massachusetts connector. The American Economic Review 102: 493-497.

9. Gigerenzer G, Todd PM, the ABC research group (1999) Simple heuristics that make us smart. New York: Oxford University Press.
10. Kaiser Family Foundation. (2011) A profile of health insurance exchange enrollees. Available: http://www.kff.org/healthreform/8147.cfm. Accessed 21 August 2012
11. Harrison GW, List JA (2004) Field experiments. Journal of Economic Literature 42: 1009-1055.
12. Reutskaja E, Nagel R, Camerer CF, Rangel A (2011) Search dynamics in consumer choice under time pressure: An eye-tracking study. The American Economic Review 101: 900-926.
13. Johnson EJ, Hershey J, Meszaros J, Kunreuther H (1993) Framing, probability distortions, and insurance decisions. Journal of Risk and Uncertainty 7: 35-51.
14. Sydnor J (2010) (Over) insuring modest risks. American Economic Journal: Applied Economics 2: 177-199.
15. Kunreuther H, Pauly M (2005) Insurance decision-making and market behavior. Foundations and trends in microeconomics 1: 63-127.
16. Barseghyan L, Prince J, Teitelbaum JC (2011) Are risk preferences stable across contexts? Evidence from insurance data. The American Economic Review 101: 591-631.

Figure Legends

Figure 1. A decision display used in experiments 4. Respondents saw either 8 (pictured) or 4 options.

Figure 2. The percentage of choices of the most cost effective option (above zero, in blue,) and the average error made by respondents (in red, below the zero line). A dashed line for each condition represents the performance of a random chooser, and the error bars represent 95% confidence intervals. Darker shades denote the provision of calculators. Panel (A) represents the results of Experiments 1-4 collapsing across other manipulations (see SM). Panel (B) represents the results of a sample of highly educated MBA students (Experiment 5), and of individuals from the target population, when given different choice architecture interventions. For (b) the random response threshold ($1264) exceeds the lower limit of the graph.

Figure 3. Premiums, deductibles, and co-payments, both without calculator (blue) and with calculator (red). The decline in odds of being chosen for each increase in $100 in annual cost for the three cost components in experiments.

Footnotes

[1] Many of the lessons also apply to the presentation of plans by private employers. There are many differences between employers sponsored insurance and the health exchanges. Health exchanges typically have more choices, and prices are set by the market rather than negotiation. However, they share the critical role of consumer choice in finding plans that are appropriate for consumer needs and cost-effective and in having many design decisions that will affect choice.

[2] While risk considerations are, no doubt, important, they are likely to make performance worse, not better than we observe, and perhaps make our interventions more effective. That is, To select the most cost-effective health insurance policy unaided consumers must, for each plan:

1. Consider the total premiums for the year
2. Combine the copayments and the expected number of visits,
3. Include the minimum of the deductible and their out-of-pocket costs, and
4. Find the option with the lowest annual cost. For equal monthly premiums this is

   (12*Monthly Premium) + (N of visits * Copay) + min(Out of Pocket Costs, Deductible).

[3] These issues are likely to be even more important on the new health exchanges, since many, 97% according to some estimates [7], will be buying health insurance for the first time and may lack experience and relevant knowledge.

[4] In addition, Experiments 1-2 varied, between respondents, the number of visits, while Experiments 3-5 varied the level of out of pocket costs. For the sake of brevity, we will not discuss these results here.

[5] The average cost is a non-continuous measure, and might be affected by the specific costs associated with the offered plans. We model all choices using a logistic model with

indicator variables for categorical variables, and an Analysis of Variance to test significance for the error variable. Please see Supporting Information for more details.

[6] Including the lottery, the expected value of selecting each right option was $1.88, and performance was unrelated to time spent on the task.

[7] This information was not diagnostic, since all options had the same total quality, and the choices made by respondents confirmed this.

[8] It is important to note that the performance of defaults is not simply due to their mindless selection. First note that a significant proportion of people (21%) chose to not take the default by actively selecting another option. Second, those choosing the default option did take a significant amount of time to choose a policy. Across the entire study, non-default choosers required 443 seconds to complete the study, and choosers required 348 seconds. Concentrating on only the choice screen, default choices took 58% and 65% as long as the no-default condition for the 4 and 8 option conditions, respectively.

Figure 1.

Health Plan	Monthly Premium	Doctor Visit Copay	Annual Deductible
A	$601	$21	$800
B	$408	$24	$1400
C	$583	$32	$1400
D	$593	$27	$1600
E	$578	$19	$900
F	$416	$33	$700
G	$419	$24	$1000
H	$393	$18	$1500

Figure 2.

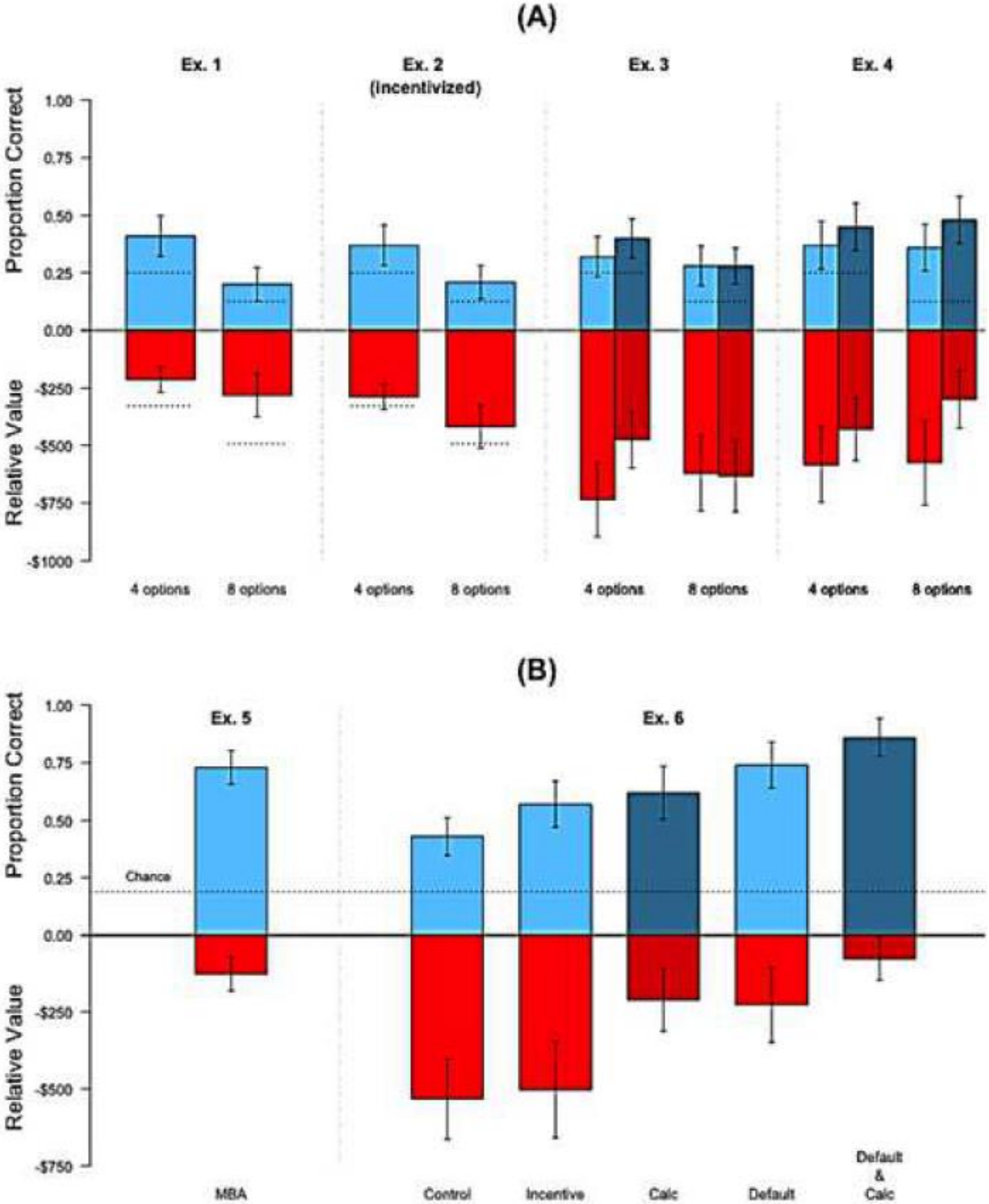
(A)
Ex. 1
Ex. 2
(incentivized)
Ex. 3
Ex. 4
Proportion Correct
Relative Value
1.00
0.75
0.50
0.25
0.00
-$250
-$500
-$750
-$1000
4 options
8 options
4 options
8 options
4 options
8 options
4 options
8 options
(B)
Ex. 5
Ex. 6
Proportion Correct
Relative Value
1.00
0.75
0.50
0.25
0.00
-$250
-$500
-$750
Chance
MBA
Control
Incentive
Calc
Default
Default
&
Calc

Figure 3.

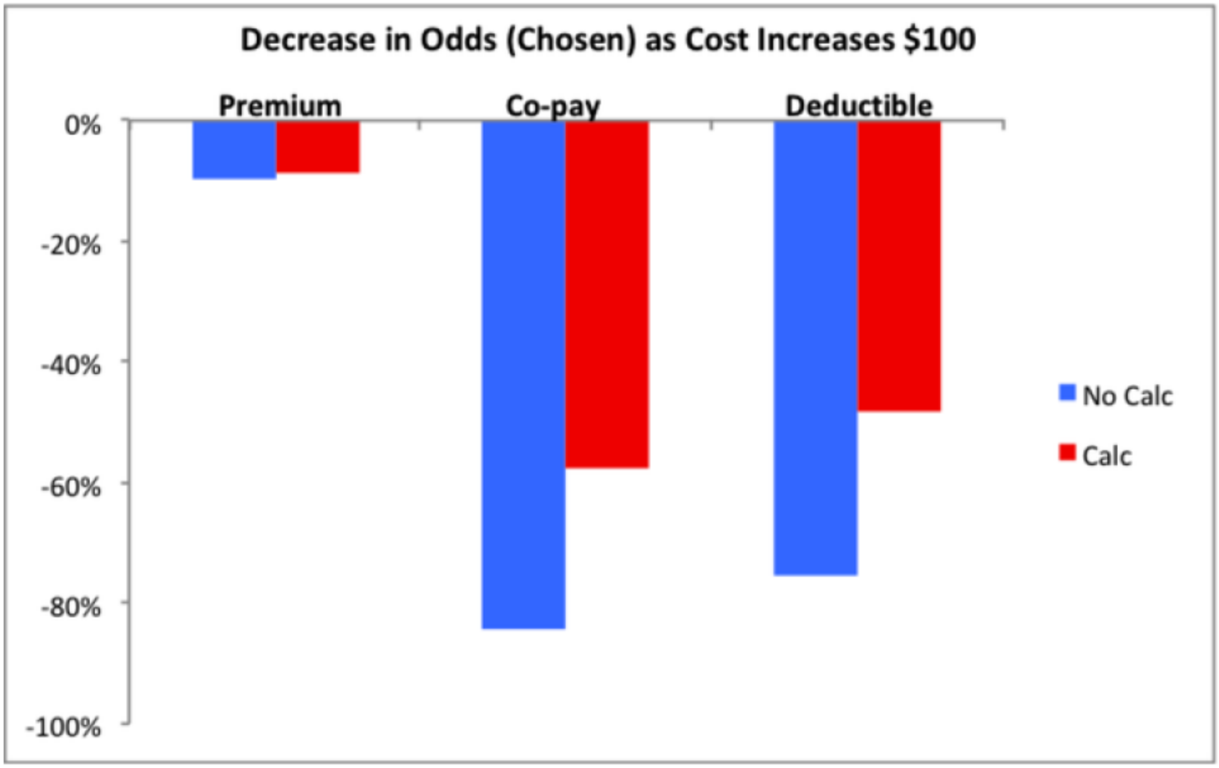
Decrease in Odds (Chosen) as Cost Increases $100
Premium
Co-pay
Deductible
0%
-20%
-40%
-60%
-80%
-100%
No Calc
Calc

Table S1.

Demographics

	Experiment 1	**Experiment 2**	**Experiment 3**	**Experiment 4**	**Experiment 5**	**Experiment 6**
N	120	131	234	177	76	330
Age (mean)	48	48	45	37	29	50
Gender	Female (72%)	Female (62%)	Female (56%)	Female (55%)	Female (46%)	Female (64%)
Marital Status	Married (56%)	Married (47%)	Married (47%)	Married (40%)	Married (32%)	Married (44%)
Children	0 – 2 (76%)	0 – 2 (75%)	0 – 2 (79%)	0 – 2 (87%)	0 – 2 (4%)	0 – 2 (74%)
Income	\$35K - \$49K (51%)	\$50K - \$99K (26%)	\$50K - \$99K (27%)	\$50K - \$99K (28%)	> \$99K (35%)	\$20K - \$49K (67%)
Education	HS diploma (35%)	HS diploma (37%)	HS diploma (42%)	Bachelor's (35%)	Master's (66%)	HS diploma (40%)
Race	White (87%)	White (83%)	White (83%)	White (81%)	White (59%)	White (87%)
Health Insurance	Yes (66%)	Yes (66%)	Yes (75%)	Yes (65%)	Yes (97%)	Yes (72%)
Political Affiliation	Democrat (32%)	Democrat (30%)	Democrat (35%)	Democrat (36%)	Democrat (46%)	Democrat (30%)

Supporting Information

Methods and Materials

Online surveys

This study was approved by and carried out under Institutional Review Board at Columbia University, and all participants in each experiment provided written informed consent. Subjects were recruited from three different populations. A total of 815 subjects were recruited for Experiments 1, 2, 3 and 6 by third party survey companies from a demographic population similar to the one likely to use online insurance exchanges. 177 subjects were recruited for Experiment 4 through Amazon Mechanical Turk (MTurk), restricted to Americans in the 97th percentile of all workers. 76 MBA students from a major research university were recruited for Experiment 5 through a class exercise. In each experiment subjects completed an online survey that included three sections: 1) an introduction to health insurance policies and comprehension quiz; 2) the choice task in which subjects were instructed to imagine they were buying health insurance for their family of three with specific expected health costs and choose one of several plans; 3) post-task individual difference questions covering strategies used, demographics and personal health insurance information.

Incentives

Two different incentive schemes were employed. In Experiment 2, subjects received an additional $1 for answering correctly on each of the choice questions. In addition, participants were given an entry into a $200 drawing for each correct answer. If participants chose the most cost effective option for both choice questions, they received $2 and two entries into the $200 drawing. In Experiment 6 we attempted to increase the incentive through loss aversion and tying payment to the cost of the plans subjects chose. A portion of the subjects were endowed with $7 which was reduced by 10 cents for each $100 difference in annual cost between the insurance plan they chose in the cheapest plan.

Preliminary Data Analysis

To ensure attention, each subject was required to pass a version of Oppenheimer and colleagues' Instructional Manipulation Check [1] before learning about health insurance policies. Subjects were removed from analysis for two reasons: 1) if they failed to demonstrate understanding by correctly answering multiple choice questions about health insurance policies in three tries, and 2) if they completed the decision task too quickly (below the 5th percentile or 10 seconds) or too slowly (above the 99th percentile or above 1100 seconds). 57 subjects (32%) were excluded from Experiment 1 and 74 (36%) from Experiment 2 for failing to demonstrate understanding. 166 subjects (41%) were excluded from Experiment 3, 34 for time, 55 for failing to demonstrate understanding and 11 for both. 18 (10%) from Experiment 4 were excluded, 10 for time, 2 for failing to demonstrate understanding and 6 for both. No subjects were excluded from Experiment 5 and 117 (26%) from Experiment 6, 17 for time, 79 for failing to demonstrate understanding and 21 for both. The variation in exclusion rates is consistent with anticipated differences in online experimental populations. Analyzing the data with the excluded respondents does not result in meaningful changes, but does add increased random error in almost all of our measures.

Experimental Designs

In all six experiments, subjects were asked to imagine they were choosing health insurance for a family of three, themselves plus a partner and one child, with an anticipated number of doctor visits and out-of-pocket health care costs over the next year. Each subject chose one plan from a set of 4 plans and one from a separate set of 8 plans. Order was counter-balanced so half of the subjects chose from the 4 plan set first and half chose from the 8-plan set first. Within each set of 4 and 8 plans the order plans were displayed was also varied. In some experiments the number of visits or anticipated costs were varied, described below.

Experiments 1 and 2 were designed to reflect the design of real insurance policies with negatively correlated premiums and deductibles. In Experiments 3, 4, 5 and 6 an orthogonal experimental design was used, in which all cost components of the plans presented are uncorrelated. This enabled us to estimate the weight given to the three cost components: premium, co-payment, and deductible. Quality information was provided in for respondents in Experiments 1-3 and half of respondents in Experiment 4. This information was designed to be independent of the price components. To simplify the choice, quality was not presented in subsequent experiments. Experiment 4 allowed us to establish that the quality measures had no effect on performance.

Data analysis

Data was analyzed in the *r* software package [2] using analysis of variance (Type III, marginal sum of squares), with the size of the error made as the dependent variable, or the equivalent model as a binomial logistic regression, with selection of the cheapest plan (yes or no) as the dependent variable. All models included as factors choice set (4 options or 8 options) and order presented (first or second). Copay usage (high or low) was included as an additional factor in Experiments 1 and 2, and out of pocket expense (high or low) was included in Experiments 3, 4, and 5. Annual cost calculator (present or not) was included in Experiments 3, 4, 5, and 6. In an alternative model incentive and default plus annual cost calculator were included as contrasts to the control group (no calculator, no default, no incentive) in Experiment 6.

Attribute Analysis

We also sought to understand the importance of the cost measures associated with each plan in driving selection of a given plan, and especially how the importance of these measures varied depending on the values of the experimental variables in a given choice scenario. In order to address these questions, we fit a logistic regression on a transformed version of the original dataset in which each choice among a choice set of k options was represented with k-1 rows and the dependent variable was defined as whether or not a given choice had been selected in that choice set. The most expensive option was the reference category; due to multicollinearity among the choice measures in the 4-option grid, analyses were restricted to the 8-opton grid data.

Detailed Results

Experiment 1

Subjects were assigned to high (15) or low (5) number of doctor visits. Overall 30% (chance = 19%) chose the most cost effective option and made an average mistake of $248 (chance = $410). Given 4 options 41% (chance = 25%) chose the most cost effective option and made an average mistake of $214 (chance = $328). Given 8 options 20% (chance = 12.5%) chose the most cost effective option and made an average mistake of $283 (chance = $492). Analysis of variance (ANOVA) with the size of error made as the dependent variable found a main effect of usage ($F(1,229) = 11.77$, $p < 0.001$) and no significant effect of grid ($F(1,229) = 1.68$) or order presented ($F(1,229) = 0.08$). High usage (15 visits), with 4 options 69% chose the most cost effective plan and made an average mistake of $89, with 8 options 38% chose the most cost effective plan and made an average mistake of $196. Low usage (5 visits), with 4 options 19% chose the most cost effective plan and made an average mistake of $311, with 8 options 6% chose the most cost effective plan and made an average mistake of $351.

Experiment 2

Experiment 2 used the same design as Experiment 1 with the added \$1 per correct answer and \$200 lottery incentive described above, in the methods section. Overall 29% chose the most cost effective option and made an average mistake of \$354. Given 4 options 37% chose the most cost effective option and made an average mistake of \$289. Given 8 options 21% chose the most cost effective option and made an average mistake of \$419. ANOVA with size of error made as the dependent variable found a main effect of usage ($F(1,249) = 12.75$, $p < 0.001$), a marginally significant effect of grid ($F(1,249) = 3.58$, $p < 0.1$) and no effect of order presented ($F(1,249) = 0.37$). High usage (15 visits), with 4 options 57% chose the most cost effective plan and made an average mistake of \$150, with 8 options 34% chose the most cost effective plan and made an average mistake of \$323. Low usage (5 visits), with 4 options 17% chose the most cost effective plan and made an average mistake of \$427, with 8 options 8% correct and made an average mistake of \$515.

Combining Experiment 1 and 2, a logistic regression with choice of the cheapest option as dependent variable found main effects of grid X^2 $(1,N = 502) = 23$, $p < 0.001$ and usage X^2 $(1,N = 502) = 90$, $p < 0.001$ but not order X^2 $(1,N = 502) = 0.1$ or incentive X^2 $(1,N = 502) = 1.3$, results further supported by binomial order regression. An ANOVA with size of error made as dependent variable found a significant (negative) main effect of the incentive added in Experiment 2 ($F(1,481) = 7.75$, $p < 0.01$) as well as usage ($F(1,481) = 24.4$, $p < 0.001$) and grid ($F(1,481) = 5.27$, $p < 0.05$) but not order presented ($F(1,481) = 0.1$). Incentives added in Experiment 2 are associated with a \$118 *increase* in losses. Though they did not help them perform better, incentives did significantly increase the time subjects spent choosing a plan from a mean of 85 second to 98 second ($t(230) = -2.1$, $p < 0.05$).

Experiment 3

In Experiment 3 doctor visits were fixed at 11, and subjects were assigned to high ($2200) or low ($900) anticipated out-of-pocket expenses. Half of subjects had the total annual cost of each plan calculated and presented along side the three cost components. Overall performance was poor with 32% choosing the most cost effective option and an average mistake of $611 (chance = $1262). Given 4 options 36% chose the most cost effective option and made an average mistake of $594 (chance = $1191). Given 8 options 28% chose the most cost effective option and made an average mistake of $627 (chance = $1337). In the high expense condition 29% chose the most cost effective option and made an average mistake of $562. In the low expense condition 35% chose the most cost effective option and made an average mistake of $654. Adding an annual calculator helped non-significantly. With a calculator 34%, compared with 30% without, chose the most cost effective option, and the average mistake was reduced by $123. A logistic regression with an ANOVA test for significance on choice of the cheapest option found a main effect of grid X^2 $(1, N = 468) = 4$, $p < 0.05$ and no significant effect of order X^2 $(1) = 0.06$, expense X^2 $(1) = 2.3$ or annual calculator X^2 $(1) = 1.4$. An ANOVA on the size of error made found no main effects of grid ($F(1,426) = 0.2$), order ($F(1,426) = 1.2$), expense ($F(1,426) = 0.8$), or annual calculator($F(1,426) = 1.8$).

Experiment 4

In Experiment 4 the design of Experiment 3 was implemented with a more highly educated and online experience MTurk population. For half of the subjects non-informative quality ratings were removed to reduce noise which had a no significant effect on performance. Though the MTurks did better than previous subjects, overall only 42% chose the most cost effective option and the average mistake was $466. Given 4 and 8 option grids subjects performed virtually the same. 41% and 42% chose the most cost effective option respectively and made average mistakes of $502 and $429. Subjects provided with annual cost calculations performed significantly better, 47% vs. 37% without the calculations and an average reduction in mistakes made of $216. Subjects in the high expense ($2200) condition, in which the deductible level was reached with every plan, performed significantly better than those in the low expense ($900) condition. 48% vs. 36% chose the most cost effective option with a $186 average reduction in mistake made. These differences in means are supported by a logistic regression with an overall ANOVA test on choice of the cheapest option that found a main effect of annual calculator X^2 (1) = 3.86, $p = 0.05$ and expense X^2 (1, p < = 0.05) but no significant effect of grid X^2 (1) = 0.2, quality ratings X^2 (1) = 2.7, p = 0.1 or order X^2 (1) = 1.3. Binomial order regressions provide further support for these findings. Similar results are found in MANOVA with mistake made. Annual calculator ($F(1,325) = 7.8$, $p < 0.01$) and expense ($F(1,325) = 6.32$, $p < 0.05$) are significant but grid ($F(1,325) = 1.27$) , quality ratings ($F(1,325) = 1.75$), and order ($F(1,325) = 1.69$) are not.

The attribute importance analysis in Study 4 revealed that participants weighted a $100-increase in annual copay cost and annual deductible cost significantly more than a similar increase in annual premium cost (Figure 3). A $100 increase in premium cost was associated with a 6% decrease in the odds of selecting a given plan (OR=.94, 95% CI: .89-.99); a $100 increase in copay cost with a 57% decrease in the odds of choosing a plan (OR=.43, 95% CI: .12-1.58); and a $100 increase in deductible cost with a 61% decrease in odds (OR=.39, 95% CI:.16-.87). Interaction effects suggested that subjects weighted premium more in the calculator condition (OR=1.89, 95% CI: 1.80-1.99, with calculator vs. OR=.94, 95% CI:.90-.99, without calculator, premium*calculator interaction *p*=.06). Copay was weighted less in the second trial (*p*=.01) and deductible was weighted less in the low expense condition (*p*=.002).

Experiment 5

To investigate if anyone can successfully perform this task, in Experiment 5 MBA students completed Experiment 4 with two changes. Subjects were instructed to “choose the most cost-effective plan” and quality ratings were removed in all conditions. The MBA students performed much better than other populations. Overall 73% chose the most cost effective option with an average loss of $126. Given 4 and 8 options subjects performed virtually the same. 72% and 75% chose the most cost effective option respectively with averages mistakes of $157 and $95. The addition of an annual cost calculator had no significant effect. This may be because 86% of the MBA students reported using Excel, a calculator or pen and paper to find the cheapest option. Subjects in the high expense ($2200) condition, in which the deductible level was reached with every plan, performed significantly better than those in the low expense ($900) condition. 83% vs. 64% chose the most cost effective option with a non-significant $65 average reduction in mistakes made. Though identifying the cheapest option appears to be easier in the high expense condition, with the use of excel and other tools the MBA students were still able to minimize their losses in the low expense condition.

A logistic regression with choice of the cheapest option and an ANOVA on the size of error made as dependent variables support these results.

Experiment 6

In an attempt to get subjects who fit the profile of people likely to use the health insurance exchanges to perform as well as MBA students, we administered a version of Experiment 5 to an online panel with incomes under $50,000. To increase power, all subjects were assigned to the high expense ($2200) condition. An incentive-compatible condition (described above) was administered to 47% of the participants. 46% were provided with annual cost calculations.

In the control condition (no calculator, no default, no incentive) 43% chose the most cost effective option, similar to overall performance in Experiment 4, with an average mistake of $533. In the incentive compatible condition (no calculator, no default) 57% chose the most cost effective option but average mistake dropped by a non-significant $30 ($t(203) = 0.28$, $p = 0.8$). Providing annual cost calculations increased choice of most cost effective option to 62% and decreased the average mistake by $323. Default selection of the most cost effective option significantly increased its selection to 74% and decreased the average mistake by $306. The combination of annual cost calculator and defaults significantly increased selection of the most cost effective option to 86% and significantly decreased the average mistake by $456.

Logistic regression on choice of the cheapest option supports these findings. Default X^2 (1) = 39, p < 0.001 and annual calculator X^2 (1) = 8.9, p < 0.01 are significant while grid X^2 (1) = 0.0, incentive X^2 (1,N = 660) = 0.9 and order X^2 (1) = 0.18 are not, results also found in binomial order regressions. ANOVA on size of error made also support significant positive impacts of the default ($F(1,640) = 17$, $p < 0.001$) and annual calculator ($F(1,640) = 27$, $p < 0.001$). Grid ($F(1,640) = 4.1$, $p < 0.05$) was also significant, but incentive (F(1,640) = 0.02) and order ($F(1,640) = 0.05$) were not.

The individual effect of defaults, annual calculators, and incentives were further confirmed through contrasts analysis in regression controlling for grid and order. Incentives improve choice of the cheapest option X^2 (1,N = 660) = 9.9, $p < 0.01$ but not mistake made ($F(1,642) = 1.5$). Smart defaults improved both choice of the cheapest option X^2 (1) = 29, p <

0.001 and size of error ($F(1,642) = 18$, $p < 0.001$). The annual calculator also improved both choice of the cheapest option X^2 (1) = 17, $p < 0.001$ and size of error made ($F(1,642) = 19.6$, p < 0.001). Defaults combined with annual calculator improved helped subjects perform even better in choosing the cheapest option X^2 (1) = 51, $p < 0.001$ and reducing mistakes ($F(1,642) = 38.5$, $p < 0.001$).

References

1. Oppenheimer DM, Meyvis T, Davidenko N (2009) Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology 45: 867-872.
2. R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Reference.