SUPPORTING STATEMENT

Part B

SelectMD 2.0 Clinician Choice Experiment

Version: November 21, 2013

Agency for Healthcare Research and Quality (AHRQ)

Table of Contents

B. STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

Respondents for this experimental study are drawn from the GfK Internet panel. GfK's Internet panel consists of over 50,000 adult panel members who are recruited by random-digit dialing (RDD) or by using address-based sampling. Typical panel members receive 3-4 invitations per month to participate in research projects.

The GfK panel is constructed to include those who do not otherwise have Internet access by providing them with free access in return for their participation on the panel and computer equipment, if not otherwise available. We do not intend to generate nationally or locally representative results or precise estimates of population parameters from this study. The sample used is best understood as a convenience sample, rather than a probability sample. The GfK panel is large and variegated enough to produce samples with a reasonable degree of diversity in key demographic characteristics. Furthermore, no legitimate weights can be constructed from non-probability samples such as the one used here. Hence, we will not in any publications emerging from this work construe this sample or the results generated from this sample as nationally or locally representative. The strength of the experimental study lies in its internal validity, on which meaningful estimates of differences across the experimental exposures (e.g., granularity of information) can be produced and generalized.

2. Information Collection Procedures

Study subjects will be randomly assigned to one of several experimental arms (see Part A Section 1 and the related Attachment C for a description of the arms) that vary according to the type and complexity of performance information included in the web-based report. Participants will complete the experiment through a secure online connection from their homes. Data will be derived from pre-choice and post-choice questionnaires and from server logs that record the web pages visited and viewing times spent on each web page.

3. Methods to Maximize Response Rates

The response rate is estimated at about 75% based on results obtained from the past projects conducted by GfK. Procedures for maximizing response rates include:

- Field period of 3 to 4 weeks
- Use of the Federal agency or University/College name in the email invitation
- Email reminders

- Telephone reminder calls to non-responders
- Incentive payments equivalent to approximately $25 conditional on participants completing all three stages of the experiment (pre-choice survey, choice of physician and post-choice survey).

Given the experimental design and with an average of approximately 175 respondents per condition (as described in Supporting Statement A), comparisons between individual conditions are sufficiently powered to detect moderately small effect sizes (equal to a difference between mean scores equaling at least .3 standard deviations). These calculations presume a type I error rate of .05, type II error rate of .2, equal sample sizes, and two-sided hypothesis tests. We observed effect sizes of this magnitude for cross-arm comparisons of a number of key outcome variables from our previous physician choice experiment (Select MD 1.0). Moreover, differences of this magnitude are likely to be viewed by both policymakers and decision-makers within the health care system as substantively meaningful, since they match the size of previously documented disparities in choice making that have been deemed important to remedy, including (1) differences between high- and low-income patients, (2) differences between patients with substantial education (college and above) and those with limited education (high school or less), and (3) those experienced and inexperienced in health care choices (those with chronic conditions vs. those who are not).

4. Tests of Procedures

To achieve the aims of the experiment, the following four analyses will be performed:

1. McFadden's (1974) conditional logit model will be used to model choice of doctors that are of higher-quality, as measured by quantifiable performance metrics (CAHPS, HEDIS and patient safety scores). Conditional logistic regression models multi-categorical choice as a function of the characteristics of the choices themselves. It also allows for examination of the influence of characteristics that do not vary within a given set of choices for a single experimental participant (such as characteristics of the person making the choice or of the choice set) through interactions with characteristics of the choices.

In this experiment, each participant chooses among 12 doctors whose characteristics depend on the experimental condition. The choices vary according to CAHPS scores (five strata: one to five stars), HEDIS (The Healthcare Effectiveness Data and Information Set) scores (five strata: one to five stars), scores on efforts to reduce medical errors (five strata: one to five stars), and the modal affect of patient comments (strongly positive/negative, weakly positive/negative, or mixed). CAHPS, HEDIS, and medical-error scores are presented either as individual, standardized performance metrics (our distinct measures in each of these three domains of quality), or "rolled-up" into broader categories of measures (a single roll-up measure in each of these three domains of quality), or both. Individual experimental arms also introduce different methods for grouping and labeling patient comments, and allow respondents to choose which (and how much) information they are shown. In one arm, respondents will have access to live

telephone assistance when making choices.

The presence or absence of roll-up scores, patient comments, patient comment grouping, information choice, and telephone assistance each changes the characteristics of the choices, so these factors will be incorporated by using parallel conditional logit models for choice sets that do or do not include each of these types of information.

In certain experimental conditions, the quality of choices can also be measured by HEDIS scores or modal affect of anecdotal information. Within these conditions, conditional logit can also be used to examine these outcomes.

2. The effects of demographic variables, such as age, gender, and education, and of other individual difference variables, such as perceived health status and decision-making style, will be examined by testing for interactions with choice characteristics.

3. Analysis of variance will be used to examine the effects of type and amount of information presented on how the respondents use the web site, including which pages they visit and how much time they spend viewing them.

4. Responses to post-experimental questionnaires will be analyzed using descriptive summaries and analysis of variance to provide insight into the strategies participants employ in using the information presented to choose a doctor or health plan and how they view this information.

5. Statistical Consultants

Statistical expertise in analyzing the results of the experiment will be available from Marc Elliott, PhD, of RAND