Framing effects of risk communication in health-related decision making. Learning from a discrete choice experiment

Florence Nguyen, Marie-Odile Carrère, Nora Moumjid

GATE
Groupe d'Analyse et de Théorie Économique
UMR 5824 du CNRS

DOCUMENTS DE TRAVAIL - WORKING PAPERS

W.P. 09-21

Framing effects of risk communication in health-related decision making. Learning from a discrete choice experiment

Florence Nguyen, Marie-Odile Carrère, Nora Moumjid

Novembre 2009

GATE Groupe d'Analyse et de Théorie Économique
UMR 5824 du CNRS
93 chemin des Mouilles – 69130 Écully – France
B.P. 167 – 69131 Écully Cedex
Tél. +33 (0)4 72 86 60 60 – Fax +33 (0)4 72 86 60 90
Messagerie électronique gate@gate.cnrs.fr
Serveur Web : www.gate.cnrs.fr

Titre de l'article

Framing effects of risk communication in health-related decision making. Learning from a discrete choice experiment

Nom des auteurs:

Florence Nguyen [1], Marie-Odile Carrère [1], Nora Moumjid [1]

[1]GATE (Groupe d'Analyse et de Théorie Economique), UMR 5824 – CNRS ; Université Lyon1 ; Université Lyon2 ; ENS LSH ; Centre Léon Bérard, Lyon, France

Adresse complète de l'auteur

Florence NGUYEN

GRESAC- Centre Léon Bérard

28 rue Laënnec

69008 Lyon France

Email: florence_thi_nguyen@hotmail.com

Fax: (00 33) 4 78 78 28 04

Tel: (00 33) 4 78 78 27 80

Abstract

Background How to communicate uncertainty is a major concern in medicine and in health economics. We aimed at studying the framing effects of risk communication on stated preferences in a discrete choice experiment (DCE) performed to elicit women's preferences for Hormone Replacement Therapy. **Methods** Two versions of the questionnaire were randomly administered to respondents. Multiple risks were expressed as natural frequencies using either a constant reference class (Design 1) or variable reference classes (Design 2). We first tested whether Design 1 would impose a lower cognitive burden than Design 2. We then examined whether the two designs resulted in different utility model estimates. **Results** Design 1 improved consistency (monotonicity and stability). However, rates of dominance or intransitive responses did not differ across designs. Design 1 decreased women's sensitivity to the risk of fractures and increased their sensitivity to the risk of breast cancer as compared to all other attributes. **Discussion** Framing effects of risk communication on stated preferences may be a major problem in the design of DCEs. More research is needed to determine whether our findings are replicable and to further investigate the normative question of how to improve risk communication in health-related decision-making.

Keywords: Framing effects, Risk communication, Discrete choice experiment

Classification JEL: C12; I19; D83

1. Introduction

Thirty years ago, Tversky and Kahnemann stated that the psychological principles that govern the perception of decision problems produce predictable shifts of preference when the same problem is framed in different ways (Tversky and Kahnemann, 1981). They also specified that "the frame that a decision maker adopts is controlled partly by the formulation of the problem and partly by the norms, habits and personal characteristics of the decision maker". Thus, variations in these factors may result in different decisions on the same problem, a phenomenon which is referred to as the framing effect.
Since then, framing effects have drawn considerable attention in various academic fields such as psychology, economics, sociology and neurosciences, and have been tested in quite different areas of human decision making such as management, finance, environment, law and medicine. A considerable amount of theoretical and empirical literature has explored the impact of framing effects on decision under uncertainty, knowing that uncertainty is a major component of most decision problems. Framing effects, either extrinsic or intrinsic to individuals, challenge their ability to make accurate decisions when they are faced with a decision problem under uncertainty. Considering the formulation of the problem, framing effects raise the essential normative question of how to communicate uncertainty so that a well informed decision can be made.

In decision-making processes involving peoples' health, the framing effects of communication of uncertainty are particularly important to consider (Politi *et al.,* 2007). Indeed, different formulations of the same uncertainty, which are logically equivalent, can lead to different decisions regarding life or death. Depending first on the available evidence, which can be very poor (British Medical Journal Clinical Evidence, 2007), uncertainty may be communicated in a number of different ways. Using verbal methods leads to highly variable interpretation (Mazur and Hickam, 1991; Bogardus *et al.*, 1999; Mazur and Merz, 1994). Quantitative information on risk (probabilities or likelihood) using for example absolute risks, relative risks or frequencies, allows for more precision. However, the framing effects of communicating relative risks compared to absolute risks have been shown to affect decisions since using relative risks emphasizes the perception of both benefits and risks (Edwards *et al.,* 2002; Hux and Naylor, 1995; Forrow *et al.,* 1992; Bucher *et al.,* 1994).
Regarding numeracy, *i.e.* the ability to understand quantitative information, Gigerenzer and Hoffrage (1995) have pointed out that it could be improved if the information format was adapted to the cognitive process. They have demonstrated that human Bayesian inference is better performed when risks are presented as natural frequencies rather than as standard probabilities (Gigerenzer and Hoffrage, 1995). Indeed, the proportion of correct responses strongly increases when natural frequencies are used instead of standard probabilities. Several experiments in medicine and law have confirmed these findings and the authors conclude that "statistics expressed as natural frequencies improve the statistical thinking of experts and non experts alike" (Hoffrage *et al.,* 2000).

It is worth noting that these marked advances in risk communication guidance were based on experiments where respondents' answers could be compared with a right answer. However, in most choice situations, there is by definition no right or wrong answer, since choice is supposed to result

from the individual's preferences. In that case, there is a lack of evidence about the best way to communicate risk information, which is all the more true when information is complex. For example, in many cases, multiple risks and benefits must be considered simultaneously. Risk information can then exceed the cognitive capacity of decision makers who may, in turn, use heuristics and other simplifying strategies (Tversky and Kahneman, 1974) or behave according to ambiguity aversion (Einhorn and Hogarth, 1986). All these strategies may result in biased decisions.

If the question of how to communicate uncertainty is a major concern in medicine, it is also of major importance in health economics, where eliciting the preferences of economic agents who make health-related decisions under uncertainty is a key issue. However, the impact of uncertainty communication on stated preferences has received little attention. Regarding the contingent valuation method, we have identified only one published study in which willingness to pay for a longevity benefit has been found sensitive to the framing of the benefit, either as a gain in life expectancy or a reduction in annual mortality risk (Morris and Hammitt, 2001). Overall, the life expectancy format performs better than the risk reduction format regarding the validity of contingent valuation.
Considering the discrete choice experiment (DCE) method, we have also identified only one very recent study testing the effects of different frames of information about risk in colorectal cancer screening (Howard and Salkeld, 2008). Risk attributes were framed either as gains or losses (e.g. cancers found versus cancers missed). The authors conclude that framing of attributes significantly influences the estimation of willingness to pay and marginal rates of substitution between attributes. However, no guidance on risk communication is given. Instead, the authors emphasize that we are at a very early stage of examining framing effects in health-related DCE and firmly encourage further research on this issue.

We aimed at studying the framing effects of risk communication in a health-related DCE. We intended not only to determine whether framing effects are at work, but also to test the impacts of different risk presentations on preference properties, in order to provide some guidance for risk communication in health-related decision making.

2. Methods

The DCE was carried out to analyse women's preferences for Hormone Replacement Therapy (HRT) - a treatment used by 20 to 30% French post-menopausal women (Allemand *et al.*, 2008). HRT alleviates menopausal symptoms such as climacteric troubles, and decreases the risk of fractures due to osteoporosis. However, it is also associated with negative effects, such as an increased risk of developing breast cancer. Besides, HRT may also entail a monetary cost for users, since some treatments are covered by the public health insurance system and others are not. Those characteristics may lead women to make trade-offs and influence their attitude towards treatment.

2.1 Choice experiment design

In order to establish which characteristics of HRT are important to women, we first searched the medical literature (WHI (Rossouw *et al.,* 2002), MWS (Beral *et al.,* 2003) and E3N (Fournier *et al.,* 2005) studies), then we organized three meetings with a panel of medical experts (three gynaecologists, 1 rheumatologist and 1 cardiologist). Finally, we organized a focus group interview with post-menopausal women, of whom some had already taken HRT and some had not.
Seven attributes were selected, six related to health and one to cost. In order to reduce respondents' cognitive burden, we decided to assign only two levels to each attribute (Table I). For each of the 6 health risks, the two levels corresponded to the situation of post-menopausal French women with or without treatment. More precisely, the levels of risk of climacteric troubles were derived from reports by Grady (2003) and by the French Association for the Study of Menopause (AFEM). The levels of other health risks were derived from a report by the French Agency for the Safety of Health Products (AFSSAPS, 2004). The two cost levels were chosen as either 0, corresponding to no treatment, or the mean total cost of treatment in France whoever the payer is.

Table I

Hypothetical scenarios were created by combining attribute levels. There were 128 possible scenarios (2^7). We developed a level balanced, orthogonal main effect fractional factorial design of 8 scenarios (Zwerina *et al.,* 1996). Applying the rules of minimal overlap and utility balance, we created eight choice sets by pairing each initial scenario with its exact opposite, *i.e.* the one which differed in all attribute levels.

In order to test for a framing effect of risk communication, two different versions of the survey instrument were developed, then randomly administered to respondents. In the two versions, the 6 health risks were expressed using natural frequencies, according to previously mentioned guidance (Hoffrage *et al.,* 2000). However, the two versions differed in the choice of reference classes. In the first version (referred to as « Design 1 »), natural frequencies were expressed using a constant reference class. We chose a 100 000 people reference class because (i) two of the 6 health risks involved cancer (breast and colorectal) and (ii) cancer incidence is usually expressed per 100 000 population (International Agency for Research on Cancer, 2008). The second version (referred to as « Design 2 ») slightly differed from the first one: risks were also expressed in natural frequencies but the reference class could vary from one attribute to another. Reference population sizes were still expressed as powers of ten but as small as possible, so that numbers of cases were still expressed as whole numbers. As a consequence, natural frequencies referred to 10, 1 000 or 10 000 people.
An example of a choice set as presented with both DCE designs is given in Table II.

Table II

2.2 Survey

The survey was conducted in December, 2006 and January, 2007. Participants randomly selected from the list of telephone subscribers in the Lyon area, France, were recruited by phone. Women who accepted to participate were sent a questionnaire and a prepaid return envelope, and were asked to return their questionnaire once it had been completed.

Of the 669 women who were sent a questionnaire, 470 returned it (70% response rate). After exclusion of 8 respondents with large amounts of missing data, 462 subjects were finally used in the analysis.

Subjects randomly received one of the two versions of the questionnaire: 229 completed the constant reference class survey instrument (Design 1) and 233 completed the variable reference class survey instrument (Design 2). Considering menopausal status, marital status, education, household income, professional or extraprofessional activities and complementary health insurance, Chi-square analyses did not detect any significant differences between respondents of the two groups.

Each respondent was presented the eight choice sets of the two scenarios and was asked which scenario she preferred in each choice set.

2.3 Model estimation

The statistical analysis in our DCE consisted in estimating a preference model by comparing two utility functions. Individual utility function is not directly observable and has to be estimated indirectly by analysing the respondents' choices.

McFadden's Random Utility Theory (McFadden, 1974) states that the utility of individual *i* for scenario *r* consists of an observable systematic component (V_{ir}) and an unobservable random component (ε_{ir}).

$$U_{ir} = V_{ir} + \varepsilon_{ir}, \quad (1)$$

The observable utility depends on the utility associated with each attribute. Assuming a linear and additive utility function, we obtain:

$$V_{ir} = \beta_i X_r \quad (2)$$

where X_r is a vector of the levels of *m* HRT attributes in scenario *r*. Thus:

$$V_{ir} = \sum_m \left(\beta_{im} X_{rm}\right) \quad (3)$$

Due to the random component, the utility function is probabilistic. The statistical model is based on the probability that individual *i* chooses scenario *A* rather than scenario *B* if:

P_{Ai}= Prob ($U_{Ai} > U_{Bi}$) (4)

Leading to $P_{Ai}(V_{Ai} - V_{Bi} \succ \varepsilon_{Bi} - \varepsilon_{Ai})$ (5)

Assuming a normal distribution of the random component, the calculation of the probability that individual *i* will choose scenario *A* is based on a *probit* specification:

$$P_{Ai} = \int_{-\infty}^{V_{Ai}-V_{Bi}} f(\varepsilon_{Ai}) d\varepsilon_{Ai} = \Phi\left(\frac{V_{Ai} - V_{Bi}}{\sigma}\right) \quad (6)$$

where *f*(.) represents the density function and Φ(.) the standardised normal distribution. The scale parameter, σ, is usually set to 1. Nevertheless, when observations are obtained from two different survey instruments as was the case here, possible differences in scale parameters between the two data sets have to be taken into account (*cf.* infra).

As each woman was presented with eight choice sets, a random effects *probit* model was applied in order to account for the individual-specific variation, α_i:

$$U_{ir} = V_{ir} + \varepsilon_{ir} + \alpha_i \quad (7)$$

This error term eliminates the correlation between ε_{ir} and explanatory variables, and improves the efficiency of the estimates.

2.4 Hypotheses and methods

We wanted to determine whether two slightly different ways of communicating risk information through DCE could result in differences in preference properties on the one hand and in preference structure on the other hand. Thus, we used a two-step approach as done by Maddala *et al.* when comparing DCE designs which differed in overlap of attribute levels (Maddala *et al*., 2003).
First, we tested the overall hypothesis that the constant reference class survey instrument (Design 1) would impose a lower cognitive burden on respondents than the variable reference class survey instrument (Design 2). This hypothesis was suggested by Gigerenzer and Edwards who stated that "[…] the ultimate source of confusion is the reference class […]", specifying that "conditional probabilities such as sensitivity and specificity refer to different classes […], which makes their mental combination difficult" (Gigerenzer and Edwards, 2003). We first compared response consistency (*i.e.* monotonicity and stability), dominance and transitivity across the two survey groups. Second, we examined whether the two different risk presentations resulted in significantly different utility model estimates, which would confirm the hypothesis that risk communication yields framing effects.
Our hypotheses and the methods used to test them are summarized in Table III and detailed hereafter.

Table III

As said before, the first group of hypotheses relates to whether the constant reference class survey instrument would impose a lower cognitive burden than the variable reference class survey instrument.

Hypothesis 1a: The constant reference class survey instrument results in more consistency

Besides the eight choice sets created to elicit women's preferences, another choice set was added in which scenario A dominated scenario B with better values on 3 attributes (climacteric risk, breast cancer risk, cost) without any difference on the other 4 attributes. Monotonicity was violated when women chose scenario B rather than scenario A.

In order to test stability, one of the eight initial choice sets was repeated at the end of the questionnaire. Stability was violated when women did not give the same answer to the two same questions.

Hypothesis 1b: The constant reference class survey instrument results in fewer dominant responses

Assuming that questionnaire complexity could be one of the reasons for dominant preferences (Scott 2002), we tested for differences in this property between the two designs. Respondents were asked whether they had taken into account one or several HRT attributes when making their choices. We considered responses as dominant when women who claimed they had taken only one attribute into account had actually always chosen the scenario in which this attribute reached its best level.

Hypothesis 1c: The constant reference class survey instrument results in fewer intransitive responses

As was the case for dominance, we assumed that questionnaire complexity could yield intransitive responses. Transitivity was assessed by creating a choice set based on one of the eight initial choice sets, with scenario A remaining the same but the new scenario B being dominated by the initial one. Thus, women choosing scenario A in the initial choice set were again expected to choose scenario A in the new choice set.

The second group of hypotheses relates to whether the two different risk presentations resulted in significantly different utility model estimates. We decided to test these hypotheses among women whom responses did not violate consistency, non-dominance and transitivity. Indeed, one may reasonably assume that inconsistent responses may result from limitation in cognitive capacities, then that they may not reflect respondents' preferences. The question is more complex regarding dominance and intransitivity, which can reflect either limitation in cognitive capacities or real preferences. However, we excluded the corresponding answers because, even when they do reflect preferences, they cannot be modelled using a utility function since this requires both non dominance and transitivity.

Hypothesis 2a: The overall vector of choice parameters differs across the two survey instruments

Two separate random parameter probit models were estimated for the two survey instruments. Using a likelihood ratio (LR) test we tested the hypothesis of equal overall vectors of parameters in the two designs, using either constant or variable reference classes. However, pooling the two datasets was problematical as the estimated parameters were confounded with the corresponding scale parameters (σ in Eq.6). That is why we adjusted for a possible difference in scales between the datasets using the Swait and Louvière procedure (Swait and Louvière, 1993). It consisted in setting

the scale of the dataset in Design 1, σ_1, to 1 and searching for the scale of the dataset in Design 2, σ_2, which would maximise the log-likelihood function of the scaled-pooled model.

A first LR test compared the two separate models and the scaled-pooled model to test the hypothesis of equal overall vectors of parameters in the two datasets. Rejection of this hypothesis would indicate significant differences in preferences between the two survey designs. In case of non rejection of the hypothesis, a further likelihood ratio between the scaled-pooled and the simple pooled model would be required to test the hypothesis of equality between the scale parameters σ_1 and σ_2, in the two datasets.

Hypothesis 2b: Some choice parameters vary depending on the survey instrument, while others do not

First, estimated parameters could have statistical significance in one of the two separate models and not in the other. Second, in case of difference in the scale parameters, significant coefficients of the two estimated models cannot be compared directly. So we compared marginal rates of substitution (MRS). Since MRS are expressed as the ratios of two coefficients (when significantly different from zero) they do not depend on scale parameters. MRS variances were estimated using the Delta method (Greene, 2003). Possible MRS differences between the two survey designs were tested using the Student *t*-test.

3. Results

Statistically significant differences in consistency were observed between the two survey groups (Table IV). More precisely, rates of non-monotonic and unstable preferences were both higher in the variable reference class survey group (Design 2) than in the constant one (Design 1). Proportions were respectively 6.9% (*n*=16) versus 2.1% (*n*=5) for non-monotonicity (p=0.0157, Chi-sq test), and 14.6% (*n*=34) versus 9.1% (*n*= 21) for instability (p=0.0720, Chi-sq test).

We observed no statistically significant difference in dominance between the two survey groups (Table IV): 10% (*n*= 24) in Design 1 and 12% (*n*=27) in Design 2 (p=0.700, Chi-sq test).

Table IV

Transitivity was studied in the 219 women who had chosen scenario A in the initial choice set. No statistical difference was observed between the two survey groups, since the same number of women made intransitive choices (1.4%, n=3).

As described above, women's preferences were then modelled among the 338 women who did not exhibit inconsistent, dominant or intransitive preferences, of whom 176 had completed the Design 1 questionnaire and 153 the Design 2 questionnaire.

Table V presents the results of separate *probit* models for each of the two experimental designs, *i.e.* the estimated choice parameters in the utility function. Of note, risks had been previously recoded so that they were expressed in the same units in the two data groups. Most of the parameters were significant, except for the risk of fractures in Design 1. As expected, all significant parameters were negative, since attributes were expressed as either risks or costs.

Table V

By setting σ_1 to 1 and maximizing the log-likelihood function of the scaled pooled model, we estimated the scale parameter σ_2 to be 0.95.
Using a likelihood-ratio test, we compared each separate model with a scaled pooling of the two datasets (Table VI). The Swait-Louviere (Swait and Louviere, 1993) likelihood ratio test statistic was LR= 2[-3809.77-(-1995.02 + -1816.23)] = 2.96. Knowing that the critical value of the chi-square distribution is 14.06 at the 0.05 significance level on 11 degrees of freedom, the hypothesis that the overall vectors of attribute parameters were equal across the two data sets was not rejected.

To test whether the scale parameters were significantly different across data sets we performed a further likelihood ratio test comparing the scaled pooled and the simple pooled models (Table VI). In the simple pooled model, the relative scale parameter was no longer allowed to differ across data

sets. The likelihood-ratio statistic for this test was LR= -2[-3811.25-(-3809.77)] = 2.96. Knowing that the critical value at the 0.05 significance level is 2.70, this result approaches statistical significance. However, the hypothesis that the scale parameters do not differ across the two datasets could not be rejected.

Table VI

Whilst the above results indicate no significant difference in the overall vectors of choice parameters across designs, some choice parameters or MRS did vary, while others did not. First, the risk of fractures had no statistically significant weight in Design 1 whereas it had in Design 2 (Table V). Then, MRS estimates were derived from the two separate models for all combinations of 2 of the 7 attributes, except when the risk of fractures was involved in Design 1 (Table VII). As a result, 15 MRS, could be compared across the two models referring to the 6 remaining attributes.

Five MRS were found significantly different. Hence, preference structures resulting from the two designs were different and these differences were not induced by possible scaling effects. All 5 MRS involved the risk of breast cancer. It seems that Design 1 made women even more sensitive to this risk than to the other attributes, either other health risks or cost.

To achieve a given reduction in the risk of breast cancer, women were willing to accept an increase in all other health risks, and this increase was higher when preferences were elicited using Design 1 rather than Design 2: 1.63 higher for climacteric troubles (10.063/6.149), 2.66 for colorectal cancer (0.568/0.213), 1.81 for thromboembolism (0.886/0.488) and 2.06 for cardiac risk (1.857/0.901).

As far as cost was concerned, women were willing to pay 1.72 times more for a given breast cancer risk reduction when presented with Design 1 rather than Design 2 (0.398/0.231).

While risk communication impacted women's preferences through the risk of fractures according to the significance of choice parameter and through the risk of breast cancer according to the related MRS, it did not seem to have any influence on MRS between the other 5 attributes, *i.e.* risks of climacteric troubles, colorectal cancer, thromboembolism or cardiovascular disease, and cost.

Table VII

4. Discussion

Our study aimed at analysing the possible framing effects of risk communication in a health-related DCE. Respondents were randomly assigned to one of two questionnaires which only differed in the way how information on risk was provided.

We first hypothesized that expressing multiple risks using a constant reference class rather than variable reference classes would impose a lower cognitive burden on respondents. This hypothesis was partly confirmed. An important finding was that preference consistency significantly improved

when risks were presented using a constant reference class. Indeed, the monotonicity test exhibited 2.1% of non-monotonic preferences when the constant reference class design was used, a proportion which increased to 6.9% with the other design. Moreover, according to the stability test, the constant reference class risk presentation generated 9.1% of unstable preferences, whereas this rate increased to 14.6% when variable reference classes were used.

The pattern of intransitive preferences was not statistically different across survey designs. Rates were very low (1.4%), which can be due to the transitivity test used which compared two choice sets only. Considering the test of dominant preferences, the corresponding rates were about 10% of respondents and they did not differ across designs. As stated by Scott (2002), dominant preferences can have various explanations such as the complexity of the choice task and the use of heuristics in the choice process. But they can also reflect real strong preferences. It was beyond the scope of this study to analyze the reasons for dominant preferences. However, dominant preferences deserve specific analysis given that, as they result in non compensatory decision making, they should not be modeled using a multiattribute utility function.

Considering our hypothesis that the two survey instruments would result in different utility model estimates, it is worth noting that this hypothesis was not tested on the whole sample. Indeed, women whose responses violated consistency (*i.e*. monotonicity and stability), non-dominance and transitivity had to be excluded from the utility function modelling. First, one may reasonably assume that inconsistent responses may result from limitation in cognitive capacities, then that they may not accurately reflect respondents' preferences. Second, regarding dominance and intransitivity, and whatever the explanations given, corresponding preferences cannot be modelled using a utility function since this would require both non dominance and transitivity.

Our findings confirm that preferences structure depends on the presentation of risk. More precisely, the attribute level utility weights resulting from the two risk presentations did not exhibit significantly different scale parameters at the 0.05 level, which means that the amount of unexplained response variability was the same across the two designs. However, while the overall vectors of choice parameters were not significantly different, significant differences in some choice parameters and/or MRS were observed across designs. Two major changes occurred when risks were presented using a constant reference class instead of variable reference classes. On the one hand, women were no more sensitive to the risk of fractures. On the other hand, women's sensitivity to breast cancer risk increased. More precisely, all MRS involving breast cancer risk changed towards a significantly greater sensitivity to this risk. The front page of the questionnaire presented all possible effects of HRT, either positive such as a decreased risk of fractures, or negative such as an increased risk of breast cancer. According to Fischoff *et al.* (1978), judgments of risk and benefit are negatively correlated, *i.e.* the greater the perceived benefit, the lower the perceived risk and vice versa. As highlighted by Slovic *et al.* (2003), "this negative relationship […] occurs even when the nature of the gains or benefits from an activity is distinct, and qualitatively different, from the nature of the risks". Our own findings suggest that, since the constant reference class design improved response consistency,

the preferences elicited accordingly better conformed to the expected negative relationship. Indeed, the perception of an increase in breast cancer risk, a negative effect of HRT, increased while the perception of a decrease in the risk of fractures, a positive effect of HRT, decreased or even disappeared.

Let us now consider the increased sensitivity to breast cancer risk in more details. In their study of people's perception of the frequency of lethal events, Lichtenstein *et al.* (1978) have shown that cancer, due to sensationalization, is one of the most overestimated causes of death, as opposed to "undramatic, quiet killers" such as stroke and heart disease. Indeed, breast cancer has been greatly sensationalized in France over the past three years, due to a national information campaign for breast cancer screening. As a consequence, since the constant reference class design improves response consistency, preferences elicited accordingly better conform to the expected strong sensitivity to breast cancer risk.

Considering MRS between health risks, women would accept a higher increase in all other health risks in return for a given reduction in breast cancer risk when preferences are elicited using constant reference than when using the other design. While this is obviously true for the risk of fractures to which women are no more sensitive, this is also the case for all other health risks, for which the corresponding MRS are multiplied by a factor ranging from 1.63 to 2.66.

As far as cost is concerned, women would be willing to pay 1.72 times more for a given breast cancer risk reduction when preferences are elicited using the constant reference class risk presentation rather than the other design. Importantly, this finding means that a slight change in risk communication may yield a major change in willingness to pay.

Our results demonstrate that the framing effects of risk communication on stated preferences may be a major problem in the design of DCEs. As mentioned, we have identified only one recent report of framing effects of risk communication in a DCE on colorectal cancer screening (Howard and Salkeld, 2008). Risk attributes were presented either as gains or losses but no guidance regarding the best way to proceed was provided. However, other framing effects have been demonstrated in DCE studies. Various levels of overlap of attributes have been compared (Maddala *et al.,* 2003) and a possible ordering effect of the price attribute has been tested (Kjaer *et al*., 2006).

Analyzing framing effects is not an easy task. But answering the essential normative question of how to overcome framing problems is even more difficult. In the presence of multiple risks, as was the case in our study, our findings are in favor of a presentation using constant reference parameters, given that the stated preferences are more consistent, thus confirming the suggestion made by others regarding risk communication in medicine (Gigerenzer and Edwards, 2003; International Patient Decision Aid Standards Collaboration, 2008).

However, more research is needed to determine whether these findings are replicable and, more generally, to investigate the normative question of how to improve risk communication in health-related decision-making.

Acknowledgements

The authors gratefully acknowledge financial support from the Fondation de France. We thank Marie-Dominique Reynaud for editorial assistance.

References

Allemand H, Seradour B, Weil A, et al. 2008. Baisse de l'incidence des cancers du sein en 2005 et 2006 en France : un phénomène paradoxal. *Bulletin du Cancer* 95(1) :11-15.

Beral V, Million Women Study Collaborators. 2003. Breast cancer and hormone-replacement therapy in the million women study. *Lancet* 362: 419-427.

British Medical Journal Clinical Evidence. http://www.clinicalevidence.com/ceweb/about/guide.jsp [14 March 2007].

Bogardus ST, Holmboe E, Jekel JF. 1999. Perils, pitfalls, and possibilities in talking about medical risk. *The Journal of the American Medical Association* 281:1037–41.

Bucher HC, Weinbacher M, Gyr K. 1994. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *British Medical Journal* 309(6957):761–4.

Edwards A, Elwyn G, Mulley A. 2002. Explaining risks: turning numerical data into meaningful pictures. *British Medical Journal* 324:827–30.

Einhorn HJ, Hogarth RM. 1986. Decision making under ambiguity. *Journal of Business* 59(4):S225–50.

Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B. 1978. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences* 8: 127-152.

Forrow L, Taylor WC, Arnold RM. 1992. Absolutely relative: how research results are summarized can affect treatment decisions. *The American Journal of Medicine* 92:121–4.

Fournier A, Berrino F, Riboli E, *et al*. 2005. Breast cancer risk in relation to different types of hormone replacement therapy in the E3N-EPIC cohort. *International Journal of Cancer* 114: 448-454.

French Agency for the Safety of Health Products (AFSSAPS). 2004. *Traitements hormonaux substitutifs de la ménopause*. Rapport d'orientation.

French Association for the study of Menopause (AFEM). http://www.menopauseafem.com [1 January 2005]

Gigerenzer G, Edwards A. 2003. Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal*, 327: 741-4.

Gigerenzer G, Hoffrage U. 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102 (4): 684-704.

Grady D. 2003. Postmenopausal hormones-therapy for symptoms only. *The New England Journal of Medicine* 348 (19):1835-7.

Greene W. 2003. *Econometric Analysis*. Prentice Hall, New Jersey.

Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. 2000. Medicine. Communicating statistical information. *Science* 290(5500):2261-2.

Howard K, Salkeld G. 2008. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. *Value in Health* 12 (2).

Hux JE, Naylor CD. 1995. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Medical Decision Making* 15:152–7.

International Agency for Research on Cancer. http://www.iarc.fr [1 November 2008].

International Patient Decision Aid Standards Collaboration *IPDAS Collaboration Background Document*. O'Connor A, Llewellyn-Thomas H, Stacey D. editors 2005. http://www.ipdas.ohri.ca/IPDAS_background.pdf. [1 February 2008].

Kjaer T, Bech M, Gyrd-Hansen D, Hart-Hansen K. 2006. Ordering effect and price sensitivity in discrete choice experiments: need we worry? *Health Economics* 15(11):1217-28.

Lichtenstein S, Slovic P, Fischhoff B, Layman M, Combs B. 1978. Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory* 4: 551–578.

Maddala T, Phillips KA, Johnson FR. 2003. An experiment on simplifying conjoint analysis designs for measuring preferences. *Health Economics*. 12 (12):1035-47.

Mazur DJ, Hickam DH. 1991. Patients' interpretations of probability terms. *The Journal of General Internal Medicine* 6(3):237–40.

Mazur DJ, Merz JF. 1994. Patients' interpretations of verbal expressions of probability: implications for securing informed consent to medical interventions. *Behavioral Sciences & The Law* 12(4):417–26.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behaviour, In *Frontiers of Econometrics.* P. Zarembka ed. London, U.K: Academic Press, 105-142.

Morris JM, Hammitt JK. 2001. Using Life Expectancy to Communicate Benefits of Health-Care Programs in Contingent-Valuation Studies. *Medical Decision Making* 21(6):468-478.

Politi MC, Han PKJ, Col NF. 2007. Communicating the uncertainty of harms and benefits of medical interventions. *Medical decision making* 27(5):681-95.

Rossouw JE, Anderson GL, Prentice RL, *et al.* 2002. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association* 288(3): 321-368.

Scott A. 2002. Identifying and analysing dominant preferences in discrete choice experiments: An application in health care. *Journal of Economic Psychology* 23(3): 383-398.

Slovic P, Finucane M, Peters E, McGregor DG. 2003. The affect heuristic. In T. Gilovich, D. Griffin, D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment*. New York : Cambridge University Press.

Swait J, Louviere J. 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* 30:305-314.

Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185:1124–31.

Tversky A, Kahneman D. 1981. The framing of decisions and the psychology of choice. *Science 211*: 453-458.

Zwerina K, Huber J, Kuhfield W. 1996. *A general method for constructing efficient choice designs*. Durham: Duke University.

Tables

Table I Attributes and levels used in the study

Attributes	Levels
Risk of climacteric troubles [a]	20 000, 70 000
Risk of fractures [b]	400, 600
Risk of colorectal cancer [b]	40, 60
Risk of breast cancer [b]	250, 350
Risk of thromboembolism [b]	150, 350
Risk of cardiovascular disease [b]	150, 200
Cost [c]	0, 250

[a] per 100 000

[b] per 100 000 per year

[c] € per year

Table II Example of a choice set as presented with both DCE designs

Design 1 [a]	Scenario A	Scenario B
Risk of climacteric troubles	70 000 per 100 000	20 000 per 100 000
Risk of fractures	600 per100 000 per year	400 per 100 000 per year
Risk of colorectal cancer	60 per 100 000 per year	40 per 100 000 per year
Risk of breast cancer	250 per 100 000 per year	350 per 100 000 per year
Risk of thromboembolism	150 per 100 000 per year	350 per 100 000 per year
Risk of cardiovascular disease	200 per 100 000 per year	150 per 100 000 per year
Cost	0€ per year	250€ per year

Design 2 [a]	Scenario A	Scenario B
Risk of climacteric troubles	7 per 10	2 per 10
Risk of fractures	6 per 1 000 per year	4 per 1 000 per year
Risk of colorectal cancer	6 per 10 000 per year	4 per 10 000 per year
Risk of breast cancer	25 per 10 000 per year	35 per 10 000 per year
Risk of thromboembolism	15 per 10 000 per year	35 per 10 000 per year
Risk of cardiovascular disease	20 per 10 000 per year	15 per 10 000 per year
Cost	0 € per year	250€ per year

[a] Risks are presented in natural frequencies, using a constant reference class in Design 1 and variable reference classes in Design 2.

Table III. Hypotheses and associated methods

Hypotheses	**Methods used**
1. The constant reference class survey instrument (Design 1) imposes a lower cognitive burden on respondents than the variable reference class survey instrument (Design 2)	
(a) Design 1 results in more consistency	(a) Chi-sq test of the proportion of monotonicity and stability in each group
(b) Design 1 results in fewer dominant responses	(b) Chi-sq test of the proportion of dominant responses in each group
(c) Design 1 results in fewer intransitive responses	(c) Student test of the proportion of intransitive responses in each group
2. The two survey instruments result in different utility model estimates	
(a) The overall vector of choice parameters differs across the two survey instruments	(a) Estimation of 2 separate models for the 2 survey instruments, and LR tests of separate and pooled, models (simple and scaled)
(b) Some choice parameters vary depending on the survey instrument, while others do not	(b) Comparison of choice parameters and marginal rates of substitution between attributes in the two survey instrument groups

Table IV. Tests for consistency and dominance differences between the two survey groups

	Hypothesis 1(a)				Hypothesis 1(b)	
	Monotonicity		Stability		Dominance	
Risk presentation	Incorrect	Correct	Incorrect	Correct	Yes	No
Design 1 [a] *n*=229	0.02	0.98	0.09	0.91	0.10	0.90
Design 2 [a] *n*=233	0.07	0.93	0.15	0.85	0.12	0.88
Chi²	5.84	(*p*=0.0157)	3.24	(*p*=0.0720)	0.14	(*p*=0.700)

[a] Risks are presented in natural frequencies, using a constant reference class in Design 1 and variable reference classes in Design 2.

Table V Probit estimates for each separate model

	Design 1 [a]		Design 2 [a]	
Variable	Coefficient	P > | t |	Coefficient	P > | t |
Intercept	-0.180728	0.021	-0.127621	0.125
Risk of climacteric troubles	-0.001133	0.000	-0.001193	0.000
Risk of fractures	0.000040	0.923	-0.000976	0.037
Risk of colorectal cancer	-0.020075	0.000	-0.034503	0.000
Risk of breast cancer	-0.011396	0.000	-0.007334	0.000
Risk of thromboembolism	-0.005559	0.000	-0.006497	0.000
Risk of cardiovascular disease	-0.010266	0.000	-0.013577	0.000
Cost	-0.002635	0.000	-0.002917	0.000
Observations	1406		1280	
Respondents	176		162	
Log likelihood	-1995.02		-1816.23	

[a] Risks are presented in natural frequencies, using a constant reference class in Design 1 and variable reference classes in Design 2.

Table VI Probit estimates for the pooled data sets

	Scaled pooled model		Simple pooled model	
Variable	Coefficient	P > | t |	Coefficient	P > | t |
Intercept	-0.163227	0.004	-0.166958	0.003
Risk of climacteric troubles	-0.000011	0.000	-0.000012	0.000
Risk of fractures	-0.000508	0.086	-0.000513	0.091
Risk of colorectal cancer	-0.025723	0.000	-0.026130	0.000
Risk of breast cancer	-0.009235	0.000	-0.009535	0.000
Risk of thromboembolism	-0.005757	0.000	-0.005895	0.000
Risk of cardiovascular disease	-0.011464	0.000	-0.011722	0.000
Cost	-0.002635	0.000	-0.002702	0.000
Observations	2686		2686	
Respondents	338		338	
Log likelihood	-3809.77		-3811.25	

Table VII Differences in marginal rates of substitution (MRS) [a] between the two designs [b]

[a] MRS of attribute x_1 compared to attribute x_2 is defined as the ratio of absolute compensating level variations with x_1 as numerator and x_2 as denominator.

[b] Risks are presented in natural frequencies, using a constant reference class in Design 1 and variable reference classes in Design 2.

[c] MRS involving the risk of fractures are not presented since the corresponding parameter is not significantly different from zero in Design 1.

MRS	Design 1 [b]	Design 2 [b]	Prob (MRS_1-MRS_2>0)
Troubles, fractures	--	0.818	--
Troubles, colorectal cancer	17.726	28.931	0.117
Troubles, breast cancer	**10.063**	**6.149**	**0.017**
Troubles, thromboembolism	4.909	5.448	0.302
Troubles, cardiac risk	9.065	11.384	0.138
Troubles, cost	2.327	2.446	0.419
Fractures, colorectal cancer	-- [c]	35.362	--
Fractures, breast cancer	-- [c]	7.516	--
Fractures, thromboembolism	-- [c]	6.659	--
Fractures, cardiac risk	-- [c]	13.915	--
Fractures, cost	-- [c]	2.990	--
Colorectal cancer, breast cancer	**0.568**	**0.213**	**0.016**
Colorectal cancer, thromboembolism	0.277	0.188	0.147
Colorectal cancer, cardiac risk	0.511	0.394	0.267
Colorectal cancer, cost	0.131	0.085	0.172
Breast cancer, thromboembolism	**0.488**	**0.886**	**0.001**
Breast cancer, cardiac risk	**0.901**	**1.857**	**0.001**
Breast cancer, cost	**0.231**	**0.398**	**0.028**
Thromboembolism, cardiac risk	1.847	2.090	0.219
Thromboembolism, cost	0.474	0.449	0.405
Cardiac risk, cost	0.257	0.215	0.200