Title

Choosing a Doctor: Does presentation format affect the way consumers use health care performance information?

Running title

Presenting performance information for consumers

Authors

Patricia Kenny[1]

Stephen Goodall[1]

Deborah J. Street[1]

Jessica Greene[2]

Affiliation

1. Centre for Health Economics Research & Evaluation

University of Technology Sydney

PO Box 123

Broadway 2007 NSW

Australia

2. Marxe School of Public and International Affairs

Baruch College, City University of New York

Corresponding author

Patricia Kenny

patsy.kenny@chere.uts.edu.au

Telephone +41 2 95144750, Fax +41 2 95144730

Abstract

Background

Choosing a new health service provider can be difficult and is dependent on the type and clarity of the information available. This study examines if the presentation of service quality information affects the decisions of consumers choosing a general medical practice.

Objectives

To examine the impact of presentation format on attribute level interpretation and relative importance.

Methods

A discrete choice experiment eliciting preferences for a general medical practice was conducted using four different presentation formats for service quality attributes: 1) frequency and percentage with an icon array, 2) star ratings, 3) star ratings with a text benchmark, and 4) percentage alone. A total of 1208 respondents from an online panel were randomised to see two formats, answering nine choices for each, where one was a dominated choice. Logistic regression was used to assess the impact of presentation format on the probability of choosing a dominated alternative. A generalised multinomial logit model was used to estimate the relative importance of the attribute levels.

Results

The probability of incorrectly choosing a dominated alternative was significantly higher when the quality information was presented as a percentage relative to a frequency with icon array, star rating or bench-marked star rating. Preferences for a practice did not differ significantly by presentation format, nor did the probability of finding the information difficult to understand.

Conclusions

Quantitative health service quality information will be more useful to consumers if presented by combining the numerical information with a graphic, or using a star rating if appropriate for the context.

Key points for decision makers

Information about service quality such as that collected in patient experience surveys would be useful for consumers choosing a new general practice.

Presentation of performance quality information for consumers will be most useful if presented as frequency and percent with icons or start ratings.

1. Introduction

Consumer choice of health service provider has been seen as an important element of many health care systems and, in Australia is formally recognised in health policy documents such as the Health Performance and Accountability Framework [1]. Despite increasing interest in public reporting of health service quality in high-income countries, like many countries, Australia does not make health care quality information on individual providers or practices available for consumer decision-making [2]. Even when comparative information is available, relatively few patients use it to choose a provider, due to a number of barriers including insufficiently clear presentation of the information [3]. This paper compares several formats for the presentation of health service quality information for consumers choosing a provider in the context of primary care so that such information can be presented in the most useful manner.

There is a large literature on the presentation of quantitative information for consumer or patient health care decision making and a number of published literature reviews. A review of the impact of visual display format found that pictographs (visual representation of a proportion) and tables (containing words, numbers and/or icons such as stars) were easier to understand than bar charts and tables led to more accurate choices, while pictographs increased risk perception leading to more risk averse choices [4]. It is not clear which symbols (in tables or pictographs) are most effective as few studies examined the same formats and many used small sample sizes [5]. Two reviews focussing on the presentation of risk information found no single visual presentation format to be consistently superior and recommended the use of both numeric and nonnumeric formats concurrently [6, 7]. Numeracy levels were also found to impact on the understanding of different formats. Risk descriptions alone were associated with poorer risk comprehension with a tendency to overestimate the risk [8], while the combination of percentage and a risk label resulted in better comprehension than either numeric or label alone [9]. In the context of presenting comparative health performance information, the use of a simplified framework and plain language to describe medical practitioner quality indicators was found to improve consumers' understanding as well as ratings of the information's usefulness [10]. Minimising the amount of information presented was also found to improve understanding of hospital quality information, particularly for those with low numeracy [11].

The above studies were mainly quantitative experiments employing hypothetical choices in which study participants were randomised to view one version of a specific comparative scenario. These were analysed by comparing the percentage choosing the best or correct answer and using mean scores on rating scales or on indices constructed from multiple questions based upon the

presentation version viewed. A discrete choice experiment (DCE) is a useful approach to examine presentation format because it allows for the assessment of the impact of presentation format on choice over multiple choices where the quality attributes are varied, rather than a single scenario. The modelling approach allows for the comparison of presentation formats, conditional on the levels of attributes of the choice. It also allows for the assessment of accuracy of understanding through the analysis of responses where one alternative is objectively better or equivalent across all attributes.

Harrison et al [12] reviewed the presentation and analysis of risk as an attribute in DCEs and identified the need for experimental studies of the impact of different methods of presentation of numerical information. We found only one study which analysed the impact of presentation format in the context of a health related DCE [13]. It found that words were preferred by respondents, graphics did not improve attribute comprehension and that presentation format impacted on preferences for infant vaccination. However this study compared the use of either words or graphics to depict the levels of all attributes, not just potentially challenging quantitative information such as rates and percentages, and did not use numeric and graphical information concurrently. It is likely that graphically depicting familiar quantitative information such as cost, or qualitative information such as location, may have made the graphical representations more challenging for responders.

Motivated by the Australian Government's recent interest in a primary health care policy involving general practice enrolment [14], we investigated the presentation of performance information for choosing a general medical practice. The Australian health care system has previously not required general practice enrolment for any patients, but is currently conducting a trial of health care homes for people with chronic disease. We used a DCE to investigate the information consumers would find useful when choosing a new general practice and the best way of presenting the information.

2. Methods

This study used a DCE to elicit preferences when choosing a general medical practice and to evaluate different formats for presenting practice quality information. It recruited 1,208 adults aged 18 years or more from an Australian panel, with age and gender quotas set to be representative of the Australian adult population. Respondents completed an online survey which included 18 choice questions as well as health, GP utilisation, subjective numeracy and socio-demographic questions, and questions about the choice question presentation format.

2.1 Choice experiment

Each choice question presented two hypothetical general medical practices (Practice A and Practice B) and respondents were asked which they would choose if they had to find a new practice. The hypothetical practices were described in terms of 10 attributes of a medical practice (Table 1), which were developed based on an earlier study [15] involving a population survey where respondents rated the importance of a list of 36 attributes of general practice identified from a literature review. That analysis identified attributes most frequently rated as important, as well as key attribute groupings from which attributes were selected for the current study. The survey was reviewed by a consumer focus group and subsequently modified before pilot testing with the first 200 online respondents who were invited to provide feedback.

The first five attributes related to service quality and covered communication, decision making, examinations, patient satisfaction and the availability of urgent appointments (see Table 1). These attributes were presented as information arising from surveys of patients attending the practice, explained in the introductory material as well as in the further detail available by hovering on the attribute name. Four different presentation formats (or styles) were used for the levels of the five service quality attributes and these are illustrated in Table 2. The formats were 1) Frequency and percentage with an icon array, 2) star ratings, 3) star ratings with a text benchmark (e.g. above average), and 4) Percentage alone. The levels represented as percentages were chosen to reflect the range of values observed in practice, based on Australian data [16] where this was available and on UK data [17] when Australian data was not available. The levels were ordered categories and respondents were not required to interpret the star ratings as different representations of the percentages, but merely to interpret a lower number of stars as worse than a higher number of stars and a lower percentage as worse than a higher percentage.

The other five attributes covered practice ownership, out-of-hours appointments, travel time, whether or not someone recommended the practice and the typical out-of-pocket cost for a visit. The levels for these attributes were presented as text or numbers only as per Table 1. Each respondent saw two of the four presentation formats which were randomly assigned in pairs such that each respondent saw one numeric format (either Style 1 or 4) and one star rating format (either Style 2 or 3). The presentation format combinations are shown in Figure 1, along with the respondent numbers completing each combination. Nine choice questions were seen for each format as a block and the order of the presentation formats was also randomised. Figure 2 presents a sample choice set for one of the presentation styles (style 1).

The experiment of ten attributes (8 by 3-level and 2 by 2-level) produces $3^8 \times 2^2 = 26{,}244$ possible attribute level combinations, which is too large to include in a single experiment. We therefore varied the attribute levels using a Street and Burgess experimental design of 216 pairs, which was replicated for each presentation style[18]. The design was produced from a starting design with 108 options, using 2 generators to give 216 pairs. The 216 choice scenarios were blocked into 27 versions of 8 which were replicated for each of the 4 presentation styles. Each respondent was presented with 2 versions of 8 choices from the design, one version for each of the two presentation formats seen. The version and presentation format combinations were constructed so that each combination included 2 different versions (i.e. 16 different design choice scenarios). A further 2 choice sets were added to include a dominated choice (see next paragraph) for each presentation format, to make the total number of choices per respondent 18. The study aimed for a sample size of 1,200 to enable separate analyses by presentation style sub-group. This sample size would allow for approximately 600 respondents per sub-group and for more than 20 respondents per design choice set within each sub-group. An empirical study [19] investigating design performance for different numbers of respondents per design choice set found that designs of the sort used in this study were always able to estimate model parameters with 10 respondents per choice set, although standard errors do decrease as the number of respondents increases, with a quadrupling of respondents resulting in a halving of the standard error across the range that they investigated. As we are particularly interested in estimating choice probabilities, we also calculated the sample size with a view to being able to estimate the true proportion to within 5% with a probability of 0.95 for values of p between 0.1 and 0.9 (see [20] equation 2 (but corrected so that α_2 is replaced by $(1-\alpha_2)$)). This gave a required sample size of 600 (for the most extreme p).

2.2 Evaluation of presentation format

Multiple approaches were used to examine the impact of the presentation styles and to determine if there was a single best or preferred format for presenting general practice quality information to consumers. The first choice of each block of nine was a dominated choice, meaning that one alternative was better (or equivalent) on all attributes. The five service quality attribute levels were all dominated (worse) for one alternative, while the remaining attribute levels were equal across the alternatives. The dominated choices were included for assessment of a choice error rate for each presentation format. Logistic regression was used to estimate the impact of the presentation style on the probability of choosing the worst alternative, using robust standard errors to account for clustering by respondent as each respondent answered two dominated choices. The model was adjusted for age, gender, time taken to complete the choice questions, whether the response

related to the first or second dominated choice seen by that respondent and subjective numeracy. Subjective numeracy was measured with the 3 item version of the Subjective Numeracy Scale (SNS-3) [21] which includes 3 questions about an individual's self-assessed ability with numerical information and perceptions of its usefulness, measured on a Likert-type response scale of 1 to 6. The SNS-3 was scored as the mean of the responses to the three questions where a higher score represents better numeracy.

The time taken to complete each choice question was recorded as an indicator of the ease or difficulty of completion. The Kruskal-Wallis test was used to test for differences between the presentation styles for the time taken to complete the 9 choice questions. This was done separately for the first and second set of 9 choice questions because of the expectation of learning effects as respondents became more familiar with the DCE choice scenarios.

The survey included questions about how easy or difficult it was to understand the information in the choice questions and which format respondents found best for choosing a practice. At the end of each set of 9 choices, respondents were asked to rate how easy or difficult the information was to understand on a 5-point scale from 1 "Extremely easy" to 5 "Extremely difficult". As the responses were highly skewed, they were dichotomised for analysis by combining 4 and 5 and combining 1, 2 and 3. Logistic regression was used to estimate the impact of presentation style on the probability of rating the choice information as difficult (4 or 5), while accounting for age, gender, time taken to complete the choice questions, whether the response related to the first or second presentation style seen by that respondent and subjective numeracy. Robust standard errors were used to account for clustering by individual as each respondent completed this question twice (once for each style seen). After completion of all choice questions, respondents were asked which presentation style they thought was better for choosing between medical practices (of the two styles they saw). The proportion of respondents who saw the presentation style and selected it as better is presented for each style and the chi-squared statistic was used to test if this differed between respondents with high and low subjective numeracy.

2.3 Analysis of preferences

The choice data (excluding dominated choices) were analysed using the generalised multinomial logit (G-MNL) model [22] which models both preference heterogeneity and scale heterogeneity . The data were first analysed by combining all presentation formats in a single model. A second model was then estimated which included terms for the presentation format interacted with each attribute level, to test for differences in preferences by presentation format. In addition, separate models

were estimated for each presentation format in order to examine choice consistency across the presentation formats, using the scale parameter tau of each G-MNL model [23]. This parameter represents the standard deviation of the person-specific scale of the error term and provides an indication of the extent to which individuals are more or less variable in their choices [22]. The analyses were conducted using Stata version 14; the G-MNL model estimation was by maximum simulated likelihood using 1,000 Halton draws, assuming uncorrelated errors [24]. The probability of choosing a hypothetical medical practice as predicted by the model was estimated by simulation with 10,000 draws using Stata's gmnlpred command [24].

3. Results

Response rates are difficult to estimate with online panel recruitment, because it is not known how many potential respondents saw the online invitation while the survey was live and declined to respond. Of the 1,522 people who started the survey, 1,208 respondents (79.4%) completed all questions, 253 (16.6%) were over the quota for their demographic group and 61 (4%) chose not to complete the survey. For each presentation style, the number of respondents completing nine choice questions ranged from 598 to 610 (see Figure 1).

The survey respondents were similar to the Australian population on most characteristics (see Table 3). The national statistics are for persons aged 15 or more while our respondents were aged 18 or more, consequently the sample includes slightly fewer respondents in the youngest age group. In addition, a higher proportion of respondents were born in Australia, were not participating in the paid workforce and had lower household income relative to the Australian population. A higher proportion of respondents reported having at least one on-going medical condition and a lower proportion reported being in very good or excellent health relative to the Australian population (Table 3). The mean subjective numeracy score was 4.44 (standard deviation (SD) 1.16) which is consistent with the mean levels reported in the patient samples used to develop and test the SNS-3 [21, 25], the median score was 4.67 (range 1-6).

3.1 Evaluation of presentation style

Overall, 10.4% of respondents chose the worse alternative in at least one of the two dominated choices seen. The worse alternative was chosen in 8.5% of choices when percentages (Style 4) were seen compared with 4.4% to 5.3% for the other styles. The probability of choosing the worse alternative when the choice was dominated (making an error) is presented in Table 4 (model 1). The

probability of choosing the worse alternative was significantly higher when the presentation format was Style 4 (percentage only). The odds of an error when seeing Style 4 were almost 1.7 times that when seeing Style 1 (p=0.035), 1.88 times that when seeing Style 2 (p=0.01) and 2.20 times that when seeing style 3 (p=0.003). Respondents who spent less than 3 minutes answering the 18 choice questions (21% of the sample) had a higher probability of choosing the worse alternative, relative to those spending longer, and the probability of choosing the worse alternative was lower for respondents with a high subjective numeracy score (see Table 4). Subjective numeracy was classified as high for respondents scoring at or above the median of 4.67 on a scale of 1 to 6. The probability of choosing the worse alternative did not differ significantly by the order of the choice (whether it was the first presentation style seen or the second). The odds of an error decreased with age and were higher among males (see Table 4), after accounting for the other variables in the model. The Hosmer-Lemeshow (H-L) test supports the model fit (Table 4).

The distribution of the time taken to complete the survey was highly skewed, with a long right tail. The median total time spent completing the survey was 11.0 minutes (interquartile range (IQR) 7.7-16.4). When considering only the time spent responding to the choice questions, the median time taken for the 18 choice questions was 5.6 minutes (IQR 3.3-8.7). Respondents spent 1.5 minutes longer on average completing the first nine choice questions relative to the second nine (median 0.9, IQR 0.2-1.8, Wilcoxon signed-rank test $p<0.001$). There were no statistically significant differences between presentation formats for the time taken to complete the first or second set of choices (Kruskal-Wallis test, p=0.99 and p=0.84 respectively). For the four presentation formats, the median time taken for the first nine choices ranged from 3.29-3.34 minutes, while for the second nine choices it ranged from 2.09-2.15 minutes.

In terms of consumer ranking, Style 1 (frequency and percentage with an icon array) was more frequently identified as the better presentation format for choosing between medical practices, selected as better by 64% of respondents who saw this style compared with 50% for Style 4 (percentage only), 45% for Style 2 (star rating) and 41% for Style 3 (star rating with benchmark). This did not differ significantly by subjective numeracy score for any of the presentation style combinations seen (combinations involving Style 1 chi-squared (2) = 4.79, p=0.09; Style 2 chi-squared (2) = 1.84, p=0.40; Style 3 chi-squared (2) = 4.98, p=0.08; Style 4 chi-squared (2) = 1.20, p=0.55). When comparing only the pairs of presentation styles seen by individuals, Style 1 remained the most favoured. Of the respondents seeing Styles 1 and 2, 62% thought Style 1 was better while 66% thought Style 1 was better among those seeing Styles 1 and 3. The differences were less clear

when Style 1 was not seen; of those seeing Styles 2 and 4, 49% thought Style 4 was better while 52% thought Style 4 was better among those seeing Styles 3 and 4.

The majority of respondents reported that they found the information easy to understand and this did not differ substantially by presentation style. On a 5-point scale where 1=extremely easy and 5=extremely difficult, the mean rating by presentation style ranged from 2.19 to 2.32. For the different presentation styles, the percentage of respondents rating the information as difficult to understand (4 or 5) ranged from 15% to 19% when the style was seen first and from 14% to 18% when it was the second style seen. The probability of rating the information as difficult to understand is presented in Table 4 (model 2). The odds of finding the choice information difficult to understand did not differ significantly by presentation style (p=0.95), numeracy, gender or the order in which it was seen. The odds of finding the choice information difficult to understand were significantly higher among respondents who completed the choice questions in less than 3 minutes relative to those who took longer, and significantly lower among older respondents relative to those aged less than 30 years (see Table 4). The H-L test supports the model fit (Table 4).

3.2 Preferences for a medical practice

Table 5 presents the G-MNL model for all data regardless of presentation style (Model 1) and the same model with interactions between presentation style and each attribute level mean effect (Model 2).The dominated choices were excluded from this analysis. In general, respondents preferred lower cost, shorter travel time and higher quality GP care, whilst practice ownership, opening hours and recommendations were less relevant. The presentation style had little impact on the preferences for a general practice. Although the main effects were statistically significant for all attribute levels, very few interaction terms were. Only three attribute levels by style interaction terms were statistically significant at the 5% level; a travel time of 30 minutes was less important when presentation Styles 3 or 4 were used and recommendation by someone you know was less important when Style 3 was used (although the different presentation styles were not used for these attributes). For the remaining attribute levels, none of the differences between presentation styles were significantly different from zero. Overall, the addition of the interaction terms did not improve the model fit as both Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were larger in Model 2 (with presentation style interactions) relative to Model 1 (no interactions), see Table 5. Both models showed evidence of both scale and preference heterogeneity, with tau significantly different from zero and a statistically significant standard deviation for the mean coefficient present for most attributes (including cost). Both models also included a statistically significant constant for Alternative B, although it has no interpretation beyond a tendency to favour

the right or left side on average regardless of the attribute levels. The measures of model fit indicated that the G-MNL model was a better fit than the simple conditional logit model for both Model s; AIC 22614 and BIC 22756 for the conditional logit model without interactions and AIC 22643 and BIC 23209 for the conditional logit with interactions, all measures are considerably larger than those for Models 1 and 2 in Table 5.

Separate G-MNL models for each presentation style were used to examine choice consistency (models available from the authors upon request). This found that the scale parameter tau was significantly different from zero for all presentation formats and there were some differences between the four formats which are illustrated in Figure 3. Style 4 showed the largest degree of scale variability (tau=1.99, 95% confidence interval 1.06 to 2.92) and Style 3 showed the lowest (tau=0.61, 95% confidence interval 0.23 to 0.99). The 95% confidence intervals for tau overlap for Styles 1, 2 and 4 but not for Style 3; respondents appeared less variable in their preferences when Style 3 (star rating with national benchmark) was used.

The impact of each attribute level on the probability of choosing a general practice was estimated from Model 1. Figure 4 presents the probability of choosing Alternative A at the given attribute level, when Alternative B and the other attributes are set to the base level (level 1 as per Tables 1 and 2). The red vertical line in the graph represents the probability of choosing alternative A with all attributes at the base level for both alternatives. The impact of a given attribute level can then be assessed in terms of the change from this base. The practice attributes with the largest impact on the choice of a practice across all presentation styles were the typical out-of-pocket cost for a visit and the quality attributes, particularly having GPs who listen and explain clearly. A typical out-of-pocket cost of $25 or $50 substantially reduced the probability of choosing a practice, relative to a practice where there is typically no out-of-pocket cost (Figure 4). There was a substantial increase in the probability of choosing a practice at higher levels of patient ratings (90%) of the practice GPs on listening and explaining, discussing treatment options and thorough examinations as well as overall satisfaction with the practice, relative to the lowest levels of these attributes.

4. Discussion

The results of this study suggest that people choosing a new medical practice would like to see information about service quality such as that collected in patient experience surveys. It also suggests that in order to be most useful for the general population, a graphical format such as icons or stars should be used to present the information. The most important attributes for choosing a medical practice were out-of-pocket costs for a visit and quality measures, particularly patient

ratings of the practice GPs with regard to communication and thorough examinations. The presentation format affected the error rate (choosing the worst alternative in a dominated choice), which was highest when the information was presented as a percentage without a graphical representation and not significantly different between the numeric (frequency and percentage) with icons and the star rating systems. The majority of respondents preferred the numeric with icons over either star rating system, although a substantial minority preferred stars. The star rating with a national benchmark appeared to result in the most consistent choices.

Our findings regarding the preferred attributes for choosing a medical practice were consistent with previous DCE studies which identified technical quality of care [26] and doctor communication [27-29] as important for choosing a GP appointment. These studies used textual information to represent the service quality attribute levels, unlike the current study which used different representations of quantitative levels. Of the practice attributes in our study which were presented using textual levels, practice ownership, after hours appointments and a recommendation from someone you know had minimal impact on the choice of a medical practice, while cost and travel time had a substantial impact, along with the five service quality attributes presented with the different formats. This provides a good indicator of the type of information which would be useful for consumer choice, if provided at the practice level.

Our findings also have implications for DCE methods. They suggest that the presentation of unfamiliar numerical information should use a graphical representation rather than just a numeric one, in order to minimise the likelihood of error. However, the differences identified between the presentation formats did not impact on preferences for a medical practice systematically, unlike the findings of Veldwijk et al [13] where substantial differences in preferences were identified between graphical and non-graphical presentation formats in a study of parental preferences for newborn rotavirus vaccination. These differences may relate either to the very different choice context, or the differences in the approaches to presenting the formats or both. Our study only used graphical representation for the service quality attributes which had numerical attribute levels and were thought likely to present challenges for respondents, rather than for all attributes as was done in Veldwijk et al [13] and Townsend and Kahn [30].

Although we found that the error rate did not differ between the star formats (Styles 2 and 3) and numeric with icons (Style 1), there were substantial numbers expressing a preference for each format. We expected respondents reporting lower subjective numeracy to prefer a star rating system over the numeric format seen, however this was not necessarily the case. Subjective numeracy was associated with making an error in choosing a dominated alternative (regardless of

format) but not with the best style identified after completion of the choice questions or with rating the format as difficult. One implication for the presentation of information for both consumers and for DCE methods is that using the format preferred by most respondents, will ignore the preferences of approximately one third of respondents. An alternative might be to allow individuals to select their preferred presentation format prior to accessing online practice information or prior to commencing the choice questions in an online survey, see for example [30]. However, this raises several additional issues. First, the most preferred format is not necessarily the most easily interpreted [4, 5] and in the current study percentages were equally preferred to star ratings but resulted in a higher error rate. Second, there will be many DCE contexts in which a star rating is not appropriate, such as when the numeric attribute level represents a risk. Third, even when either format might be appropriate, this raises the issue of whether or not the different representations address the same question. The lack of impact of the presentation format on preferences in our study would suggest that this might not be an issue. A further issue for consumer choice is the need to choose from a large number of providers, where a star rating system might make the highest quality provider easier to distinguish.

In the context of recent interest in Australia in primary care policies which include patient enrolment with a general medical practice, facilitating better informed choice of practice becomes more important. The findings from this research contribute to understanding the nature of the information which would be most useful to consumers choosing a medical practice for enrolment and how best to present the information. It also makes a methodological contribution, informing the presentation of attribute levels for DCE research.

5. Conclusion

The study found that health care performance information was important for choosing a new general practice and that the use of a graphic to present quantitative information resulted in fewer errors and more consistent choices. Quantitative health service quality information will be more useful to consumers if presented by combining the numerical information with a graphic, or using a star rating if appropriate for the context.

Compliance with ethical standards

The research was part of a program of research which was approved by the University of Technology Sydney Human Research Ethics Committee (Approval 2015000135). Participants were adult volunteers with an online panel and indicated consent by completing the online survey. Patricia Kenny has no potential conflicts of interest to declare. Stephen Goodall has no potential conflicts of interest to declare. Deborah Street has no potential conflicts of interest to declare. Jessica Greene has no potential conflicts of interest to declare.

Financial support

The research was funded by the Research in the Finance and Economics of Primary Health Care (ReFinE-PHC) Centre of Research Excellence under the Australian Primary Health Care Institute's Centres of Research Excellence funding scheme which was supported by a grant from the Australian Government's Department of Health. The paper does not necessarily reflect the views of APHCRI or the Australian Government.

Author contributions

Stephen Goodall and Patricia Kenny conceptualised and designed the study with input from Deborah Street and Jessica Greene. Deborah Street produced the design for the discrete choice experiment. Patricia Kenny analysed the data and drafted the manuscript. All authors edited the manuscript and approved the final version.

References

1. Australian Institute of Health and Welfare. Performance and Accounability Framework. Available from: http://www.aihw.gov.au/health-performance/performance-and-accountability-framework/. Accessed 26th October 2016.
2. Rechel B, McKee M, Haas M, Marchildon GP, Bousquet F, Blümel M, et al. Public reporting on quality, waiting times and patient experience in 11 high-income countries. Health Policy. 2016;120(4):377-83.
3. Victoor A, Delnoij DM, Friele RD, Rademakers JJ. Determinants of patient choice of healthcare providers: a scoping review. BMC Health Serv Res. 2012;12(272).
4. Hildon Z, Allwood D, Black N. Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review. Int J Qual Health Care. 2012;24(1):55-64.
5. Kurtzman ET, Greene J. Effective presentation of health care performance information for consumer decision making: A systematic review. Patient Educ Couns. 2016;99(1):36-43.
6. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PK, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med Inform Decis Mak. 2013;13(2):1-15.
7. West SL, Squiers LB, McCormack L, Southwell BG, Brouwer ES, Ashok M, et al. Communicating quantitative risks and benefits in promotional prescription drug labeling or print advertising. Pharmacoepidemiol Drug Saf. 2013;22(5):447-58.
8. Peters E, Hart PS, Tusler M, Fraenkel L. Numbers Matter to Informed Patient Choices: A Randomized Design across Age and Numeracy Levels. Med Decis Making. 2014;34(4):430-42.
9. Sinayev A, Peters E, Tusler M, Fraenkel L. Presenting Numeric Information with Percentages and Descriptive Risk Labels: A Randomized Trial. Med Decis Making. 2015;35(8):937-47.
10. Hibbard JH, Greene J, Daniel D. What Is Quality Anyway? Performance Reports That Clearly Communicate to Consumers the Meaning of Quality of Care. Med Care Res Rev. 2010;67(3):275-93.
11. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less Is More in Presenting Quality Information to Consumers. Med Care Res Rev. 2007;64(2):169-90.
12. Harrison M, Rigby D, Vass C, Flynn T, Louviere J, Payne K. Risk as an Attribute in Discrete Choice Experiments: A Systematic Review of the Literature. The Patient - Patient-Centered Outcomes Research. 2014;7(2):151-70.
13. Veldwijk J, Lambooij MS, van Til JA, Groothuis-Oudshoorn CGM, Smit HA, de Wit GA. Words or graphics to present a Discrete Choice Experiment: Does it matter? Patient Educ Couns. 2015;98(11):1376-84.
14. Primary Health Care Advisory Group. Better outcomes for people with chronic and complex health conditions. Final Report. Canberra: Commonwealth of Australia; 2015.
15. Kenny P, De Abreu Lourenco R, Wong CY, Haas M, Goodall S. Community preferences in general practice: important factors for choosing a general practitioner. Health Expect. 2016;19(1):26-38.
16. Australian Bureau of Statistics. Patient Experiences in Australia: Summary of Findings, 2013-14. Canberra: Australian Bureau of Statistics; 2014.
17. NHS England. GP Patient Survey - National summary report. London: Ipsos MORI Social Research Institute; 2015.
18. Street D, Burgess L. The construction of stated choice experiments. Hoboken NJ: John Wiley & Sons Inc; 2007.
19. Burgess L, Street DJ, Wasi N. Comparing Designs for Choice Experiments: A Case Study. Journal of Statistical Theory and Practice. 2011;5(1):25-46.
20. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. The Patient - Patient-Centered Outcomes Research. 2015;8(5):373-84.

21. McNaughton CD, Cavanaugh KL, Kripalani S, Rothman RL, Wallston KA. Validation of a Short, 3-Item Version of the Subjective Numeracy Scale. Med Decis Making. 2015.
22. Fiebig D, Keane M, Louviere J, J., Wasi N. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Market Sci. 2010;29:393-421.
23. Fiebig DG, Viney R, Haas M, Knox S, Street D, Weisberg E, et al. Complexity and doctor choices when discussing contraceptives. Working Paper 15/14. York: Health, Econometrics and Data Group, University of York; 2015 September 2015.
24. Gu Y, Hole AR, Knox S. Fitting the generalized multinomial logit model in Stata. Stata Journal. 2013;13(2):382-97.
25. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. Med Decis Making. 2007;27(5):672-80.
26. Cheraghi-Sohi S, Hole AR, Mead N, McDonald R, Whalley D, Bower P, et al. What Patients Want From Primary Care Consultations: A Discrete Choice Experiment to Identify Patients' Priorities. Ann Fam Med. 2008;6(2):107-15.
27. Longo MF, Cohen DR, Hood K, Edwards A, Robling M, Elwyn G, et al. Involving patients in primary care consultations: assessing preferences using discrete choice experiments. Br J Gen Pract. 2006;56(522):35-42.
28. Philips H, Mahr D, Remmen R, Weverbergh M, De Graeve D, Van Royen P. Predicting the place of out-of-hours care--a market simulation based on discrete choice analysis. Health Policy. 2012;106(3):284-90.
29. Scott A, Watson MS, Ross S. Eliciting preferences of the community for out of hours care provided by general practitioners: a stated preference discrete choice experiment. Soc Sci Med. 2003;56:803-14.
30. Townsend C, Kahn BE. The “Visual Preference Heuristic”: The Influence of Visual versus Verbal Depiction on Assortment Processing, Perceived Variety, and Choice Overload. Journal of Consumer Research. 2014;40(5):993-1015.

Table 1: Attributes and levels

Attribute text	Dictionary text shown by hovering on attribute name	Short Name	Level code	Level text
The GPs listen and explain clearly	Whether the GPs from this practice listen to patients and explain the diagnosis and treatment clearly as reported in patient surveys (percentage who said this or average rating on this).	Listen	1 2 3	50%[a] 70% 90%
The GPs involve patients in discussions	Whether the GPs from this practice involve patients in discussions about their treatment as reported in patient surveys (percentage who said this or average rating on this).	Discuss	1 2 3	50%[a] 70% 90%
The GPs conduct a thorough examination when necessary	Whether the GPs from this practice conduct a thorough examination when necessary as reported in patient surveys (percentage who said this or average rating on this).	Exam	1 2 3	50%[a] 70% 90%
Patients are satisfied with the practice overall	Whether patients were satisfied with the care they received at this practice as reported in patient surveys (percentage who said this or average rating on this).	Overall	1 2 3	50%[a] 70% 90%
Availability of urgent appointments on the same day	Whether people usually get to see a GP at this practice on the same day when they request an urgent appointment (percentage seeing GP on same day or rating on this). In Australia, 75% of people requesting an urgent appointment see the GP on the same day.	Day	1 2 3	50%[a] 70% 90%
The practice is owned by	Medical practices can be owned by either the GPs who work there or by a company which employs the GPs. The companies which own practices can be either charitable (not-for-profit) organisations or commercial (for-profit) organisations with shareholders.	Owner	1 2 3	A private company A not-for-profit organisation The practice GPs
Appointments offered evenings and Saturdays	The availability of GPs from the practice for appointments outside of standard office hours (evenings and Saturday morning).	Hours	1 2	No Yes
Travel time to the practice	The time it will take you to travel to the practice.	Travel	1 2 3	10 minutes 20 minutes 30 minutes
Someone you know recommended the practice GP/s	Someone you know (such as a family member, friend or colleague) recommended one or more of the GPs at this practice.	Recom	1 2	No Yes
The typical cost for a visit (after refunds) is	This is the amount you pay minus the Medicare refund (eg if the refund is $36 and the GP charges $76, then the cost to you is $40). Many patients are bulk-billed and do not pay. When patients are charged, the average cost is about $30 (after refund).	Cost	1 2 3	$0 $25 $50

a. See Table 2 for the presentation format

Table 2: Presentation formats used for practice quality attributes

Level	Style 1	Style 2	Style 3	Style 4
1	 5 in 10 (50%)		 Below National average	50%
2	 7 in 10 (70%)		 Average	70%
3	 9 in 10 (90%)		 Above National average	90%

Table 3: Health and socio-demographic characteristics of sample

	Sample % (n=1,208)	Australia[a] %
Age: 18-29	22.2	25.6
30-39	17.6	17.2
40-49	17.6	16.8
50-59	16.5	15.7
60-69	13.2	12.5
70 or more	13.0	12.3
Female	51.1	50.6
Aboriginal or Torres Strait Islander	2.6	2.3
Australian born	75.1	68.8
Speak English at home	92.3	N/A
Married/have partner	56.4	60.6
Educational qualifications:		
No post- high school	32.5	38.8
Post-high school non-degree	33.8	36.6
University degree or higher	33.7	24.6
Median household income per week	$960-1,149	$1,548
Employment status:		
Employed fulltime	37.9	42.8
Employed part-time	16.6	21.4
Unemployed	6.7	4.2
Retired	21.2	19.2
Other (not in labour force)	17.6	12.5
Residential area: Major cities	74.6	71.9
Inner regional	17.9	18.1
Outer regional/Rural/remote	7.5	10.0
Health in general: Excellent	8.9	20.3
Very good	32.4	37.0
Good	36.8	27.0
Fair	17.8	11.2
Poor	4.1	4.5
On-going medical condition	64.1	56.5
Usually: Attend same practice	90.8	N/A
See same GP	83.6	75.3
Private Health Insurance	53.7	55.0

a National statistics are for persons aged 15 or more. Source: Australian Bureau of Statistics (ABS), Australian demographic statistics June 2014; General social survey: Australia 2014; Patient Experiences in Australia 2014-15; Household income and wealth Australia 2013-14.

Table 4: Logit models for evaluating presentation styles: choosing the worst alternative (Model 1) and finding the choice information difficult[a] to understand (Model 2)

	Model 1 Choosing worst alternative				Model 2 Difficult to understand			
	Odds Ratio	95% CI		p	Odds Ratio	95% CI		p
Style 2	0.891	0.519	1.530	0.677	0.931	0.714	1.214	0.597
Style 3	0.763	0.463	1.256	0.287	0.976	0.745	1.278	0.860
Style 4	1.678	1.038	2.713	0.035	0.980	0.720	1.336	0.900
Choice time[b]: <3 min	4.282	2.756	6.655	<0.001	2.328	1.701	3.188	<0.001
High numeracy[c]	0.573	0.377	0.870	0.009	0.905	0.685	1.196	0.484
Second style seen	1.113	0.784	1.581	0.550	0.988	0.848	1.150	0.874
Age: 30 to 44	0.620	0.404	0.949	0.028	0.661	0.463	0.943	0.022
45 to 59	0.297	0.161	0.548	<0.001	0.495	0.336	0.730	<0.001
60 to 74	0.256	0.123	0.534	<0.001	0.470	0.304	0.728	0.001
75 or more	0.441	0.139	1.406	0.166	0.502	0.223	1.129	0.096
Male	1.602	1.040	2.469	0.032	1.107	0.835	1.468	0.481
Constant	0.055	0.032	0.095	<0.001	0.253	0.176	0.365	<0.001
Hosmer-Lemeshow $\chi^2_{(8)}$	6.00				9.90			
p	0.65				0.27			

a. Rated 4 or 5 where 1=extremely easy & 5=extremely difficult. b. Time spent on all 18 choice questions. c. Above sample median numeracy score of 4.66 on scale 1 to 6.

Robust standard errors were used to account for clustering by individual (2416 observations for 1208 respondent).

Model 1: Wald chi-squared test for 3 style coefficients p=0.01; Wald chi-squared test for 4 age coefficients p<0.001.

Model 2: Wald chi-squared test for 3 style coefficients p=0.95; Wald chi-squared test for 4 age coefficients p=0.002.

Table 5: Generalised multinomial models for choosing a medical practice with and without presentation style interactions

Attribute	Level[a]	Model 1				Model 2									
										Model 2 Interaction terms					
		Mean	p	SD	p	Mean	p	SD	p	Style 2	p	Style 3	p	Style 4	p
GPs listen and explain clearly	2	0.842	<0.001	0.011	0.911	0.817	<0.001	0.041	0.656	0.169	0.209	-0.008	0.952	0.124	0.348
	3	1.464	<0.001	0.767	<0.001	1.540	<0.001	0.784	<0.001	0.010	0.945	-0.092	0.503	0.151	0.310
GPs involve patients in discussions	2	0.467	<0.001	0.034	0.696	0.465	<0.001	0.017	0.862	0.194	0.151	-0.034	0.796	-0.106	0.419
	3	0.929	<0.001	0.364	<0.001	0.897	<0.001	0.333	<0.001	0.147	0.294	0.041	0.764	0.082	0.564
Thorough examinations	2	0.629	<0.001	0.096	0.242	0.700	<0.001	0.120	0.199	0.021	0.878	-0.103	0.417	-0.047	0.718
	3	1.223	<0.001	0.840	<0.001	1.285	<0.001	0.909	<0.001	-0.033	0.812	0.090	0.506	0.008	0.957
Patients satisfied with practice	2	0.540	<0.001	0.001	0.992	0.621	<0.001	0.209	0.033	-0.085	0.516	-0.059	0.655	-0.014	0.913
	3	1.002	<0.001	0.624	<0.001	1.140	<0.001	0.680	<0.001	-0.045	0.749	-0.078	0.561	-0.130	0.371
Same day urgent appointments	2	0.315	<0.001	0.126	0.157	0.332	0.001	0.188	0.032	0.093	0.498	-0.036	0.786	-0.090	0.496
	3	0.653	<0.001	0.321	0.001	0.801	<0.001	0.305	0.039	-0.093	0.491	-0.110	0.410	-0.262	0.057
Practice is owned by	Not-for-profit	0.325	<0.001	0.248	0.070	0.405	<0.001	0.281	0.004	0.002	0.985	-0.204	0.100	-0.022	0.862
	Practice GPs	0.247	<0.001	0.642	<0.001	0.290	0.004	0.705	<0.001	0.043	0.749	-0.135	0.303	0.113	0.417
Appointments evenings/Saturdays	Yes	0.176	<0.001	0.496	<0.001	0.170	0.011	0.489	<0.001	0.029	0.744	0.048	0.579	-0.037	0.686
Travel time to practice - minutes	20	-0.480	<0.001	0.000	0.999	-0.623	<0.001	0.070	0.490	0.187	0.149	0.165	0.209	0.068	0.593
	30	-0.933	<0.001	0.836	<0.001	-1.216	<0.001	0.873	<0.001	0.170	0.199	0.349	0.011	0.289	0.036
Recommended to you	Yes	0.138	<0.001	0.142	0.151	0.221	0.003	0.084	0.533	-0.014	0.897	-0.253	0.021	-0.060	0.571
Typical cost for a visit	$	-0.068	<0.001	0.062	<0.001	-0.072	<0.001	0.068	<0.001	-0.005	0.295	0.004	0.369	-0.002	0.767
Alternative B constant		0.152	<0.001	0.631	<0.001	0.159	0.002	0.629	<0.001	0.178	0.010	0.033	0.627	-0.237	0.001
tau		0.827	<0.001			0.885	<0.001								
gamma		0.090	0.107			0.046	0.342								
Log likelihood		-10542				-10500									
AIC		21159				21184									
BIC		21458				21908									

a. Levels 2 and 3 according to presentation style as per Table 2, level 1 is always the lowest rating and level 3 is always the highest rating.
AIC=Akaike's information criterion; BIC=Bayesian information criterion.

Figure 1: Numbers of respondents completing each pair of presentation styles and number completing 8 choice sets of each style

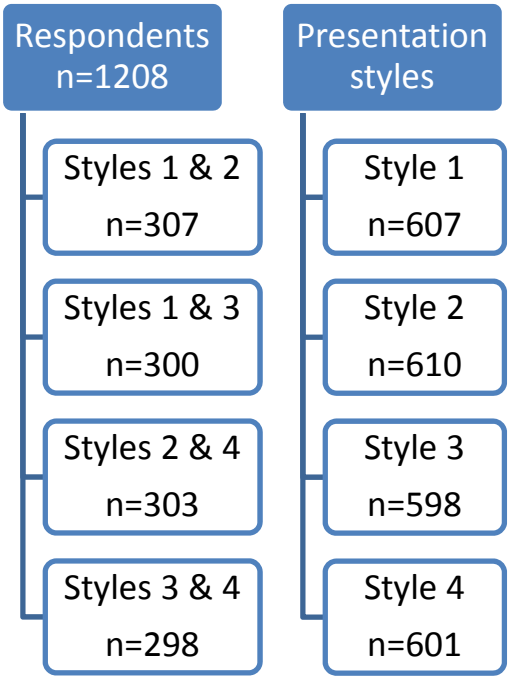
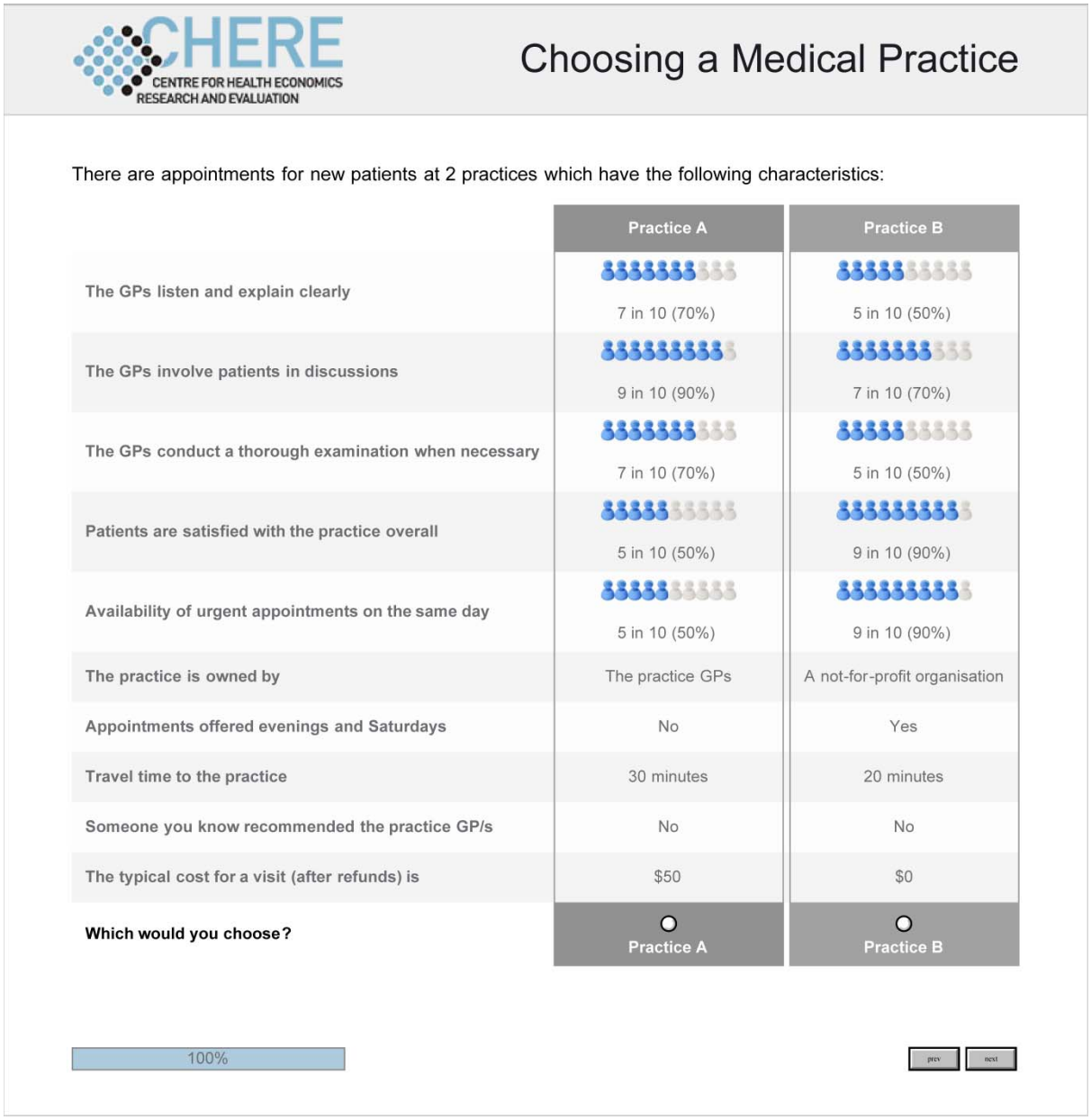


Figure 2: Sample choice for presentation style 1



Note: additional explanatory information was available by placing the cursor on the attribute text.

Figure 3: Scale parameter (tau) and 95% confidence intervals for the four presentation styles

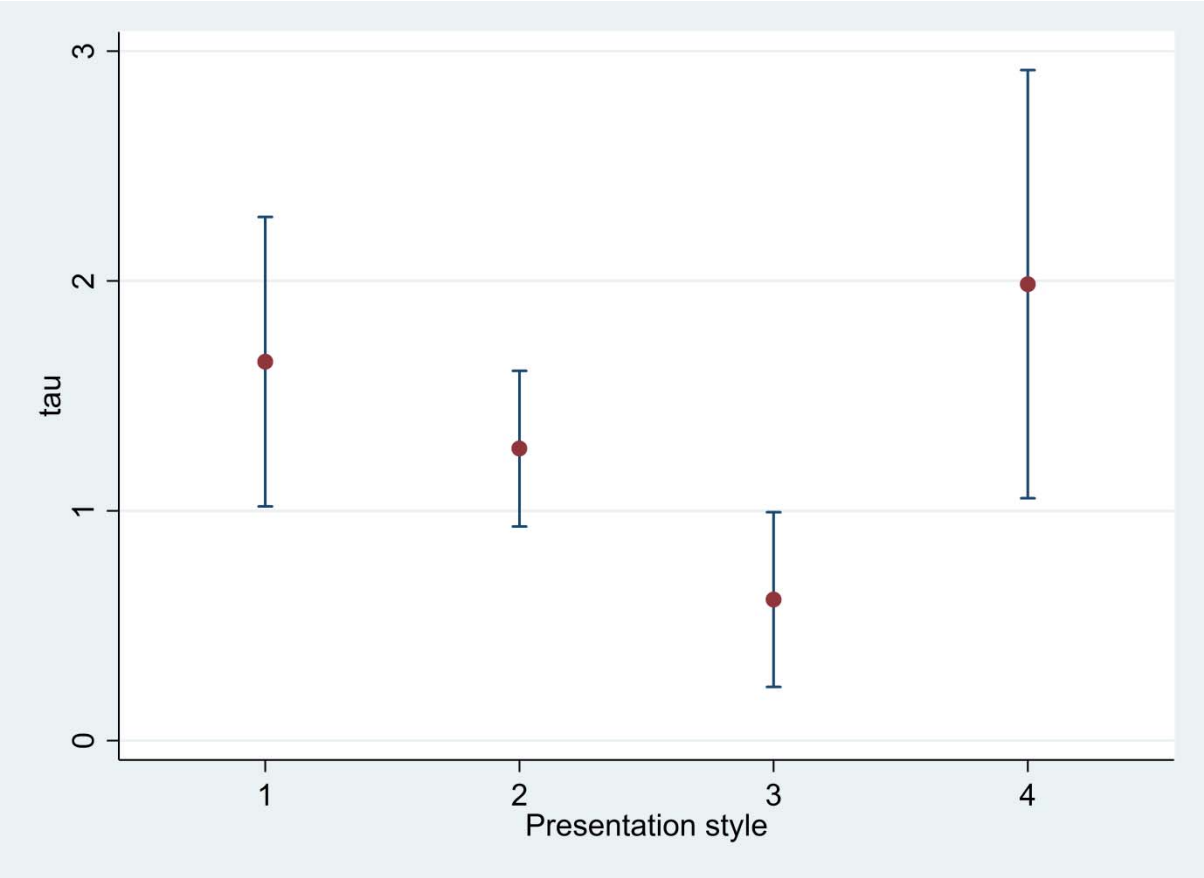
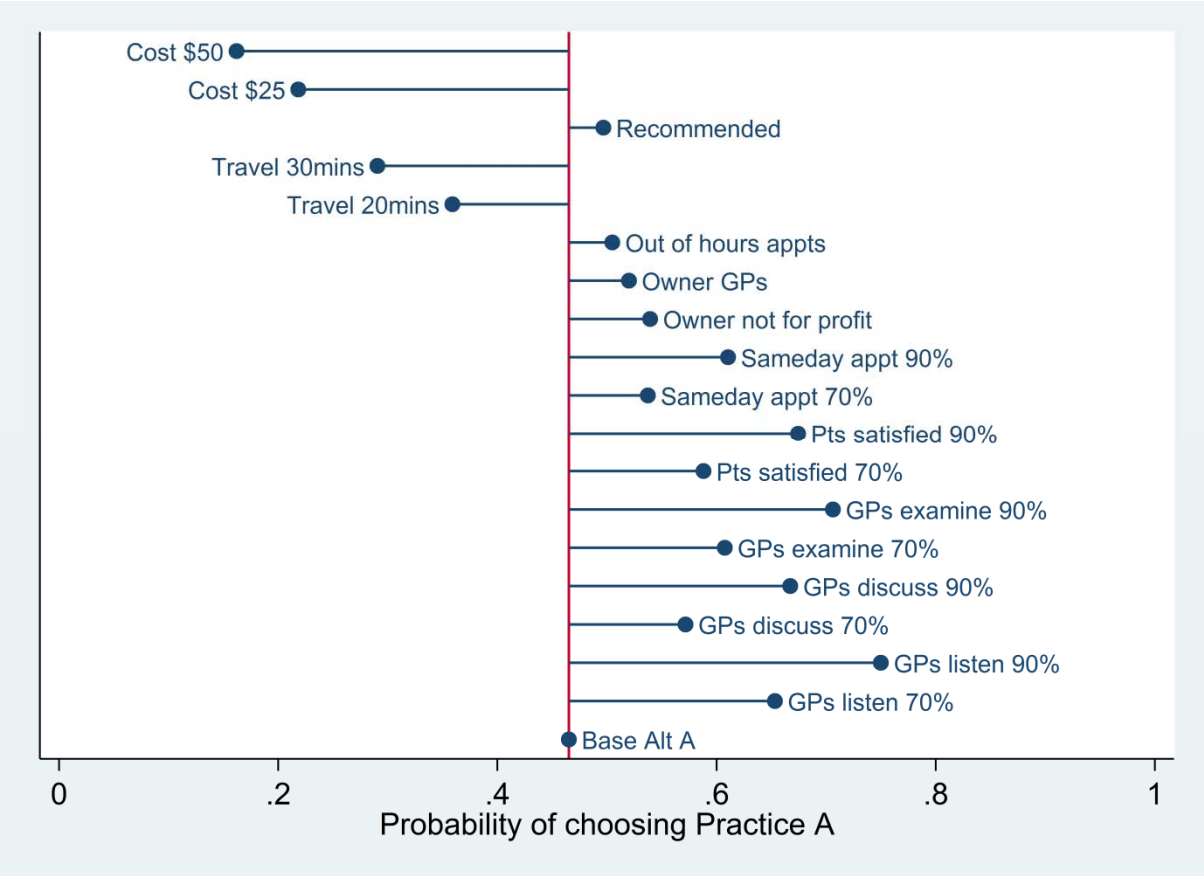


Figure 4: Predicted probability of choosing a general practice at each attribute level, when other attributes are set to the base level



The red line at 0.47 represents the probability of choosing alternative A when all attributes are set to the base level for both alternatives. Base levels: Cost=$0; Travel time=10 minutes; Owner=Private company; Recommended by someone you know=No; Appointments evenings or Saturdays=No; Five quality attributes=50% or 2.5 stars. Quality: 1.GPs listen and explain, 2.GPs involve patients in discussions, 3.GPs conduct thorough examinations, 4.patients satisfied with practice, 5.urgent appointments on same day.