How do physician assessments of patient preferences for colorectal cancer screening tests differ from actual preferences? A comparison in Canada and the United States using a stated-choice survey

Deborah A Marshall, PhD, MHSA[1,2], **F Reed Johnson, PhD**[3], **Nathalie A Kulin, MSc**[1,2], **Semra Özdemir, MS**[3], **Judith ME Walsh, MD MPH**[4,5], **John K Marshall, MD MSc**[6], **Stephanie Van Bebber, MSc**[7], and **Kathryn A Phillips, PhD**[7]

[1]Dept. of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario

[2]Centre for Evaluation of Medicines, St. Joseph's Hospital, Hamilton, Ontario

[3]Research Triangle Institute, North Carolina

[4]Division of General Internal Medicine, University of California, San Francisco, California

[5]Women's Health Clinical Research Center, University of California, San Francisco, California

[6]Department of Medicine (Division of Gastroenterology), McMaster University, Hamilton, Ontario

[7]School of Pharmacy, University of California, San Francisco, California

Abstract

Background—Patient preferences can affect colorectal cancer screening test use. We compared utility-based preferences for alternative CRC screening tests from a stated-preference discrete-choice survey of the general population and physicians in Canada and the United States.

Methods—General population respondents (Canada, n=501; US, n=1087) participated in a survey with twelve choice scenarios and nine CRC screening test attributes. Physicians (n=100, both Canada and US) reported expected patient preferences. We estimated relative importance of attributes using bivariate probit regression analysis and calculated willingness-to-pay for various CRC screening tests.

Results—In 28% and 31% of scenarios, Canadian and US respondents, respectively, chose no screening over a hypothetical test. Canadian (45%) and US (46%) physicians expected patients to choose no screening more often.

For all groups the most important attribute was sensitivity, but physicians' perception of patients' preferences are significantly different from actual preferences. Other key attributes are those related to test performance or the testing process. Fecal DNA, colonoscopy, and virtual colonoscopy were the most preferred tests by all groups, but respondents were willing-to-pay more than physicians predicted.

Conclusion—Physicians' perception of patients' preferences are quite different from those of the general population. However, among general population and physicians, Canadian and US preferences were similar.

Introduction

As the third most commonly diagnosed cancer and with a long premalignant phase and a more favourable prognosis when detected early (American Cancer Society, 2007; Canadian Cancer Society and National Cancer Institute of Canada, 2007), colorectal cancer (CRC) is a promising target for population-based screening. CRC screening is recommended in both the US and Canada (McLeod et al., 2001; U.S. Preventive Services Task Force, 2002). Increasing the number of individuals who receive CRC screening has been identified as a health priority in both Canada and the US. Several CRC screening modalities are available and recommended, including fecal occult blood testing (FOBT), flexible sigmoidoscopy (SIG), colonoscopy (COL), and double-contrast barium enema (DCBE). These strategies vary considerably in aspects of their process (e.g., preparation and procedure) and performance (sensitivity and specificity). Given this wide range of options, physician and patients must appreciate and weigh multiple dimensions in choosing a CRC screening test (Marshall, 2000). A better understanding of general population preferences for alternative CRC screening tests can help to inform future programs and policies since a key factor driving the success or failure of any screening program is the patient's willingness to undergo the test.

The objectives of this study were to measure and compare preferences for attributes of CRC screening tests (both those currently in use and on the horizon) and the alternative of not undergoing screening (Louviere et al., 2000; Ryan et al., 1998; Ryan and Farrar, 2000b). It is critical to distinguish here between preferences and attitudes (Phillips et al., 2002a). Preferences in this paper reflect economic theory, wherein utility is maximized subject to a budget constraint (Drummond et al., 2005). Attitude surveys rate respondents' opinion about alternatives but provide no insight into how they trade off risks and benefits. This study followed an empirical, utility-based preference approach, using a stated-preference, discrete-choice survey to measure preferences. We also examined the relationship between the demographic variables (gender, income, age) on the choice to opt out of screening.

Increased emphasis on patient involvement in healthcare decisions underscores the importance of understanding preferences (Marshall, 2000; Walsh and Terdiman, 2003; Woolf, 2000). Patient preferences have been found to have a significant impact on their willingness to use health care services (Dominitz and Provenzale, 1997; Gyrd-Hansen and Sogaard, 2001; Ling et al., 2001; Marshall, 2000; Woolf, 2000). Previous studies have described attitudes to CRC screening (Greisinger et al., 2006; Ling et al., 2001; Messina et al., 2005; O'Malley et al., 2004; Salkeld et al., 2003) but there has been little empiric measurement of preferences for cancer screening. A systematic review published in 2006 identified only 8 studies that measured preferences for cancer screening programs, of which only three assessed CRC screening and none were conducted in North America (Phillips et al., 2006). To date, no study has evaluated physicians' preferences or physicians' perceptions of patients' preferences. As physicians' recommendations are strong predictors of screening utilization (Federici et al., 2006; Lewis et al., 2006; Ross et al., 2006; Shokar et al., 2006; Zapka et al., 2004), their perception of their patients' preferences may be important in influencing whether or not they recommend screening to their patients. Discrepancies between the physicians' perception of patient preferences and actual patient preferences are relevant for clinical and public communications intended to promote

screening uptake. Thus, we studied preferences in samples of both the general population and physicians.

Health care systems and CRC screening programs differ significantly between the US and Canada. CRC screening is recommended by US and Canadian professional societies (Leddin et al., 2004; Smith et al., 2003; U.S. Preventive Services Task Force, 2002; Winawer et al., 2003) using any of several available methods. Despite promoting and reimbursing all of these modalities, however, uptake of CRC screening remains poor – only 50% of Americans over 50 years of age have ever been screened (Nagamine et al., 2005). Health Canada established a National Committee on CRC screening and recommended FOBT testing every two years as the modality of choice in 50- to 74- year olds (Coombs et al., 2002) Nonetheless, only Ontario has announced province-wide population-based (FOBT) screening (Ontario Ministry of Health and Long-Term Care, 2007). Several other Canadian provinces are considering screening programs, and all endorse FOBT as the preferred modality (Barkun et al., 2004; BC Cancer Agency, 2003; Cancer Care Ontario and Action Cancer Ontario, 2004; Cancer Care Ontario and Canadian Cancer Society, 2003).

Thus, in this study, we have used an empirical, utility-based preference approach with discrete-choice stated-preference survey to measure and compare general-population and physician preferences in Canada and the US. These findings highlight differences in these groups and can help to inform policy decisions on the implementation of CRC screening programs.

Methods

Stated-Preference Approach to Elicit Preferences

Conventional strategies for CRC screening vary considerably in aspects of both process (e.g., invasiveness and preparation) and outcome (e.g., accuracy and complications) (Lieberman, 1995; Marshall, 1996a; Marshall, 1996b; Marshall, 2000; Thomas et al., 1995; Vernon, 1995; Vernon, 1997). Process attributes are not captured by more conventional preference-based outcome measures such as the quality-adjusted life year, which has been used previously to evaluate CRC screening (Dominitz and Provenzale, 1997). Stated-preference methods can model preferences for individual attributes, estimate relationships among attributes, and estimate utility of combinations of attributes that represent real or hypothetical CRC screening programs. The key advantage of stated-preference methods (Bryan et al., 2000; Ryan et al., 1998; Ryan and Farrar, 2000a) is that it is possible to quantify preferences based on the trade-offs made amongst multiattribute choices.

Undertaking a choice-format stated-preference survey entails several steps: (a) identifying key attributes and assigning levels to the attributes; (b) selecting the format of the survey and developing scenarios that describe services defined by various combinations of attributes and levels; (c) eliciting preferences; and (d) analyzing choice data to quantify the benefit of offering more favourable testing options (Louviere et al., 2000; Maddala et al., 2003a; Phillips et al., 2002b; Phillips et al., 2002a; Ryan and Farrar, 2000b). These steps provide a framework for our approach (Phillips et al., 2002b; Phillips et al., 2002a) and are summarized briefly below in the context of our study.

This study received approval from the St Joseph's Healthcare Research Ethics Board, Hamilton, Ontario, Canada (04-2369) and the UCSF Committee on Human Research (H10284-2479-01).

Identifying Key Attributes of CRC Screening Tests and Assigning Levels—A literature review, focus groups and the results of a Canadian stated-preference discrete-

choice survey (Marshall et al., 2007), were used to select 18 potential attributes for inclusion in the survey. This list was further refined based on clinical and methodological input that balanced the completeness of the test attribute descriptions with survey feasibility. The final survey included nine attributes (Table 1): test process, test frequency, requirement for a follow-up test if the initial screening test is positive, test-related pain or discomfort, preparation for the test, risk of complications, and test accuracy as measured by sensitivity and specificity. To compare subjects' willingness to accept tradeoffs amongst these attributes using a common metric, cost was included as an attribute to estimate willingness-to-pay for each attribute and attribute combination. The survey was pilot-tested on a sample of patients and physicians attending clinics at the University of California at San Francisco (UCSF) before fielding.

Survey Format and Scenario Development—The survey used a conditional two-alternative discrete-choice format with binary responses (Savage and Waldman, 2004). Given the large number of possible combinations of attributes and levels, we used a fractional factorial design. There were no *a priori* restrictions regarding allowable combination of attribute levels, including implausible combinations (e.g., FOBT requiring sedative for pain).

The experimental design was constructed using an adaptation of the search algorithm outlined in Huber and Zwerina (1996). The algorithm conducts a randomized search over the candidate design space for a design that maximizes D-efficiency. The search is constrained by the requirement that each choice set include 2 overlaps and that the larger choice probability in each profile pair never exceed 0.85 (Kanninen B, 2002). Probabilities were based on beta priors derived from a previous study. Overlaps reduce the statistical efficiency of the design but also reduce measurement error related to the difficulty of simultaneously evaluating 9 attributes (Maddala et al., 2003b). This inefficiency is offset by avoiding nearly dominated pairs that often occur using other design methods.

After screening dominated and nearly dominated profile pairs out of the candidate design space, the program evaluates 10,000 designs (the 1000 design number in the manuscript was a typographical error), and provides diagnostics for the best 6 designs. These designs had very similar D-scores. The final design was selected on the basis of: a) number of larger choice probabilities less than 0.85; b) smallest imbalance in level counts; c) smallest pairwise correlations; and d) smallest number of scenarios with technically implausible combinations. The resulting 33-row design was randomly divided into three blocks of 11 choice sets to reduce the number of choice tasks each respondent was required to evaluate. The sequence of questions was randomized, with the first choice set repeated as the last choice set for a total of 12 choice tasks. The repeated question provided a check of response consistency.

For each choice task, respondents were first asked to identify which of two hypothetical screening tests they would prefer (Choice 1 of A vs. B) (Johnson et al., 2000). Respondents were then asked to choose between that preferred screening test (A or B) and no screening (Choice 2). This design maximizes the information obtained on both the marginal rates of substitution among test attributes and conditional uptake rates. An example of a choice task from the survey is provided in Appendix A. Physicians were asked about their expectations of preferences for a "typical patient, who is an otherwise healthy 50-year old with no family/personal history of colorectal cancer". We did not have preference data from the physician's actual patients, and thus all of our comparisons are based on independent samples.

Survey Administration to Elicit Preferences—This survey was conducted using an online format during February and March 2005. Harris Interactive Inc. (Harris), an online

research company administered the survey. Harris maintains panels of potential participants who have previously agreed to be surveyed. Panels include a U.S. general population panel and a specialty panel of physicians (http://www.harrisinteractive.com/partner/hpolpanel.asp). Although the panel members have agreed to participate in Harris surveys, consent was specifically obtained from all respondents for this study.

The Harris general-population is panel recruited from over 35 sources that can be weighted to produce nationally representative estimates. Panel members from the general-population panel were eligible if they were Canadian or US residents aged 45-70 years, able to read and understand English and with no history of CRC. The eligible age range started at 45 years to capture individuals who were starting to think about screening. The target sample size was 1000 US and 500 Canadian respondents from the general population. The survey was sent to general population and physician panel members, and the survey was subsequently closed after the quota was met.

Harris also maintains a panel of US physicians (over 23,000 in 2003), appended to the AMA master files. Survey invitations were emailed randomly to general practitioners in this panel. Physicians were asked to report their expectations of patients' preferences, not what the physician himself preferred. Since Harris does not have a Canadian physician panel, invitations to the online survey were mailed to a random sample of general practitioners listed in the Canadian Medical Directory. Physician panel members were eligible if they currently practiced in the US or Canada for at least one year, were general practitioners, internists, or practiced family medicine, and saw at least 6 patients/week who are 50 years of age or older. In the US, respondents included internists and family practitioners. However, in Canada internal medicine is not a primary-care specialty, so all the Canadian physicians were general practitioners.

Respondents did not receive incentives directly from either UCSF or McMaster, but did receive the usual incentives that Harris offers. General population panel members receive points that can be accumulated and used to "purchase" goods and a chance to win a monthly prize drawing. Both US and Canadian physicians were sent invitations offering a US $50 honorarium upon completion of the survey.

Consistency Tests—Testing the consistency of respondents' responses to the series of trade-off questions establishes the internal validity of the stated-preference data. Based on our findings, we retained all respondents in the final analysis, and dropped the responses to the first scenario, which served as a warm-up question.

We assessed consistency in the following ways. First, we tested individual responses for monotonic preferences, which simply requires that respondents prefer more of any positive test feature. Thus if respondents prefer scenario B to scenario A in one question, they should prefer scenario C to scenario A in another question if C is better than B for at least one attribute, and at least as good as B for all other attributes. Our design with an optout alternative allowed us to test monotonic preferences across choice sets. If someone had a preference of optout>A>B in one choice set, but then had a preference of C>optout>D while C is worse than A, then he fails the monotonicity test. In short, when you have an optout alternative, you are only comparing A and C. If you didn't have an optout alternative, then you would need to compare all four alternatives.

Second, individual responses were assessed for stability of their preferences. Stability of preferences assumes that respondents will have same preferences at any two points during the survey. In our survey, the first scenario was repeated as the final scenario in each survey.

This particular approach is a weak test of stability because answers to the first question generally are noisy as respondents learn how to do the trade-off task.

In addition to monotonicity and stability, valid stated-preferences assume non-satiation—that is, respondents are willing to accept some decrease in one attribute for an improvement in some other attribute for all attributes in the experimental design. We identified respondents who always or nearly always chose the alternative with the better level of one attribute (dominant attribute). A respondent was considered unwilling to trade if they made the same choices in at least nine (that is, nine or more) of twelve scenarios with that attribute. We assessed the influence of dominant preferences by creating and interacting a dummy variable with the dominant attribute and running a regression analysis.

Model Estimation—Stochastic utility maximization theory provides a well-established conceptual framework for modeling individual preferences. Random utility models were used to define the utility of choice alternatives defined as a function of the attributes (McFadden, 1981). Respondent's utility or satisfaction from testing is specified as linear in test features and the utility of not testing is an alternative-specific constant:

$$
\begin{aligned}
V_{\text{test}} = &\beta_{\text{stool}} + \beta_{\text{scope}} + \beta_{\text{barium enema}} \\
&+ \beta_{\text{once}} + \beta_{\text{q 10 yr}} + \beta_{\text{q 5 yr}} + \beta_{\text{no\_ followup}} \\
&+ \beta_{\text{no\_ pain}} + \beta_{\text{mild}} + \beta_{\text{diet}} + \beta_{\text{laxative}} \\
&+ \beta_{\text{enema}} + \beta_{\text{no\_ risk}} + \beta_{1/10001/10000} \\
&+ \beta_{\text{spec=90\%}} \\
&+ \beta_{\text{spec=80\%}} \\
&+ \beta_{\text{sens=90\%}} \\
&+ \beta_{\text{sens=70\%}} + \beta_{\text{cost}} + \varepsilon_{\text{test}}
\end{aligned}
\tag{1}
$$

The V_{test} β parameters can be interpreted as relative importance weights.

We estimated the parameters using bivariate probit because the choice between two hypothetical tests followed by a test/no test choice question are correlated. Because individual-specific characteristics are constant for any pair of test alternatives, they vanish from the utility differences that determine choices among competing tests unless they are interacted with one or more test features. In this model specification, demographic variables (defined below) were interacted with D_{notest}, a dummy variable indicating that the respondent picked the no-test option. These parameters indicate the effect of individual characteristics on testing uptake. Additional interactions were tested, but were not significant.

$$
V_{\text{notest\_ general population}} = (\beta_{\text{o}} + \beta_{\text{male}} + \beta_{\text{income}} + \beta_{\text{age=5160}} + \beta_{\text{age>60}}) \cdot D_{\text{notest}} + \varepsilon_{\text{notest}} \tag{2a}
$$

$$
V_{\text{notest\_ physician}} = (\beta_{\text{o}} + \beta_{\text{years\_ of\_ experince}} + \beta_{\text{number\_ of\_ patients per week}}) \cdot D_{\text{notest}} + \varepsilon_{\text{notest}} \tag{2b}
$$

Responses to the forced-choice and the test/no-test questions clearly are correlated. Following Savage and Waldman, 2004, the variance of $\varepsilon_{\text{test}}$ in equation 1 is assumed to be 1, the variance of $\varepsilon_{\text{notest}}$ in equations 2a and 2b is assumed to be λ, and their covariance is estimated as $1/(2\lambda)$.

Categorical test characteristics were effects coded,[1] while cost was treated as a continuous variable. Standard maximum-likelihood techniques were used to estimate the parameters for this model.

The following personal characteristics were included in the final regression model: an indicator variable for gender, a continuous annual household income variable, two age dummy variables (age < 51 years-omitted category, age = 51-60 years, age >60 years) in the general population model; and years of experience and number of patients per week variables in physicians model. These variables were interacted with D_{notest}, a dummy variable indicating that the respondent selected the no-test option.

Models were evaluated for goodness of fit using the likelihood-ratio chi-square statistic for the global test of zero model coefficients and McFadden's pseudo R-square. The model results are expressed as estimates, corresponding 95% confidence intervals, and p-values. All statistical analyses were performed using GAUSS version 7.0.

The final model results were used to evaluate the relative importance of the attributes. Based on our previous findings (Marshall et al., 2007), it was expected that the most important attributes (given how the levels of each attribute were defined) would be sensitivity and specificity, followed by pain, preparation and cost and that respondents would prefer CRC screening modalities that had high sensitivity and specificity, were non-invasive with no pain or preparation, and had a small cost.

For ease of presentation and interpretation, the model results were rescaled from 0 to 10 using a linear transformation of β coefficients from 0 (the least desirable attribute level) to 10 (the most desirable attribute level). These rescaled results are presented in Figure 1 to illustrate relative attribute-level importance.

Calculating Willingness to Pay

The final model coefficients were used to calculate the respondents' predicted willingness-to-pay (WTP) for various CRC screening tests. The WTP of commonly available CRC screening modalities (FOBT, DCBE, SIG and COL) and two modalities that may soon be commonly available (DNA stool tests and virtual COL) was estimated using the model coefficients for attribute levels that most closely approximated those testing modalities. Following standard methods for stated-preference discrete-choice surveys, the available attribute levels in the survey are not always complete and accurate clinical descriptions of the actual CRC screening tests due to the practical limitations of the survey format. The attribute levels in the survey therefore focus on the most important features that distinguish the alternative CRC screening tests, and do not provide an exhaustive and complete description of all aspects of the tests. In consultation with our clinical collaborators, we mapped the characteristics for each CRC screening test to the closest attribute levels available in our survey.

For estimating WTP, the primary role of the price attribute is to obtain an accurate measure of the marginal utility of money for scaling non-price utility differences in money equivalents.

We calculated unconditional, expected WTP (EWTP) to incorporate the probability that survey respondents actually would select one of the test options. In a choice model with more than two alternatives, expected utility E(U) is:

---

[1]Choice-format stated-preference studies conventionally employ effects coding rather than dummy coding of categorical attribute levels. Dummy coding combines all the omitted categories in a single constant or intercept that is confounded with the grand mean of the latent utility function. Effects coding avoids this problem by replacing the 0 code for omitted categories with -1. Thus the coefficient for each omitted category is the negative sum of the included categories and the mean effect is 0 (Bech and Gyrd-Hansen, 2005; Hensher et al., 2005).

$$E(U) = \ln\left[\sum_{K=1}^{N} \exp(\sum_i X_{Ki} \cdot \beta_i) + \exp(\beta_{opt}^*)\right] + C \quad (3)$$

where C is a constant indicating that U is defined up to an affine[2] transform. EWTP for a new test is the difference in expected utility with and without a given test alternative in the choice set, scaled by the absolute value of the price coefficient. C falls out of the utility difference.

$$EWTP(K=N+1) = \frac{\ln\left[\sum_{K=1}^{N+1} \exp(\sum_i X_{Ki} \cdot \beta_i) + \exp(\beta_{opt}^*)\right] - \ln\left[\sum_{K=1}^{N} \exp(\sum_i X_{Ki} \cdot \beta_i) + \exp(\beta_{opt}^*)\right]}{-\beta_P} \quad (4)$$

That is, EWTP is the unconditional value of having alternative N+1 in the choice set relative to not having it in the choice set, whether the test actually is chosen or not.

EWTP for general-population respondents and physician estimates of patients' EWTP were compared using a Wald Test to determine if differences were statistically significant. Confidence intervals were estimated from 10,000 Krinsky-Robb draws on the joint parameter distribution (Krinsky and Robb, 1986).

Results

Consistency Tests

The proportion of respondents in the Canadian and US general population that passed the test for monotonicity was 98% and 99%, respectively, and it was 100% for both Canadian and US physicians. Including the inconsistent observations did not affect the parameter estimates, so we retained all observations in the final analysis.

The stability tests confirmed that the responses differed between the first and last question. We treated the first question as a warm-up question and dropped the responses to the first choice question.

The dummy-variable interaction estimates for dominant preferences were statistically insignificant, even using our conservative definition for unwillingness to accept tradeoffs. Thus, we did not include interaction terms to control for dominance in the final analysis and we included all respondents and observations.

Model Fit

Overall, the final model fit is good. The likelihood ratio chi-square statistic for global test of zero model coefficients is -6,003 and -12,688 for Canadian and US sample of the general populations sample, respectively, and -1,218 and -1,261 for Canadian and US physician sample respectively. The chi-square statistic for likelihood ratio is $p < 0.0001$ for all samples.

Respondent Demographics

The survey was completed by 501 Canadian and 1,087 US general population respondents with average age 57 years (SD = 6.6 y) and 55 years (SD = 6.7 y), respectively (Table 2). The respondents were highly educated with 15 and 11 years of education, respectively, and

---

[2]An affine transform is the sum of any linear transformation of utility arguments and a constant. Ordinal utility is unaffected by adding a constant and or any linear transformation of the arguments.

high mean income ($51,000 and $48,000, respectively). The majority of US respondents (53%) had at least some private health insurance. A family history of CRC was reported by 12% and 14%, respectively.

One-half (48%) of Canadian and 70% of US respondents reported having ever been screened for CRC, with FOBT being the most common test in both the US (51%) and Canada (31%). Seventy six and 71 per cent, respectively, stating they were "somewhat" or "very" likely to get screened when due.

The 100 participating physicians in Canada and the US had a mean ages of 42 and 45 years, respectively, and were mostly male (67% and 78%, respectively) (Table 3). The average number of years in practice was 14 for both groups, and mean numbers of patients over 50 years of age seen per week were 64 and 74, respectively. Approximately half of physicians (52%) ordered FOBT and COL 11-40 times in a typical month (48% and 52% respectively). In contrast, 81% of physicians reported they ordered or performed SIG ten or fewer times per month.

Stated-Preference Survey Results

Preferences for No Colorectal Cancer Screening—No screening was preferred over the screening test described in 28% and 31% of scenarios presented to Canadian and US general population respondents respectively (p<0.00001). This means that about 28% and 31% of the time respectively, respondents would rather have no screening than undergo their preferred screening test from the initial choice task. Physicians expected their patients to choose no screening more often (55% in Canada vs. 46% in the US p <0.0001). This difference between physicians and general population respondents was also statistically significant within each country (p<0.05).

General Population Sample Preferences for Colorectal Cancer Screening – Bivariate Probit Regression Model Results—In general, respondents preferred test processes that were non-invasive (e.g., CT scan), did not require repeated measurements over time, were associated with no pain, required no preparation, caused no complications, and had high accuracy (sensitivity 90% and specificity 100%) (Table 4, Figure 1). These results are consistent with expected directions of effect.

The attribute levels are generally well ordered and most are statistically significant (p<0.05). Only some levels for the frequency (every 5 years) and follow-up (need for an additional test to confirm a positive test result, no follow up needed) attributes, were not statistically significant. These findings were consistent between Canadian and US respondents.

Parameter estimates for interactions between income and no screening, and age and no screening, were statistically significant (p < 0.05) for both Canadian and US respondents. Respondents with a higher income level and older than 60 years were less likely to choose no screening. Among US respondents (but not Canadians), males were also less likely to choose no screening (p < 0.001).

Figure 1 illustrates the relative importance of the attribute levels on a standardized scale, given the levels specified in this study. The utility of moving from 40% sensitivity to 90% sensitivity exceeds that for any other change between attribute levels. The smallest utility difference was between needing vs. not needing an additional confirmatory test. The difference between 40% sensitivity and 90% sensitivity is about three times greater than the difference between laxative vs. no preparation and between annual testing vs. one-time testing. Utility differences for pain (mild pain level and no pain level), process (DCBE and CT scan), specificity (50% level and 90% level) and complication risk (1 in 100 and none)

are about 70-80% of the difference between 40% and 90% sensitivity. This was consistent between Canadian and US respondents in the general population.

Physician Estimates of Patient Preferences for Colorectal Cancer Screening – Bivariate Probit Regression Model Results—Physicians' estimates of patient preference for attribute levels were generally well ordered and statistically significant (p <0.05), with the exception of the need for a confirmatory test (Table 4).

The parameter estimates for interactions between physicians' years in practice and choice of no screening were statistically significant (p < 0.05) in both Canada and the US. There was also a significant interaction between the volume of patients seen aged over 50 years and choice of no screening in Canada, where physicians with greater volume were more likely to choose screening.

Physicians in Canada and the US valued a change in sensitivity from 40% to 90% more than any other change between attribute levels (Figure 1). The smallest utility difference was between needing or not needing an additional confirmatory test. Differences for pain (mild pain level and no pain level), process (DCBE and CT scan), specificity (50% level and 90% level), complication risk (1 in 100 and none), preparation (diet or laxative and no preparation) and testing frequency (annually and once only) were 50 to 80% of the difference between 40% and 90% sensitivity levels.

Predicted Willingness to Pay for Alternative CRC Screening Tests

Sample of the General Population Estimates—EWTP among general population respondents for various CRC screening tests was estimated using the results from the regression model analyses (Table 4). Estimates of EWTP were similar between Canadian and US respondents (p<0.05) (Figure 2). For both Canadian and US respondents, DNA stool testing, COL, and virtual COL were the most preferred screening tests. The most valued test among Canadian respondents (assuming average values for all personal characteristics) was virtual COL with estimated EWTP $232. The most valued test among US respondents was fecal DNA testing, with EWTP $222, but this differed by only a few dollars from EWTP for virtual COL. The high sensitivity of COL and virtual COL appeared to drive preferences despite their requirements for preparation with laxatives and some pain. The low sensitivity (40%) of DNA stool testing is offset by positive attributes such as lack of pain, lack of complication risk and lack of preparation.

We also explored EWTP for CRC screening tests (relative to no screening) within specific subpopulations. The EWTP estimates for females under 51 years old with low income ($25K) were lower than those for males over 60 years old with high income ($80K). However, willingness-to-pay values were not significantly different (p>0.05).

Physician Estimates—Physicians' predictions of EWTP for CRC screening tests were estimated using the result of regression analyses in Table 4. For both Canadian and US physicians, the most valued tests are colonoscopy, virtual COL, and DNA stool testing (Figure 2). Colonoscopy with an estimate of EWTP of $167 for Canadian physicians and $148 for US physicians. All estimates of EWTP for physicians were lower than equivalent values for the general population.

We also investigated the influence of physician characteristics on their estimates of patients' EWTP for alternative CRC screening tests. The EWTP estimates for physicians with an average of 20 years of experience and an average of 40 patients over 50 years old per week were lower than those for physicians with an average of 8 years of experience and an average of 100 patients over 50 years old per week. Willingness-to-pay values were not

significantly different (p>0.05), except for the DNA test, for which Canadian physicians thought their patients would be willing to pay significantly less ($57) than US physicians ($93).

Comparison of Estimates of EWTP—Significant differences in EWTP for all CRC screening tests were noted between general population responses and physicians' estimates (Table 5). General population respondents were willing to pay much more than physicians predicted for each CRC screening test (p<0.05), except for COL in the Canadian sample (p=0.1940).

No significant differences in EWTP were found between Canadian and US respondents in the general population or the physician sample except for fecal DNA testing, which was valued more by US physicians.

Discussion and Conclusions

Given the poor uptake of CRC screening, the increased focus on the patients' role in health care decisions, and the variety of screening choices available, this study sought to measure preferences for CRC screening tests to inform health policy decisions about the implementation of screening programs. We assessed and compared preferences among members of the general population and physicians in Canada and the US, using a stated-preference discrete-choice survey administered through the internet.

One key finding of this study is that physician assessments of patients' willingness to undergo CRC screening appear to be closer to actual uptake rates (approximately 44%) (Meissner et al., 2006) than respondents in our general population sample indicated. This is consistent with the stated preferences literature where estimates of stated preference have been shown to differ from estimates of revealed preferences based on actual behaviours (Carson et al., 1996). Two other stated-preference surveys conducted in general population samples have estimated willingness to undergo CRC screening of 66% for Danes and 88% for Australians (Gyrd-Hansen and Sogaard, 2001; Salkeld et al., 2003). This has serious implications for CRC screening uptake. Given that physicians influence decisions about health care treatment, and especially screening programs (Federici et al., 2006; Lewis et al., 2006; Ross et al., 2006; Shokar et al., 2006; Zapka et al., 2004), physicians need to help convert patients' apparent willingness to be screened into actual behavior. It may help physicians motivate their patients to be screened if they better understand what features of CRC screening are important to patients.

Another key finding of this study is that physicians' perceptions of patients' preferences are often inaccurate. Both physicians and the general population rated test accuracy (sensitivity and specificity) as the most important attribute. There has been little previous research on this topic (Phillips et al., 2006). Ling et al. used a qualitative survey to assess physician and patient attitudes (not utility-based preferences), as well as physicians' perceptions of patient attitudes (Ling et al., 2001). While patient and physician attitudes were similar, physicians underestimated the importance of test accuracy and overestimated that of discomfort. In that study, patients preferred FOBT the most (43%), followed by COL (40%), FOBT combined with SIG (12%), DCBE (3%), and SIG alone (2%) (Ling et al., 2001). Despite this, physicians recommended FOBT and SIG most often (54%), followed by FOBT (23%), SIG (15%), and COL (3%) (none recommended DCBE). A US study of attitudes towards CRC screening (n = 2,119) found that respondents who prefer to let their physicians make all decisions (n = 316) were more likely to consider COL and SIG too uncomfortable than those who prefer to make their own decisions (n = 300) (Messina et al., 2005). In our study,

physicians believed that their patients would be willing-to-pay the most for COL and the least for FOBT.

We also found that for all patient and physician groups, sensitivity is the key driver of preference for CRC screening tests. Attributes regarding the frequency of testing and the need for follow up testing are least important. These results are consistent with previous studies of preferences for CRC screening for patients (Gyrd-Hansen and Sogaard, 2001; Marshall et al., 2007). In a previous discrete-choice preferences study, we found that Canadian patients valued accuracy attributes (sensitivity and specificity) over preparation, process, and pain (Marshall et al., 2007). Using a simple survey with responses on a Likert scale, Australian respondents would consider a CRC screening tests if it had 99.9% sensitivity (92%), they would not experience pain (76%), they would need to avoid certain foods and medications for 3 days prior to the test (77%), and it required a stool sample (74%) (Salkeld et al., 2003).

Given the preferences regarding the individual attributes of CRC screening tests, it was not surprising that fecal DNA testing, COL and virtual COL are the most preferred tests when we use the model results to estimate preferences for a variety of actual CRC screening tests. Consistent with our findings regarding the differences in preferences between physicians and general population respondents, estimates of ex-ante WTP were remarkably similar for the analogous Canadian and US groups, but produced significantly higher estimates (by a factor of 6 to 8 times) for the general population samples than the physician samples.

Our results suggest that both physicians and members of the general population prefer CRC screening tests with high sensitivity. Hence, programs based on FOBT may not optimize uptake in a fully informed public. Alternatives with better performance characteristics could be offered (e.g., COL), and emerging techniques like fecal DNA testing and virtual COL should be pursued as attractive options for future.

We collected these data successfully using an online panel through the internet. Use of internet for survey research in health care is growing in popularity (Chen et al., 2005; Granger et al., 2006; Johanson and Kralstein, 2007; Lembo and Lembo, 2004; Pasternak et al., 2004; West et al., 2006). Online panels have the advantages of being easier, cheaper, and quicker to conduct compared to mail-based or telephone surveys (Dillman, 2007). However, there are concerns raised about the use of the internet in research, primarily because of the biases introduced by the selection of the sample. Weighting was used to make these panels more representative of the national populations and decrease the effects of non-sampling error, (Terhanian et al., 2000) but this does not guarantee that there are no biases since the weighting only reduces biases if the observed characteristics are correlated with the preferences.

In this study, we report estimates unconditional (ex ante) willingness-to-pay (EWTP) because there is uncertainty about whether respondents would actually choose one of the two CRC screening tests offered in each discrete-choice scenario if they could choose no screening. The difference between conditional and unconditional estimates of WTP can be explained by an example by Freeman et al. (Freeman, 1993). Suppose air pollution increases the risk of chronic lung disease. An unconditional analysis would ask healthy respondents their WTP to reduce the risk of chronic lung disease at some time in the future, that is, unconditional WTP analysis. However, a conditional analysis would ask respondents their WTP to be restored to a health state if they are already suffering from chronic lung disease.

Lancsar and Savage (2004) advocate an unconditional perspective to be consistent with welfare and random-utility theory, although there is currently debate in the health economics literature about whether conditional (Ryan, 2004; Santos Silva, 2004) or unconditional WTP

results should be reported. Which perspective is appropriate depends on the nature of the problem being analyzed. In the case of optional screening tests, the societal policy context requires accounting for the likelihood that a person will choose a given test or no test at all. Unconditional values, accounting for both the effects of perceived utility and the probability of testing, are the relevant results for this study, and are reported as the main analysis.

There are several possible policy implications of this study. Similar preferences amongst physicians and general population respondents in Canada and the US were observed despite the differences in the respective health care systems and level of diffusion of CRC screening. Both physicians and the general population need to be motivated to increase CRC screening uptake, given the low participation rate in current programs. Physicians generally exert substantial influence on decisions about health care treatment, and especially participation in screening. Understanding what is important to patients for choosing CRC screening through such quantitative stated-preference research methods as this study may help physicians better motivate their patients to undertake screening.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

American Cancer Society. Cancer Facts & Figures 2007. 2007

Barkun AN, Jobin G, Cousineau G, Dubé S, Lahaie R, Paré P, et al. The Quebec Association of Gastroenterology position paper on colorectal cancer screening - 2003. Can J Gastroenterol. 2004; 18:509–519. [PubMed: 15372115]

BC Cancer Agency. Power to the people: Population-based screening saves lives. Cancer Matters: Cancer Care and Research. 2003; 5

Bech M, Gyrd-Hansen D. Effects coding in discrete choice experiments. Health Econ. 2005; 14:1079–1083. [PubMed: 15852455]

Bryan S, Gold L, Sheldon R, Buxton M. Preference measurement using conjoint methods: an empirical investigation of reliability. Health Econ. 2000; 9:385–395. [PubMed: 10903539]

Canadian Cancer Society, National Cancer Institute of Canada. Canadian Cancer Statistics, 2007. Toronto, Canada: 2007.

Cancer Care Ontario, Action Cancer Ontario. Ontario Cancer Plan 2005-2008. 2004

Cancer Care Ontario, Canadian Cancer Society. Targeting Cancer: An Action Plan For Cancer Prevention & Detection. Cancer 2020 Summary Report. 2003

Carson RT, Flores NE, Martin KM, Wright JL. Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods. Land Economics. 1996; 72:80–99.

Chen K, Chiou C-F, Plauschinat CA, Frech F, Harper A, Dubois R. Patient satisfaction with antihypertensive therapy. J Hum Hypertens. 2005; 19:793–799. [PubMed: 15951740]

Coombs A, JonesMcLean E, LePetit C, Flanagan W, White K, Berthelot JM, et al. Technical Report for the National Committee on Colorectal Cancer Screening. 2002

Dillman, DA. Mail and Internet Surveys: The Tailored Design Method. 2. Hoboken NJ: John Wiley & Sons, Inc.; 2007.

Dominitz JA, Provenzale D. Patient preferences and quality of life associated with colorectal cancer screening. Am J Gastroenterol. 1997; 92:2171–2178. [PubMed: 9399747]

Drummond, MF.; Sculpher, MJ.; Torrance, GW.; O'Brian, BJ.; Stoddart, GL. Methods for the Economic Evaluation of Health Care Programmes. 3. Oxford: Oxford University Press; 2005.

Federici A, Giorgi RP, Bartolozzi F, Farchi S, Borgia P, Guastcchi G, et al. The role of GPs in increasing compliance to colorectal cancer screening: a randomised controlled trial (Italy). Cancer Causes Control. 2006; 17:45–52. [PubMed: 16411052]

Freeman, AMI. The Measurement of Environmental and Resource Values Theory and Methods. 1. Washington DC: RFF Press; 1993.

Granger AL, Fehnel SE, Hogue SL, Bennett L, Edin HM, Granger AL, et al. An assessment of patient preference and adherence to treatment with Wellbutrin SR: A web-based survey. J Affect Disord. 2006; 90:217–221. [PubMed: 16360216]

Greisinger A, Hawley ST, Bettencourt JL, Perz CA, Vernon SW, Greisinger A, et al. Primary care patients' understanding of colorectal cancer screening. Cancer Detect Prev. 2006; 30:67–74. [PubMed: 16458449]

Gyrd-Hansen D, Sogaard J. Analysing public preferences for cancer screening programmes. Health Econ. 2001; 10:617–634. [PubMed: 11747045]

Hensher, DA.; Rose, JM.; Greene, WH. Applied Choice Analysis: A Primer. Cambridge, UK: Cambridge University Press; 2005.

Huber J, Zwerina K. The importance of utility balance in efficient choice designs. Journal of Marketing Research. 1996; 33:307–317.

Johanson JF, Kralstein J. Chronic constipation: A survey of the patient perspective. Alimentary Pharmacology & Therapeutics. 2007; 25:599–608. [PubMed: 17305761]

Johnson FR, Banzhaf MR, Desvousges WH. Willingness to pay for improved respiratory and cardiovascular health: a multiple-format, stated-preference approach. Health Econ. 2000; 9:295–317. [PubMed: 10862074]

Kanninen BJ. Optimal Design for Multinomial Choice Experiments. Journal of Marketing Research. 2002; 39:214–227.

Krinsky I, Robb A. On approximating the statistical properties of elasticities. Rev Econ Stat. 1986; 68:715–719.

Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: Inconsistency between current methods and random utility and welfare theory. Health Econ. 2004; 13:901–907. [PubMed: 15362181]

Leddin D, Hunt R, Champion M, Cockeram A, Flock N, Gould M, et al. Canadian Association of Gastroenterology and the Canadian Digestive Health Foundation: Guidelines on colon cancer screening. Can J Gastroenterol. 2004; 18:93–99. [PubMed: 14997217]

Lembo A, Lembo A. Irritable bowel syndrome medications side effects survey. J Clin Gastroenterol. 2004; 38:776–781. [PubMed: 15365404]

Lewis CL, Kistler CE, Amick HR, Watson LC, Bynum DL, Walter LC, et al. Older adults' attitudes about continuing cancer screening later in life: a pilot study interviewing residents of two continuing care communities. BMC geriatr. 2006; 6:10. [PubMed: 16887040]

Lieberman DA. Cost-effectiveness model for colon cancer screening. Gastroenterology. 1995; 109:1781–1790. [PubMed: 7498642]

Ling BS, Moskowitz MA, Wachs D, Pearson B, Schroy PC. Attitudes toward colorectal cancer screening tests. J Gen Intern Med. 2001; 16:822–830. [PubMed: 11903761]

Louviere, JJ.; Hensher, DA.; Swait, J. Analysis and Application. New York: Cambridge University Press; 2000. Stated Choice Methods.

Maddala T, Phillips KA, Johnson FR. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Economics. 2003b; 12(12):1035–1047. [PubMed: 14673812]

Maddala T, Phillips KA, Johnson FR. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Econ. 2003a; 12:1035–1047. [PubMed: 14673812]

Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value in Health. 2007; 10:415–430. [PubMed: 17888107]

Marshall KG. Prevention. How much harm? How much benefit? 2. Ten potential pitfalls in determining the clinical significance of benefits. Can Med Assoc J. 1996a; 154:1837–1843. [PubMed: 8653643]

Marshall KG. Prevention. How much harm? How much benefit? 3. Physical, psychological and social harm. Can Med Assoc J. 1996b; 155:169–176. [PubMed: 8800074]

Marshall KG. Population-based fecal occult blood screening for colorectal cancer: will the benefits outweigh the harm? Can Med Assoc J. 2000; 163:545–546. [PubMed: 11006766]

McFadden, D. Econometric Models of Probabalistic Choice. In: Manski, CF.; McFadden, D., editors. Structure Analysis of Discrete Data with Econometric Application. Cambridge MA: MIT Press; 1981. p. 198-272.

McLeod RS. Canadian Task Force on Preventive Health Care., McLeod RS, Canadian Task Force on Preventive Health Care. Screening strategies for colorectal cancer: A systematic review of the evidence. Can J Gastroenterol. 2001; 15:647–660. [PubMed: 11694901]

Meissner HI, Breen N, Klabunde CN, Vernon SW. Patterns of colorectal cancer screening uptake among men and women in the United States. Cancer Epidemiol Biomarkers Prev. 2006; 15:389–394. [PubMed: 16492934]

Messina CR, Lane DS, Grimson R, Messina CR, Lane DS, Grimson R. Colorectal cancer screening attitudes and practices preferences for decision making. Am J Prev Med. 2005; 28:439–446. [PubMed: 15894147]

Nagamine, M.; Phillips, K.; Haas, J.; Liang, S-Y. Predictors of utilization of colorectal cancer screening tests (CRC) and adherence to the screening guidelines among low-income population. Academy Health Annual Research Meeting; Boston MA. 2005. abstract

O'Malley AS, Beaton E, Yabroff KR, Abramson R, Mandelblatt J, O'Malley AS, et al. Patient and provider barriers to colorectal cancer screening in the primary care safety-net. Prev Med. 2004; 39:56–63. [PubMed: 15207986]

Ontario Ministry of Health and Long-Term Care. McGuinty government launches first colorectal cancer screening program of its kind in Canada. Jan 23. 2007 http://www.health.gov.on.ca/english/media/news_releases/archives/nr_07/jan/nr_012307.html

Pasternak RC, McKenney JM, Brown WV, Cahill E, Cohen JD, Pasternak RC, et al. Understanding physician and consumer attitudes concerning cholesterol management: results from the National Lipid Association surveys. Am J Cardiol. 2004; 94(9A):9F–15F. [PubMed: 15219500]

Phillips KA, Johnson FR, Maddala T. Measuring what people value: A comparison of "attitude" and "preference" surveys. Health Serv Res. 2002a; 37:1659–1679. [PubMed: 12546291]

Phillips KA, Maddala T, Johnson FR. Measuring preferences for health care interventions using conjoint analysis: An application to HIV testing. Health Serv Res. 2002b; 37:1681–1705. [PubMed: 12546292]

Phillips KA, Van Bebber S, Marshall D, Walsh J, Thabane L. A review of studies examining stated preferences for cancer screening. Prev Chronic Dis. 2006; 3:A75. [PubMed: 16776876]

Ross LE, Richardson LC, Berkowitz Z, Ross LE, Richardson LC, Berkowitz Z. The effect of physician-patient discussions on the likelihood of prostate-specific antigen testing. J Natl Med Assoc. 2006; 98:1823–1829. [PubMed: 17128693]

Ryan M. Deriving welfare measures in discrete choice experiments: A comment to Lancsar and Savage (1). Health Econ. 2004; 13:909–912. [PubMed: 15362182]

Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. Br Med J. 2000a; 320(7248):1530–1533. [PubMed: 10834905]

Ryan M, McIntosh E, Shackley P. Methodological issues in the application of conjoint analysis in health care. Health Econ. 1998; 7:373–378. [PubMed: 9683097]

Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. Br Med J. 2000b; 320:1530–1533. [PubMed: 10834905]

Salkeld GP, Solomon MJ, Short L, Ward J. Measuring the importance of attributes that influence consumer attitudes to colorectal cancer screening. ANZ J Surg. 2003; 73:128–132. [PubMed: 12608975]

Santos Silva J. Deriving welfare measures in discrete choice experiments: A comment to Lancsar and Savage (2). Health Econ. 2004; 13:913–918. [PubMed: 15362183]

Savage S, Waldman DM. United States demand for internet access. Review of Network Economics. 2004; 3:228–247.

Shokar NK, Carlson CA, Shokar GS, Shokar NK, Carlson CA, Shokar GS. Physician and patient influences on the rate of colorectal cancer screening in a primary care clinic. J Cancer Educ. 2006; 21:84–88. [PubMed: 17020519]

Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2003. Cancer J Clin. 2003; 3:27–43.

Terhanian G, Bremer J, Smith R, Thomas R. Correcting data from online survey for effects of nonrandom data selection and nonrandom assignment. Harris Interactive. 2000 White Paper.

Thomas W, White CM, Mah J, Geisser MS, Church TR, Mandel JS. Longitudinal compliance with annual screening for fecal occult blood. Minnesota Colon Cancer Control Study. Am J Epidemiol. 1995; 142(2):176–182. [PubMed: 7598117]

U.S.Preventive Services Task Force. Screening for colorectal cancer: Recommendation and rationale. Ann Intern Med. 2002; 137:129–131. [PubMed: 12118971]

Vernon SW. Adherence to colorectal cancer screening. A brief overview. Ann NY Acad Sci. 1995; 768:292–295. [PubMed: 8526370]

Vernon SW. Participation in colorectal cancer screening: A review. J Natl Cancer Inst. 1997; 89:1406–1422. [PubMed: 9326910]

Walsh JM, Terdiman JP. Colorectal cancer screening: Scientific review. J Am Med Assoc. 2003; 289:1288–1296.

West R, Gilsenan A, Coste F, Zhou X, Brouard R, Nonnemaker J, et al. The ATTEMPT cohort: A multi-national longitudinal study of predictors, patterns and consequences of smoking cessation; introduction and evaluation of internet recruitment and data collection methods. Addiction. 2006; 101:1352–1361. [PubMed: 16911736]

Winawer S, Fletcher R, Rex D, Bond J, Burt R, Ferrucci J, et al. Colorectal cancer screening and surveillance: Clinical guidelines and rationale - Update based on new evidence. Gastroenterology. 2003; 124:544–560. [PubMed: 12557158]

Woolf SH. The best screening test for colorectal cancer: A personal choice. N Engl J Med. 2000; 343:1641–1643. [PubMed: 11096175]

Zapka JG, Lemon SC, Zapka JG, Lemon SC. Interventions for patients, providers, and health care organizations. Cancer. 2004; 101(5 Suppl):1165–1187. [PubMed: 15329892]
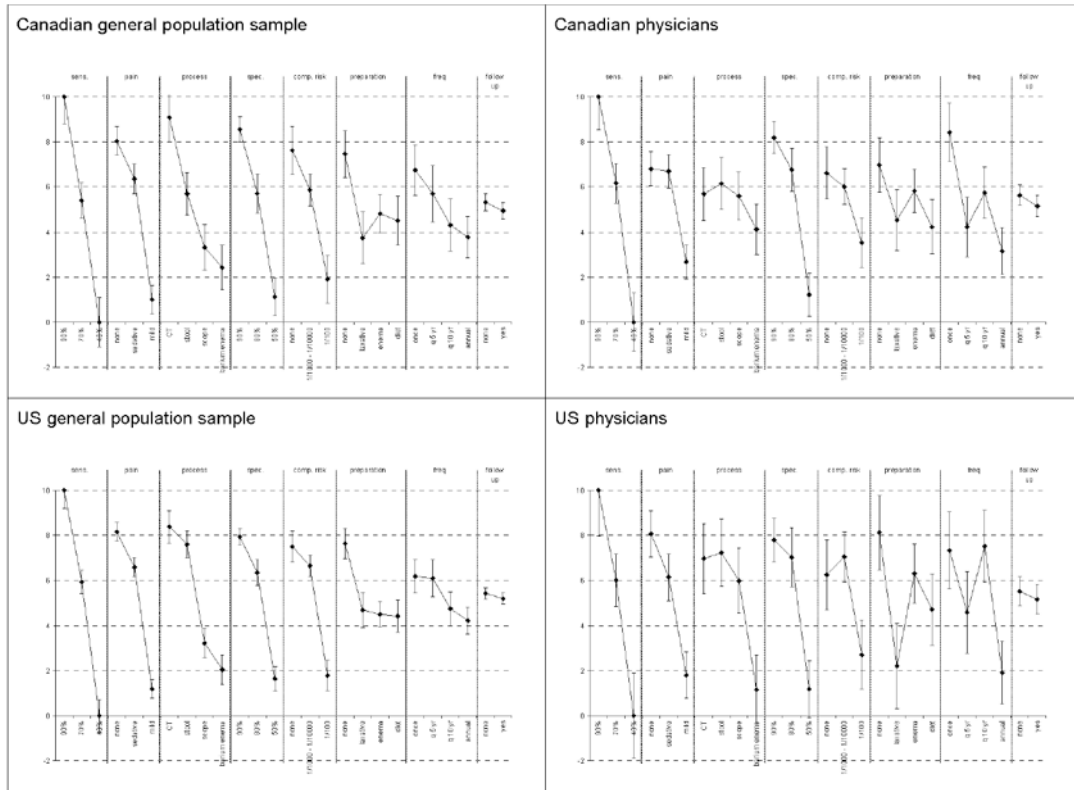
Figure 1.
Relative importance of CRC screening test attributes and attribute levels for sample of the general population and physician respondents.
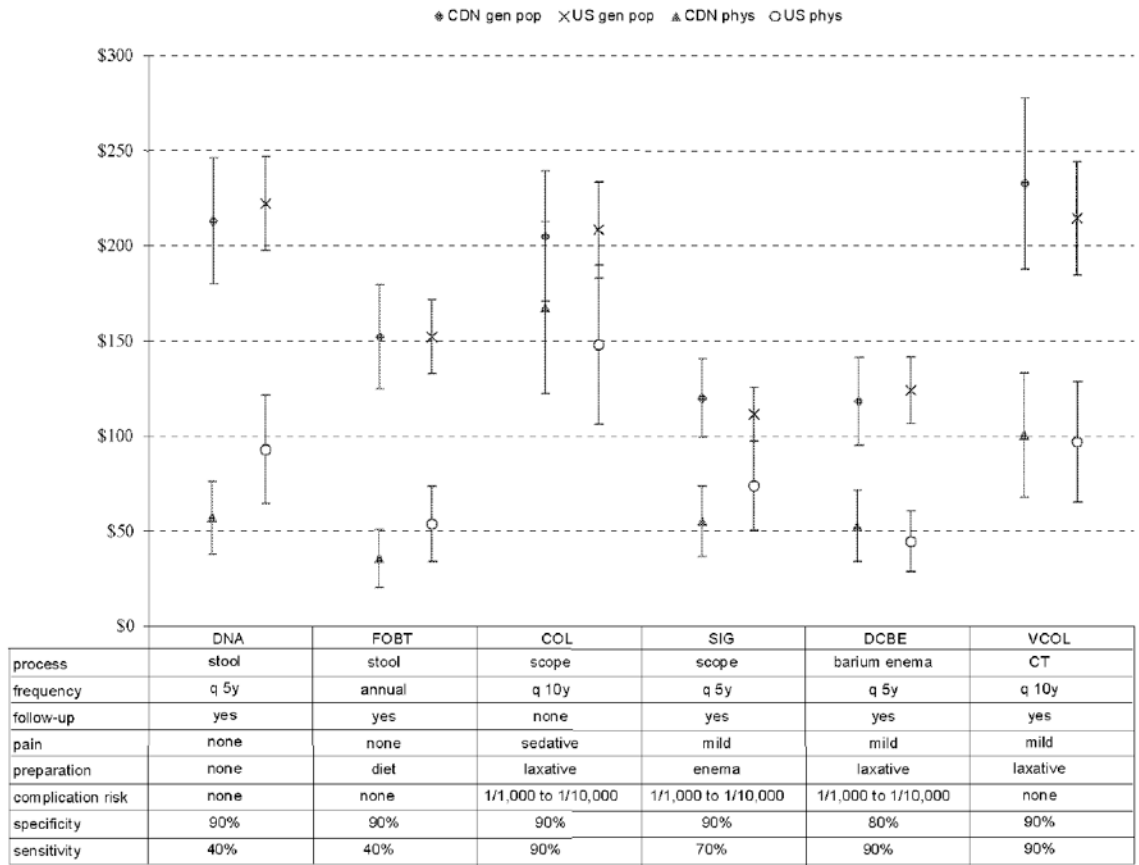
Figure 2.
Predicted Willingness to Pay for Alternative Colorectal Cancer Screening Tests

Table 1

Attributes and levels used in stated-preference survey.

Attribute, as presented to respondent (analysis short form)	Levels, as presented to respondent	Analysis short form of level name
A. How is it done? (process)	You place 2 stool samples on to special cards for 3 consecutive days and return them to your doctor for testing.	stool
	A flexible tube is inserted into your rectum and through your colon. This is done at the hospital.	scope
	Air and a white liquid are injected into your colon through a tube inserted into your rectum. X-rays are taken as the liquid moves through your colon. This is done at the hospital.	barium enema
	You have a CT scan. This is done at the hospital.	CT
B. Is there pain or discomfort? (pain)	You feel no pain or discomfort.	none
	You may feel pain or discomfort like intense cramps.	mild
	You will be given a sedative so that you should only feel pressure or mild discomfort, but you will feel sleepy and need to relax for the rest of the day.	sedative
C. How often will the screening test be done? (frequency)	You will have the screening test only once.	once
	You will have the screening test every 10 years.	q 10y
	You will have the screening test every 5 years.	q 5y
	You will have the screening test every year.	annual
D. What do I do to prepare? (preparation)	For 5 days you must avoid certain foods (such as red meat and some vegetables) and medications (such as aspirin and vitamin C).	diet
	Before the test you must take a laxative to clean your colon. After the laxative, you can only have clear liquids until after your test.	laxative
	One hour prior to the test you must have an enema to empty the last part of your colon.	enema
	No preparation is required.	none
E. If this screening test result is abnormal, will an additional test be needed to confirm whether you have cancer? (follow-up)	No.	none
	Yes, you will undergo another test on a different day.	yes
F. If 10 people without cancer get this screening test, how many of them will the test say do have cancer? (specificity)	The screening test will say that 5 out of 10 people without cancer do have cancer.	50%
	The screening test will say that 2 out of 10 people without cancer do have cancer.	80%
	The screening test will say that 1 out of 10 people without cancer do have cancer.	90%
G. If 10 people with cancer get this screening test, how many of them will the test say do not have cancer? (sensitivity)	The screening test will say that 6 out of 10 people with cancer do not have cancer.	40%
	The screening test will say that 3 out of 10 people with cancer do not have cancer.	70%

Attribute, as presented to respondent (analysis short form)	Levels, as presented to respondent	Analysis short form of level name
	The screening test will say that 1 out of 10 people with cancer do not have cancer.	90%
H. How many people who get this screening test have a complication? (complication risk)	No one who gets this screening test will have a complication.	none
	1 out of 100 people who get this screening test will have a complication.	1/100
	1 out of 1,000 people who get this screening test will have a complication.	1/1,000 to 1/10,000 [*]
	1 out of 10,000 people who get this screening test will have a complication.	
I. What is the cost of the screening test to me? (cost)	$25	$25
	$100	$100
	$500	$500
	$1000	$1000

[*] Respondents ranked 1/1,000 higher than 1/10,000, therefore we merged these two levels

Table 2

Self-reported characteristics and history of general population respondents (n=1588)

characteristic	distribution	
	Canada	**US**
number of general populations (n)	501	1087
% white	97%	86%
% male	45%	48%
age, mean (SD)	55 ± 6.7 y	57 ± 6.6 y
years of education, mean (SD)	14 years ± 2.2	15 years ± 2.6
annual income, mean (SD)	CA $48,214 ± $24,925	US $50,972 ± $26,071
health history		
non-cancer chronic health condition [*]	82% yes	83% yes
cancer (non-CRC)[†]	7% yes	9% yes
relatives with colorectal cancer	14%	12%
test history		
any genetic test	4% yes	6% yes
ever screened for CRC[‡]	48% yes	70% yes
FOBT	31% yes	51% yes
sigmoidoscopy	19% yes	34% yes
colonoscopy	24% yes	46% yes
plans to have CRC screening when due		
very unlikely	16%	12%
somewhat unlikely	13%	12%
somewhat likely	36%	29%
very likely	35%	47%
health insurance[§]		
public[‖]	88%	31%
private	52%	74%
none	2%	8%
not sure	0%	2%

[*] Includes: hypertension, high cholesterol, depression, asthma, ulcer, arthritis, diabetes, and chronic back pain

[†] General populations reporting having had CRC were excluded from participating in the survey.

[‡] Reported ever having FOBT, sigmoidoscopy, or colonoscopy.

[§] Totals sum to more than 100% since some general populations had both public and private insurance

[‖] For Canada, includes universal health care and federal or provincial government prescription drug plans. For US, includes: Medicaid, Medicare, Medigap, military, and Indian health service.

Table 3

Physicians' Characteristics

Characteristic	Distribution	
	Canada	**US**
number of physicians	100	100
% male	67%	78%
age (years, mean (SD))	42 ± 8.9	45 ± 8.3
years of experience	14 ± 9.0	14 ± 7.6
number of patients 50 years seen in a week	64 ± 33.2	74 ± 27.1
primary medical specialty	91% family medicine 9% general practitioner	47% internist 53% family medicine

Table 4

Preferences for Canadian and US General Population and Physician Respondents - Bivariate Probit Regression model results

Attribute	Level	Canadian General Population		US General Population		Canadian Physicians		US Physicians	
		parameter estimate	SE	parameter estimate	SE	parameter estimate	SE	parameter estimate	SE
	# observations	9,988		21,142		2,200		2,178	
	# respondents	454		961		100		99	
Process	stool	0.0357		0.1543	***	0.0903		0.1591	*
	scope	-0.1145	***	-0.1414	***	0.0240		0.0548	
	barium enema	-0.1704	***	-0.2208	***	-0.1496	*	-0.3510	***
	CT†	0.2492	***	0.2079	***	0.0353		0.1371	*
Frequency	once	0.1019	**	0.0589	*	0.3585	***	0.1674	*
	q 10 yr	-0.0518		-0.0381		0.0422		0.1831	**
	q 5 yr	0.0357		0.0531		-0.1371		-0.0630	
	annual†	-0.0858	**	-0.0739	***	-0.2636	***	-0.2875	***
Follow-up	none	0.0118		0.0080		0.0279		0.0148	
	yes†	-0.0118		-0.0080		-0.0279		-0.0148	
Pain	none	0.1830	***	0.1925	***	0.1667	***	0.2293	***
	mild	-0.2603	***	-0.2788	***	-0.3210	***	-0.2976	***
	sedative†	0.0773	***	0.0863	***	0.1543	***	0.0683	
Preparation	diet	-0.0393		-0.0606	*	-0.1373		-0.0525	
	laxative	-0.0879	*	-0.0420		-0.1016		-0.2632	**
	enema	-0.0198		-0.0548	**	0.0513		0.0811	
	none†	0.1470	***	0.1574	***	0.1876	**	0.2346	***
Complication risk	none	0.1572	***	0.1481	***	0.1448	*	0.0767	
	1/1000 to 1/10000	0.0464	*	0.0902	***	0.0744		0.1439	**
	1/100†	-0.2036	***	-0.2383	***	-0.2192	***	-0.2206	***
Specificity	90%	0.2161	***	0.1781	***	0.3305	***	0.2070	***

Attribute	Level	Canadian General Population		US General Population		Canadian Physicians		US Physicians	
		parameter estimate	SE	parameter estimate	SE	parameter estimate	SE	parameter estimate	SE
	80%	0.0364		0.0700	***	0.1627	**	0.1421	*
	50%†	-0.2525	***	-0.2481	***	-0.4932	***	-0.3491	***
Sensitivity	90%	0.3071	***	0.3170	***	0.5448	***	0.3916	***
	70%	0.0167		0.0421	*	0.0920		0.0564	
	40%†	-0.3238	***	-0.3591	***	-0.6368	***	-0.4480	***
	cost constant	-0.0010	***	-0.0010	***	-0.0019	***	-0.0018	***
marginal utility of money		0.0085		0.0095		0.0134		0.0199	
	NO TEST	-0.9087	**	-0.4592	***	0.1801		-0.1886	
	ln(LAM)	0.8482	**	0.9289	***	0.4486	**	0.4296	**
Personal characteristic interactions									
	male	-0.1570		-0.3247	***				
	income‡	-0.3849	*	-0.5944	***				
	age 51-60 years	-0.0444		-0.1097					
	age > 60 years	-0.4343	*	-0.5039	***				
Physician characteristic interactions									
	experience (y)					0.0199	*	0.0162	*
	n pts >50 y/wk					-0.0068	**	-0.0041	
Model goodness of fit statistics									
	-2 Log likelihood, full model:	12005.42		25376.52		2436.10		2521.93	
	-2 Log likelihood, restricted model:	12794.07		27710.83		3003.14		2983.58	
	likelihood ratio χ^2	788.64	**	2334.31	**	567.04	**	461.65	**
	degrees of freedom	24		24		22		22	
	Maddala's pseudo-R^2	0.1461		0.1981		0.4028		0.3455	
	McFadden's pseudo-R^2	0.0616		0.0842		0.1888		0.1547	
	Craig & Uhler's pseudo-R^2	0.2724		0.3589		0.6461		0.5741	

SE = standard error;

*
= p<0.05,

**
= p<0.01,

***
= p<0.001

[†] calculated as the negative sum of the parameter estimates of all other levels for the same attribute

[‡] $ CDN for CDN general population and $ US for US general population

Table 5

Comparison of Willingness to Pay, by Country and by Respondent Group

CRC screening test*	estimate (95% CI)				p-value			
	Canada		US		general population vs. physician		Canada vs. US	
	general population	physician	general population	physician	Canada	US	general population	physician
DNA	$213 ($180 to $246)	$57 ($38 to $76)	$222 ($197 to $247)	$93 ($64 to $121)	***	***	NS	*
FOBT	$152 ($125 to $179)	$35 ($20 to $51)	$152 ($133 to $172)	$54 ($34 to $74)	***	***	NS	NS
COL	$205 ($171 to $239)	$167 ($122 to $212)	$208 ($183 to $233)	$148 ($106 to $190)	NS	*	NS	NS
SIG	$120 ($99 to $140)	$55 ($36 to $73)	$111 ($97 to $125)	$74 ($50 to $97)	***	**	NS	NS
DCBE	$118 ($95 to $141)	$53 ($34 to $71)	$124 ($106 to $141)	$44 ($29 to $60)	***	***	NS	NS
VCOL	$232 ($187 to $278)	$100 ($68 to $133)	$214 ($184 to $244)	$97 ($65 to $128)	***	***	NS	NS

*
Attribute levels for CRC screening tests are defined in Figure 2.

†
* = p<0.05,

**
= p< 0.01,

***
= p<0.001,

NS = not statistically significant

‡
$ CDN for CDN general population and $ US for US general population