

Measuring Preferences for Health Care Interventions Using Conjoint Analysis: An Application to HIV Testing

Kathryn A. Phillips, Tara Maddala, and F. Reed Johnson

Objective. To examine preferences for HIV test methods using conjoint analysis, a method used to measure economic preferences (utilities).

Data Sources. Self-administered surveys at four publicly funded HIV testing locations in San Francisco, California, between November 1999 and February 2000 ($n = 365$, 96 percent response rate).

Study Design. We defined six important attributes of HIV tests and their levels (location, price, ease of collection, timeliness/accuracy, privacy/anonymity, and counseling). A fractional factorial design was used to develop scenarios that consisted of combinations of attribute levels. Respondents were asked 11 questions about whether they would choose “Test A or B” based on these scenarios.

Data Analysis. We used random effects probit models to estimate utilities for testing attributes. Since price was included as an attribute, we were able to estimate willingness to pay, which provides a standardized measure for use in economic evaluations. We used extensive analyses to examine the reliability and validity of the results, including analyses of: (1) preference consistency, (2) willingness to trade among attributes, and (3) consistency with theoretical predictions.

Principal Findings. Respondents most preferred tests that were accurate/timely and private/anonymous, whereas they had relatively lower preferences for in-person counseling. Respondents were willing to pay an additional \$35 for immediate, highly accurate results; however, they had a strong disutility for receiving immediate but less accurate results. By using conjoint analysis to analyze new combinations of attributes, we found that respondents would most prefer instant, highly accurate home tests, even though they are not currently available in the U.S. Respondents were willing to pay \$39 for a highly accurate, instant home test.

Conclusions. The method of conjoint analysis enabled us to estimate utilities for specific attributes of HIV tests as well as the overall utility obtained from various HIV tests, including tests that are under consideration but not yet available. Conjoint analysis offers an approach that can be useful for measuring and understanding the value of other health care goods, services, and interventions.

Key Words. Conjoint analysis, discrete choice experiment, patient preferences, patient acceptance of health care

Increasing the number of individuals knowing their HIV status continues to be a major public health goal (Centers for Disease Control 2001). Although the importance of early HIV testing has increased as new and effective treatments for HIV have become available, it has been estimated that one-third of infected people in the United States still do not know their HIV status (Sweeney et al. 1997; Bozette et al. 1998). Several new methods of testing have been developed, including home specimen collection tests ("one week home tests"), where blood samples are collected at home and then sent to a laboratory (Phillips et al. 1995; Phillips et al. 2000), rapid tests (Kassler et al. 1997), urine tests (Urnovitz et al. 1999), and oral fluids (swab) tests (Ferri 1998). These new testing methods may decrease barriers to testing and be more cost-effective than standard methods. However, there has been tremendous controversy over these new methods; for example, there was almost 10 years of bitter debate before the FDA (Food and Drug Administration) approved the home collection test (Phillips et al. 1998).

The controversy over new HIV testing methods is likely to accelerate as new technologies are developed that could result in the approval of instant home tests. In the United States, as of March 2002, one home collection test (CONFIDE, manufactured by Home Diagnostics, Inc.) and one rapid test (SUDS, manufactured by Murex) had been approved by the FDA. However, the home collection test does not provide instant results and the rapid test requires trained personnel and laboratory facilities and is for screening purposes only. Several other rapid tests have been developed and are being used in other countries and for investigational studies in the United States. None of these companies have stated that they plan to seek approval for use of these tests as instant home tests. However, several companies are in the process of seeking approval for rapid tests that are accurate, easy to use, do not require specialized equipment or lab technicians (Orasure Technologies 2001), and that could in the future be marketed as instant home tests.

This work was supported by an NIH R01 grant to Dr. Phillips from the National Institutes on Allergies and Infectious Diseases (AI43744).

Address correspondence to Kathryn A. Phillips, Ph.D., Associate Professor of Health Services Research and Health Economics, School of Pharmacy, Institute for Health Policy Studies, and Center for AIDS Prevention Studies, University of California-San Francisco, 3333 California St. Suite 420, Box 0613, San Francisco, CA 94143 (Fedex 94118). Tara Maddala, Ph.D., is Senior Biostatistician with Clinimetrics Research Inc., San Jose, CA. (Dr. Maddala conducted this research while at the University of California-San Francisco.) F. Reed Johnson, Ph.D., is RTI Fellow with Research Triangle Institute in North Carolina.

Although several studies have examined barriers to testing using attitude surveys (e.g., Phillips 1993; Phillips et al. 1995), to our knowledge there have not been any preference-based surveys of HIV testing methods (i.e., surveys measuring utilities based on economic theory). In this study, we use a multiattribute, stated preference method called “conjoint analysis.” Stated preference techniques include the methods of preference weighting (e.g., rating scales, standard gamble, and time trade-off), willingness to pay as measured by contingent valuation surveys, and “conjoint analysis.” Conjoint analysis is an approach to measuring preferences (utilities) that estimates both overall preferences for a good or service as well as preferences for its specific attributes. Conjoint analysis surveys involve comparing hypothetical scenarios by ranking, rating, or choosing a particular scenario. For example, respondents may be asked to choose from “Test A” and “Test B,” where each test is described using a series of attribute levels (e.g., a test at a doctor’s office costing \$100 or a test at a public clinic costing \$25). Conjoint analysis has been widely used in several fields of economics as well as in marketing research (e.g., Viscusi, Magat, and Huber 1991; Adamowicz, Louviere, and Williams 1994). However, although the use of conjoint analysis to examine health care interventions is increasing, it is not yet widely used or understood by many health care researchers (e.g., Ratcliffe 2000; Ryan and Farrar 2000; Bryan et al. 1998; Singh et al. 1998; Farrar and Ryan 1999; Ratcliffe and Buxton 1999; Bryan et al. 2000; Johnson and Lievense 2000; Ryan, McIntosh, and Shackley 1998, Vick and Scott 1998; Ryan and Hughes 1997; Ryan 1999).

The purpose of this study is to measure preferences for HIV testing using conjoint analysis and to demonstrate how conjoint analysis can be used in other contexts. Our study adds to the literature by: (1) analyzing HIV testing methods, a topic that has substantial clinical and policy interest, (2) using extensive analyses to examine the reliability and validity of the results, and (3) demonstrating how conjoint analysis can be applied to other health care interventions.

We chose conjoint analysis to analyze HIV testing methods for several reasons. First, conjoint analysis is strongly rooted in economic theory. It is derived from key assumptions of welfare economics—that economic agents, when presented with a choice, will prefer one bundle of goods over another, and that agents will attempt to maximize their satisfaction or “utility” when making choices. This basis in utility theory allows one to use powerful statistical techniques to model preferences and their interrelationships. Thus, we were able to develop new insights about preferences for HIV testing.

Second conjoint analysis quantifies the contribution of individual attribute levels to utility as well as the utility of complete test profiles. Thus, we were able to examine the utility provided by, for example, fingerstick samples, as well as the utility of tests composed of a combination of attributes that includes fingerstick samples.

Third, conjoint analysis allows quantification of the utility obtained from process attributes as well as outcome attributes. HIV tests provide utility not only through its impact on health *outcomes*, but also through factors related to the *process* of care such as where and how the test is conducted. The measurement of multiple process and outcome utilities using other preference measurement approaches such as contingent valuation is problematic (Donaldson et al. 1995; Donaldson and Shackley 1997).

Fourth, conjoint analysis allows utility estimation for any combination of attributes, including combinations that represent goods or services that may not currently be available. This feature is particularly important for examining HIV tests because of the need to develop new test methods that will encourage individuals to be tested.

Another reason we chose conjoint analysis is that it can be used to estimate willingness to pay for use in economic evaluations. In contrast to contingent valuation surveys (O'Brien and Gafni 1996), conjoint analysis surveys do not ask respondents directly for their willingness to pay. Rather, if price is included as an attribute, respondents make trade-offs between price and other attributes. Therefore, willingness to pay can be indirectly estimated, which may elicit more realistic estimates.

Lastly, conjoint analysis may better mimic actual decision making because it requires respondents to make trade-offs in a choice context. In contrast, attitude surveys do not impose a resource constraint; for example, respondents can rate all attributes as "extremely important" without having to evaluate trade-offs.

METHODS

Conceptual Framework

Conjoint Analysis is derived from early work in mathematical psychology (Luce and Tukey 1964) and is consistent with a random utility framework (Louviere, Hensher, and Swait 2000). Conjoint analysis surveys involve comparing hypothetical scenarios by ranking, rating, or choosing a particular scenario. These choices indicate the relative importance of the product attributes and provide data for estimating utility functions. Parameter

estimates indicate the marginal rates of substitution among attributes. When price is included as an attribute, its parameter can be used to rescale utility differences to calculate money-equivalent values or willingness to pay. The technique is based on three concepts:

1. Each good or service is a bundle of potential attributes.
2. Each individual has set of unique relative utility weights for attribute levels.
3. Combining the utilities for different attributes provides an individual's overall relative utility (Singh et al. 1998).

Survey Procedures

We developed a survey to examine both attitudes and preferences (utilities) about HIV tests. A previous paper (Skolnik et al. 2001) reported on the results from the attitude component as well as details of the survey procedures and sample population, while the accompanying paper compares the results from the attitude and conjoint analysis survey components as well as from focus groups (Phillips et al. 2001).

Three pilots of the instrument were conducted to determine whether the survey language and the chosen attributes and their levels were appropriate, and to test the survey procedures. Surveys were fielded at four publicly funded HIV testing locations in San Francisco, California, between November 1999 and February 2000. Of the 380 HIV testers approached, 365 agreed to complete the survey (96 percent response rate). Similar procedures for testing and counseling were in place at all sites: tests were anonymous, free of charge, and included counseling.

The two part, 10–15 minute, self-administered survey was designed to be completed while respondents waited for their counseling and testing sessions. Respondents were paid \$5 upon completion. We excluded 11 respondents who either did not get tested after completing the survey or who were missing large amounts of key data. Additionally, we excluded one respondent whose choices were arbitrary (chose Test A each time).

Steps in Survey Development and Analysis

Conjoint Analysis has been defined as consisting of five steps (Louviere, Hensher, and Swait 2000), which we use as a framework for describing our methods:

1. Defining attributes
2. Assigning attribute levels
3. Creating scenarios

4. Determining choice sets and obtaining preference data
5. Estimating model parameters.

The appendix provides a brief primer on the methods of conjoint analysis.

Step 1: Defining Attributes

We defined eight key features of an HIV test based on literature review, previous work, a focus group, review by experts, and pilot surveys (Table 1). The attributes chosen were found to be meaningful to test users and relevant for developing new tests and policies. The attributes of timeliness and accuracy, as well as privacy and anonymity, were combined because they are inherently linked and in order to avoid implausible combinations of attribute levels. For example, it is illogical to have a test that takes one to two weeks but is less accurate and it is impossible to have a test that is completely private (i.e., no one knows the results) but that is linked with the tester’s name. The attribute of privacy/anonymity thus captures both the method by which someone receives his or her test result (i.e., face-to-face contact or not) and what happens with test results, that is, whether they are anonymous or confidential (names are recorded and then reported if the individual is HIV-positive).

Table 1: Attributes and Levels Used in the Survey

<i>Attributes</i>	<i>Levels (Choices)</i>
Location	(Public clinic), doctor’s office, home
Price	(\$0), \$10, \$50, \$100
Sample collection	(Draw blood), swab mouth/oral fluids, urine sample, prick finger/fingerstick
Timeliness/accuracy	(Results in 1–2 weeks, almost always accurate), immediate results almost always accurate, immediate results less accurate
Privacy/anonymity of test results	(Results given in person—not linked to name), “only you know that you are tested”—results not linked, results given by phone—not linked, results given by phone—linked, results given in person—linked
Counseling	(Talk to a counselor), read brochure then talk to counselor

Levels in parentheses will be referred to as the “baseline scenario.”

Table 2: Sample Characteristics

<i>Sample Characteristics</i>	<i>(N= 354)</i>
Age (mean, s.d.)	34.0 (8.07)
Gender (% male)	77.4
White (%)	63
African American (%)	12
Latino/Hispanic (%)	10
Asian/Pacific Islander (%)	8
Other ethnicity (%)	7
Heterosexual (male or female) (%)	44
Gay or bisexual male (MSM) (%)	49
Bisexual female or lesbian (%)	7
Education	
Less than high school (%)	6
High school diploma (%)	11
Some college (%)	25
College degree (%)	32
Graduate school (%)	25
Income (annual household)	
Less than \$20,000 (%)	29
\$20,000–\$40,000 (%)	28
\$40,000–\$60,000 (%)	16
\$60,000 or more (%)	22

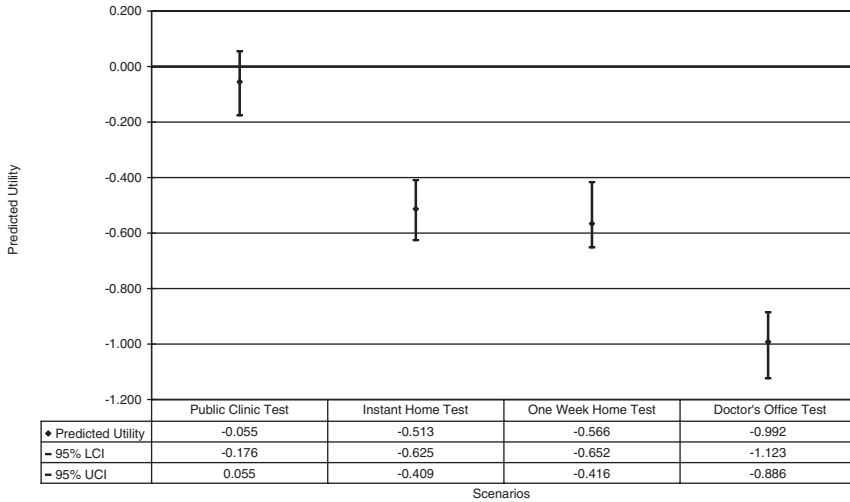
Step 2: Assigning Levels

Attribute levels were identified and reviewed for appropriateness and wording by experts in HIV testing and by a focus group of HIV testing counselors (Table 1). These levels represent the most relevant possibilities given current and expected future developments in HIV testing methods.

Step 3: Creating Scenarios

In this study, as in most conjoint analysis studies, the large number of possible combinations of attributes and levels made it implausible to generate a design based on all possible combinations. Thus, we used a fractional factorial design (based on an algorithm by [Zwerina, Huber, and Kuhfeld 1996]) to reduce the

Figure 1: Predicted Utilities for Representative Testing Scenarios



Public Clinic Test: clinic location, draw blood, \$0, accurate results in 1–2 weeks, in-person linked or unlinked results, in-person counseling.

Instant Home Test: home location, finger prick, \$50, immediate but less accurate results, only you know whether tested, brochure/phone counseling.

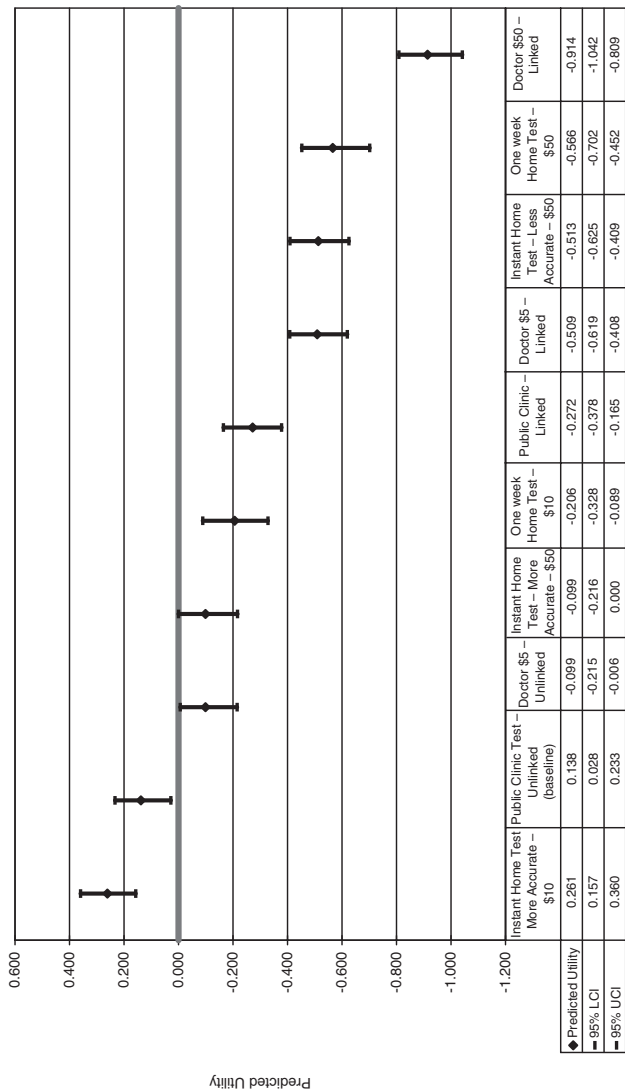
One Week Home Test: home location, finger prick, \$50, accurate results in 1–2 weeks, by phone unlinked results, brochure/phone counseling.

Doctor's Office Test: doctor's office location, draw blood, \$5, co-pay or \$50 self-pay, accurate results in 1–2 weeks, in-person or by phone linked or unlinked results, in-person or brochure/phone counseling.

number of paired comparisons to the smallest number necessary for efficient estimation of utility weights (Dey 1985). This design maximized the four properties of efficient designs (Zwerina, Huber, and Kuhfeld 1996; Huber and Zwerina 1996):

- Level balance: levels of an attribute occur with equal frequency;
- Orthogonality: the occurrences of any two levels of different attributes are uncorrelated;
- Minimal overlap: cases where attribute levels do not vary within a choice set should be minimized;
- Utility balance: the probabilities of choosing alternatives within a choice set should be as similar as possible.

Figure 2: Predicted Utilities for Testing Scenarios of interest



Scenario

Instant Home Test/More Accurate/\$10: home location, finger prick, \$10, immediate/accurate results, only you know whether tested, brochure/phone counseling.
Public Clinic Test/Unlinked Results: clinic location, draw blood, \$0, accurate results in 1-2 weeks, in-person unlinked results, in-person counseling.
Doctor's Office Test/\$5/Unlinked Results: doctor's office location, draw blood, \$50, accurate results in 1-2 weeks, linked results, in-person or brochure/phone counseling.
Instant Home Test/More Accurate/\$50: home location, finger prick, \$50, immediate/accurate results, only you know whether tested, brochure/phone counseling.
One Week Home Test/\$10: home location, finger prick, \$10, accurate results in 1-2 weeks, by phone unlinked results, brochure/phone counseling.
Public Clinic Test/Unlinked Results: clinic location, draw blood, \$0, accurate results in 1-2 weeks, in-person linked results, in-person counseling.
Doctor's Office Test/\$5/linked Results: doctor's office location, finger prick, \$5, accurate results in 1-2 weeks, linked results, in-person or brochure/phone counseling.
Instant Home Test/Less Accurate/\$50: home location, finger prick, \$50, immediate/less accurate results, only you know whether tested, brochure/phone counseling.
One Week Home Test/\$50: home location, finger prick, \$50, accurate results in 1-2 weeks, by phone unlinked results, brochure/phone counseling.

Step 4: Determining Choice Sets and Obtaining Preference Data

Our survey employs two-alternative discrete choice questions with no opt-out or reference alternative. This discrete choice approach was preferred because it mimics many real-life decisions and is consistent with random utility theory. Scenarios were placed into choice-set pairs (choice sets) by maximizing the D-efficiency score (Zwerina, Huber, and Kuhfeld 1996). Based on this criterion, 11 choice tasks were created for each survey version. Each version included two consistency checks, one question to address stability of preferences and another to address monotonicity (discussed further below). Respondents were randomized to receive one of six survey instruments.

Step 5: Estimating Model Parameters

Survey responses are interpreted as utility differences between the choice (Johnson et al. 1998). Each time respondents make a choice, they select the alternative that leads to a higher level of utility. The dependent variable is the response choice and the differences in levels for each attribute are the independent variables included in the model. Thus, we estimated the utility function as:

$$U_{ijr} = V_i(X_{ijr}, p_{ijr}, Z_i; \beta_i, \delta_i) + e_{ij} \tag{1}$$

where

U_{ijr} is individual i 's utility for alternative r in choice set j ($j = 1, \dots, 10$), and alternative r represents choosing either scenario A or B

$V_i(\cdot)$ is the nonstochastic part of the utility function

X_{ijr} is a vector of attribute levels (except for price) in choice set j for alternative r

p_{ijr} is a scalar representing the price level attribute in choice set j for alternative r

Z_i is a vector of personal characteristics

β_i is a vector of attribute parameters and

δ_i is the price parameter

Then, for person i , the utility difference in the attribute profile between scenario A and scenario B for question j is defined as:

$$\begin{aligned} \Delta U_{ij} &= V_{ijA} - V_{ijB} + (e_{ijA} - e_{ijB}) \\ &= \left[\sum_{m=0}^{17} (\beta_m X_{mAj}) + \delta p_{Aj} + \sum_{n=1}^2 \gamma_n p_{Aj} Z_n + \sum_{n=3}^7 \gamma_n X_{(n+8)Aj} Z_3 \right] \\ &\quad - \left[\sum_{m=0}^{17} (\beta_m X_{mBj}) + \delta p_{Bj} + \sum_{n=1}^2 \gamma_n p_{Bj} Z_n + \sum_{n=3}^7 \gamma_n X_{(n+8)Bj} Z_3 \right] + (e_{ijA} - e_{ijB}) \end{aligned} \tag{2}$$

Personal characteristics do not vary between alternatives for each individual. However, we may hypothesize that attribute importance is related to personal characteristics of the respondent. Such nonrandom variation in preferences can be modeled by constructing interaction terms between a personal characteristic and an attribute already in the model. The last two summations of V_{ijA} and V_{ijB} represent the interactions of interest between attributes already in the model and personal characteristics specific to our model.

The discrete choice nature of the questions suggests that either a logit or probit model be used to model the data. We assume normally distributed errors and thus use probit to obtain parameter estimates. A random-effects model is necessary to account for the potential correlation introduced by repeated observations on each respondent.

Effects coding was chosen over dummy coding of categorical variables. Using effects coding, one can compute an effect size for each attribute level, whereas with dummy coding the parameter estimate for the baseline (omitted) category cannot be recovered. When using effects coding, parameter estimates sum to zero whereby the parameter value for the base category is equal to the negative sum of the parameter values for all other categories of that variable (Appendix). Therefore, unlike the dummy coded model in which the base category is incorporated into the intercept, in an effects-coded model the intercept is a reflection of other attributes not included in the model and statistical significance is evaluated relative to the mean effect, which is normalized at zero, rather than relative to the omitted category.

STATA, version 6.0, (www.stata.com) was used for the majority of the modeling and was supplemented by bootstrapping analysis in SAS to calculate willingness-to-pay and predicted utility confidence intervals (Appendix). We estimated a main effects, attributes-only model and an interaction model that

tested differences in preferences due to education, income, and sexual orientation (interactions with gender and testing and drug use history were not significant). We also estimated the propensity to choose a testing scenario by converting probit estimates to probabilities. From the main effects model, we averaged the absolute values of the attribute level coefficients to determine which attributes have the largest impact on utility. This method uses the information from all coefficients, rather than other approaches that use the range of coefficients (Ryan 1996). The coefficient for price was multiplied by \$5 to reflect the most likely scenario in order to compare coefficient magnitudes.

Willingness-to-Pay Estimation and Comparison of Scenarios

Willingness to pay for each attribute level relative to the baseline level was calculated by dividing the coefficient differences by the price coefficient and subtracting from this the appropriate baseline attribute level willingness to pay (see Table 1). We calculated predicted utilities for four testing scenarios that are the most representative of current testing alternatives, and for nine testing scenarios that are of particular interest (Figures 1 and 2). Confidence intervals for the willingness-to-pay estimates and predicted utilities were calculated using bootstrapping. Bootstrapping allows us to determine the confidence limits through repeated sampling, using parameter point estimates and their estimated variance–covariance matrix.

Validity Issues

If subjects' stated preferences violate welfare-theoretic principles, we cannot impute valid welfare values for these subjects. However, conjoint analysis tasks are cognitively challenging and even the most attentive subjects with well-behaved preferences may report some inconsistent responses. Thus, the challenge is to evaluate whether consistency failures are serious enough to invalidate the welfare-theoretic validity of a subject's responses. We thus examined the validity of responses in detail rather than simply excluding inconsistent respondents.

We used three approaches to measuring validity: (1) consistency of preferences, (2) willingness to trade, and (3) consistency with theoretical predictions.

(1) Internal consistency was measured in two ways. The first is a test of monotonicity, which postulates that a subject should prefer more rather than less of any good. In the survey we included a "dominant pair" comparison in

which all attributes of one scenario were the same as the other with the exception of price. We expected that subjects would prefer the lower price test, all other attributes held equal. The second method is a measure of stability by which subjects are asked to consider the same discrete-choice comparison twice (early and late in the survey instrument). We would expect subjects to make the same choice both times the question is offered.

Inconsistent respondents were defined as those who did not answer either one or both of the two consistency checks as expected. Overall, 32 percent of respondents were inconsistent (25 percent answered the repeated questions and 10 percent answered the dominant pair questions incorrectly). These levels are similar to those found in other studies, which range from 9 percent to 39 percent (Ratcliffe and Buxton 1999; Ryan and Hughes 1997). We also estimated three models that included varying levels of consistent responses to determine if preferences changed with degree of consistency. The first model included only consistent respondents; the second added respondents who were inconsistent on either one of the consistency measures; the third included all respondents (consistent, inconsistent on one measure, and inconsistent on both measures). Results from these models indicate that there are no apparent differences between the overall fit and the individual attribute level coefficients of the three models.

(2) Respondents who are not willing to trade (“dominant preferences”) may violate one of the tenets of utility theory. We classified respondents as exhibiting dominant preferences if they chose a specific attribute level every time it was offered, given that it was offered at least five times and excluding choices with the same level in both (Ratcliffe and Buxton 1999). Dominance can be distinguished from the stricter criteria of lexicographic preferences, which implies no substitution between any attributes (Ratcliffe and Buxton 1999). Using this definition, 28 percent of respondents exhibited dominant preferences. However, we did further analyses to determine if including respondents with dominant preferences biased the results rather than simply excluding them from analyses (Bryan et al. 2000). Inspection of the willingness-to-pay point estimates and confidence intervals for the model consisting of all respondents compared to the model consisting of nondominant respondents suggested that the results were not biased, with the exception of phone-not-linked results.

(3) Theoretical validity was explored by examining the sign and significance of parameter estimates. Results were all as expected, including the negative coefficient for price.

In sum, we chose to include all respondents in our analysis because there was little evidence that doing so biased our results. In addition, we obtained valuable information even from respondents who were less consistent or willing to trade. Since lower socioeconomic status was associated with higher levels of inconsistency, we would have excluded a population of interest if these respondents had been excluded.

RESULTS

Respondents had relatively high levels of education and income levels compared to the U.S. population (Table 2). The majority of respondents were male and there were a high proportion of gay males. The sample was diverse in terms of race/ethnicity, with one-third of the sample from minority groups.

Based on the sign and significance of the regression coefficients, we see that respondents preferred testing at a public clinic, swabbing of the mouth or urine test samples, immediate and highly accurate results, and testing where only they know the results or results are given either in-person or by phone without linking of names (Table 3). Conversely, respondents did not prefer testing at a doctor's office, having blood drawn or a fingerstick to obtain the test sample, having to wait one to two weeks to get accurate results or getting immediate but less accurate results, results by phone or in-person with linking of names to results, and a higher price. Although respondents preferred in-person counseling, this coefficient was not significant.

We also averaged the absolute value of coefficients across *levels* of attributes (including the baseline coefficients) in order to estimate which *attributes as a whole* are most important to respondents. Privacy/anonymity is the most important attribute, followed by timeliness/accuracy, method of sample collection, location, price, and availability of counseling (Table 3).

Compared to the baseline scenario (the "public clinic" scenario with \$0 price and other attributes as defined above), increasing the price of testing to \$5 would decrease the probability of choosing the baseline-testing scenario by 2 percent (not shown). Increasing the price to \$25 would decrease the probability of choosing the baseline scenario by 9 percent. Conversely, changing from unlinked results and in-person counseling to linked results and phone counseling would decrease the probability of choosing the baseline scenario by 13 percent.

Table 3 also shows the willingness to pay for attribute levels as compared to the baseline levels. For example, the attribute level for immediate, highly

Table 3: Results from the Random Effects Probit Model—Effects Coding

Variable	Coefficient	Standard Error	WTP		
			Compared to baseline	95% CI (Lower)	95% CI (Upper)
Testing location					
Public	0.110***	0.023			
Doctor's office	-0.082**	0.024	-\$21	-\$28	-\$15
Home	-0.029	0.021	-\$15	-\$22	-\$9
Sample collection method					
Draw blood	-0.139***	0.030			
Swab/oral fluids	0.116***	0.030	\$28	\$19	\$36
Urine	0.088**	0.030	\$25	\$15	\$33
Finger prick	-0.065*	0.031	\$8	-\$1	\$17
Timeliness/accuracy					
1-2 weeks > accuracy	-0.074***	0.022			
Immediate > accurate	0.244***	0.024	\$35	\$28	\$41
Immediate < accurate	-0.170***	0.023	-\$11	-\$17	-\$3
Privacy/anonymity of test results					
Only you know	0.225***	0.036	\$1	-\$8	\$10
Results in person—not linked to name	0.217***	0.034			
Results by phone—not linked to name	0.076*	0.037	\$16	\$6	\$25
Results in person—linked to name	-0.193***	0.037	-\$46	-\$56	-\$34
Results by phone—linked to name	-0.326***	0.036	-\$60	-\$71	-\$47
Availability of counseling					
In-person counseling	0.024	0.030			
Brochure	-0.024	0.030	-\$5	-\$13	\$3
Test price					
Constant	-0.009***	0.000	NA	NA	NA
Number of observations	3366				
Number of respondents	339				
Log-likelihood	-1975.4				
Chi-square	739.18				
	($p < 0.001$)				

$P < 0.001$ ***, $p < 0.01$ ** , $p < 0.05$ *

Note: Willingness to pay (WTP) is calculated as $(\beta_x - \beta_0)/(-\delta)$ for each attribute level where β_x is the attribute level WTP of interest and β_0 is the baseline scenario attribute level.

accurate results has the largest positive willingness-to-pay value, indicating willingness to pay an additional \$35 to take a test with immediate, highly accurate results.

In the model including interactions with income, sexual orientation, and education, the interaction of (price * income) was significant, with higher income individuals ($\geq \$50,000$ income) willing to pay more ($p = 0.009$) (not shown). The interaction of (price * sexual orientation) was also significant, with men-who-have-sex-with-men willing to pay less ($p < 0.001$). The interactions of education with the privacy/anonymity variables were also significant, with higher educated respondents preferring unlinked in-person results ($p = 0.03$) and disliking linked results by phone ($p = 0.03$). Parameter estimates in the main effects and interaction models were similar, except that the levels of “phone, not linked” and “person linked” became nonsignificant in the interactions model.

Comparison of Testing Scenarios

Figure 1 shows the predicted utilities for four test scenarios that represent “typical” testing scenarios. The most preferred scenario is a public clinic test, which assumes a blood draw, \$0 price, accurate results in one to two weeks, in-person linked or unlinked results, and in-person counseling. Note that the predicted utility for even this test scenario is negative because it is less preferred than a public clinic test that has unlinked results (see Figure 2). The next preferred test is an instant home test, followed by a one-week home test and a doctor’s office test.

Several interesting results emerged when we examined how changes to representative testing scenarios could make tests more or less preferred (Figure 2). The most striking example is that instant home tests had the highest utility if they could be both highly accurate and cost only \$10 (instead of less accurate and \$50, as assumed in the representative scenario). Respondents were willing to pay \$39 for such a test, which is over twice the amount they were willing to pay for the baseline scenario (\$15 for public clinic tests). Another example is that of one-week home tests. If they cost \$10 (instead of \$50) they would have similar utility as free public clinic tests with linked results.

When we examined the results assuming a \$0 price, we found fewer differences in preferences among tests (not shown). On the one hand, the instant home test with highly accurate results was still the most preferred scenario. However, preferences became statistically equivalent (based on

overlapping confidence intervals) for the following test scenarios: instant home tests with less accurate results, one-week home tests, unlinked testing at doctors' offices, and linked public testing.

DISCUSSION

To our knowledge, this is the first study to use a theoretically based and statistically rigorous approach to measure economic preferences (utilities) for HIV testing methods. We found that respondents value several HIV test attributes, particularly immediate/highly accurate, and private/anonymous testing. Interestingly, respondents disliked immediate but less accurate results, which is a feature of the rapid test currently being used in the United States and may be a feature of future rapid tests, including those developed for home use. Similarly, respondents disliked fingerstick tests, which is a feature of the only available home collection test in the United States and has also been proposed as a possible feature of future home tests.

The results suggest that the baseline testing scenario, which reflects the typical publicly funded clinic in the United States, provides relatively high utility levels for these respondents and that the value of such testing, as measured by willingness to pay, was greater than is typically charged. This is not surprising, given that these respondents chose to be tested at a public clinic.

However, by using conjoint analysis results to estimate utilities for new combinations of test attributes, we found that instant home tests would become at least as preferred as the baseline scenario (public clinic tests) if they were highly accurate and cost \$10. These results are striking, given that these tests are not currently available and respondents were surveyed while waiting to obtain an HIV test at a public clinic. In a statewide survey of Californians, we found that more than one-third (39 percent) of respondents would consider using an instant home HIV test, including many respondents who had never been tested for HIV (Phillips and Chen, Working Paper). These results, along with the results from the conjoint analysis survey, suggest that making instant home tests available may result in significant welfare gain. However, we found that instant home tests will gain wider acceptance if they have highly accurate results and are relatively inexpensive—but these are features that are not currently available in any tests being proposed and there are technical, regulatory, and policy barriers to providing such tests.

Conversely, there was a surprising lack of interest in home collection tests as described in our survey, given earlier estimates before the tests were approved that many people would use them (Phillips et al. 1995). By using conjoint analysis, we were able to tease apart possible explanations for the lack of interest. Our results suggest that these tests are less valued because of the longer wait for results and the higher price.

This study illustrates how conjoint analysis can be used to examine other health care goods and services. Conjoint analysis is particularly useful when economic preferences and willingness to pay are relevant, when it is important to estimate relative preferences for individual attributes of goods and services, when the processes involved in obtaining the good or service are important, and when it is necessary to estimate preferences for goods and services that are currently unavailable.

Our study suggests additional areas for future research. There is a need for more research on how individuals value HIV testing and how this varies in different populations. In addition, it would be useful to understand how individuals value the outcomes provided by HIV testing (i.e., whether one is negative or positive). One challenge in examining this issue is that interventions such as HIV testing are not pure private goods, as people may value the provision of such services to others regardless of whether they plan to be tested. However, there has been little work to date on using conjoint analysis to measure social preferences (Ratcliffe 2000).

LIMITATIONS AND CONCLUSION

Our sample was not generalizable to the U.S. population or the population of individuals at-risk for HIV, and our respondents may have had more well-defined preferences than these groups. However, a general population survey would not have been as relevant and a survey of untested but high-risk individuals would have been infeasible, given the difficulties associated with locating such individuals. Perhaps most importantly, studies have demonstrated the importance of repeated testing among individuals at high-risk (Phillips et al. 1995), and therefore preferences for testing among our sample of tested individuals are still highly relevant to policy concerns. It should also be noted that our sample had a similar percentage of men-who-have-sex-with-men (MSM) as the number of HIV+ MSM in the United States and a high percentage of minorities, similar to the distribution of HIV+ individuals in

San Francisco. Lastly, we had an extraordinarily high response rate, particularly for a self-administered conjoint analysis survey.

To conclude, the method of conjoint analysis enabled us to estimate utilities for specific attributes of HIV tests as well as the overall utility obtained from various HIV tests, including tests that are under consideration but not yet available. Conjoint analysis offers an approach that can be useful for measuring and understanding the value of other health care goods, services, and interventions.

APPENDIX: METHODS PRIMER ON CONJOINT ANALYSIS

We provide an introduction to the five key steps used in conducting a conjoint analysis. Readers should refer to (Louviere, Hensher, and Swait 2000; Ryan 1999; Ryan and Farrar 2000) for further information.

1. Defining attributes.

The first step is to identify the key characteristics of the good or service being evaluated. Literature reviews and interviews are often used for this purpose.

2. Assigning attribute levels.

The second step is to assign levels to each attribute that should be realistic ranges under which an attribute may vary. Levels can be cardinal, ordinal, or categorical.

3. Creating scenarios.

The third step is to develop hypothetical scenarios that combine different levels of attributes. In studies with small numbers of attributes and levels, scenarios can be created that include all possible combinations (“full factorial design”). However, in most studies, such designs are impractical because subjects’ cognitive limitations and time constraints do not allow consideration of a large number of profiles. Thus, fractional factorial designs, which assume no interactions between attributes and that ensure the absence of multicollinearity, are used to reduce the number of scenarios. Several approaches, including proprietary and canned software packages (e.g., SPEED, SPSS ORTHOPLAN), have been used to create scenarios and place them into choice sets.

4. Determining choice sets and obtaining preference data.

The fourth step involves determining questions (choice sets) and then obtaining preference data using the choice sets. Preferences may be established using ranking, rating, or discrete choice exercises. The

discrete choice approach is often chosen because it more closely resembles many real life decisions and is consistent with random utility theory (Louviere, Hensher, and Swait 2000).

5. Model estimation.

The final step in the conjoint analysis method is estimating preferences (utilities). Probit or logit analysis is used to estimate the utility function. Random effects models are commonly used to account for the multiple observations obtained from each individual.

For two-alternative choice sets, the dependent variable is the response choice and the differences in levels for each attribute are the independent variables included in the model. A dataset is thus created where the dependent variable is the person’s choice and the independent variables are the difference in the attribute levels in the choice that the respondent saw. For example, one question could be:

	<i>Location</i>	<i>Price</i>	<i>Sample</i>	<i>Timeliness</i>	<i>Anonymity</i>	<i>Counseling</i>
Question 1						
Choice A	Public clinic	\$50	Swab mouth	Immediate accurate	Only you know Phone not linked	Brochure
Choice B	Home	\$0	Draw blood	Immediate less accurate	Phone linked In-person linked	Talk to counselor

The choices would then be effects coded as follows:

	L1	L2	P	S1	S2	S3	T1	T2	A1	A2	A3	A4	C
Choice A	1	0	50	0	0	1	0	1	1	0	0	0	-1
Choice B	-1	-1	0	-1	-1	-1	1	0	0	1	0	0	1

(Note that, when using effects vs. dummy coding, the baseline category is coded as -1.)

If we subtract attributes of Choice B from Choice A we get a vector of attribute values for person *i*, question 1. $X_{i7} = [2 \ 1 \ 50 \ 1 \ 1 \ 2 \ -1 \ 1 \ 1 \ -1 \ 0 \ 0 \ 2]$

For a series of three different questions, for which respondent 101 chose scenarios A, B, and A respectively, the data set would be as follows:

ID	Y	X _{L1}	X _{L2}	X _P	X _{S1}	X _{S2}	X _{S3}	X _{T1}	X _{T2}	X _{A1}	X _{A2}	X _{A3}	X _{A4}	X _C
101	1	2	1	50	1	1	2	-1	1	1	-1	0	0	2
101	0	1	2	-40	-1	0	1	1	2	0	0	1	-1	0
101	1	-1	-1	100	1	0	-2	-1	-1	-2	-1	1	-1	-2

Survey responses are interpreted as utility differences between the choice sets (Johnson et al. 1998). Each time a respondent makes a choice, they are choosing the alternative that leads to a higher level of utility (Ratcliffe and Buxton 1999). Therefore, we can estimate an individual *i*'s utility for a commodity profile described by question *j* as

$$U_{ijr} = V_i(X_{ijr}, p_{ijr}, Z_i; \beta_i, \delta_i) + e_{ij} \tag{1}$$

where

U_{ijr} is individual *i*'s utility for commodity profile described by question *j* ($j = 1, \dots, 10$), and choice *r* represents choosing either scenario *A* or *B*

For an individual *i*'s utility profile,

V_i is the nonstochastic part of the utility function

X_{ijr} is a vector of attribute levels (except for price) in question *j* for choice *r*

p_{ijr} is a scalar representing the price level attribute in question *j* for choice *r*

Z_i is a vector of personal characteristics

β_i is a vector of attribute parameters and

δ_i is the price parameter

Then, for person *i*, the utility difference (with no personal characteristics included) in the attribute profile between scenario *A* and scenario *B* for question *j* is defined as:

$$\begin{aligned} \Delta U_{ij} &= V_{ijA} - V_{ijB} + e_{ij} \\ &= \sum_{m=0}^{17} (\beta_m X_{mAj}) + \delta p_{Aj} - \left[\sum_{m=0}^{17} (\beta_m X_{mBj}) + \delta p_{Bj} \right] + e_{ij} \end{aligned} \tag{2}$$

We do not observe ΔU_{ij} directly. Instead, we observe choice C_{ij} , which is discrete and related to the latent variable U_{ij} . In using a random effects model, we incorporate an individual specific error term such that $e_{ij} = \eta_{ij} + \lambda_i$ where η_{ij} is the common error term and λ_i is the individual specific error term

(Greene 1993). Personal characteristics do not appear in equation 2 because they do not vary between choice sets within each individual. However, we may hypothesize that attribute importance is related to a personal characteristic of the respondent. Such nonrandom variation in preferences can be modeled by constructing interaction terms between a personal characteristic and an attribute already in the model.

From the regression model, we can estimate:

- The utility from individual attribute levels based on the magnitude, sign, and statistical significance of their coefficients.
- The total utility obtained from a combination of attribute levels based on predicted utility scores (calculated from estimated coefficients).
- The relative utilities obtained from attribute levels based on their relative size (adjusting for differences in measurement).
- Willingness to trade between attribute levels based on the ratio of any two parameter coefficients (the marginal rate of substitution).
- Willingness to pay based on the ratio of a parameter coefficient to the cost coefficient (compared to the baseline level when relevant).

WILLINGNESS-TO-PAY CONFIDENCE INTERVALS

Bootstrapping is used to determine willingness-to-pay confidence intervals because willingness to pay is not a linear transformation of the coefficients' confidence intervals. The bootstrapping technique involves drawing a sample of 1,000 observations from a multivariate normal distribution with mean vector and variance–covariance matrix from the parameter estimates and variance–covariance matrix of the parameter estimates obtained from the main-effects only random-effects probit model. Each of these observations can then be used to compute corresponding attribute level willingness to pay. The 50th and 950th values represent the 95% confidence intervals for the respective attribute level willingness to pay.

VALIDITY ISSUES

Several approaches have been used to examine the validity of responses to conjoint analysis surveys, including tests of internal consistency and reliability; willingness to trade among attributes (analyses of dominant or lexicographic

preferences); test-retest reliability; and tests of criterion, convergent, and theoretical validity. See (Ryan 1996) for an overview.

REFERENCES

- Greene, W. H. 1993. *Econometric Analysis*, 2d ed. New York: Macmillan.
- Johnson, F., W. Desvousges, M. Ruby, D. Stieb, and P. De Civita. 1998. "Eliciting Stated Health Preferences: An Application to Willingness to Pay for Longevity." *Medical Decision Making* 18 (2): S57-66.
- Louviere, J. J., D. Hensher, and J. D. Swait. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge University Press.
- Ratcliffe, J., and M. Buxton. 1999. "Patient's Preferences Regarding the Process and Outcomes of Life-Saving Technology: An Application of Conjoint Analysis to Liver Transplantation." *International Journal of Technology Assessment in Health Care* 15 (2): 340-51.
- Ryan, M. 1996. "Using Consumer Preferences in Health Care Decision Making: The Application of Conjoint Analysis." London: Office of Health Economics, Association of the British Pharmaceutical Industry.
- . 1999. "A Role for Conjoint Analysis in Technology Assessment in Health Care?" *International Journal of Technology Assessment in Health Care* 15 (3): 443-57.
- Ryan, M., and S. Farrar. 2000. "Using Conjoint Analysis to Elicit Preferences for Health Care." *British Medical Journal* 320 (7248): 1530-3.

REFERENCES

- Adamowicz, W., J. Louviere, and M. Williams. 1994. "Combining Revealed Preference and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economic Management* 6 (6): 271-92.
- Bozzette, S. A., S. H. Berry, N. Duan, M. R. Frankel, A. A. Leibowitz, D. Lefkowitz, C. A. Emmons, J. W. Senterfitt, M. L. Berk, S. C. Morton, and M. F. Shapiro. 1998. "The Care of HIV-Infected Adults in the United States. HIV Costs and Services Utilization Study Consortium." *New England Journal of Medicine* 339 (26): 1897-904.
- Bryan, S., M. Buxton, R. Sheldon, and A. Grant. 1998. "Magnetic Resonance Imaging for the Investigation of Knee Injuries: An Investigation of Preferences." *Health Economics* 7 (7): 595-603.
- Bryan, S., L. Gold, R. Sheldon, and M. Buxton. 2000. "Preference Measurement Using Conjoint Methods: An Empirical Investigation of Reliability." *Health Economics* 9 (5): 385-95.
- Centers for Disease Control. 2001. *HIV Prevention Strategic Plan through 2005*. Atlanta: Centers for Disease Control and Prevention.
- Dey, A. 1985. *Orthogonal Fractional Factorial Designs*. New York: Halstead Press.

- Donaldson, C., and P. Shackley. 1997. "Does 'Process Utility' exist? A Case Study of Willingness to Pay for Laparoscopic Cholecystectomy." *Social Science and Medicine* 44: 699-707.
- Donaldson, C., P. Shackley, M. Abdalla, and Z. Miedzybrodska. 1995. "Willingness to Pay for Antenatal Carrier Screening for Cystic Fibrosis." *Health Economics* 4 (6): 439-52.
- Farrar, S., and M. Ryan. 1999. "Response-Ordering Effects: A Methodological Issue in Conjoint Analysis." *Health Economics* 8 (1): 75-9.
- Ferri, R. 1998. "Oral Mucosal Transudate Testing for HIV-1 Antibodies: A Clinical Update." *Journal of the Association of Nurses in AIDS Care* 9 (2): 68-72.
- Huber, J., and K. Zwerina. 1996. "The Importance of Utility Balance in Efficient Choice Designs." *Journal of Marketing Research* 33 (August): 307-17.
- Johnson, F., W. Desvousges, M. Ruby, D. Stieb, and P. De Civita. 1998. "Eliciting Stated Health Preferences: An Application to Willingness to Pay for Longevity." *Medical Decision Making* 18 (2, Supplement): S57-66.
- Johnson, F., and K. Lievense. 2000. "Stated-Preference Indirect Utility and Quality-Adjusted Life Years." Durham, NC: Triangle Economic Research.
- Kassler, W., B. Dillon, C. Haley, W. Jones, and A. Goldman. 1997. "On-site, Rapid HIV Testing with Same-Day Results and Counseling." *AIDS* 11 (8): 1045-51.
- Louviere, J., D. Hensher, and J. Swait. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.
- Luce, D., and J. Tukey. 1964. "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement." *Journal of Mathematical Psychology* 1: 1-27.
- O'Brien, B., and A. Gafni. 1996. "When Do the 'Dollars' Make Sense? Toward a Conceptual Framework for Contingent Valuation Studies in Health Care." *Medical Decision Making* 16 (3): 288-99.
- Orasure Technologies I. 2001. "Orasure Technologies Files for Pre-market Approval of Oraquick Rapid HIV Test." Available at http://www.orasur.com/news/default.asp?art_id=145. Accessed on September 27, 2002.
- Phillips, K. A. 1993. "Subjective Knowledge of AIDS and HIV Testing Use." *American Journal of Public Health* 83 (10): 1460-2.
- Phillips, K. A., B. Branson, S. Fernyak, R. Bayer, and S. Morin. 1998. "The Home Collection HIV Test: Past, Present, and Future. A Background Paper Prepared for the Kaiser Family Foundation Forum on 'Understanding the Impact of New Treatments on HIV Testing'." January 28-30, 1998. San Francisco: University of California.
- Phillips, K. A., T. Coates, R. Eversley, and J. Catania. 1995. "Who Plans to Be Tested for HIV or Would Get Tested if No One Could Find out the Results?" *American Journal of Preventive Medicine* 11 (3): 158-64.
- Phillips, K. A., S. J. Flatt, K. R. Morrison, and T. J. Coates. 1995. "Potential Use of Home HIV Testing." *New England Journal of Medicine* 332 (19): 1308-10.
- Phillips, K. A., D. Flick, and J. Chen. "Willingness to Use Instant Home HIV Tests: Data from the California Behavioral Risk Factor Surveillance Survey." *Surveillance Survey*. Under review, 2002.

- Phillips, K. A., F. Johnson, and T. Maddala. 2002. "Measuring What People Value: A comparison of 'Attitude' and 'Preference' Surveys." *Health Services Research* 36 (7): 1659–79.
- Phillips, K. A., S. Morin, T. Coates, S. Fernyak, A. Marsh, and L. Ramos-Irizarry. 2000. "Home Sample Collection for HIV Testing [letter]." *Journal of the American Medical Association* 283 (2): 198–9.
- Phillips, K. A., J. Paul, S. Kegeles, R. Stall, C. Hoff, and T. Coates. 1995. "Predictors of Repeat HIV Testing among Homosexual and Bisexual Men." *AIDS* 9 (7): 769–75.
- Ratcliffe, J. 2000. "The Use of Conjoint Analysis to Elicit Willingness-to-Pay Values." *International Journal of Technology Assessment in Health Care* 16 (1): 270–90.
- Ratcliffe, J., and M. Buxton. 1999. "Patient's Preferences Regarding the Process and Outcomes of Life-Saving Technology: An Application of Conjoint Analysis to Liver Transplantation." *International Journal of Technology Assessment in Health Care* 15: 340–51.
- Ryan, M. 1999. "A Role for Conjoint Analysis in Technology Assessment in Health Care?" *International Journal of Technology Assessment in Health Care* 15 (3): 443–57.
- . 1996. *Using Consumer Preferences in Health Care Decision Making: The Application of Conjoint Analysis*. London: Office of Health Economics, Association of the British Pharmaceutical Industry.
- Ryan, M., and S. Farrar. 2000. "Using Conjoint Analysis to Elicit Preferences for Health Care." *British Medical Journal* 320 (7248): 1530–3.
- Ryan, M., and J. Hughes. 1997. "Using Conjoint Analysis to Assess Women's Preferences for Miscarriage Management." *Health Economics* 6 (3): 261–73.
- Ryan, M., E. McIntosh, and P. Shackley. 1998. "Methodological Issues in the Application of Conjoint Analysis in Health Care." *Health Economics* 7 (4): 373–80.
- Skolnik, H., K. A. Phillips, D. Binson, and J. Dilley. 2001. "Deciding Where and How to Be Tested for HIV: What Matters Most?" *Journal of Acquired Immune Deficiency Syndromes* 27 (3): 292–300.
- Singh, J., L. Cuttler, M. Shin, J. Silvers, and D. Neuhauser. 1998. "Medical Decision-Making and the Patient: Understanding Preference Patterns for Growth Hormone Therapy Using Conjoint Analysis." *Medical Care* 36 (8, Supplement): AS31–45.
- Sweeney, P., P. Fleming, J. Ward, and A. Karon. 1997. "Minimum Estimate of HIV Infected Persons Who Have Been Tested in the U.S." Interscience Conference on Antimicrobial Agents and Chemotherapy, Toronto, Abstract 017–I.
- Urnovitz, H., J. Sturge, T. Gottfried, and W. Murphy. 1999. "Urine Antibody Tests: New Insights into the Dynamics of HIV-1 Infection." *Clinical Chemistry* 45 (9): 1602–13.
- Vick, S., and A. Scott. 1998. "Agency in Health Care. Examining Patients' Preferences for Attributes of the Doctor–Patient Relationship." *Journal of Health Economics* 17 (1): 587–605.
- Viscusi, W., W. Magat, and J. Huber. 1991. "Pricing Environmental Health Risks: Survey Assessments of Risk–Risk and Risk–Dollar Trade-offs for Chronic Bronchitis." *Journal of Environmental Economics and Management* 21: 32–51.
- Zwerina, K., J. Huber, and W. Kuhfeld. 1996. *A General Method for Constructing Efficient Choice Designs*. Durham, NC: Fuqua School of Business, Duke University.