

We All Want It, but We Don't Know What It Is: Toward a Standard of Affordability for Health Insurance Premiums

Peter Muennig
Bhaven Sampat
Nicholas Tilipman
Lawrence D. Brown
Sherry A. Glied
Columbia University

Abstract The 2010 Patient Protection and Affordable Care Act (P.L. 111-148), or ACA, requires that U.S. citizens either purchase health insurance or pay a fine. To offset the financial burden for lower-income households, it also provides subsidies to ensure that health insurance premiums are affordable. However, relatively little work has been done on how such affordability standards should be set. The existing literature on affordability is not grounded in social norms and has methodological and theoretical flaws. To address these issues, we developed a series of hypothetical vignettes in which individual and household sociodemographic characteristics were varied. We then convened a panel of eighteen experts with extensive experience in affordability standards to evaluate the extent to which each vignette character could afford to pay for one of two health insurance plans. The panel varied with respect to political ideology and discipline. We find that there was considerable disagreement about how affordability is defined. There was also disagreement about what might be included in an affordability standard, with substantive debate surrounding whether savings, debt, education, or single parenthood is relevant. There was also substantial variation in experts' assessed affordability scores. Nevertheless, median expert affordability assessments were not far from those of ACA.

The 2010 Patient Protection and Affordable Care Act (P.L. 111-148), or ACA, requires that U.S. citizens purchase health insurance or pay a fine, but it also provides subsidies to make health insurance premiums affordable for the less affluent (White House 2010). *Affordability*, as used in

Thanks to the Robert Wood Johnson Foundation (and particularly Brian Quinn and Andrew Hyman) for research support. Thanks to Richard Frank and Sharon Lise Normand for advice on the Delphi method. Thanks as well to all our panelists for giving us so much insight into this process.

Journal of Health Politics, Policy and Law, Vol. 36, No. 5, October 2011
DOI 10.1215/03616878-1407640 © 2011 by Duke University Press

public policy, is a normative concept describing how much people in a given society typically believe others should pay for a good or service. This concept necessarily accounts for the financial resources needed to maintain a minimally acceptable standard of living. But defining a minimal standard of living is an inherently subjective process.

Because affordability is a normative standard, it can be challenging to assess. To date, two main strategies have been used to assess the affordability of health insurance: behavioral analysis and surveys of the uninsured population (Bundorf and Pauly 2006; Blumberg et al. 2007; KFF 2009). These approaches, though, do not adequately capture differences in subjective views of affordability, particularly in the context of a reformed health care system.

We report on the results of an analysis using an alternative strategy—a Delphi experiment—that elicited the views of a panel of experts.

Existing Methods to Evaluate Affordability

Behavioral Approach

There are two main strategies used to assess health insurance affordability. One strategy is to examine data on the current behavior of similarly situated people. This approach, initially developed by Bundorf and Pauly (2006), who call it a behavioral definition, suggests that health insurance is affordable to a family if 50 percent or more of similarly situated families purchase insurance. Their definition of *similarly situated* incorporates family income, family composition, and measures of the price of health insurance (either prices themselves or expected expenditures, unadjusted for geography). Blumberg and colleagues (2007) adopted a similar standard in their analysis of affordability, which contributed to the development of the standard used in Massachusetts's health reform law.

The behavioral approach, however, has three problems. First, unobserved heterogeneity in the prices faced by different families (both those in the group and nongroup health insurance markets) makes this calculation empirically suspect (Bradley 2008). For example, people who do not hold coverage but who appear to be able to afford it may work in jobs that do not provide health insurance coverage. They may also work in jobs that offer coverage with very high premium sharing. On the other hand, those who hold coverage may be employed in jobs that offer coverage with low premium sharing. For these reasons, it is particularly challenging to identify insured “nonafforders” using available data. Information on insurance

premiums can be used to impute a price of insurance for those who are uninsured, but no available data provide information on the price of insurance paid by those who currently hold coverage.

Second, the behavioral approach elides the normative nature of the question. Some people who could (normatively) afford to purchase coverage may not be willing to pay the full price of such coverage. People's willingness to pay might be depressed for many reasons, including the local availability of free care or because they are healthy and do not believe that they might one day become ill.

On the other hand, some of those people who purchase insurance may well be unable to afford it according to normative notions of affordability (Glied 2009). Such people may be willing to do without other merit goods in order to buy health insurance, though society as a whole may not view those individual decisions as optimal. There is no reason to believe that uninsured "afforders" and insured "nonafforders" balance out at the population median.

Finally, individual willingness to pay for coverage under the current health insurance system may not reflect societal norms about affordability under a reformed system. For example, the current health insurance system incorporates risk rating and does not constrain medical loss ratios. Some might feel that people cannot "afford" to pay premiums that vary with risk or that contribute significantly to insurance company profits, whether or not they are willing to do so. Alternatively, some people might feel that an affordability standard under a system in which people are compelled to purchase a defined health insurance package should be well below what they would pay if they were choosing coverage freely.

Survey approach

A second strategy to determine standards of affordability is to survey members of the public. Most surveys to date focus on perceptions of affordability among the uninsured. In the late 1990s, Yegian and colleagues conducted a survey of the nonpoor uninsured in California, asking questions about willingness to pay for coverage and about the cost of coverage. They found that there was a considerable mismatch between respondents' assessments of the cost of coverage and the actual cost of coverage and that willingness to pay for different types of plans did not match well to differences in premiums (Yegian et al. 2000).

The Kaiser Family Foundation as part of its tracking polls also surveyed uninsured Americans on how much they could afford to pay for

coverage (KFF 2009). The median respondent in the foundation's survey reported that he or she could afford to pay about \$100 per month for coverage. Unfortunately, these data do not provide enough information to see how affordability varies with income or with health plan characteristics.

Moreover, the survey estimates suffer from some of the same problems as the behavioral estimates. They do not provide insight into society's normative preferences, they do not consider affordability under a reformed system, and they are likely to be quite inaccurate because most respondents have very little understanding of the price of health insurance (Yegian et al. 2000). Some of these issues can be tackled by eliciting the respondents' opinion of others' ability to pay (e.g., by using vignettes) and by varying plan characteristics.

Conceptual Framework

While the "ideal" democratic public policy practice privileges the voice of the people, democracy highly values an informed and deliberative citizenry. It is under these conditions of increasing the public's knowledge, as well as that of the policy makers who represent them, that subject matter experts can be viewed as most useful. The traditional policy-analysis view is that experts should base their policy advice on an assimilation of existing literature and, after thorough analysis and assessment of existing evidence, offer an informed "objective" opinion (Weimer and Vining 2005).¹ While many scholars question such claims of objectivity, they at the same time acknowledge how a host of policy questions can benefit from expert analysis and accumulated knowledge. One view of such benefits is that experts can "speak truth to power." Some argue that as outsiders experts may be able to bring a more neutral perspective to muddled or politically charged concepts (Price 1965). Examining the process of expert deliberation and decision making provides insights on the extent to which this is an accurate characterization of what experts do.

In the area of setting affordability thresholds, bureaucrats will heavily influence how the individual mandate and subsidies are implemented and in turn shape higher-level decisions about affordability thresholds. Like policy makers, bureaucrats often seek the advice of subject matter experts. While this is not a study of bureaucratic behavior and decision making, understanding how experts think through this issue will help illuminate their potential influence on the policy process.

1. In the case of affordability standards, this literature would include studies of behavioral affordability and of subjective preferences.

Methods

We recruited an interdisciplinary panel of experts in social policy who could provide assessments of affordability (see technical documentation).² We used a modified policy Delphi process to elicit the panel's opinions on coverage affordability (Linstone and Turoff 1975; Fowles and Fowles 1978). In our implementation of the Delphi method, experts were asked to respond to a series of vignettes designed to elicit responses to complex problems (see technical documentation for a sample of vignette materials). The results of the initial responses to these vignettes were then summarized and returned to the expert panelists, who were asked to reassess their responses in light of this feedback (Fowles and Fowles 1978). The usefulness of this method depends on the quality of the experts who form the panel and on the extent to which they have appropriate information that can help them shape their opinions (Linstone and Turoff 1975; Fowles and Fowles 1978).

To help ensure a quality expert panel, we first generated a broad list of researchers and practitioners who have grappled with the concept of affordability. The authors and a research assistant obtained names from publications, from public and private agencies, and from other colleagues familiar with affordability debates. Once our team had obtained a broad list of approximately forty researchers and practitioners, our research team subjectively sorted the experts by political leaning (where applicable), practical experience, theoretical experience, and academic discipline. Ultimately, twenty-five experts were invited, and eighteen attended. To increase the likelihood that experts devoted adequate time and attention to the task, they were provided an honorarium of \$1,000 each. The participant list included people with widely recognized expertise in studying and delivering social services including housing, food, and income support services, as well as health insurance.

The Delphi approach allows for assessments of which sociodemographic characteristics might be important in considering such a standard. This allows for direct comparisons with the ACA affordability standard, which ranges from 3 percent of household income for families at 150 percent of the federal poverty level (FPL) to a maximum of 9.5 percent of household income for families earning more than 400 percent of the FPL. The meeting of experts was held after the rough outline of ACA had been drafted and deliberated but before passage of the law.

We provided each of the eighteen experts with a series of sixty-one

2. The technical documentation is available online at www.pceo.org/Affordability.pdf.

vignettes briefly describing the living situations of fictitious households (see table 1). Vignettes were used to facilitate proxy responses from each participant. Proxy responses are particularly important for making normative judgments (Alves and Rossi 1978), allowing the participant to consider how different characteristics of a household might influence his or her sense of what that household might afford to pay. By providing a referent or proxy, vignettes can help facilitate decision making by outlining parameters around which subjective assessments can be made. Prior research suggests that experts can reasonably rate 120–150 permutations of vignettes (Normand, Frank, and McGuire 2002). The vignettes are presented in the technical documentation;³ however, one sample vignette follows:

George is fifty-five years old. He is married and has grown children. His household's annual income is \$28,600. He lives outside Iowa City and pays rent of \$475 monthly.

These vignettes incorporated characteristics of the hypothetical households, including age, household income, marital status, presence of children, health conditions, the presence of debt or savings, and housing costs. The vignettes varied within and across these characteristics. For instance, characters were presented as either twenty-five or fifty-five years of age (the typical postcollege and preretirement period), single or married, and with no children or two children. In all vignettes with children, we specified that the children received free coverage through the Children's Health Insurance Program, as would be the case for most recipients of government subsidies.

We also defined two different health insurance plans. The first plan (A) has a \$2,000 annual deductible, 15 percent cost sharing, and a \$3,000 out-of-pocket maximum. The second plan (B) has no deductible, a \$20 office visit co-pay, and a \$50 hospital co-pay with a \$1,000 out-of-pocket maximum. In both plans, all preventive services were provided free of charge.

After sending out the background information, we created a Web survey in which we asked each expert to score his or her responses to each of the sixty-one vignettes (organized in random order) for each of the two health insurance plans (a total of 122 affordability scores per respondent). The experts were asked to specify how much the fictional household in each vignette could afford to pay monthly for each plan. For single-person

3. Even though plan administrators will be told they cannot implement risk rating, some experts may advise states to set affordability levels based on risk-related attributes.

Table 1 Characteristics of the Vignette Characters

Characteristic ^a	Values	Frequency
Age	25	29
	55	32
Marital status	Single	29
	Married	32
Children	No children	40
	Two children	21
Income	\$1,470/mo.–\$6,960/mo.	—
Debt	\$0	54
	\$7,750	7
Savings	\$0	53
	\$35,000	8
Child from previous relationship?	No	59
	Yes	2
Spouse does not have an income-earning job	No	59
	Yes	2
Did not complete high school	No	59
	Yes	2
Graduated college	No	59
	Yes	2
Excellent health	No	56
	Yes	6
Asthma	No	56
	Yes	5
Hypertension	No	55
	Yes	6
Lives in Chicago	No	57
	Yes	4
Lives in Iowa City	No	52
	Yes	9
Lives in Austin	No	57
	Yes	4
Has a 40-mile drive to work	No	58
	Yes	3
Rent/mortgage	\$0/mo.	44
	\$380/mo.	3
	\$475/mo.	6
	\$875/mo.	4
	\$1,450/mo.	4

Note: See technical appendix for the vignettes. For instance, one vignette took the following form: “James is 55 years old. He lives alone. His household income is \$35,600.”

^aThese characteristics were used to ascertain the impact of each parameter on experts’ perceived affordability of two hypothetical health insurance policies.

households, experts could select a score ranging from \$0 to \$800. For married households, experts could select scores ranging from \$0 to \$1,600 to reflect the cost of coverage for the couple. The experts were also given the option of describing any rules they used in calculating their affordability scores and were offered the chance to explain their scoring for each vignette.

In October 2009 the panel and the Columbia University researchers met in New York City to present, explain, and debate what affordability means. This discussion was in part generalized and in part guided by respondents' first round of scoring data. Each panelist received a statistical summary of the results of the full panel and of their own responses to each vignette. They then discussed discordant cases in the survey and expressed their views about which characteristics should affect any standard of affordability. After the discussion, the panelists were asked to individually and independently rerate their scores using the new information and to describe how their decision rules had changed. The hope was that the final rating scores would be enhanced by the deliberations among the experts in the interdisciplinary panel.

By asking the experts to describe how their decision rules had changed, in theory, we will not only be able to obtain a formal quantitative assessment of the determinants of an affordability standard using the Delphi method but also a qualitative understanding of what leading experts think about when considering such a standard. This qualitative analysis is based on transcript data from discussions about the affordability scores at the half-day conference.

Statistical Analyses

With respect to the numerical data generated, we assess factors affecting affordability determinations using median regression models to reduce skewing due to outliers. The standard errors for within-respondent correlations were estimated via bootstrapping. To capture how affordability assessments varied with income, the regression analyses interacted income with income categories relative to the FPL (150–200 percent FPL, 200–250 percent FPL, etc). The analysis was then repeated including all other household characteristics incorporated in the vignettes in addition to income. We also conducted ordinary least squares regression to highlight how the mean may differ from the median. The mean and median values provide useful information surrounding magnitude and distribution of expert-opinion output. Regressions were run with individuals as fixed effects.

Results

Rules of Thumb

Many of the panelists found the task of assessing the affordability of coverage for each of the sixty-one households described in the vignettes rather daunting. Several spent over five hours determining the appropriate standards. Ultimately, each panelist was asked to derive and write down some rule of thumb that would ensure consistency across the responses. In many cases, these rules were quite complex:

Respondent A: Rent at 30% of income unless otherwise specified; taxes at 20–30% of income depending on family size and poverty level; food costs adjusted . . . Plan A—put aside \$250/mo to handle deductible; Plan B—put aside \$83/month to handle deductible; transportation costs . . . [including] gas—usually at \$500–\$600/month; included misc item—which accounted for clothing, recreation, entertainment, pharmacy (if not separately noted).

Respondent B: I used a formula that accounted [for] income, taxes, numbers of adults and children at home, debt, savings, child support payments, location, maximum out-of-pocket medical expenses, and my normative views on minimum consumption of housing, food, and other goods.

Respondent C: I divided the cases into four groups, defined by the household income divided by the number of people to be covered . . . I then divided these groups into those with household heads over 40 and those under 40 and “charged” 1% more for those over 40 on the grounds that medical costs go up as one ages. . . . I followed the same principle (age grading and one versus two covered persons) for the second income category (4% for single people under 40, 5% for a couple under forty, 5% for a single over 40 and 6% for a couple over 40). And so on, up the income continuum.

Others used less complex rules and relied on data, similar to the approach provided by Blumberg et al. (2007) described above. The idea is to base the estimates on the amount households spend on health insurance relative to their overall income. At the meeting held at Columbia University, some of the expert rules of thumb were collectively discussed.

In-Person Discussion

What Does Affordability Mean? The discussion started in earnest with a review of the vignettes, beginning with Sarah, a vignette character who earns \$1610 a month, just under 200 percent of the FPL. In the case of characters with very low earnings, such as Sarah, there was considerable agreement among the panelists. All but one of the experts thought that she should pay something for health insurance, and all agreed that it should be very little.

The discussion heated up when participants considered Jessica, who earns \$2,210 per month. Here it was revealed that some participants interpreted affordability as what the person *should* pay and others as what they *could* pay. This provided a jumping-off point to a broader discussion of what affordability means.

One participant suggested that Jessica could “literally” afford \$800 per month, and denounced affordability as “a lousy policy where you impoverish people so they can buy a minimal set of other goods after they end up paying for health insurance.” Another respondent who also gave a very high affordability threshold noted, “I’m not saying it would be worth it for her to pay this much, but she could still manage reasonably well if she [did] . . . so.” Many others disagreed with this literal interpretation, noting, “I don’t think literal affordability is an interesting policy.” One respondent described affordability as “asking *really*, ‘How much could one reasonably pay for health insurance.’” Several others argued that a policy would be affordable for beneficiaries if there would “be no adverse consequences by pushing them to that limit.” (However, the specific adverse consequences mentioned in this context were not actually discussed.)

Those who disagreed with the literal definition generally emphasized the importance of a variety of nonhealth-related “merit” goods that public policy makers would want low-income families to be able to purchase. For example, one respondent noted, “There are a number of things I want people to save for besides having to worry about paying for health insurance.” Another stated, “In theory, if there’s a minimum amount of resources you need to live, we should look at [an] amount over and above that minimum amount.” One participant, who argued that the standard should be \$100 to \$150, said, “I’m a low-ball guy because [I take into account] the real costs of living, the fact that the poverty line is too low, etc.” Some specific examples of merit goods were discussed, but these were generally only discussed to highlight the need to be comprehensive in one’s assessment. For example, a car for transit to work was considered a merit good by some experts.

Participants then noted that even the boundaries between *could* and *should* blurred. The group discussed various definitions, including, respectively, what one is “able” to pay and what one “ought” to pay given his or her financial circumstances. Perhaps the closest approximation to a normative model came from a panelist who advanced what Carla Saenz (2010) calls a “reasonable trade-off” definition of affordability, to wit, “what one ought to be paying in relation to the bundle of other minimal goods.” As noted above, such minimal goods need not be limited to those basic to survival but rather those needed to maintain an acceptable standard of living.

Different notions of affordability arose because panelists were given the opportunity to define affordability in their own way prior to the meeting. They were not asked to devise lists of what might be called a merit good and what might not. Therefore, the Delphi round during the meeting provided the opportunity to refine each expert’s own personal notions of affordability standards. Still, attempts to bridge the gap by formulating general principles that might guide the definition and application of affordability were surprisingly tentative. One panelist insisted that “skin in the game” made for better consumers. In other words, the panelist felt that those who actually made investments in health care out of their own pockets would choose more cost-effective insurance options than those who were fully subsidized. The panelist further noted that the idea that individuals above 200 percent FPL might legitimately be required to pay more than a negligible amount, while those below that threshold should pay little or nothing, had broad intuitive appeal. However, this panelist did not state how steep the gradient after this threshold should be.

At one point, participants considered the question of how high the threshold for subsidies should be drawn, noting that the median income is about 400 percent of the FPL. Some questioned whether we really want to be subsidizing half of the population, and others suggested that the threshold should be even higher because insurance is so expensive. One participant noted, “In Massachusetts, there’s a problem where many people whose incomes are above 300 percent FPL can’t buy health insurance and by appealing they’ve been deemed that they can’t afford it and [can] get a waiver.” This participant’s observation suggests that assessments of affordability should be evaluated on an ongoing basis.

Another dimension of affordability that arose in the responses and discussion was the time frame over which decisions are made. One participant emphasized that avoiding undue dislocation and disruption in people’s financial lives was important: “In the real world people make financial

commitments associated with their inflows and they are not changeable immediately, and we're talking about an immediate change and the notion that they could adjust all other expenses that they've already locked in for their standard of living and they now should back off of that for health insurance; that's not affordability."

All panelists made adjustments for family size and composition, but the method of adjustment and the adjustment size varied substantially among panelists. When panelists spoke of adjustments for family size, one panelist noted, "I didn't charge twice as much [when there was another adult who could pay] but just adjusted a fixed rate for the additional person." Another noted that adjustments could be nuanced, pointing out that her preferred schedule allows more income for families with children under age five (i.e., a lower affordability threshold) than for those with children over five years of age, because there is a burden associated with caring for children who are not yet in school. This, in turn, raises questions about whether to adjust for the receipt of other social benefits, such as Head Start.

Complex Family Circumstances. There was even less agreement about how to address other dimensions of the vignettes. Many respondents did not incorporate variation in housing costs in their estimates. Some argued that these variations were discretionary, while others claimed that adjusting for such variations would be politically unpalatable. But several respondents asserted that housing costs should be reflected in affordability calculations for public policy reasons, noting, for example, "you don't want to force people into more dangerous housing just to afford health insurance."

About one-third of respondents thought that savings were an important consideration, but fewer thought that debt was important. The general tenor of discussion, as well as the premeeting responses, held that debt should not be taken to diminish ability to pay, in part because policy makers should not be encouraging debt accrual. However, possible reasons for debt gave some respondents pause. For example, one respondent stated, "I did not view credit card debt as a worthy factor—unless there were an indication it was due to medical out-of-pocket expenses (expenses directly paid for medical care from the recipient)." While no one specifically followed up on this comment, many experts seemed to feel that this was an important consideration. Similarly, most respondents did not alter their overall assessment that savings should not be taken into account, arguing, "I would not punish her for having savings." After the meeting, however, a few respondents altered their responses, with respondents who did not

provide a score for savings and debt indicating that they would in fact incorporate such considerations. One such respondent noted that in the real world “debt and savings would have an effect on affordability.”

Others suggested that other variables, such as the individual’s age, might be considered. People in their fifties use three times more health care than do those in their twenties, it was pointed out, and should therefore perhaps pay more.³ Other participants, however, emphasized the need for older households to save for retirement and for later medical expenses. One respondent pointed out, “A couple needs more than \$100,000 just to pay Medicare premiums . . . so I took savings into account for those [nearing] Medicare age.” Other panelists agreed, citing the danger of unintended consequences from uncritical applications of affordability. “If fifty-five-year-olds are charged more [than younger people], then they’ll suddenly show up for food stamps,” said one. One participant noted the impracticality of incorporating savings, age, or other such considerations: “I can’t imagine a plan that uses individual debt in determining affordability and what someone should pay.” Another noted, “I tried to completely minimize discretionary behavior in calculating this, but some could say that [even] having kids is a discretionary decision.” Presumably, the first respondent was referring to the administrative barriers to implementing a complex standard of affordability, and the second respondent was referring to the political barriers to doing so.

Plan Choices and Affordability. Each vignette asked participants to decide between the basic and the more comprehensive health plan. This led to a discussion about how affordability related to choice of plan design. Clearly, a person’s health determined whether it was logical to pay more for a plan without a deductible. For example, one respondent argued that generous subsidy of the higher deductible plan would be preferred even for people with chronic conditions because such cost-sharing provides incentives for good chronic-disease maintenance: “I feel that the policy will push [vignette character] to better control her asthma and avoid hospitalization and reduce M.D. visits.” Others disagreed, asserting that the affordability thresholds for the two plans should be set so as to encourage enrollment in the more generous, lower cost-sharing plan. Such an incentive would encourage less risk taking with respect to unforeseen catastrophic medical expenses. For instance, one expert noted, “For low-income people [the basic plan] would be a waste of money because it increases exposure to out-of-pocket expenses, [raising concerns about affordability].”

Determining affordability thresholds for different plan choices raised questions again about the role of debt and savings in these determinations. Many respondents argued that households with more savings would be better able to afford the basic plan with higher deductibles and cost sharing: “For me, the price differential was smaller if I felt someone had savings to serve as their insurance to protect against larger out-of-pocket costs.” Debt in relation to plan cost was not discussed.

Expert Affordability Scores

After the meeting, the participants were asked to reevaluate their affordability assessments for each of the vignettes. Despite the lively discussion and the divergent opinions voiced, most participants made only minor revisions to their premeeting affordability scores. It should be noted that core beliefs surrounding what is and is not normative with respect to affordability assessments are unlikely to be altered in a half-day meeting. However, one might expect greater variation in second-round scoring given that some respondents changed their own definitions of affordability that they had brought with them to the meeting. The final scores were used in our affordability estimates.

While most experts thought subsidy levels should be progressively rated with income, there was substantial divergence in expert opinion with respect to exact quantification of the affordability standard (see fig. 1). If we look at the median score offered among experts, we find that affordable monthly health insurance payments were considered to be about \$50 for households at 150 percent of the FPL to around \$325 at 400 percent of the FPL for the more-basic plan A (with substantial cost sharing) and to a high of roughly \$425 per month for the more-comprehensive plan B. This corresponds to 3.4 percent of income at 150–200 percent of the FPL and 7.4 percent of income at 350–400 percent of the FPL.

Note, however, that these levels simply reflect the median meaning that half of the experts thought affordable payments should be below these levels, while the remaining half believed affordable monthly payments could be set higher. There was substantial divergence between the 25th and 75th percentile scores, especially for vignette characters with higher earnings. The mean value was significantly higher than the median value—skewed much closer to the 75th percentile score—reflecting the wide variation in affordability opinions for persons earning around \$88,000 annually (400 percent of FPL for a family of four in 2010). Given that median household income in 2009 was about \$50,000 and median family income for a fam-

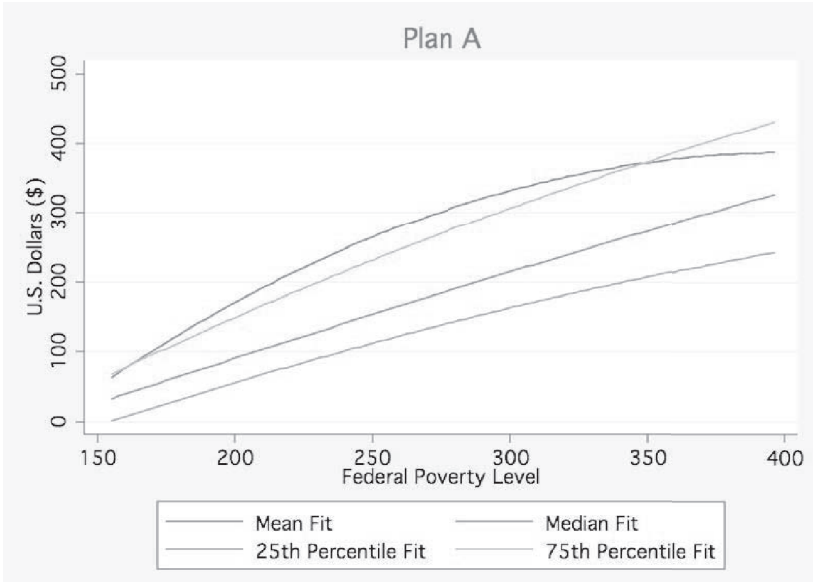


Figure 1A Expert-Generated Affordability Scores by Federal Poverty Level, Plan A

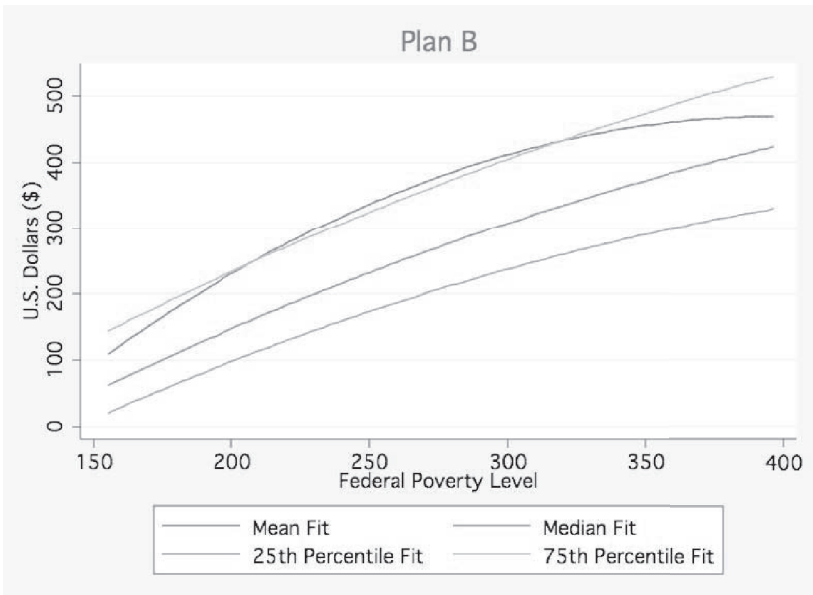


Figure 1B Expert-Generated Affordability Scores by Federal Poverty Level, Plan B

ily of four in 2010 was \$72,000⁴ (a value that is substantially below 400 percent of the FPL for a family of four), this means some experts believe government subsidies should remain fairly generous for middle-income Americans, while others believe that subsidies should be set very low.

To provide a general picture of how experts thought affordability scores should be adjusted according to increases in income as well as the other household characteristics discussed qualitatively above, we conducted regressions of the mean and median affordability scores on income by FPL category (see table 2). Median affordability scores suggest that each additional dollar in income should produce a \$0.12 increase in payments for health insurance. Experts, at the median, also felt that singles should be penalized with a \$47 additional payment for insurance, but that having children should reduce payments by \$99. Moreover, the “median expert” opined that marital status, family size, type of insurance, and spousal employment were all relevant in affordability determinations. Moreover, the variation in affordability scores (e.g., the higher subsidies for those who live in expensive cities, such as Chicago), while not statistically significant, suggests that at least some experts felt strongly about many of the other sociodemographic factors presented in the vignettes.

Table 3 provides the median regressions relative to the ACA and Massachusetts affordability standards. The median affordability scores were not far from those stipulated by the ACA and were within range of those used in Massachusetts. While the expert panel felt that the average person living at between 100–150 percent of the FPL could afford to spend 3.4 percent of his or her income on health insurance, the corresponding range for 100–150 percent of the federal poverty level for the Massachusetts plan is 2.5 percent–3.7 percent and 4 percent to 6.3 percent for the ACA. Both the Massachusetts and ACA thresholds were somewhat higher than the respondent assessments for households with higher incomes.

While regressions on the median scores are useful to grasp a general picture of expert opinion, these median values also mask variation in the factors that matter among experts. To illuminate this variation, we conducted regression analyses for each individual expert and then counted the number of experts whose score for a given vignette characteristic was statistically significant. This allows us to determine not only which individual characteristics experts view as important when assessing affordability but also the extent to which experts vary with respect to afford-

4. The exact figures are \$49,777 and \$72,352, respectively, according to the U.S. census extracted on March 2, 2011.

Table 2 The Effect of Various Vignette Characteristics on the Mean and Median Affordability Scores (robust t statistics)

	Monthly Amount Affordable	
	Mean regression	Median regression
Income	0.17 (0.04)**	0.12 ^a (0.02)**
Single	8.72 (27.79)	46.62 (14.01)**
Two children?	-75.63 (22.40)**	-98.50 (30.83)**
Single × two children	-45.99 (31.62)	-5.52 (20.11)
Income squared	-5.73E-06 (4.03E-06)	-9.67E-07 (2.41E-06)
Plan	-71.41 (9.50)**	-60.00 (13.31)**
Debt?	-7.33 (6.10)	3.96 (9.13)
Savings?	18.67 (5.97)**	12.37 (6.86)
Child from previous relationship	-32.40 (8.64)**	-6.31 (12.52)
Spouse does not have job	19.95 (10.19)	46.65 (15.35)**
Did not complete high school	21.91 (15.93)	4.92 (14.65)
Graduated from college	30.15 (18.79)	17.996 (9.46)
Excellent health?	29.94 (13.29)*	22.23 (11.34)
Asthma?	0.47 (9.54)	-10.24 (8.71)
Hypertension?	-1.96 (6.71)	-3.61 (12.55)
Lives in Chicago	-0.28 (163.47)	-319.36 (182.88)
Lives in Iowa City	18.77 (49.83)	-80.21 (54.05)
Lives in Austin	29.38 (109.80)	-185.65 (108.24)
Drives to work	14.92 (17.95)	-8.04 (15.44)
Owns home	-0.02 (0.12)	0.23 (0.13)
Constant	-132.20 (41.34)**	-140.51 (42.66)**
Observations	2195	2195
R-squared	0.18	

Note: Values are in US\$ and errors are clustered.

^aFor every additional dollar increase in income, the experts felt that the vignette characters should pay 12 cents more for their insurance premiums.

*Significant at 5 percent; **significant at 1 percent

Table 3 Median Regression Score Relative to the Massachusetts Affordability Standard for 2010 and the Patient Protection and Affordable Care Act Standard for 2014

Income as a Share of the Federal Poverty Level (by %)	Median Regression	Massachusetts Affordability (2010) ^a	Patient Protection and Affordable Care Act (2014) ^b
150–200	0.034 (0.011)*	0.025–0.037	0.04–0.063
200–250	0.047 (0.011)*	0.038–0.056	0.063–0.0805
250–300	0.057 (0.008)*	0.047–0.07	0.0805–0.095 ^c
300–350	0.065 (0.007)*	0.07(a)	0.095 ^c
350–400	0.074 (0.007)*	—	0.095 ^c
Constant	9.813 (23.222)		
Observations	2195		

^aPremium contributions based on analysis of Massachusetts Health Care Reform Law and Patient Protection and Affordable Care Act (P.L. 111-148) as amended.

^bCalculated as share of least costly health plan offered by Connector as share of income at 300 percent of the federal poverty level.

^cIndividuals are not subject to mandate penalties if premium costs exceed 8 percent of household income.

*Significant at $p < 0.001$

ability scores — and how (increasing or decreasing the score) — for each characteristic.

Of the eighteen experts, the majority believed income, marital status (single), and plan choice should be used to determine affordability levels (see table 4). There was almost complete agreement (sixteen out of eighteen) that payments should increase as income increases and that monthly premium payments should be lower if individuals choose the basic plan (A). On average, these sixteen experts believed payments should increase 20 cents for every \$1 increase in earnings, whereas the basic plan A should be about \$80 cheaper than the comprehensive plan (B). While the majority of experts thought marital status should be considered in establishing affordability levels, there was more variation in opinion: six experts thought it should not be used, and two experts believed monthly payments should be substantially lower for singles.

Table 4 Variation in Expert-Specific Affordability Scores across 61 Vignette Cases.

	Experts with Significant Positive Coefficient		Experts with Significant Negative Coefficient	
	Number of experts	Effect on average monthly affordability score	Number of experts	Effect on average monthly affordability score
Income (dollars)	16	0.2	0	
Single	10	67.7	2	-266.4
Two children	0	0	9 ^a	-198.5
(Single) × (two kids)	3	74.4	5	-151.6
Plan A	0	0	16	-80.9
Debt	1	23.6	2	-52.0
Savings	4	32.2	1	-19.3
Child from previous relationship	0	0	6	-54.5
Spouse does not work	3	47.0	1	-19.2
Did not graduate HS	1	41.2	2	-30.1
Graduated college	2	237.7	0	0
Excellent health	2	131.8	0	0
Asthma	1	49.1	4	-35.8
Hypertension	1	48.0	1	-56.4
Chicago	1	390.7	1	-769.1
Iowa City	1	124.8	1	-174.0
Austin	1	233.3	1	-487.4
Drive to work	2	213.0	3	-48.5
Owens home	1	0.5	0	0

Notes: Each column presents the number of experts with a significantly positive or negative coefficient and the overall magnitude of the effect on the affordability score. We show only those experts whose scores were statistically significant. It should be kept in mind that this does not mean that the other experts did not assign a higher or lower score to any given vignette characteristic.

^aFor instance, nine experts strongly felt that the presence of two children in the household means that the vignette character's health insurance payments should be lower than a household with no children.

While half of the experts did not adjust payments to account for children, the remaining nine experts thought the presence of children in the household should be used as a major determinant of plan affordability. All nine thought monthly payments should be reduced according to the presence of children, with the average reporting that individuals with two children should pay \$199 less than those with no children. When

single marital status is interacted with children, three experts felt that a single parent with two children should be penalized (i.e., payment should increase), whereas five experts believed payments should be reduced.

Only a small number of experts believed the following factors should be taken into account in setting monthly payment levels: debt (three experts), savings (five), educational attainment (two), children from previous marriage (six), unemployed spouse (four), a commute to work requiring a car (five), the presence of preexisting conditions (asthma, five, hypertension, two), geographic location cost-of-living adjustments (two), and home ownership (one). Some of these scores reflect subsidies that are particularly large in magnitude (large subsidies or large payment increases). As illustrated above in reporting qualitative findings from the in-person discussion, when experts discussed the use of these factors to determine affordability levels, where considered appropriate, such factors were typically seen as incentives for good social behavior (through rewards or penalties) or as behavior-neutral, with the factors simply placing individuals in disadvantaged positions, thereby making them worthy of increased subsidies. For instance, graduation from high school or college can be seen as an indicator of future prosperity. One respondent thought that high school dropouts should pay more and two thought that they should pay less.

Table 5 builds on table 4 to show the number of significant coefficients per expert respondent. Thus, while table 4 shows the variation in characteristics considered significant, table 5 shows the variation in characteristics seen as significant by each expert. For instance, expert number 17 felt that eleven characteristics were important to consider, but expert 18 felt that only two were important to consider. This table also shows the direction of the response. For example, expert 1 felt that six characteristics should merit a higher subsidy and four characteristics should lower the subsidy. There was a wide range of variation in this respect as well, with five experts providing more positive scores, four providing an equal number of positive and negative scores, and nine providing more negative scores than positive ones.

However, when one looks at the frequency of factors that experts felt should be significant, a different picture emerges (see fig. 2). On average, experts believe 6.5 adjustments should be taken into account when setting subsidy levels, and a clear majority of experts (fifteen out of eighteen) believe a standard should include five adjustments or more (up to a maximum of eleven adjustments). Thus, while the significance for any one adjustment (see table 4) suggests that most experts do not believe many

Table 5 Number of Statistically Significant Affordability Scores (Positive and Negative) Assigned to Vignette Character Characteristics for Each Expert.

Expert	Number of Positive Scores	Number of Negative Scores
1	6	4
2	3	5
3	3	4
4	4	4
5	8	2
6	1	4
7	3	4
8	2	4
9	4	2
10	2	2
11	2	5
12	4	3
13	2	3
14	1	1
15	4	3
16	2	4
17	3	8
18	1	1

Notes: The values reflect the number of positive coefficients for any given vignette trait (e.g., income) based on individual regressions. Thus the affordability scores from respondent 18 were only significant for two vignette characteristics. The expert felt that the affordability score should be higher for one of these traits and lower for the other trait.

adjustments (aside from income and plan design) should be taken into account, figure 2 suggests otherwise. In fact, these experts believe quite a few adjustments should be taken into account in setting subsidy levels, but disagree about which particular factors should be used.

Discussion

Among of the expert panel there were substantial variations in assessments of the affordability of health insurance and the mix of elements that should be included in setting an affordability standard. The expert panel even disagreed about what *affordability* actually means, with some interpreting it as what one “could” pay and others what one “should” pay. The expert panel also held relatively fixed notions of what the affordability

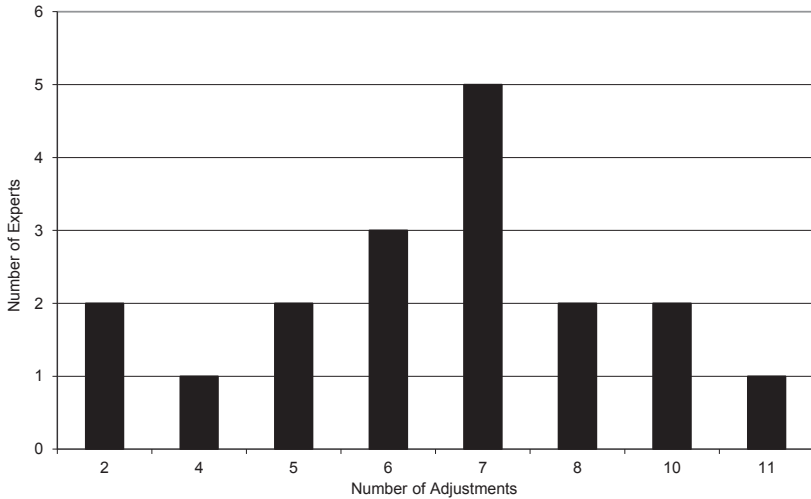


Figure 2 Frequency Distribution of the Number of Significant Adjustments Experts Made

standard should be; after discussing the need to define affordability as a normative concept, most experts did not change their rating of how much the vignette characters should pay for each of the different health insurance plans.

In their discussions, experts were sparing in taking into account most living circumstances other than household income and family size. Those who thought additional factors should be considered were highly divided about whether policies should incentivize good behavior or simply redistribute income to people in various disadvantaged positions. In discussions at the in-person meeting, respondents voiced concern about both the potential for unintended distortionary effects from taking into account factors such as debt or living expenses and the administrative complexity of doing so. Likewise, in the formal Delphi analysis scoring process, only a minority of respondents saw such factors as significant considerations in their affordability scoring. However, these data mask the fact that many experts did in fact want additional adjustments (see fig. 2), but the specific adjustments desired varied considerably. It is tempting to speculate that such adjustments varied according to this same trade-off between individual incentives and net social-welfare benefits.

For most experts, fundamental differences in what affordability means,

as well as their assessment of the basic bundle of goods that public policy should assure low-income people have, trumped nuanced arguments about particular behavioral dimensions. This, combined with the substantial variation in affordability scores and affordability determinations beyond household income relative to the FPL, suggests that affordability is at its core a political concept.

Given that notions of affordability are inherently normative, our findings raise questions surrounding the extent to which mainstream public values align with those of experts. In cost-effectiveness analysis used in medicine, it is necessary to elicit preferences to quantify the morbidity associated with a given health state, such as blindness (Gold et al. 1996). For this task, a U.S. government panel, employing a Rawlsian framework, decided it was best not to elicit such preferences from experts or even from patients suffering from the condition but rather from a representative sample of the population. The idea behind this decision was that, since cost-effectiveness analysis is used to allocate tax dollars for the benefit of everyone, everyone in society should decide how different health states are valued. Indeed, the scores estimated by experts look quite different from those elicited from the general public (Gold and Muennig 2002). However, experts bring to the table a much more nuanced understanding of the implications of competing alternative policies, and there is no gold standard to assess which choice is best under any given set of circumstances. Indeed, data from our public sample of affordability scores suggest that, among other issues, most people have very little idea what their employers pay for their premiums (Muennig et al.). A separate public sample was obtained and is discussed elsewhere (*ibid.*).

Limitation

This study was limited by the sample size and composition; we present the opinions and affordability scores of just eighteen experts. While these experts were chosen from a broad range of disciplinary backgrounds, it is difficult to assess the extent to which their ideas are representative of other expert panels used to make public policy.

Conclusion

There is not a natural, objective standard of affordable health insurance coverage. Of course this is true for most (if not all) policies that require

government subsidies. Reasonable people disagree about how much a given household can be expected to put aside to spend on health insurance and health care.

That there is considerable disagreement even among distinguished experts in social policy on this issue suggests that the concept is mainly a subjective one and that any specific affordability threshold or subsidy structure is likely to be contentious and contested. In this area, as with many others, even highly qualified experts are unlikely to provide scientific answers to guide policy.

References

- Alves, W. M., and P. H. Rossi. 1978. Who Should Get What? Fairness Judgments of the Distribution of Earnings. *American Journal of Sociology* 84:541–564.
- Blumberg, L. J., J. Holahan, J. Hadley, and K. Nordahl. 2007. Setting a Standard of Affordability for Health Insurance Coverage. *Health Affairs* Web Exclusive 26:W463–W73. content.healthaffairs.org/content/26/4/w463.full.html.
- Bradley, R. 2008. Comment—Defining Health Insurance Affordability: Unobserved Heterogeneity Matters. *Journal of Health Economics* 27:1129–1140.
- Bundorf, M. K., and M. V. Pauly. 2006. Is Health Insurance Affordable for the Uninsured? *Journal of Health Economics* 25:650–673.
- Fowles, J., and R. B. Fowles. 1978. *Handbook of Futures Research*. Westport, CT: Greenwood Press.
- Glied, S. 2009. Mandates and the Affordability of Health Care. *Inquiry* 46:203–214.
- Gold, M. R., and P. Muennig. 2002. Measure-Dependent Variation in Burden of Disease Estimates: Implications for Policy. *Medical Care* 40:260–266.
- Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein. 1996. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.
- Kaiser Family Foundation (KFF). 2009. Public Opinion on Health Issues. November. Menlo Park, CA: KFF. www.kff.org/kaiserpolls/upload/8020.pdf
- Linstone, H. A., and M. Turoff, eds. 1975. *Delphi Method: Techniques and Applications*. Reading, MA: Addison-Wesley. is.njit.edu/pubs/delphibook (accessed March 11, 2009).
- Muennig, P., N. Tilipman, B. Sampat, and S. A. Glied. The Public's Perceptions of the Affordability of Health Insurance. Unpublished manuscript.
- Normand, S. L. T., R. G. Frank, and T. G. McGuire. 2002. Using Elicitation Techniques to Estimate the Value of Ambulatory Treatments for Major Depression. *Medical Decision Making* 22:245.
- Price, D. K. 1965. *The Scientific Estate*. Boston: Harvard University Press.
- Saenz, C. 2010. What Is Affordable Health Insurance? The Reasonable Tradeoff Account of Affordability. *Kennedy Institute of Ethics Journal* 19:401–414.

- Weimer, D. L., and A. R. Vining. 2005. *Policy Analysis: Concepts and Practice*. 4th ed. Upper Saddle River, NJ: Prentice Hall.
- White House. 2010. The President's Proposal for Health Reform. www.whitehouse.gov/Issues/health-care (accessed February 26, 2010).
- Yegian, J. M., D. G. Pockell, M. D. Smith, and E. K. Murray. 2000. The Nonpoor Uninsured in California, 1998. *Health Affairs* 19(4): 171–177.

