University of Alberta

Valuation of Health Risk Reductions from Municipal Drinking Water Treatment

by

Jing Zhang ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Agricultural and Resource Economics

Department of Rural Economy

Edmonton, Alberta
Spring 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-45639-2
Our file Notre référence
ISBN: 978-0-494-45639-2

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

This thesis investigates Canadians' preferences for different municipal water treatment technologies that differ in their effectiveness in reducing microbial risk versus cancer risk based on their responses to a series of hypothetical choice questions. The thesis consists of three papers and each deals with one important empirical issue in the valuation of health risk reductions from municipal drinking water, and they are: uncertainty in willingness-to-pay (WTP) estimates due to the uncertainty in model selection; valuation of public risk reductions when altruism is present, and effects of choice format on preference elicitation in a stated choice survey.

Results of the three papers indicate that Canadians prefer a water treatment technology that reduces both microbial risks and cancer risks from their drinking water, although effectiveness in reducing microbial mortality risk is more important. In the first paper, we show how to derive model weighted WTP estimates using a model averaging approach in a random utility framework. It is found that among a variety of estimated models, capturing unobserved heterogeneity in preferences improves model fit the most. Our results suggest that it is important to capture the way heterogeneity enters a model (preferences versus scale) in model estimation, and to control the way decision complexity affects preferences or scale through experimental design. In the second paper, we distinguish an individual's WTP by motivations based on actual self-protection expenditure data and provide our value of statistical life (VSL) estimates in a public good provision context. Our results confirm that individuals are willing to pay for other

people's health risk reductions. We report different VSL estimates conditional on the assumptions about the nature of altruism. In the third paper, based on context-variable augmented random utility models, we reconcile preference differences inferred from two different choice formats. It appears that choice format affects preference elicitation, but the effect can be controlled and predicted. The paper shows how to derive preferences averaged over two different choice formats, which is one step closer toward deriving context-free preferences using stated choice surveys.

ACKNOWLEDGEMENTS

I know I could not have finished my thesis alone. I owe many thanks to the people who have helped me along the way.

The first person who I must thank is my supervisor: Wiktor (Vic) Adamowicz. For five years, Vic has provided me tremendous guidance, support, encouragement and inspiration. He is an academic supervisor, a mentor and a friend. His exemplary professionalism has influenced me profoundly. I would also like to thank Ellen Goddard. She has been an important source of encouragement and has provided many challenges and opportunities. Thank you, Ellen.

Many other faculty members in the department have helped me during my studies. I would like to explicitly thank Sean B. Cash, Peter Boxall, Chokri Dridi, Bodo Steiner, Scott Jeffrey and Mel Lerohl. Also, to all other faculty members who have either taught or helped me in one way or another, I am very grateful! After eating lunch in the department lounge for five years, they have all become my friends. The Department's administrative staff members are the ones who make a place pleasant to study. My thanks go to them: Robin McClelland, Dawn Zrobok, Sarah Fatemi, Mary Rivard and Randy Page.

I would like to express my deep appreciation to Mel McMillan and Fredrik Carlsson, my external committee members, for their insightful comments. I am also grateful to Diane Dupont and Alan Krupnick, who are co-investigators of the project that this thesis originated from, and who have provided me with ongoing comments and suggestions for my thesis research.

I came to Edmonton five years ago completely alone. I've found a home here. My friends in the Department of Rural Economy have become like family to me. To enumerate the names of all the people with whom I have shared friendship would take far deep appreciation to you all, dear friends. One name I have to mention, however, is Robert Boswell, for his love and support.

My final thanks go to my family, father Decheng Zhang, mother Jian Wang and younger brother Hang Zhang. Despite being separated by the Pacific Ocean, they know every detail of my life in Edmonton through my weekly phone "report". It has been my favorite routine of every week.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Background

Life cannot exist without water. A person consumes approximately 2.4 litres of water each day (Anonymous 2006) and every government strives to ensure the safety of drinking water. In Canada, about ninety percent of the population receives their tap water through public water systems (Environmental Canada 2004). Over the past few years, there has been increasing concern about the safety of drinking water, especially after the outbreaks of waterborne diseases in Walkerton, Ontario and North Battleford, Saskatchewan in 2000 and 2001 respectively. In the case of Walkerton, about 2300 residents became seriously ill and seven died from exposure to microbially contaminated drinking water (Hrudey et al. 2003). A study reported that the health impact (human suffering) from the Walkerton tragedy amounted to 91 million Canadian dollars (CTV 2001). In response to the tragic events, drinking water has been identified as a public health issue, and the federal government, provincial governments and territorial governments have started to review their regulations, policies and programs related to drinking water (Federal/Provincial/Territorial Advisory Committee on Population Health and Health Security (F/P/T ACPHHS) 2003). Some governments have made their regulations more stringent (F/P/T ACPHHS 2003).

In a recent summary of the *Guidelines for Canadian Drinking Water Quality* (Federal/Provincial/Territorial Committee on Drinking Water 2004), the identified harmful substances include both pathogens (microbes such as E. coli, cryptosporidium, giardia, etc.) and potentially carcinogenic chemical disinfection by-products (DBPs), such as Total Trihalomethanes (TTHMs). TTHMs are formed when chlorine, the most effective disinfectant for destroying nearly all microorganisms, reacts with other chemicals present in the water. A survey conducted by the U.S. Environmental Protection Agency (USEPA) discovered that "TTHMs are present in virtually all chlorinated water supplies" (Capece 2003, p. 1). Several studies on humans have found a link between long-term exposure to high levels of chlorination by-products and a higher risk of cancer (e.g., King 1995). In Canada, according to a national survey of chlorinated DBPs in drinking water conducted in 1995, the TTHMs levels in the majority of treatment facilities were relatively low (<50 μg/L[1]) and a small number of facilities had relatively high TTHMs values (>100 μg/L) (Health Canada 1995). Health risks from pathogenic microorganisms far exceed those potential health risks associated with chemical DBPs. It is therefore suggested that "the solution to any problems with high concentrations of DBPs is not to reduce disinfection since this would pose an unacceptable health risk" (Health Canada 1995, p. 10). It has to be noted that these TTHMs levels were from samples measured at the treatment plant rather than at the consumption tap. It is possible that health risks imposed by TTHMs from drinking water are higher since a certain level of chlorine has to be kept at the distribution system to maintain the effectiveness of disinfection. Health Canada recently suggested a maximum acceptable concentration for TTHMs in drinking water of 0.1 milligrams per litre (or 100 μg/L) (F/T/P CDW 2004). In contrast, the health risks imposed by TTHMs from drinking water appear to be more serious in the United States. The Environmental Working Group (EWG) in the U.S. analyzed TTHMs tests reported by 28,082 public water suppliers in 41 states and showed that between 1998 and 2003, 170 million people in 14,685 communities drank water contaminated with TTHMs (Environmental Working Group 2006). In 6,975 of these communities, tap water was contaminated at levels above health-based thresholds. Consequently, one of the USEPA's current priorities for regulation development is to balance the risks from microbial pathogens and DBPs (U.S. Environmental Protection Agency 2006). According to the USEPA's the *Stage 1 Disinfectants/Disinfection Byproducts Rule*, the new maximum allowable annual average level of TTHMs is set to be 80 μg/L for large surface water public water systems as well as for small surface water and all ground water systems, replacing the old standard of 100 μg/L (U.S. Environmental Protection Agency 1998).

In Canada, due to the widespread dread of microbially contaminated drinking water, the health risks imposed by TTHMs may be overlooked. As a matter of fact, the current criterion in choosing treatment technology is aimed at "minimizing the microbial-related health risk without any compromise" (Health Canada 1995, p. 10). However, with increasing awareness of health risks from DBPs, some concerns about cancer risks from

---

[1] μg/L is the abbreviation for microgram/litre. 1 microgram (μg) = 0.001 milligram (mg).

TTHMs in drinking water have been raised as well. In response to these concerns, F/P/T CDW prepared a document: *Trihalomethanes in Drinking Water* for public comment (F/P/T CDW 2004).[2]

Ideally, a preferred water treatment technology would be one that reduces both types of health risk: microbial risk from pathogens and cancer risk from DBPs such as TTHMs. Unfortunately, the few available alternative disinfection methods that produce fewer carcinogenic DBPs are not only generally more expensive, but also may not be as effective as chlorine-based methods at reducing microbial contaminants. With increasing awareness of cancer risk from TTHMs, it is possible that there is a tradeoff between reducing TTHMs and reducing microbial contaminants. Thus, information on values of risk reductions in a risk-risk tradeoff context is needed. These values can be used to inform choices of technologies for treating drinking water at the plant level and may also be used to help evaluate policy options at the provincial or federal level (Adamowicz, Dupont and Krupnick 2005).

There is a sizable literature on the valuation of health risk reductions. Depending on whether the risk is fatal or not, values of risk reductions are categorized into value of risks to life (mortality risks) and value of risks to health (morbidity risks) (Viscusi 1993). Compared to the value of morbidity risks, the value of mortality risks is more heavily studied and reported since there is a clear-cut definition associated with mortality risks: one death is counted as one death regardless of causes. The value of mortality risks is therefore widely used to inform risk management decisions (Viscusi 1993; Adamowicz 2004). The policy implication of the value of morbidity risks are usually much smaller, in part due to the fact that the human sufferings differ substantially across causes and types of injuries or diseases, which makes it difficult to define and standardize the value and then subsequently transfer it to other studies.[3]

A commonly used measure of the value of mortality risk reductions is the value of statistical life (VSL). Previous VSL estimates have a large variation. According to the *Canadian Handbook on Health Impact Assessment*, VSL estimates from a list of studies

[2] The public comment was closed on January 2005.

[3] Although "the loss of a day of work" is one popular measurement (Viscusi 1993, p. 1934), it ignores human suffering that does not involve days away from work, and it also precludes human suffering outside of the workforce.

ranged from $1.0 to $22.6 million (2000 Canadian dollars), with a median value of $6.8 million, a mean value of $8.4 million and a standard deviation of $5.8 million (Health Canada 2004). Out of 25 studies referenced, twenty were derived from wage-risk models, i.e., a hedonic wage model, and the other five were based on contingent valuation models. The variation in these values mainly comes from studies using a wage-risk model. The five CVM studies had a mean value of $4.42 million (a median of $4.2 million) with a standard deviation of $1.78 million.

While these values provide us with a good reference for health impact assessment, it has to be acknowledged that these values are somewhat outdated. All studies were conducted before 1993. It is likely that the value of risk reductions has increased over the last decade as per capita income increases (Costa and Kahn 2003). It is also likely that there have been some changes in the public preferences for water related risk reductions, especially after the Walkerton incident.

Moreover, most of previous studies examine VSL in a private good context. That is, most VSLs assess the value that an individual places on reducing his or her personal risk level rather than the value that an individual places on reducing risk levels that the public, including him or her, face. It is clear that the VSL in our context should be treated as a public good. A few studies have shown that individuals are willing to pay significantly more, sometimes up to five or six times as much, to reduce health risks at a public level relative to reducing the risks for themselves (Strand 2004; Bergstrom 2006). The information on the magnitude of altruistic effects, thus, has important policy implications (Viscusi, Magat and Forrest 1988).

Furthermore, most of the studies derive the VSL estimates from a risk-dollar tradeoff, which might not be suitable for project evaluations involving risk-risk tradeoffs. In fact, daily-life human decisions probably involve more risk-risk tradeoffs compared with risk-dollar tradeoffs (Johnson 1991). What's more, while an incident is likely to impose both mortality risk and morbidity risk on humans, there have been very few studies estimating the values of both risks within a single context (Bosworth, Cameron and DeShazo 2005). This is probably because it is difficult to disentangle the values associated with both types of risk reductions in a single context, since the correlation between fatal and nonfatal risk measures is generally strong (Viscusi 1993). However, in a study aiming to calculate willingness to pay to avoid skin cancer in the demand for a skin product involving joint production (to avoid skin cancer, to prevent premature aging of skin and to prevent sun burning and tanning), Dickie and Gerking (1996) used a choice experiment method to disentangle willingness to pay for each attribute of the product. Willingness to pay for morbidity and mortality risk reductions revealed in the demand for drinking water quality improvement programs can be similarly separated.

In benefit cost analysis with health implications, values associated with mortality risk are the most influential values in terms of policy implications in the area of environmental valuation (Krupnick 2002; Adamowicz 2004; Kochi, Hubbell and Kramer 2006). With an increased use of benefit transfer techniques, a mean value of VSL estimates from various studies and contexts is often employed to infer relevant benefits or costs of a project, due to the fact that VSL estimates are relatively generic. A review of previous studies indicates that most estimates are derived from a limited range of contexts.[4] The applicability of these VSL estimates in the drinking water context is doubtful (Raucher 2004). If VSLs are context dependent (e.g., VSLs depend on an initial risk level, and/or type of risk), this conventional practice might be problematic. This study will provide VSL estimates when the risk level is very small, of long latency and in a risk-risk tradeoff drinking water quality context.

Another important application of these VSL estimates is to inform risk management decisions when cost-effectiveness information of a public project is available. According to Johnson (1991), the mean value per premature death averted revealed in eighteen USEPA regulations was US $6.9 billion, with the costs ranging from $200,000 for initiating the TTHMs drinking water standards to $92 billion for the atrazine/alachlor drinking water standard. An estimated VSL of $6 million, for example, will draw a line between projects requiring cost outlays larger than the benefits and those that do not.

---

[4] For example, an environmental valuation database, ENVALUE (last updated 2004), maintained by the Department of Environmental & Climate Change, New South Wales, Australia, does not include any human health impact studies using media other than air. In another larger and more up-to-dated database (last updated 2007), the Environmental Valuation Reference Inventory (EVRI), maintained by Environment Canada, about 145 human health impact studies are based on air quality, and 95 are based on water quality, of which 46 studies are on drinking water quality (Appendix 1.1 Table A1.1).

This dissertation attempts to explore three different but closely related issues in the valuation of health risk reductions. While the research is motivated by a real issue, i.e., to reveal Canadians' risk preferences in the demand for drinking water treatment; the dissertation also aims to address some important methodological issues related to risk valuation, which have implications for risk valuation beyond the specific context.

Current valuation studies are mainly conducted within a random utility theory (RUT) framework, especially for stated choice data. Variants of models can be developed within a RUT framework, but each has different behavioural implications. While researchers embrace this flexibility, a model selection issue often arises. Many studies report that estimated values from alternative random utility models (RUM) were very different (Hensher 2001; Train 2003). It is therefore important to estimate a model with different specifications and to assess the robustness of welfare estimates. In the first paper, values of health risk reductions are estimated using alternative RUMs. A model averaging approach is then employed to synthesize these estimates based on the relative goodness-of-fit from these models. These RUMs differ in the assumptions about the error structure of the data, in the ways to incorporate preference heterogeneity and scale heterogeneity, and in the specifications of functional forms for income. A model averaging approach is used to deal with uncertainties in willingness-to-pay (WTP) estimates derived from different models with similar statistical performance (Layton and Lee 2006).

The second paper examines the degree of altruism in the valuation of municipal health risk reductions. The studies reviewed in the *Canadian Handbook on Health Impact Assessment* were mostly undertaken in the marketplace where values of risk reductions were treated as private goods. In this study, values of risk reductions are clearly public goods. It is likely that individuals' WTPs contain elements of altruism. Therefore, special attention has to be paid in deriving the aggregate social value of the risk reductions to avoid double counting (McConnell 1997). In this paper respondents are first differentiated according to indicator variables that identify whether an individual has taken some self-protection measures against health risks. Then, the demand with different motivations for the public good aiming at health risk reduction becomes distinguishable. An individual's total WTP for community health risk reductions is partitioned into two parts: self-interested value and altruistic value. Once the altruistic effect on the demand for the public good is identified, it is then used to calibrate the total social value of the risk reductions.

Despite the increased popularity of the stated choice method in the area of environmental valuation, marketing and transportation, its application is not without criticisms. One of them is the possibility of context effects on choice decisions. While some scholars claim "everything is context", which makes economic analysis and policy making essentially impossible. Adamowicz (2004) argued that "... what we should strive for is a more structured representation of choice behaviour in which systematic relationships between contexts, incentives, constraints and the decision structure are developed" (p. 432). The third paper is therefore an attempt to examine one type of context effect in the elicitation of risk preferences in the demand for drinking water quality in Canada. The paper attempts to explain one of the phenomena found in the first two papers. That is, datasets from different choice formats (two versus three alternatives) cannot be pooled. In other words, risk preferences revealed in different choice formats are different. If the choice format indeed matters and is not controlled for in subsequent estimation, the validity of the derived welfare estimates might be jeopardized. Therefore, the paper aims to develop a model that incorporates the effect of choice format on preference elicitation so that unified preference parameters can be derived from datasets of different choice formats. This investigation will be guided by behavioural decision theory and economic theory. Recommendations about how to incorporate or control for these factors will be made at the end of the paper.

In summary, this dissertation is motivated by three issues related to the valuation of municipal risk reductions within a RUT framework. More specifically, these issues are: 1) the impact of model specification on welfare estimates; 2) the impact of altruism on individuals' willingness to pay for community health risk reductions in drinking water; and 3) the impact of choice format on stated choice decisions.

1.2 Survey and Data

To investigate public preferences for multiple risk reductions in drinking water related health risks, an internet-based survey was conducted across Canada during the summer of 2004 (hereafter the water survey, Appendix A). It was funded by Health Canada, the Canadian Water Network and the USEPA. The water survey employs the Attribute Based Stated Choice Method (ABSCM) (Adamowicz, Louviere and Swait 1998) to obtain information about consumer preferences and tradeoffs relating to household water bill increases and morbidity and mortality health risks associated with the consumption of drinking water. In this survey, four types of drinking water related health risk are identified: *microbial illnesses*, *microbial deaths*, *cancer illnesses* and *cancer deaths*. The level of risk is defined as the number of morbidity and mortality cases related to drinking water quality in a community of 100,000 people over a 35-year period. See Appendix A for an explanation of these four types of health risks (Adamowicz, Dupont and Krupnick 2005). The survey design features eight different versions varying in elicitation methods (a CVM or an ABSCM), the number of alternatives in a choice set (for the ABSCM elicitation method), and levels of attributes (proportional attributes or non-proportional attributes). These variations enable researchers to examine the framing or context effects of different survey designs (Adamowicz, Dupont and Krupnick 2005). Individuals were asked to make tradeoffs between reducing microbial risks and reducing cancer risks as well as between reducing mortality risks and morbidity risks. With such a design, the data collected using this survey can be used to examine a range of issues that are not adequately addressed in the current risk valuation literature. A total of 32 choice sets were generated using a D-optimal design with restrictions imposed on combinations of attribute levels (e.g., to exclude choice sets containing dominating or dominated alternatives). These choice sets were then blocked into eight groups. A respondent was randomly assigned to one of the eight blocks.

In addition to an explanation of the baseline scenario of health risks in tap water and improved scenarios depicting an improved level of drinking water quality, the water survey also collected information about respondents (Appendix A). The information can be placed into four categories. The first category is socio-demographic information about respondents. This includes, for example, household income, gender, age, education, spoken language(s), health status and so forth (Appendix 1.2 Table A1.2 panel a). The information in this category is widely used to explain heterogeneity in preferences. The second category includes information about respondents' experiences or concerns associated with drinking water consumption (Table A1.2 panel b). An example is whether a respondent has had any unpleasant experiences related to tap water consumption at home. Another example is, whether a respondent has taken some averting measures against undesirable water quality at home. Information on respondent experience and averting behaviour are important for us to control for endogeneity issues that arises in preference elicitation (Cameron and Englin 1997; Louviere et al. 2005). The third category (Table A1.2 panel c) is information on respondents' attitudes towards paying for improved water quality, their opinions about current public expenditure levels in various areas (e.g., education and health care services), and other related attitudinal questions (Table A1.2 panel d). This information can be used to identify protest responses as well as "yea-saying" responses. The fourth category is information on respondents' understanding of survey information (debriefing questions), such as whether they understood the described health risk levels.

This dissertation uses data collected from two versions of the attribute based stated choice questionnaires with non-proportional attributes. One adopts a 2-alternative conjoint design (a status quo and an alternative) and the other adopts a 3-alternative conjoint design (a status quo and two alternatives). For the purpose of presentation, the sub-sample using a 2-alternative conjoint design is called CE2[5], and the sub-sample using a 3-alternative design is called CE3, and a dataset that pools CE2 and CE3 is called CE23. Summary statistics of socio-demographic information of individuals in each sub-sample is provided in Appendix 1.3 Table A1.3.[6] Note that the proposed alternatives (relative to a status quo option) are generic (unlabelled) and the status quo option is always the same across choice formats with its price level set at zero (Table 1.1).

[5] Since a conjoint choice design is often referred to as a choice experiment in the non-market valuation literature, CE is used as an abbreviation for Choice Experiment.

[6] Sample descriptive statistics for the pooled dataset CE23, including household annual income, gender, age, marital status, household size, and proportion of respondents speaking English are similar or close to Canadian population values. See details in Appendix 1.3 Table A1.3.

1.3 Organization

The thesis is organized as follows. Chapters 2 through 4 each address one of the three research topics outlined earlier in this chapter: the impact of model specification on welfare estimates in a random utility framework; the impact of altruism on individuals' willingness to pay for community health risk reductions; and the impact of choice format on choice decisions. Chapter 5 is a concluding chapter that summarizes the contributions of this thesis to the literature and outlines directions for future research on related issues.

References

Adamowicz, W. L. 2004. What's it worth? An examination of historical trends and future directions in environmental valuation. *The Australian Journal of Agricultural and Resource Economics* 48(3): 419-443.

Adamowicz, W. L., D. Dupont and A. Krupnick. 2005. *Willingness to Pay to Reduce Community Health Risks from Municipal Drinking Water: A Stated Preference Study*, Report for the U.S. Environmental Protection Agency. 62pp.

Adamowicz, W. L., J. Louviere and J. Swait. 1998. *Introduction to Attribute-Based Stated Choice Methods*. Final Report to Resource Valuation Branch, Damage Assessment Center, NOAA, U.S. Department of Commerce.

Anonymous. 2006. *Why Do We Need to Drink Water?* Web page, [accessed 19 February 2006]. Available at http://www.harmonikireland.com/index.php?topic=needwater.

Bergstrom, T. C. 2006. Benefit-cost in a benevolent society. *American Economic Review* 96(1): 339-351.

Bosworth, R., T. A. Cameron and J. R. DeShazo. 2005. Advances in evaluating the demand for risk prevention policies. *Working paper,* [accessed 12 October 2006]. Available at http://www2.bren.ucsb.edu/~kolstad/events/OccWkshp/Bosworth_-Cameron_DeShazo.pdf.

Cameron, T. A. and J. Englin. 1997. Respondent experience and contingent valuation of environmental goods. *Journal of Environmental Economics and Management* 33: 296-313.

Capece, J. 2003. *Trihalomethanes and Our Water Supply.* Web page, [accessed 19 February 2006]. Available at http://www.southerndatastream.com/thm/.

Costa, D. L. and M. E. Kahn. 2003. The rising price of nonmarket goods. *American Economic Review* 93(2): 227-232.

CTV. 2001. *Walkerton Chronology.* Web page, [accessed 19 February 2006]. Available at http://www.ctv.ca/servlet/ArticleNews/story/CTVNews/11035592658839896-8465/.

Dickie, M. and S. Gerking. 1996. Formation of risk beliefs, joint production and willingess to pay to avoid skin cancer. *The Review of Economics and Statistics* 78(3): 451-463.

Environment Canada. 2004. Municipal water use: 2001 statistics. *2004 Municipal Water Use Report.* [accessed 21 February 2005]. Available at http://www.ec.gc.ca/Water/en/info/pubs/sss/e_mun2001.pdf.

Environmental Working Group. 2006. *Contaminants Database, National Tap Water Quality Database.* Web page, [accessed 16 February 2006]. Available at http://www.ewg.org/tapwater/contaminants/contaminant.php?contamcode=2950.

Federal/Provincial/Territorial Committee on Drinking Water. 2004. *Summary of Guidelines for Canadian Drinking Water Quality.* Health Canada. [accessed 16 February 2005]. Available at http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/doc_sup-appui/sum_guide-res_recom/index_e.html.

Federal/Provincial/Territorial Advisory Committee on Population Health and Health Security. 2003. *Public Health Initiatives Related to Drinking Water Quality in Canada.* Health Canada. [accessed 16 February 2005]. Available at http://www.hc-sc.gc.ca/ewh-semt/alt_formats/hecs-sesc/pdf/water-eau/health-initiatives-sante_e.pdf.

Health Canada. 1995. *A National Survey of Chlorinated Disinfection By-Products in Canadian Drinking Water.* Health Canada. [accessed 16 February 2005]. Available at http://www.hcsc.gc.ca/ewhsemt/alt_formats/hecssesc/pdf/pubs/water-eau/byproducts-sousproduits/byproducts-sousproduits_pt1_e.pdf.

———. 2004. *Canadian Handbook on Health Impact Assessment.* Vol.3. Health Canada.

Hensher, D. 2001. The sensitivity of the valuation of travel time savings to the specification of unobserved effects. *Transportation Research Part E* 37: 129-142.

Hrudey, S. E., P. Payment, P. M. Huck, R. W. Gillham and E. J. Hrudey. 2003. A fatal waterborne disease epidemic in Walkerton, Ontario: comparison with other waterborne outbreaks in the developed world. *Water Science and Technology* 47(3): 7-14.

Johnson, G. W. 1991. Recognizing risks and paying for risk reduction. *Risk: Health, Safety & Environment* 2.

King, W. 1995. *Great Lakes Water and Your Health. A Summary of "Great Lakes Basin Cancer Risk Assessment: A Case Control Study of Cancers of the Bladder, Colon and Rectum".* Health Canada.

Layton, D. F. and S. T. Lee. 2006. Embracing model uncertainty: strategies for response pooling and model averaging. *Environmental and Resource Economics* 34(1): 51-85.

Louviere, J., K. Train, M. Ben-Akiva, C. Bhat, D. Brownstone, T. A. Cameron, R. Carson, J. DeShazo, D. Fiebig, W. Greene, D. Hensher and D. Waldman. 2005. Recent progress on endogeneity in choice modeling. *Marketing Letters* 16(3): 255-265.

McConnell, K. E. 1997. Does altruism undermine existence value? *Journal of Environmental Economics and Management* 32: 22-37.

Raucher, R. S. 2003. Benefit-cost analysis and drinking water regulation. In: *Drinking Water Regulation and Health,* (ed.) P. Frederick, 225-249. John Willey & Sons.

Strand, J. 2004. Public and private-good values of statistical lives: results from a combined choice-experiment and contingent valuation survey. *Working Paper,* [accessed 20 May 2006]. Available at http://folk.uio.no/jostrand/lifepaper02.pdf.

Train, K. E. 2003. *Discrete Choice Methods with Simulation.* Cambridge University Press: NY.

U.S. Environmental Protection Agency. 1998. *Stage 1 Disinfectants and Disinfection Byproducts Rule.* Web page, [accessed 16 May 2006]. Available at http://www.epa.gov/safewater/mdbp/dbp1.html.

U.S. Environmental Protection Agency. 2006. *Setting Standards for Safe Drinking Water.* Web page, [accessed 21 May 2006]. Available at http://www.epa.gov/safewater/standard/setting.html.

Viscusi, W. K. 1993. The value of risks to life and health. *Journal of Economic Literature* 31(4): 1912-1946.

Viscusi, W. K., W. A. Magat and A. Forrest. 1988. Altruistic and private valuations of risk reduction. *Journal of Policy Analysis and Management* 7(2): 227-245.

Table 1.1 Definition of Attributes and Attribute Levels

Attribute	Definition	Level
MICI	Number of microbial infections over a 35-year period from drinking tap water in the community	7500, 15000, 23000[a], 30000
MICD	Number of deaths due to microbial infection over a 35-year period from drinking tap water in the community	5, 10, 15[a], 20
CANI	Number of cancer cases over a 35-year period from drinking tap water in the community	50, 75, 100[a], 125
CAND	Number of cancer deaths over a 35-year period from drinking tap water in the community	10, 15, 20[a], 25
BILL	Annual increase in the current water bill in 2004 Canadian dollars	0[a], 25, 75, 125, 150, 250, 350

Note: [a] indicates the status quo level of attributes.

Appendix 1.1

Table A1.1 Basic Statistics of the Environmental Valuation Reference Inventory Database

Category			Number of Studies	Percentage of Total Studies (%)
Total number of studies			2721	
By medium				
	Air		234	8.60
	Water		732	26.90
	Land		598	21.98
	Other		1157	42.52
By method				
	Contingent Valuation Method (CVM)		1258	46.23
	Choice Experiment (CE)		15	0.55
	Travel Cost Model (TCM)		335	12.31
	Hedonic Price Method (HP)		152	5.59
	Other		961	35.32
By country				
	U.S.A.		932	34.25
	Canada		206	7.57
	Australia		50	1.84
	U.K.		245	9.00
	Other		1288	47.34
Human health impact studies			399	14.66
	By medium			
		Air	142	5.22
		Water	95	3.49
		Land	45	1.65
		Other	117	4.30
	By method			
		Contingent Valuation Method (CVM)	183	6.73
		Choice Experiment (CE)	2	0.07
		Travel Cost Model (TCM)	3	0.11
		Hedonic Price Method (HP)	22	0.81
		Other	189	6.95
Drinking water studies			113	4.15
	Of human health impact		46	1.69
		U.S.A.	20	0.74
		Canada	4	0.15
Last updated	July 2007			
Host	Environment Canada			

Appendix 1.2

Table A1.2 Description of the Water Survey Questionnaire

	Variable	Description
		a. Demographic information
1	Household income	Annual household income in Canadian dollars.
2	Gender	Male or female.
3	Age	Age in number of years.
4	Marital status	Categorical variable indicating a respondent's marital status, such as, single, married, domestic partnership, divorced, widowed or separated.
5	Household size	Number of individuals in a household.
6	Education	Categorical variable indicating a respondent's education level, ranging from 1 (primary school or less) to 6 (university graduate degree).
7	Employment	Categorical variable indicating a respondent's employment status, ranging from 1 (employed full time) to 7 (currently unemployed).
8	English	A respondent's main communication language, English or French.
9	Kids information	Categorical variable indicating number of kids in different age groups in a household.
10	Assets	Total value of a household's financial assets in Canadian dollars.
11	Urban	Categorical variable indicating the size of the city where a respondent lives in, ranging from 1 (> 1 million) to 6 (under 1499).
12	Health status	Types of disease (e.g., asthma, heart disease, cancer, chronic bronchitis, and so forth) a respondent has experienced.
		b. Experience/concern related to drinking water consumption
13	Unpleasant experience	Whether a respondent has unpleasant experience of consumption of tap water at home, such as rusty color, sediment, unpleasant smell and so forth.
14	Averting behavior	Information on whether a respondent undertakes any averting behavior against drinking water related health risks, such as boiling water, purchasing water filter system at home.
15	Concern about water quality	Whether a respondent has heard about various drinking water quality concerns in general and in the community where he currently lives in; such as E. Coli, fluoride, Trihalomethanes and so forth.
16	Filter expenditure	Household expenditures on installing water filter system and on the monthly replacement at home.
17	Bottled water expenditure	Monthly expenditures on purchasing bottled water consumed at home.

Note: This table is continued on the next page.

Table A1.2 Description of the Water Survey Questionnaire (Continued)

	Variable	Description
		<u>b. Experience/concern related to drinking water consumption</u>
18	Primary reason for using purchased water at home	Categorical variable indicating various primary reason of using perceived current tap water quality: convenience, taste, health concern and so forth.
19	Perception of quality of bottled water compared to tap water	Information on respondents' perceived bottled water quality: much more safe, a little safer, the same as, a little less safer and much less safer than tap water.
20	Expectation of future tap water quality	Respondents' expected tap water quality in two years: worse than today, same as today, better than today.
		<u>c. Attitudinal information</u>
21	Belief scientists' information on drinking water related health risks	Information on whether a respondent believe scientists are certain about health risks (microbial illnesses, microbial deaths, cancer illnesses, and cancer deaths) arising from drinking tap water.
22	Belief about the appropriateness of public expenditure on various items	Respondents' opinion on the appropriateness of the level of public expenditure on various items, such as health care service, education and environmental protection.
23	Yea-saying	Respondents' opinion on the following statement, "I am willing to see my household water bill increase by as much as it takes to reduce deaths and illnesses from drinking tap water.": strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree and strongly disagree.
24	Protest response	Respondents' opinion on the following statement, "The public should not have to pay for new water treatment options": strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree and strongly disagree.
		<u>d. Information on survey understanding</u>
25	Belief survey information on drinking water related health risks	Information on whether a respondent believe survey information on health risks arising from drinking tap water is true.
26	Survey understanding 1	When you looked at the numbers of health effects from drinking your home's tap water, did you understand that these numbers were for a 35-year period?
27	Survey understanding 2	Did you understand that your water bill would increase?

Appendix 1.3

Table A1.3 Demographic Statistics of CE2, CE3 and CE23 Datasets

Variable	Canadian Population[a]	CE2	CE3	CE23
INCOME (per household)	58360	60290 (37069)	57797 (35865)	59030 (36437)
MALE (% people who are male)	49.5%	56.35% (0.49)	50.81% (0.50)	53.55% (0.50)
AGE65 (% people over 65 years old)	13.0%	14.36% (0.35)	10.81% (0.31)	12.57% (0.33)
HHSIZE (number of persons in a household)	2.6	2.53 (1.25)	2.52 (1.37)	2.53 (1.31)
MARRY (% people married)	48.6%	50.82% (0.50)	47.03% (0.49)	48.90% (0.50)
HIGHSCHL (% people who complete high school)	55.4%	80.11% (0.40)	82.70% (0.38)	81.68% (0.39)
ENGLISH (% people whose first language is English)	73.2%	74.03% (0.44)	75.68% (0.43)	74.86% (0.43)
URBAN (% people who live in city of more than 10,000)	79.6%	71.27% (0.45)	70.27% (0.46)	70.77% (0.46)
ILLNESS (% people whose household members have ever become sick from drinking water)	n.a.	3.01% (0.17)	2.21% (0.15)	3.78% (0.19)
Number of individuals		181	185	366

Notes: Standard deviations are in parentheses. "n.a." denotes data are not available. [a]Sample descriptive statistics for the pooled dataset CE23, including household annual income, gender, age, marital status, household size, and proportion of respondents speaking English are similar or close to Canadian population values. Two sample statistics that differ from the population values are the proportion of Canadians receiving education equal to or above high school and the proportion of Canadians residing in an urban area that has a population greater than 10,000. The 2001 Census estimate on the proportion of people with more than high school education is 55 percent, while the corresponding value for our sample, collected in 2004, is 81.7 percent. A comparison of the 1996 and 2001 Census values shows that, over that five year period, the percentage of people educated beyond high school increased 5 points. So, the 2004 percentage is likely to exceed 55 percent. The Census definition for urban population is more encompassing than ours. It defines an urban area if the population is more than 1000. We used 10,000 to better capture locations with municipally supplied water.

Chapter 2 A Model Averaging Approach to Pooling Willingness to Pay Estimates from Different Model Specifications

2.1 Introduction

Drinking water treatment reduces microbial mortality and morbidity risks, but disinfectant by-products (DBPs) raise concerns about cancer mortality and morbidity. A necessary component of treatment option policy analysis is the value of these risk reductions. Information on the magnitude of these values and the differences associated with risk context (i.e., cancer versus microbial illness) would help assess the efficiency of alternative water treatment options. However, estimates of the value of risk reductions vary according to specification, functional form and a host of other factors. This paper uses a model averaging approach as a systematic method of assessing the variation in the estimated values of risk reductions.

There has been tremendous progress in the econometric modelling of discrete choice data. With advances in computing technology, it has become much easier to estimate complex models that might better approximate "true" preferences. At the same time, however, model specification or model selection has become more of an issue. We have more models to choose from yet the true model is still "unknown". Willingness-to-pay (WTP) estimates are often sensitive to model specifications (Herriges and Kling 1999; Haab and McConnell 2002; Layton and Lee 2006). Layton and Lee (2006) state that good research generates estimates that are not sensitive to a particular model specification and provides an analysis that is "sufficiently transparent and robust so that readers or policy makers can believe the results" (p. 53). Therefore, communicating the process by which the WTP estimates are derived is an important part of empirical analysis.

Recently, there is increasing interest in using a model averaging approach to improve the transparency of the derivation of WTP estimates to ensure their validity and robustness (Buckland et al.1997; Burnham and Anderson 2004; Layton and Lee 2006). A model averaging approach acknowledges uncertainty in model selection, and reports a weighted averaging WTP estimate based on a range of models. These synthesized WTP estimates are likely to be more robust (Burnham and Anderson 2004).

To implement a model averaging approach, we first need to choose alternative specifications. We identify four types of major specification issues that are relevant when modelling discrete choices. These include the error term structure, preference heterogeneity, scale heteroscedasticity and the income effect. Unlike previous research where only issues related to preference functions are examined (e.g., Herriges and Kling 1999; Layton and Lee 2006), this paper examines some major specification decisions or assumptions researchers make when estimating random utility models. In addition, a hierarchical structure is built into alternative specifications so that main and interaction effects can be examined. Apart from providing WTP estimates to aid policy making, this paper is also a methodological exploration of how fundamental choices about specification affect model fit, and how a model averaging approach can be used to derive more robust WTP values. We believe that emphasis on the process by which WTP estimates are derived provides valid and robust estimates for policy makers.

The paper is organized as follows. Section 2.2 introduces some major specification issues in modelling discrete choice data based on random utility theory. Section 2.3 explains how to use a model averaging approach to synthesize WTP estimates from different models. Section 2.4 introduces the data, variables and alternative specifications for modelling individual choice decisions to reduce health risks from drinking water. Section 2.5 reports model estimation, examines effects of model specifications on model fit and derives weights for different models. Model weights derived by a combined hypothesis testing and model averaging approach versus a model averaging approach alone are compared and discussed. Section 2.6 reports model weighted WTP estimates. The last section concludes the paper.

2.2 Specification Issues in Modelling Individual Choices for Drinking Water Health Risk Reduction Programs

The random utility model (RUM) is the most popular model for analyzing discrete choices in the area of non-market valuation (McFadden 1974). Within the random utility theory (RUT) framework, a rational individual chooses the most preferred alternative from a choice set comprised of a finite number of exclusively defined alternatives. This chosen alternative gives him or her the highest level of utility. An individual n derives utility from an alternative j (U_{nj}), which includes two components: one is deterministic and observable (V_{nj}), and the other is stochastic and unobservable from a researcher's perspective (ε_{nj}).

(2.1) $$U_{nj} = V_{nj} + \varepsilon_{nj}$$

A task faced by a researcher is to find factors affecting the deterministic part of utility while assuming that in aggregate the individuals' stochastic component (or error term) follows a specified statistical distribution. An individual chooses alternative j over alternative k, if $V_j > V_k$, so that the probability of choosing j for individual n is

(2.2) $$\begin{aligned} P_{nj} &= P(V_{nj} + \varepsilon_{nj} > V_{nk} + \varepsilon_{nk}) \\ &= P(\varepsilon_{nk} - \varepsilon_{nj} < V_{nj} - V_{nk}) \end{aligned}$$

Let ε be the difference between the error terms, and let F_ε (a) be the probability that the random variable ε is less than a,

(2.3) $$P_{nj} = F_\varepsilon(V_{nj} - V_{nk}) = F_\varepsilon(a)$$

For this point on, a researcher has to make assumptions about the distribution of the error term ε and the functional form of the indirect utility function or preference function V_{nj} to proceed with empirical estimation (Haab and McConnell 2002). However, estimation results are likely to be dependent on these assumptions. An examination of how estimates vary across different model specifications is important. It ensures the validity of derived statistiscal inferences (Kling 1987; Hensher 2001; Scarpa, Ferrini and Willis 2005).

2.2.1 Assumptions about the Error Term Structure

Variants of RUMs can be derived when different assumptions about error term distributions are invoked (Ben-Akiva and Lerman 1985). For example, if the error terms are distributed type I extreme value, a conditional logit model (CL) results; if they are distributed generalized extreme value (GEV), a nested logit model (NL) results; and if they are normally distributed, a probit model results (or a multinomial probit model, MNP, for more than two alternatives) (Appendix 2.1).

While researchers have flexibility in assuming the distribution of error terms, one has to be aware of the behavioural implications underlying these assumptions. For example, a type I extreme value distribution implies error terms are independent and identically distributed (IID), i.e., $\varepsilon_{nj} \sim i.i.d(0, \pi^2 / 6)$. The resulting model satisfies the independence of irrelevant alternatives (IIA) property. The IIA property implies that the odds-ratio between two alternatives does not change by the inclusion (or exclusion) of any other alternatives. This IIA assumption is rather restrictive and is often violated. Bhat (1995) suggests three ways to fully or partially relax the IID assumption removing any of its essential characteristics: 1) error terms are non-independent and non-identically distributed; 2) error terms are non-independent but identically distributed; or 3) error terms are independent but non-identically distributed. A NL model, for example, allows for some correlation of error terms between alternatives within a specified nest but no correlation of error terms between alternatives in different nests. This removes the IID assumption by assuming error terms are non-independently but identically distributed. A heteroscedastic extreme value (HEV) model overcomes the IID assumption by assuming the error terms are non-identical but independently distributed. The HEV model permits a more flexible cross-elasticity structure than the NL model (Bhat 1995). A MNP model relaxes the IID assumption by assuming a multivariate normal distribution for error terms so the error terms are non-independent and non-identically distributed. While it allows for any correlation pattern between alternatives as well as non-constant variance of each alternative, the MNP model is computationally cumbersome and has other undesirable properties (Bhat 1995).

Models with more general error structures are more difficult to estimate. Some error structures, e.g., type I extreme value IID distributed or GEV distributed, result in a closed form solution, while others do not. In those cases, numerical techniques are used to facilitate estimation. Inappropriate assumptions about error term distributions may result in erroneous prediction and invalid inferences (Train 1998). As underlying distributions of error terms of actual data are usually unknown, econometric analysis of different model specifications is warranted (Hensher 2001). Appendix 2.1 Table 2.1 panel a is a list of RUM specifications with different error structures, their algebraic forms and associated underlying behavioural implications.

While choosing an appropriate assumption about the error structure is crucial in estimating a random utility model, there are two other important aspects to consider. These include incorporating heterogeneity in preferences (or taste heterogeneity) and heterogeneity in scale (scale heteroscedasticity).[7]

2.2.2 Assumptions about Preference Heterogeneity

One of the primary advantages of a model based on individual level data is its ability to incorporate preference heterogeneity. Revelt and Train (1998) and Train (1998) have shown that failure to incorporate heterogeneity in taste results in erroneous welfare estimates. Instead of assuming a fixed marginal effect of an attribute on the probability of choice across consumers, a random parameters logit (RPL) model (or mixed logit model, ML) assumes that population tastes follow a statistical distribution (e.g., a normal distribution or a lognormal distribution). Analytically, the RPL probabilities are the integrals of standard logit probabilities over a density of parameters (Train 1998). A preference function V = $f(X,\beta)$, where **X** is a vector of attributes describes the alternatives and β is a vector of parameters to be estimated. The average probability for choosing alternative j is

$$\bar{P}_{nj} = \int P_{nj} f(\beta \mid \theta) d\beta \tag{2.4}$$

where P_{nj} is the probability individual n choosing j conditional on a specific value of β, and θ is a vector of parameters that describe the distribution of β (e.g., mean and variance of a normal distribution) (Table A2.1 panel b).

If there are discrete groups of consumers with equal tastes, a latent class model (LCM) is more suitable. In that case, the integral operation in the RPL will be replaced by a summation operation over the number of groups.

Another way to incorporate taste heterogeneity is to use socio-demographic variables to differentiate consumers with different tastes. It is fairly straightforward to incorporate socio-demographic variables in RUMs. It involves specifying indirect utility

---

[7] In this dissertation scale heterogeneity and scale heteroscedasticity are used interchangeablely, although heterogeneity originally was used to describe differences among people and heteroscedasticity was a term from econometric analysis and is used to describe non-constant variance of error terms. Since both socio-demographic variables and contextual variables can lead to non-constant variance, we do not differentiate between them in this study.

functions with socio-demographic variables interacting with attributes. The resulting RUM is a mixture of CL and multinomial logit (MNL) models.[8] The RUM literature does not distinguish between the two models since most specifications include both attributes of quality-differentiated alternatives and socio-demographic variables.

The decision to incorporate preference heterogeneity is largely an empirical question. Some studies have shown that welfare estimates between a fixed coefficient specification and a random coefficient specification are similar despite the fact the there is a strong evidence of heterogeneity in preferences (Mazzanti 2003). It is generally more difficult to estimate a RPL model, especially when the number of random parameters increases. Many studies report that the welfare estimates derived from the RPL model, are sensitive to distributional assumptions about random parameters, and tend to have unreasonably large variances (Meijer and Rouwendal 2006; Train and Weeks 2005). To obtain reasonable welfare estimates, common practice is to assume a fixed price coefficient, re-introducing some inefficiency that the RPL model aims to reduce.

2.2.3 Assumptions about Scale Heteroscedasticity

There has been increased concern about issues related to scale heteroscedasticity in choice modelling since Swait and Louivere (1993) illustrated the role of the scale parameter using a multinomial logit (MNL) model. For a simple MNL model, the probability that an individual n chooses alternative j, where $j = 1, 2, ..., J$, is defined as follows.

$$P_{nj} = \exp(\mu V_{nj}) \Big/ \sum_{j \in C} \exp(\mu V_{nk}) \tag{2.5}$$

$$\text{and } V_{nj} = \alpha(Y_n - PRICE_j) + \boldsymbol{\beta}\mathbf{X}_j$$

where Y_n is income for individual n, $PRICE_j$ is the price of alternative j. The scale parameter μ is confounded with the deterministic component of utility V_{nj}, which is a linear additive function of individual n's residual income on the numeraire good j (i.e., Y_n - $PRICE_j$) and non-price attributes $\mathbf{X}_j$. Common practice is to normalize μ to one. In

[8] Strictly speaking, a CL model assumes $V_{nj} = \boldsymbol{\beta}\mathbf{X}_j$ while a MNL model assumes $V_{nj} = \boldsymbol{\beta}_n\mathbf{X}$. The former implies the indirect utility function is a function of attributes of alternatives, while the latter assumes it is a function of individual socio-demographics.

fact, μ is the inverse of variance of the IID error term. So, $\mu = \frac{1}{\sqrt{\text{var}(\varepsilon_j)}}$, and $var(\varepsilon_j)$ = $\pi^2/6$ in a simple logit model (Train 2003). Swait (2005) illustrates that as the scale parameter μ approaches infinity, the probability of choosing j goes to unity; and as the scale parameter approaches zero, each alternative has an equal probability of being chosen, i.e., $1/J$. Therefore, a large scale parameter implies a higher weight on the deterministic component of utility, thus more predictable behaviour, *ceteris paribus.* A small scale parameter implies a higher weight on the stochastic component of utility, thus less predictable behaviour. Clearly, the scale parameter has important behavioural implications. It indicates how the variance of responses varies due to changes in choice environements or in the levels of choice complexity. Homoscedasticity of error terms is one of key assumptions for the unbiased estimators in logit models, unlike ordinary least square estimators, for which heteroscedasticity in error terms does not result in biased estimators (Swait and Louviere 1993).

Bhat (1995) and Allenby and Ginter (1995) propose a heteroscedastic extreme value (HEV) model that allows for estimating $n-1$ alternative specific scale parameters for a n-alternative choice decision. The model is actually a degenerate NL model where each alterantive is a nest by itself. The model relaxes the constant variance restriction on the error term structure. Yet it does not offer any behavioural explanations for why variances differ across alternatives, and it does not attempt to capture "systematic relationships between random component variances and covariance and attributes of choice options and characteristics of individuals" (Louviere 2006, p. 185).

To explore the behavioural role of the scale parameter within a RUT framework, Swait and Adamowicz (2001a, 2001b) suggest estimating the scale parameter as a function of socio-demographic variables and contextual variables. They specify the scale parameter as an exponential function of exogenous variables:

(2.6) $\mu_{nt} = \exp(\alpha'\mathbf{S}_n + \theta'\mathbf{Z}_{nt})$,

where $\boldsymbol{S_n}$ is a vector of socio-demographic variables of individual n and $\boldsymbol{Z_{nt}}$ is a vector of contextual variables describing choice environments faced by individual n at choice task t. However, identification issues might arise when **S** is included in both the preference function and the scale function for a given dataset. Therefore, **S** often only appears in one of the functions.

Controlling for variance in responses is recommended for studies attempting to combine data from difference sources. It is now widely recognized that to test for preference homogeneity of data from difference sources, scale heteroscedasticity should be controlled (Adamowicz, Louviere and Williams 1994; Mazzotta and Opaluch 1995; Hensher, Louviere and Swait 1999; DeShazo and Fermo 2002). Since different choice formats are used for preference elicitation in the survey examined in this study, it is necessary to examine the impact of choice format on the variance of choice responses by comparing models that allow for scale heteroscedasticity with those that do not.

2.2.4 Assumptions about Income Effects

In addition to the aforementioned specification issues in estimating a random utility model, attention should also be given to the impact of functional form (of the indirect utility function) on welfare estimates (Huang and Smith 1998). Here, we focus on the implications of a linear specification of income that is widely adopted in the literature (Morey, Rowe and Watson 1993). Consider an indirect utility function assumed to be linear additive in attributes,

(2.7) $$V_{nj} = \alpha(Y_n - PRICE_j) + \beta\mathbf{X}_j$$

where Y_n, $PRICE_j$ are defined the same as in Equation 2.5; $\mathbf{X}_j$ is a vector of attributes describing alternative j; α and β are parameters to be estimated. If the preference function is linear in income, there is no income effect on the decision to choose between alternatives. Welfare estimates will thus be independent of income (Morey, Rowe and Watson 1993). Some studies have reported that people with higher income levels are willing to pay higher amounts for health risk reductions (Viscusi 1993). Therefore, a linear-in-income specification might not be appropriate in this context. A linear specification also implies that respondents' willingness to pay is unbounded (Haab and McConnell 2002). One would expect that the expected willingness to pay derived from a sample should be bounded between zero and the sample mean income for an increase in the levels of attributes of desirable goods (Haab and McConnell 2002; Cameron et al. 2002).

Therefore, it is important to investigate the impact of a linear specification on welfare estimates.

A few studies illustrate a few different approaches to constrain the expected WTP within a desired interval (Kling 1987; Herriges and Kling 1999; Cameron et al. 2002; Haab and McConnell 2002; Layton and Lee 2006). A simple form is

(2.8) $$V_{nj} = \alpha ln(Y_n - PRICE_j) + \beta X_j$$

which allows for the marginal utility of income to decrease as income increases (Haab and McConnell 2002). This specification implies that WTP estimates are bounded from above. More complicated non-linear income effects, like Generalized Leontief functions (or Diewart functions) and Translog utility functions can also be used (e.g., Herriges and Kling 1999) but they are in general more difficult to estimate as a result of high collinearity among right-hand-side variables.

2.3 A Model Averaging Approach

Often theory provides little guidance for choosing among models with different specifications, especially when models are non-nested. For models that are nested, statistical tests can be conducted to facilitate model selection. Non-nested models with competing levels of goodness of fit may provide distinctively different estimates. A researcher may find it difficult to decide which set of results should be reported. Results from any single model may be specific to its specification and are difficult to generalize (Louviere 2006). Layton and Lee (2006) recommend a model averaging approach, proposed by Buckland et al. (1997), to deal with model selection uncertainty in stated preference modelling.[9] The model averaging technique develops a weighted estimate of expected willingness to pay (*EWTP*) derived from a range of models (Buckland, Burnham and Augustin 1997):

(2.9) $$EWTP_{Mavg} = \sum_{1}^{M} w_m EWTP_m$$

[9] The model averaging approach is referred as a frequentist based approach. Layton and Lee (2006) acknowledge that this approach lacks the formal justification provided by the Bayesian framework (see Koop and Tole 2004).

where $EWTP_m$ is the expected willingness to pay derived from model m, and w_m is the weight of the willingness to pay estimates provided by model m. Akaike's Information Criterion denoted AIC (Akaike 1973) or the Bayesian Information Criterion denoted BIC (Schwarz 1978) are used to determine the weights (Layton and Lee 2006).[10]

$$\begin{aligned} AIC &= -2\ell + 2b \\ BIC &= -2\ell + b\ln(N) \end{aligned} \tag{2.10}$$

where ℓ is the log-likelihood, b is the number of parameters in the model, and N is the sample size. For both criteria, the smaller is the absolute value, the better is the goodness of fit of a model. Buckland et al. (1997) derive the weight w_m for model m as follows. Let the value of one of the criteria for the model be $Crit_m$, and $m = 1, \ldots, M$.

$$w_m = \frac{\exp(-Crit_m / 2)}{\sum_{m=1}^{M} \exp(-Crit_m / 2)} \tag{2.11}$$

The difference between the AIC criterion and the BIC criterion is that BIC penalizes the additional parameters more heavily than AIC for any reasonable sample size (as long as the sample size is greater than exp(2), i.e., 7.39). Thus BIC tends to select models with fewer parameters than AIC (Buckland, Burnham and Augustin 1997).

However, Equation 2.11 may not be always feasible for models that have large absolute values of AIC or BIC given the fact the exponential function of a positive number quickly goes to infinite as it increases. One solution is to calculate weights based on differences in AIC or BIC since only differences in AIC or BIC are a meaningful measure of relative model fit (Burnham and Anderson 2004). Let $Dcrit_m$ be the relative AIC or BIC for model m and $Crit_{min}$ be the minimum,

$$Dcrit_m = Crit_{min} - Crit_m \tag{2.12}$$

So $Dcrit_m$ is zero for the best fitting model with the lowest AIC or BIC, and is negative for models with higher AICs or BICs. Thus, Equation 2.11 becomes,

$$w_m = \frac{\exp(Dcrit_m / 2)}{\sum_{m=1}^{M} \exp(Dcrit_m / 2)} \tag{2.13}$$

[10] Please refer to Buckland et al. (1997) for reasons using AIC or BIC to derive weights.

The exponential of a negative value approaches zero as its absolute value increases. Thus the higher the relative AIC or BIC, the smaller the weight. Using the model averaging technique, Layton and Lee (2006) derive the expected WTP estimates from 25 different models, differing in the specifications of the preference function (such as interaction effects between attributes, and linear versus non-linear income effects). They found that using BIC, the weight allocated to the model with the smallest BIC is greater than 99%, so the weighted EWTP estimates are essentially the estimates provided by the best fitting model. Using AIC, although the weights are more spread out across models, only a handful of models have non-zero weights. One reason for the unbalanced weights might be that Layton and Lee (2006) did not conduct nested model selection tests before synthesizing the model results and some models are nested within others. Statistical testing could be conducted to exclude models that are inferior statistically. Then, a model averaging approach is used to synthesize results from the remaining models.

In summary, a model averaging approach is used to synthesize willingness-to-pay estimates provided by different models based on relative performance. Model weights will be calculated based on the AIC or BIC criterion. An analysis is also conducted to assess which specification, functional form and error distribution assumption contribute most to model performance.

2.4 Data and Model Specifications

2.4.1 Data

This paper uses data collected from two versions of the internet survey introduced earlier: CE2 and CE3 (i.e., the 2-alternative and the 3-alternative conjoint design datasets). Our analysis is conducted based on the pooled dataset CE23 (see sample statistics in Chapter 1 Appendix Table A1.3). This dataset excludes observations defined as "yea-saying" data (Mitchell and Carson 1989; Andreoni 1995).[11] A total of 1464 observations from 366 respondents are included in our analysis.

[11] "Yea-saying" data are defined here as those respondents who stated that they were willing to pay any amount to reduce the health risks in the surveys. It is possible that these individuals did not make tradeoffs between attributes or between an attribute and money, and therefore, inclusion of their responses in the

In this study, each water treatment program is characterized as a bundle of health risk attributes and an increase in current annual household water bill. A status quo option is included as a baseline program that does not involve any increase in the water bill. The alternative programs are characterized with at least one type of health risk reductions as well as some greater-than-zero increase in the water bill. The risk attributes are, as introduced in Chapter 1 (Table 1.1), number of microbial illnesses (*MICI*), number of microbial deaths (*MICD*), number of cancer illnesses (*CANI*) and number of cancer deaths (*CAND*). Individual n's indirect utility associated with alternative j is specified as, in the most basic form,

(2.14)

$$V_{nj} = \beta_1 SQ_j + \beta_2 MICI_j + \beta_3 MICD_j + \beta_4 CANI_j + \beta_5 CAND_j + \beta_6 BILL_j + \beta_7 SQCE3_j$$

where *SQ* is the alternative specific constant (ASC) for the status quo option. It is included to capture unobserved utility associated with staying at the status quo (Adamowicz et al. 1998; Scarpa, Ferrini and Willis 2005). An interaction term *SQCE3* between *SQ* and a version dummy variable for the 3-alternative choice data (*CE3*) is included to account for the choice format effect on preferences.[12] Table 2.1 provides definitions and levels of these attributes. Other demographic information, such as income, age, gender, major communication language, city size, marital status, number of children in a household, is summarized in Table 2.1.

2.4.2 Hierarchical Model Specifications

For a large dataset, the number of alternative specifications can be large (Weakliem 2004). Researchers can only choose a subset of these specifications. Any arbitrary subset may be just as good provided it is large enough to appropriately cover the parameter space. Careful examination of alternative specifications may provide a better understanding of how a particular specification affects model performance, holding other specifications constant. For example, Layton and Lee (2006) estimate 25 models for both rating and ranking data that are specified with different preference functions: linearity

analysis might lead to erroneous inference. In this study, about 10% of the survey responses are identified as the yea-saying data.

[12] Chapter 4 of this thesis is devoted to explain the choice format effect.

versus non-linearity in quality attributes and in income, and with various interaction effects between these attributes. Implicitly, they hold other specifications constant cross models. In this paper we attempt to examine some "major" specification issues pertaining to random utility model estimation. These are the four types of specification issues we discussed earlier: error structure, preference heterogeneity, scale heterogeneity and linearity versus non-linearity in income. Therefore, a four level hierarchical structure is built into alternative model specifications (Table 2.2).

On the top level, there are four types of specifications on scale heterogeneity, homoscedastic logit model (HLM), heteroscedastic logit type I model (HET1), heteroscedastic logit type II model (HET2) and heteroscedastic logit type III model (HET3). Recall that our dataset CE23 is a pooled dataset between CE2 and CE3. In the HLM model, the scale parameter for both datasets is assumed to be equal to one. This is a baseline scenario, which implies variance in responses is the same across all choices. For the other three heteroscedastic logit models, such a restriction is relaxed. The scale parameter is parameterized as a function of variables that might explain difference in the variance of responses. Three types of scale functions are hypothesized. The HET1 model assumes the scale parameter is a function of the dataset dummy variable (*CE3*) only, the HET2 model assumes it is a function of *CE3* and the order of choice tasks an individual is faced with (*ORDER*), where *ORDER* is a dummy variable indicates whether a choice task is the first one faced by a respondent (Table 2.1). The HET3 model assumes it is a function of *CE3*, *ORDER* and a vector of socio-demographic variables. Other things being equal, a HLM model is nested within a HET1 model, a HET1 model is nested within a HET2 model, and a HET2 model is nested within a HET3 model.

At the second level, different error term structures are specified (Table 2.2). They are the conditional logit model (CL), nested logit model (NL), random parameters logit model (RPL) and error term correlated random parameters logit model. The CL model assumes independence of irrelative alternatives (IIA), the NL model partially relaxes the assumption, and the last two random parameter models fully relax the assumption on the correlation among alternatives. The last model allows for different degrees of correlation across specified nests as well as estimating individual specific preference parameters. It is actually a nested random parameters logit model (NRPL). It captures both unobserved preference heterogeneity as well as different degrees of correlations across alternatives that are observed by researchers. For our dataset, nests only exist for the CE3 dataset. The CE3 dataset has a status quo nest that only has one alternative--the status quo option and a non-status-quo nest that has two alternative programs. We could also estimate a heteroscedastic extreme value model (HEV) which assumes error terms associated with each alternative have different variances. This is not necessary in this study because our non-status quo alternatives are designed to be generic. No uniqueness should be associated with one of the alternative programs versus the other.[13] Therefore, for the four types of models, a CL model is nested within a NL model, a RPL model and a NRPL model. The NL model is nested within the NRPL model only, as is the RPL model. The NL model and the RPL model are two non-nested models. Equation 2.14 is a CL model. For a NL model, an additional parameter ζ is estimated to measure how similar alternatives within a nest that consists of non-status-quo options versus the status quo option.

(2.15) $V_{nj}\_NL = V_{nj} + \zeta\, NEST\_NONSQ_j$

where V_{nj} is defined in Equation 2.14 and *NEST_NONSQ* is a dummy variable indicating whether an alternative is a non-status-quo option or not.

For the RPL model, we estimate *SQ, MICI, MICD, CANI, CAND* and *SQCE3* as random parameters. Since we are more interested in how a random parameter specification affects model fit compared to a fixed parameter one, only one type of distribution is assumed for a parameter.[14] For the coefficients on risk attributes, theoretically, a lognormal distribution is more appropriate because people in general prefer lower health risks. Empirically, it is more difficult to estimate a RPL model with lognormally distributed coefficients. These coefficients are assumed to be normally distributed and biases in welfare estimates are expected to be small as long as the proportion of the distribution in an undesired interval is small (Revelt and Train 1998). A

[13] A HEV model was estimated, but it is found that the null hypothesis of equal variance between the non-status-quo alternatives cannot be rejected.

[14] Another reason is that allowing different specifications about the distributions of random parameters substantially increases the number of models to be estimated because of the adoption of an experimental design approach.

normal distribution is also assumed for the coefficients on *SQ* and *SQCE3*. In addition, we assume fixed price effects to facilitate welfare estimates.[15]

Using "_SD" to denote standard deviation of each variable, the RPL model has six extra parameters compared to the CL model (Equation 2.14),

$$(2.16)\quad \begin{aligned} V_{nj}\_RPL = V_{nj} &+ \xi_1 SQ\_SD + \xi_2 MICI\_SD + \xi_3 MICD\_SD \\ &+ \xi_4 CANI\_SD + \xi_5 CAND\_SD + \xi_6 SQCE3\_SD \end{aligned}$$

Lastly, a NRPL model is specified as

$$(2.17)\quad V_{nj}\_NRPL = V_{nj}\_RPL + \xi_7 NEST\_NONSQ\_SD$$

where we estimate a parameter of the standard deviation of *NEST_NONSQ* and fix its mean at zero so that error term associated with each alternative within a nest is allowed to be correlated differently than other alternatives. This parameterization acknowledges that the RPL model is just one of two interpretations of the mixed logit model. Another interpretation is that it is an error components model which creates correlation among the unobserved utilities for different alternatives (Train 2003).

At the third level, two types of specifications are examined: linearity versus non-linearity in income. In the linear specification (V_{nj}_Y), income drops out of the preference function and marginal utility of income is equal to the negative of the marginal utility of the price effect.

$$(2.18)\quad V_{nj}\_Y = V_{nj}$$

where V_{nj} is defined in Equation 2.14. For a nonlinear specification (V_{nj}_LNY), a basic model is,

(2.19)

$$V_{nj}\_LNY = \beta_1 SQ_j + \beta_2 MICI_j + \beta_3 MICD_j + \beta_4 CANI_j + \beta_5 CAND_j + \beta_6 LNY_{nj} + \beta_7 SQCE3_j$$

where LNY_{nj} = $\ln((INCOME_n - BILL_j)/INCOME_n)$ and $INCOME_n$ is annual household income for individual *n* (Table 2.1). That is, it is the logarithm of the percentage of residual income on the numeraire good when alternative *j* is chosen.

At the lowest level, two additional types of specifications are examined: a no-covariates specification versus a with-covariates specification. A linear no-covariates

---

[15] Since a welfare estimate is calculated as the ratio between the marginal utility of an attribute and the marginal utility of income, the estimated distribution does not have a resulting distribution when both the numerator and denominator are random variables (Meijer and Rouwendal 2006). Moreover, welfare estimates often have large variances when the price effect (or income effect) is not fixed (non-random).

specification ($V_{nj}\_Y\_0$) is indicated by Equation 2.18 and a non-linear no-covariates version is indicated by Equation 2.19 ($V_{nj}\_LNY\_0$). For a with-covariates specification, linear and non-linear forms are,

(2.20) $V_{nj}\_Y\_1 = V_{nj}\_Y + \gamma SQ_j * \mathbf{S}_n$, and

(2.21) $V_{nj}\_LNY\_1 = V_{nj}\_LNY + \gamma SQ_j * \mathbf{S}_n$

respectively. Equation 2.18 is nested within Equation 2.20, and Equation 2.19 is nested within Equation 2.21. A set of candidate socio-demographic variables are *AGE65*, *INCOME*, *INCOME2* (squared income), *ENGLISH*, *CITYSIZE*, *ILLNESS*, *MALE*, *MARRY*, *KID06*, *KID612* and *KID137* (Table 2.1). To keep the modelling exercises more focused on the major specification issues, the number of socio-demographic variables is kept constant across all specifications and only interactions between the status-quo alternative specific constant (*SQ*) and socio-demographic variables are included. However, for the linear specification, interaction terms between income variables and *SQ* are included, while for the non-linear specification, the interactions are dropped to avoid collinearity.

The no-covariates and with-covariates specifications for the HET3 models are specified slightly differently:

(2.22) $V_{njt}\_HET3\_0 = \mu^i(V_{nj} + \delta SQ_j * ORDER_t),\ \ \mu^i = f(CE3)$

(2.23) $V_{njt}\_HET3\_1 = \mu^{nit} V_{nj},\ \ \mu^{nit} = f(CE3, ORDER_t, \mathbf{S}_n)$

where V_{nj} is defined in Equation 2.14 or 2.18 (a linear in income specification versus a non-linear specification); i indicates datasets: CE2 or CE3, t indicates the sequence of choice task, and t = 1,2 3, 4; other variables are defined as above. For the no-covariates specification, order effects enter the preference function rather than the scale function so that only the dataset dummy variable (*CE3*) enters the scale function. For the with-covariates specification, the scale parameter is a function of *CE3*, *ORDER* and a vector of socio-demographic variables and the preference function uses the basic specification of V_{nj}. Equation 2.22 and Equation 2.23 are non-nested models.

By building a hierarchical structure into alternative specifications, we actually have designed an experiment to examine the model specification effect. Each specification can be considered a treatment as in the standard design of experiments.

Based on the levels of each "treatment", we are going to estimate a total of 64 models ($4 \times 4 \times 2 \times 2 = 64$).[16] We name each model using abbreviations of levels of specifications. For example, HLM_CL_Y_0 indicates a homoscedastic conditional logit model with linear in income and no-covariates specification. For an example, HET2_NRPL_LNY_1 is a heteroscedastic type II, error term correlated and random parameters logit model with non-linear in income and with-covariates specification.

In summary, given the dataset, we adopt an experimental design approach to examine effect of model specifications on model performance. The experimental approach enables us to investigate main effects and interaction effects of different specifications on model performance in a systematical fashion. There are two advantages of focusing on the major specification issues. One is that we can have a large pool of alternative models (or a good coverage of the parameter space) for model averaging in the next stage. Another is that it is possible to shed some light on how model performance is affected by some popular specifications in estimating random utility models.

2.5 Model Estimation and Model Selection

2.5.1 Sensitivity Analysis of Model Specifications on Model Performance

A total of 64 models are estimated. The log-likelihood (LL) values and pseudo R^2 values of these models are reported in Table 2.3, along with calculated AIC and BIC values and the AIC and BIC model weights.[17]

The 64 models are presented in a sequence in line with the hierarchical structure of model specification (Table 2.3). The first 16 models are the HLM models, followed by 16 HET1 models, 16 HET2 models and 16 HET3 models. For every sixteen models, eight linear models are presented before eight non-linear models. For the eight-model

[16] Strictly speaking, we do not have a full factorial design because the HET3 models are specified depending on the specification about covariates.

[17] For a with-covariates linear-in-income specification, eight out of the eleven socio-demographic variables are chosen, and they are *AGE65, INCOME, INCOME2, ENGLISH, CITYSIZE, ILLNESS, MALE* and *MARRY* (Table 2.1). For a non-linear specification, in contrast, only six of the eight socio-demographic variables are chosen, and two variables on income (INCOME, INCOME2) are dropped out.

block, four no-covariates models are presented first, followed by four with-covariates models. The four models are CL, NL, RPL and NRPL models respectively. Since no-covariates models are nested within their corresponding with-covariates models except for the HET3 models, and the linear and non-linear model are non-nested models, we block the 64 model into 8 groups to facilitate analysis.

The 64 models are specified with between 7 and 24 variables (Table 2.3). HET2_NRPL_Y_1 has the highest LL value (-975.68) and it is also the most complex model (with 24 parameters) while HLM_CL_LNY_0 has the lowest LL values (-1160.19). The pseudo R^2 values range from 0.09 to 0.23. AIC values range from 1998.78 to 2336.54 and BIC values range from 2097.19 to 2394.36. The model with the lowest AIC is HET2_RPL_Y_1 and the one with the lowest BIC is HET3_RPL_Y_0. Not surprisingly, the BIC criterion prefers a simpler model. Based on AIC values, a dozen models have non-zero weights (AICW is at least greater than 0.001%) as compared to eight models based on BIC values. The fact that only a small percentage of models have non-zero weights is dictated by the formula for calculating the weights (Equation 2.13). As Burnham and Anderson (2004) explain that a zero weight is essentially assigned to a model that has an AIC or BIC value larger than that of the best-fitting model by 10. Given the fact that AICs and BICs in this study have a wide range (about 300), it is not surprising that only a handful of models have nonzero weights.

To systematically analyze the effect of model specifications on model fit, we run a simple regression of various measures of model fit on model specification. Since the relationship between model fit and model specification is deterministic, we technically have an identity relationship rather than a behavioural relationship. However, errors associated with model specifications due to misspecifications because an analyst does not know the true model may generate variation in the relationship between specification and model fit. As analysts, we often do not know what exactly drives model fit. This analysis can be considered a way to summarize the relationship between model fit and model specification. Therefore, three types of measures of model fit: LL, AIC and BIC are regressed on a series of model specifications as “treatments”. There are a total of 64 observations, which is exactly the number of models estimated. The results of these regressions are presented in Table 2.4.

Sixteen explanatory variables are included in the regressions, a constant (Intercept), eight main effect variables and seven interaction terms. The eight main effect variables are number of parameters, *HET* (allowing for scale heteroscedasticity or not, which takes 1 for the HET1, HET2 and HET3 models and zero for the HLM models), *RPL* (random parameters logit specification or not), *NL* (allowing for nesting structure or not), *Nonlinear* (nonlinear in income effect or not), *CovariateP* (1 for with-covariates specification and covariates enter preference function, and 0 otherwise), *CovariateS* (1 for covariates enter scale function, and 0 otherwise), *Order* (1 if the variable *ORDER* that defined in Table 2.1 is included in preference function). The number of parameters is included as models specified with *RPL, NL, Order, CovariateP and CovariateS* specifications necessarily involve extra parameters and the inclusion of this variable captures how these specifications contribute to model fit in addition to just bringing in extra free parameters. Then, interactions between the RPL specification and *HET*, *CovariateP* and *CovariateS* are included (*RPL*HET, RPL*CovariateP, RPL* CovariateS*), and interactions between the NL specification and *HET, CovariateP* and *CovariateS* (*NL*HET, NL*CovariateP, NL*CovariateS*), as well as an interaction term between *RPL* and *NL* (*NL*RPL*). The R^2 reaches 0.996, reflecting a nearly deterministic relationship between model specification and model fit as expected.

Since AICs and BICs are calculated based on the LL values and number of parameters, the signs and significance levels for each variable are consistent with the model of the LL. In fact, the estimated coefficients in the AIC model are exactly the same as in the BIC model except for the coefficient on number of parameters. An extra parameter decreases the AIC by -1.723 but increases the BIC by 3.566, *ceteris paribus*. So, adding an extra parameter increases model fit according to the AIC criterion, but decreases model fit according to the BIC criterion.

For main effects of alternative specifications, *number of parameters*, *RPL*, *CovariateP* and *Order* have positive and significant effects on LL. Non-linear specification, on the contrary, has a significant negative effect on model fit. This is probably due to little variation in residual income on the numeraire good (i.e., BILL is too small relative to income). The RPL specification improves model fit the most, increasing the LL by 106.02 on average, and decreases the AIC or BIC by twice as much.

This substantial improvement in model fit forecasts that only RPL models would be assigned non-zero weights. The RPL specification, compared to non-RPL specifications, only involves six more free parameters, but the improvement on model fit is significantly higher. It is also more efficient to improve model fit by letting covariates enter the preference function than the scale function (*CovariateP*). *Order* is also found to be significant when it is included in the preference function. Nesting structure does not matter much in this study, probably because we have included an SQ alternative specific constant in the preference function.

For interaction effects, the interactions between *RPL* and *HET*, *CovariateP* and *Nonlinear* are found to have significant effects on model fit. While the main effect of heteroscedastic scale is not significant, model fit improves when scale is allowed to be heteroscedastic in RPL models. Maybe heterogeneity in preferences is associated with larger variance in responses due to added sampling variation. The coefficient on the interaction between *RPL* and *CovariateP* is negative, which means that random parameters specifications can substitute partially for a with-covariates specification in explaining observed heterogeneity. The RPL specification improves model fit in non-linear in income models more than in linear models. Interactions between *NL* and *HET*, *CovariateP*, *Nonlinear* and *RPL* do not have significant effects on model fit. The coefficient for the interaction *NL*RPL* is not significant, probably because the RPL model alleviates the need for explicit specification about correlation structure since it allows for free correlation among error terms (unobserved utility) associated with each alternative.

The regressions generate some interesting insights. First, it suggests that the RPL specification is effective at improving model fit. The RPL model can capture unobserved preference heterogeneity as well as some observed preference heterogeneity that is explained by socio-demographic variables and the RPL model may alleviate the need for specifying a nesting structure explicitly. Second, when the RPL specification is employed, it is important to control for scale heteroscedasticity because heterogeneity in preferences might be associated with larger variance in choice responses. Third, in a non-linear in income model, the RPL specification contributes to model fit more than in a linear model. Fourth, there might be an order effect on preference elicitation when respondents are not familiar with the choice scenario. These findings, however, might be limited to this study only; one should be cautious in generalizing these results. For example, the fact that the nesting specification does not improve model fit may be due to the fact that the tree structure only exists for about half of the dataset and a status-quo alternative specific constant is included in the preference function.

2.5.2 Model Averaging and Classical Hypothesis Testing

Table 2.3 indicates only a handful of models have non-zero weights based on the AIC or BIC criterion (AICW or BICW). This is because analytically the weight of a model decreases substantially even for a small increase in the difference in AIC or BIC, and yet AICs and BICs of the estimated 64 models have a wide range. Model averaging is recommended by Layton and Lee (2006) as a way to reconcile WTP discrepancies from different models with similar levels of model fit when the "true" model is unknown. It can be inferred that, only models with similar statistical performance should be considered for model averaging. Presumably, model averaging should not replace hypothesis testing for model misspecification. A discriminatory analysis could be conducted to exclude "mis-specified" models from model averaging. The models "surviving" the discriminatory analysis will be more likely to have similar statistical performance to be considered for model averaging. AICs and BICs of these remaining models are more likely to have a narrower range so that model weights might be more balanced.

Table 2.5 reports results of the discriminatory analysis. Hypothesis tests are conducted based on the likelihood-ratio (LR) test, which is used to test if a model is nested by another. Since we built the hierarchical structure in specifying models, LR tests can be conducted at various levels. As mentioned earlier, the 64 models can be blocked into 8 groups based on the specification of scale heteroscadesticity and income effects except for the HET3 models. LR tests are first conducted within each block. If a null hypothesis about a simpler model is rejected, this model is excluded from model averaging analysis. Since a simple model can be nested by multiple models that are more general, multiple LR tests are conducted. A simple model is excluded from model averaging analysis if it is rejected once by any LR test.

Table 2.5 Columns 6 and 7 indicates models nesting a given model within each block and across blocks. See Appendix 2.3 for an explanation of nesting relationships. Pairwise LR tests are conducted within each block and across blocks and results are presented in Appendix 2.2 Tables A2.2.1-2.2.5. Any model rejected once by any pairwise LR test is reported as a "Reject". The last column of Table 2.5 therefore is a final result about whether a null hypothesis is either rejected from a within-block LR test or an across-block LR test. The remaining models are the ones that "survive" from both within-block and across-block LR tests or the ones that are on the top of a nesting tree. The last column of Table 2.5 indicates a total of 11 models either cannot be rejected by LR tests ("Not Reject") or cannot be tested ("-").

The last column of Table 2.5 indicates whether a model is rejected or not based on hypothesis testing. In the end, there are 11 models that cannot be rejected (or cannot be tested). AICs of these models range from 1998.78 to 2097.75 and BICs range from 2098.69 to 2206.06. Standard deviations of the AICs are BICs are 39.99 and 36.35 respectively, much smaller than those based on all 64 models (118.98 and 102.62 respectively). However, six of them are non-linear models, which have much higher AICs and BICs than the linear models. ANOVA tests indicate that the differences in AIC or BIC between the two groups are significant at the 1% level. The group mean of AICs of non-linear models is 2083.45 versus a group mean of 2010.03 of linear models. Table 2.6 reports the AIC and BIC weights recalculated based on 11 models. Based on the AIC criterion, all linear models have non-zero weights (at least 0.001%) although the top two models have a total weight over 99%. In contrast, based on the BIC criterion, one model has a weight of 100% (HET3_NRPL_Y_0). Essentially, the best model is chosen when using the BIC criterion. The AIC criterion picks more complex models than the BIC as expected. The top two models based on AIC are HET2_RPL_Y_1 and HET2_NRPL_Y_1. Both are RPL models and specified with covariates. The former has 23 parameters while the latter has 24 parameters. In contrast, HET3_NRPL_Y_0 picked by the BIC has only 16 parameters because it uses a no-covariates specification. However, all three models are heteroscedastic RPL linear-in-income models.

Table 2.6 also indicates both AIC and BIC rankings of the 11 models, compared to their corresponding rankings when all 64 models are considered for model averaging (Table 2.3). The top two models based on the AIC criterion are the same no matter whether model averaging is applied over the remaining models only (11 models) or all models (64 models). The two models, HET2_RPL_Y_1 and HET2_NRPL_Y_1 have a sum of weights over 90% in two types of ranking. The top model based on the BIC criterion of the remaining models, HET3_NRPL_Y_0, however, is ranked third based on the full ranking. The top two ranked models based on full ranking using 64 models are excluded from model averaging based on LR test results (HET3_RPL_Y_0 and HET2_RPL_Y_0). In summary, we find that AIC weights derived from a combined hypothesis testing and model averaging approach are similar to those derived from a direct model averaging over all models. However, classical hypothesis tests exclude two of the best models based on a full BIC ranking from subsequent model averaging analysis, resulting in substantially different model weights.

Therefore, a question arises, "which approach should we use – hypothesis testing and then model averaging over remaining models or direct model averaging over all models?" Before answering this question, it is informative to review general debates on model averaging versus classical hypothesis testing as a tool for model selection.[18] Supporters of model averaging suggest that classical hypothesis testing is arbitrary when it comes to choosing significance levels, have essentially no test power when sample size is large and cannot be used to provide a full ranking of different models (Weakliem 2004). Model averaging, on the other hand, depending on whether the AIC or BIC criterion is adopted, may select substantially different models. Burnham and Anderson (2004) provide a cautionary recommendation: the criterion to use depends on the purpose of the analysis. Therefore, we consider choosing between hypothesis testing, AIC model averaging or BIC model averaging rather than between hypothesis testing and model averaging, although the latter two are based on information theory. Based on results indicated by Table 2.3 and Table 2.6, we contribute to the debate as follows.

First, we believe that hypothesis testing should be conducted whenever it is appropriate. In this study, hypothesis testing is conducted based on a given dataset of a

[18] Sociological Methods & Research has a special issue on the debate (vol. 33, 2004).

moderate sample size of 1464 observations. The testing power is therefore not a function of sample size. We think it can be used to discriminate between model specifications. Since classical hypothesis testing has been routinely conducted by many applied econometricians, and nesting relationships exist among models in this study, it would be natural to use hypothesis testing to select models. Hypothesis testing alone, however, is not a satisfactory tool for model selection. In our case, LR tests cannot discriminate between models at the top of a nest tree. Non-nested tests like Vuong tests may be able to differentiate models based on statistical performance under some circumstances but provide no guidance on how to deal with different willingness-to-pay estimates derived from models with similar statistical performance. Among the 11 selected models (Table 2.6), six non-linear models are clearly inferior to other linear models based on the AIC or BIC criterion. They are also inferior to some linear models that are rejected by other LR tests. Of course, six models are assigned with zero weights subsequently when model averaging is applied.

Second, we advocate a combined approach of hypothesis testing and model averaging. Model averaging is proposed to deal with modeling uncertainty when "the true model is unknown" (Layton and Lee 2006). Implicitly, to select models with equal statistical performance is the first step. It is therefore logical to conduct hypothesis testing before applying model averaging. Note that model averaging, as proposed by Buckland et al. (1997), is a frequentist based approach to handle modelling uncertainty in estimating stated preference models. Hence, there is no philosophical inconsistency in the combined approach.

Third, for model averaging, we propose AIC model averaging because AIC's implied philosophy about models and model based inference is more consistent with the essence of model averaging. Model averaging recognizes that there is uncertainty in model selection. The AIC has a similar philosophy: models are only approximations to unknown truths but can be useful to understand the reality (Burnham and Anderson 2004). In contrast, both hypothesis testing and BIC assume the existence of a true model. Some studies report that model selection results from hypothesis testing are more consistent with those derived based on BIC than with AIC (Burnham and Anderson 2004). Our results, however, indicate the opposite. We believe our finding is not accidental given that both hypothesis testing and the AIC criterion tend to choose more complex models while the BIC criterion prefers simpler ones. Consistency in model selection between the three criteria might be an empirical issue. In addition, AIC model averaging generates more balanced model weights than BIC model averaging, which might be a preferred criterion when uncertainty in model selection increases.

Does it matter if one conducts a combined hypothesis testing and model averaging approach or a direct model averaging? It depends on which criterion is used in model averaging. Models selected by a combined approach are very similar to those selected based on AIC model averaging, but are different from those based on BIC model averaging. Given that the debate about AIC and BIC remains unsettled, it is customary to report both AIC and BIC model averaging results. Therefore, for the purpose of comparison, we will use a combined approach along with a model averaging approach in calculating weighted WTP estimates.

2.5.3 Preferred Models based on the AIC or BIC Criterion after Discriminatory Analysis

We now turn to a discussion of the preferred models. We choose to report on preferred models selected by the AIC or BIC criterion after discriminatory analysis. Table 2.7 reports two preferred models based on the AIC criterion and the best model based on the BIC criterion. The top two models based on AIC are HET2_RPL_Y_1 and HET2_NRPL_Y_1, with a model weight of 57.12% and 42.86% respectively. HET3_NRPL_Y_0 is assigned a weight of 100% based on the BIC criterion.

The estimated random parameters are listed first, with estimated mean effects followed by estimated standard deviations. These are *SQ, SQ_SD, MICI, MICI_SD, MICD, MICD_SD, CANI, CANI_SD, CAND, CAND_SD, SQCE3* and *SQCE3_SD*. All coefficients are highly significant and with consistent signs across three models. The coefficient for *SQ* is positive and large, indicating a status quo effect. Still, preferences for the status quo option are highly heterogeneous as indicated by large magnitudes of *SQ_SD*. The estimated mean effect of the interaction term between *SQ* and dataset dummy variable *CE3* is highly significant and negative. This indicates that a respondent is less likely to choose a status quo option when he or she is asked to choose between the status quo option and two other alternatives than to choose between the status quo option and another alternative. The framing effect also varies substantially across respondents.[19] Health risk attributes are found to have negative effects as expected and the effects also differ substantially across individuals. The magnitudes of estimated means and standard deviations are similar across the three models. Since all estimated standard deviations are highly significant, and most of them have a magnitude comparable to their corresponding means, it is not surprising that RPL models are preferred to models assuming fixed effects.

The estimated price effect (*BILL*) is assumed to be fixed to facilitate welfare calculation. It is negative and significant at the 1% significance level in all three models with very similar magnitude.

The estimated coefficient for nesting structure (*NEST_NONSQ_SD*) is estimated in HET2_NRPL_Y_1 and HET3_NRPL_Y_0 only, and it is found to be insignificant in both models. It is surprising that it is insignificant in HET3_NRPL_Y_0 given the fact that an LR test rejects the null that HET3_RPL_Y_0 is true against HET3_NRPL_Y_0 (that's why HET3_RPL_Y_0 is not included in model averaging). Perhaps specifying a nesting structure has affected how unobserved utilities are correlated. This might justify a model averaging approach as it reflects the fact that we cannot fully understand what drives model fit. Random parameters logit models, in particular, are more difficult to understand from a behavioural perspective because they rely on assumptions about statistical distributions and complicated correlations between random parameters make interpretation difficult.

The coefficients on interactions between *SQ* and socio-demographic variables are presented next. There are eight of them, and they are only estimated in the HET2_RPL_Y_1 and HET2_NRPL_Y_1 models. The estimates are similar across the two models. Age, city size, gender and marital status are found to have significant effects on choosing the status quo option. More specifically, individuals who are 65 years old or older (*AGE65*) are more likely to be willing to pay for a water treatment program that lowers microbial or cancer health risks from drinking water. This is also the case for

[19] This framing effect is investigated in Chapter 4.

individuals residing in smaller cities or communities (*CITYSIZE*) compared to those in larger cities. Men (*MALE*) are more likely to choose the status quo than women. A married individual (*MARRY*) is also more likely to choose a status quo option. Income effects on status quo are found to be insignificant.

The coefficient of the interaction between choice task order and SQ (*ORDER*SQ*) is only included in HET3_NRPL_Y_0. It is negative and highly significant. Individuals are more likely to choose a water treatment program that involves lower health risks and higher water bill than the status quo program in the first choice task than the other choice tasks.[20] This is another type of context effect worthy of further research.

Lastly, estimated coefficients for variables entering the scale function are reported: *SCALE_CE3* and *SCALE_ORDER*. Both can be considered as choice environment variables (see *CE3* and *ORDER* in Table 2.1). HET2_RPL_Y_1 and HET2_RPL_Y_1 estimate both coefficients while HET3_NRPL_Y_0 only estimates *SCALE_CE3* since *ORDER* enters the preference function through interactions with *SQ* (*ORDER*SQ*). When both coefficients are estimated, only the coefficient for *ORDER* is significant, and it is negative. Since the magnitude of *CE3* is very small compared to that of *ORDER*, the overall effect would be negative for *ORDER* = 1. When only *SCALE_CE3* is estimated, it is almost significant at the 10% level and it is also negative. Referring to Equation 2.6, the estimated μ is therefore less than one when *ORDER* = 1 or *CE3* = 1. This implies that there is a larger variance in responses when individuals are answering the first choice question or answering a 3-alternative choice question than answering the following choice questions or a 2-alternative choice question. It could be that learning occurs when individuals answer multiple choice questions. A 2-alternative choice question alternatively may be easier to answer than a 3-alternative question, or it is more difficult to model choice decisions involving three alternatives because more things are going on in decision-makers' minds (Adamowicz et al. 1998; Scarpa, Ferrini and Willis 2005).

All the estimated coefficients are similar with consistent signs and magnitudes across the models. However, many estimated coefficients in HET3_NRPL_Y_0 have slightly larger absolute values than the other two that use a with-covariates specification.

[20] In the water survey, individuals are asked to do four choice tasks in total.

It is possible that controlling for observed heterogeneity might improve efficiency of estimators. As some socio-demographic variables are able to explain choice decisions, including these variables enable us to better understand and predict choice decisions. Policy makers are also interested in how a policy affects different populations. A with-covariates specification may therefore be more useful from a policy making perspective. The variance in responses caused by choice environment variables also motivates us to further investigate effects of survey design on preference elicitation.

2.6 Willingness-to-Pay Estimation

We report four types of marginal WTP estimates for risk reductions in microbial illnesses, microbial deaths, cancer illnesses and cancer deaths. Due to space limitations, we only report on WTP estimates from models with non-zero weights. Table 2.8 reports WTP estimates from non-zero AIC weights based on all 64 models and based on the 11 remaining models after hypothesis testing as well as weighted WTP estimates are derived. Table 2.9 presents the BIC version of the results.

There are 12 models with non-zero AIC weights (at least 0.001%) based on all 64 models (Table 2.8). The four models that are specified with no covariates have only a total weight of less than 0.05%. The sum of weights of the top two models is about 92%. WTP estimates provided by the 12 models do not differ substantially.[21] Standard deviations in estimated WTP for risk reductions from microbial or cancer illnesses are especially small. WTP estimates for risk reduction from microbial or cancer deaths vary slightly across models, but still have standard deviations less than 10% of mean values. The AIC weighted WTP estimates are basically a weighted average of estimates provided by HET2_ RPL Y_1 and HET2_ NRPL Y_1. When model averaging is applied over the remaining models after hypothesis testing, there are 5 models with non-zero AIC weights. The estimates are very similar across models with even smaller standard deviations. The sum of weights of the top two models is over 99%, and the two top models are same as in the models examined from the full set of 64. AIC weighted WTP estimates are very

[21] Here, standard deviations reported in Tables 2.8 and 2.9 are called unconditional sampling variance (Burnham and Anderson 2004), which measure how estimates vary across different models.

similar to each other no matter whether hypothesis testing is used to discriminate model specification or not. Since we prefer a combined hypothesis testing and model averaging approach, we report AIC weighted WTP estimates derived from the 11 remaining models. Marginal WTP for drinking water risk reductions from microbial illnesses is 0.017, which means an individual/household is willing to pay 1.7 cents to reduce one case of microbial illnesses. Marginal WTP for risk reductions from cancer illness is $2.104, much larger than microbial illness reduction. An individual is willing to pay $15.858 to avoid one microbial death due to contaminated tap water quality in the community, which is one third larger than the amount they are willing to pay to avoid one cancer death ($10.588) caused by DBPs in tap water. It is interesting that people are willing to pay more to reduce one microbial death than to reduce one cancer death. This finding may be only limited to our context. Microbial contaminated water is considered to be more worrying than probably because of high profile and the acute occurrence of microbial diseases in Canada.

Table 2.9 presents four types of WTP estimates based on BIC model averaging in a similar fashion to Table 2.8. When model averaging is applied to all 64 models, eight of them have non-zero BIC weights. Only four of the eight also have non-zero AIC weights. WTP estimates provided by these models are also similar, with relatively smaller standard deviations for each type of WTP estimate. When applied over the 11 remaining models from the discriminative analysis, only one model has non-zero weights (HET3_NRPL_Y_0). Based on this model, the marginal WTP for drinking water risk reductions in microbial illnesses, microbial deaths, cancer illnesses and cancer deaths are $0.018, $15.773, $1.898 and $9.066 respectively. These estimates are similar to the AIC weighted estimates, although WTP estimates for cancer illnesses and cancer death are slightly smaller.

In this section, we report model weighted WTP estimates for four types of health risk reductions using a model averaging approach. Although WTP estimates are slightly different depending on whether we use the AIC or BIC criterion or whether we perform discriminatory analysis or not before implementing model averaging, the difference is minor. Individuals are willing to pay more to reduce mortality risks than morbidity risks, and are willing to pay significantly more to reduce microbial mortality risks than cancer mortality risks.

2.7 Conclusions

This paper is an empirical study of the valuation of Canadians' preferences for health risk reductions from drinking water using choice data collected from a stated preference survey. To obtain robust estimates for different types risk reductions, we adopt a model averaging approach to synthesize WTP estimates provided by different models. A total of 64 models are estimated that differ in specifications on scale heteroscedasticity, the error term structure, preference heterogeneity and income effect. Each level of the specifications is associated with important behavioural assumptions about choice decisions. Similar to designing an experiment, we built a hierarchical structure into the levels of alternative specifications to enable us to investigate how some of the most popular specifications affect model fit. Therefore, this paper also explores some methodological issues in modelling discrete choice data. Lastly, the paper provides some suggestions about using model averaging as a tool for model selection.

The WTP estimates indicate that Canadians are willing to pay positive amounts to reduce both microbial and cancer health risks from drinking water. On average, each household is willing to pay about $0.017 to reduce one microbial illness case from their tap water and about $2 to reduce one cancer illness caused by disinfection by-products (DBPs) in the tap water. Each household is willing to pay up to $15.8 to avoid one microbial death, and about $10 to avoid one cancer death caused by drinking water problems. Canadians are willing to pay sizable amounts to reduce mortality risks. In particular, they are willing to pay much higher amounts to reduce microbial mortality risks than cancer mortality risks. This indicates that it is important to use a water treatment technology that is efficient in treating microbial pathogens while controlling for levels of DBPs in drinking water.

The results of sensitivity analysis of model specifications on model fit indicate that the RPL specification is the most efficient way to improve model fit. The RPL specification can partially capture what is explained by a with-covariates specification or a nesting structure specification. For the RPL model, allowing for scale heteroscedasticity will further improve model fit. A non-linear in income specification fits models much worse than a linear in income specification.

In terms of model averaging, unlike Layton and Lee (2006), we prefer the AIC criterion to the BIC criterion because the philosophy about models underlying the AIC criterion is more consistent with model averaging in supporting a multi-model inference. AIC weights are also more balanced, which results in less "risky" estimates when model selection uncertainty increases. We also compare hypothesis testing and model averaging as tools for model selection. We recommend a combined hypothesis testing and model averaging approach. In particular, hypothesis testing and AIC model averaging is preferred because they are more consistent with each other in model selection.

In this study, despite a wide range of alternative specifications, WTP estimates are fairly robust. It might be not so clear what we gain by using model averaging. However, the mere fact that our WTP estimates are robust across a variety of specification is reassuring. Limited to space, we only focus on some specifications issues while holding other specifications constant. Future research could be conducted to examine other specification issues or to continue to examine similar specification issues with different datasets to see whether our results about the sensitivity of model fit to model specification can be generalized to other studies.

References

Adamowicz, W. L., P. Boxall, M. Williams and J. Louviere. 1998. Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics* 80(1): 64-75.

Adamowicz, W. L., J. Louviere and M. Williams. 1994. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management* 26: 271-292.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory.* (eds.) B. N. Petrov and F. Csaki, 267-281. Akademiai Kiado: Budapest.

Allenby, G. M. and J. L. Ginter. 1995. The effects of in-store displays and feature advertising on consideration sets. *International Journal of Research in Marketing* 12: 67-80.

Andreoni, J. 1995. Cooperation in public goods experiments: kindness or confusion? *American Economic Review* 85(4): 891-904.

Ben-Akiva, M. and S. Lerman. 1985. *Discrete Choice Analysis - Theory and Application to Travel Demand.* MIT Press: Cambridge, Mass.

Bhat, C. R. 1995. A heteroscedastic extreme value model of intercity travel model choice. *Transportation Research Part B* 29(6): 471-483.

Buckland, S. T., K. P. Burnham and N. H. Augustin. 1997. Model selection: an integral part of inference. *Biometrics* 53(2): 603-618.

Burnham, K. P. and D. R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research* 33: 261-303.

Cameron, T. A., G. L. Poe, R. G. Ethier and W. D. Schulze. 2002. Alternative non-market value-elicitation methods: are the underlying preferences the same? *Journal of Environmental Economics and Management* 44: 391-425.

DeShazo, J. R. and G. Fermo. 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44: 123-143.

Haab, T. and K. E. McConnell. 2002. *Valuing Environmental and Natural Resources: the Econometrics of Non-Market Valuation.* Edward Elgar Publishing: Cheltenham, UK.

Hensher, D. 2001. The sensitivity of the valuation of travel time savings to the specification of unobserved effects. *Transportation Research Part E* 37: 129-142.

Hensher, D., J. Louviere and J. Swait. 1999. Combining sources of preference data. *Journal of Econometrics* 89: 197-221.

Herriges, J. A. and C. L. Kling. 1999. Nonlinear income effects in random utility models. *The Review of Economics and Statistics* 81(1): 62-72.

Huang, Ju-Chin and V. K. Smith. 1998. Monte Carlo benchmarks for discrete response valuation methods. *Land Economics* 74(2): 186-202.

Kling, C. 1987. Comparing welfare estimates of environmental quality changes from recreation demand model. *Journal of Environmental Economics and Management* 15: 331-340.

Layton, D. F. and S. T. Lee. 2006. Embracing model uncertainty: strategies for response pooling and model averaging. *Environmental and Resource Economics* 34(1): 51-85.

Louviere, J. J. 2006. What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments. *Environmental and Resource Economics* 34(1): 173-188.

Mazzanti, M. 2003. Discrete choice models and valuation experiments: an application to cultural heritage. *Journal of Economic Studies* 30(6):584-604.

Mazzotta, M. and J. Opaluch. 1995. Decision making when choices are complex: a test of Heiner's hypothesis. *Land Economics* 71: 500-515.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behaviour. *Frontiers in Econometrics*. (ed.) P. Zarembka, 105-142. Academic Press: New York.

Meijer, E. and J. Rouwendal. 2006. Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics* 21(2): 227-244.

Mitchell, R. C. and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Resources for the Future. The Johns Hophkins University Press: USA.

Morey, E. R., R. D. Rowe and M. Watson. 1993. A repeated nested-logit model of Atlantic Salmon fishing. *American Journal of Agricultural Economics* 75(3): 578-592.

Revelt, D. and K. E. Train. 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *The Review of Economics and Statistics* 80(4): 647-657.

Scarpa, R., S. Ferrini and K. G. Willis. 2005. Performance of error component models for status-quo effects in choice experiments. In: *Applications of Simulation Methods in Environmental and Resource Economics* (eds.) R. Scarpa and A. Alberini, 247-

274. Springer Publisher: Dordrecht, The Netherlands.

Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6: 461-464.

Swait, J. 2005. Advanced choice models. *Valuing Environmental Amenities using Choice Experiments: A Common Sense Guide to Theory and Practice,* (ed.) B. Kanninen, 229-293. Springer Publisher: Dordrecht, The Netherlands.

Swait, J. and W. L. Adamowicz. 2001. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Process* 86(2): 141-167.

Swait, J. and J. Louviere. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* XXX: 305-314.

Train, K. E. 1998. Recreation demand models with taste differences over people. *Land Economics* 74(2): 230-239.

Train, K. E. and M. Weeks. 2005. Discrete choice models in preference space and willingness-to-pay space. In: *Applications of Simulation Methods in Environmental Resource Economics* (eds.) R. Scarpa and A. Alberini, 1-16. Springer Publisher: Dordrecht, The Netherlands.

Viscusi, W. K. 1993. The value of risks to life and health. *Journal of Economic Literature* 31(4): 1912-1946.

Weakliem, D. L. 2004. Introduction to the special issue on model selection. *Sociological Methods & Research* 33: 167-187.

Table 2.1 Definition of Variables

Variable	Definition	Level / Mean / Percentage
	Attributes and choice environmental variables	
SQ	1 if an alternative is the status quo option and 0 otherwise.	-
MICI	Number of microbial illness cases over a 35-year period from drinking tap water in the community.	7500, 15000, 23000[a], 30000
MICD	Number of deaths due to microbial illnesses over a 35-year period from drinking tap water in the community.	5, 10, 15[a], 20
CANI	Number of cancer cases over a 35-year period from drinking tap water in the community.	50, 75, 100[a], 125
CAND	Number of cancer deaths over a 35-year period from drinking tap water in the community.	10, 15, 20[a], 25
BILL	Annual increase in the current water bill in 2004 Canadian dollars.	0[a], 25, 75, 125, 150, 250, 350
CE3	1 if an individual is faced with a choice set of 3 alternatives, and 0 if faced with a choice set of 2 alternatives.	50.55%
SQCE3	Interaction between *SQ* and *CE3*.	-
NEST_ NONSQ	1 if an alternative is a non-status quo option and 0 otherwise.	-
ORDER	1 if a choice task faced by an individual is the first choice task out of a total of four choice tasks, and 0 for other three choice tasks.	-
	Socio-demographic variables	
AGE65	Dummy variable, equals 1 if an individual is equal to or over 65 years old and 0 otherwise.	12.57%
INCOME	Annual household income	59029.59
INCOME2	Squared household income.	4.80E+10
ENGLISH	1 if English is the corresponding language and 0 otherwise.	74.86%
CITYSIZE	Categorical variables from 1 to 6, ranging from 1 denoting 1,000,000 plus to 6 denoting 1499 and under.	2.56
ILLNESS	1 if an individual has ever being ill due to drinking tap water, 0 otherwise	3.01%
MALE	1 if male, and 0 otherwise.	53.55%
MARRY	1 if an individual is married and 0 otherwise.	48.90%
KID06	1 if a household has kid(s) under 6, and 0 otherwise.	5.46%
KID612	1 if a household has kid(s) aged between 6 and 12, and 0 otherwise.	7.38%
KID137	1 if a household has kid(s) aged between 13 and 17, and 0 otherwise.	6.28%

Notes: [a] indicates the status quo level of attributes. There are a total of 1464 observations.

Table 2.2 Hierarchical Model Specifications - Four Levels

Levels	Model Specifications (Abbreviations)	Comments
<u>Level 1</u>		
Scale hetero-scedasticity	Homoscedastic logit model (HLM)	$\mu^{CE2} = \mu^{CE3} = 1$
	Heteroscedastic logit model-type I (HET1)	$\mu^{i} = f(CE3)$, i = CE2, CE3
	Heteroscedastic logit model-type II (HET2)	$\mu^{it} = f(CE3, ORDER_t)$, i = CE2, CE3
	Heteroscedastic logit model-type III (HET3)	Depend on the specification about covariates. A no-covariates specification V_{njt}_HET3_0= $\mu^{i}(V_{nj} + SQ_j * ORDER_t)$, $\mu^{i} = f(CE3)$; A with-covariates specification, V_{njt}_HET3_1= $\mu^{nit} V_{nj}$, $\mu^{nit} = f(CE3, ORDER_t, \mathbf{S_n})$
<u>Level 2</u>		
Error Structure/ unobserved preference heterogeneity	Conditional Logit model (CL)	Extreme type I distributed value error terms
	Nested Logit model (NL)	Generalized extreme type I value distributed error terms
	Random Parameters Logit model (RPL)	Extreme type I value distributed error terms and individual specific parameters
	Correlated Random parameters Logit model (NRPL)	Generalized extreme type I value distributed error terms and individual specific parameters
<u>Level 3</u>		
Income effects	Linear income effect (Y)	$INCOME_n$-$BILL_j$
	Log-linear income effect (LNY)	$\ln((INCOME_n - BILL_j)/INCOME_n)$
<u>Level 4</u>		
Observed heterogeneity	Without covariates (0)	$V_{nj} = f(\boldsymbol{MAIN_j})$
	With covariates (1)	For linear income effect model, eight covariates are included, for non-linear income effect, six covariates are included.

Notes: **MAIN** is a vector of attributes that are included in every specification, it includes *SQ, MICI, MICD, CANI, CAND, BILL* (or LNY for non-linear specification), and *SQCE3.* For random parameters models, all except for *BILL* (or LNY*)* are assumed to be normal distributed. *ORDER* is a dummy variable that indicates whether a choice task is the first choice task one encounters.

Table 2.3 Model Fits and Weights for All Estimated Models

#	Model Names	# para.	Log-likelihood	Psuedo R^2	AIC	BIC	AICW (%)	BICW (%)
1	HLM_CL_Y_0	7	-1121.28	0.12	2256.56	2293.58	0.000	0.000
2	HLM_NL_Y_0	8	-1121.27	0.12	2258.55	2300.86	0.000	0.000
3	HLM_RPL_Y_0	13	-1003.45	0.21	2032.91	2101.67	0.000	0.552
4	HLM_NRPL_Y_0	14	-1000.95	0.21	2029.90	2103.94	0.000	0.057
5	HLM_CL_Y_1	15	-1087.09	0.14	2204.18	2283.51	0.000	0.000
6	HLM_NL_Y_1	16	-1086.91	0.14	2205.83	2290.45	0.000	0.000
7	HLM_RPL_Y_1	21	-983.48	0.22	2008.96	2120.03	0.324	0.000
8	HLM_NRPL_Y_1	22	-982.69	0.23	2009.37	2125.73	0.264	0.000
9	HLM_ CL_LNY_0	7	-1160.19	0.09	2334.38	2371.40	0.000	0.000
10	HLM_NL_LNY_0	8	-1160.19	0.09	2336.39	2378.70	0.000	0.000
11	HLM_RPL_LNY_0	13	-1040.12	0.18	2106.23	2174.99	0.000	0.000
12	HLM_NRPL_LNY_0	14	-1040.73	0.18	2109.45	2183.50	0.000	0.000
13	HLM_CL_LNY_1	13	-1129.12	0.11	2284.25	2353.00	0.000	0.000
14	HLM_NL_LNY_1	14	-1129.12	0.11	2286.24	2360.29	0.000	0.000
15	HLM_RPL_LNY_1	19	-1024.51	0.19	2087.02	2187.51	0.000	0.000
16	HLM_NRPL_LNY_1	20	-1023.34	0.19	2086.68	2192.46	0.000	0.000
17	HET1_CL_Y_0	8	-1118.71	0.12	2253.42	2295.73	0.000	0.000
18	HET1_NL_Y_0	9	-1118.04	0.12	2254.08	2301.68	0.000	0.000
19	HET1_RPL_Y_0	14	-998.75	0.21	2025.51	2099.55	0.000	4.566
20	HET1_NRPL_Y_0	15	-995.94	0.21	2021.88	2101.21	0.000	0.867
21	HET1_CL_Y_1	16	-1086.06	0.14	2204.12	2288.75	0.000	0.000
22	HET1_NL_Y_1	17	-1086.06	0.14	2206.11	2296.02	0.000	0.000
23	HET1_RPL_Y_1	22	-980.24	0.23	2004.48	2120.83	3.053	0.000
24	HET1_NRPL_Y_1	23	-978.93	0.23	2003.86	2125.50	4.163	0.000
25	HET1_ CL_LNY_0	8	-1159.54	0.09	2335.08	2377.39	0.000	0.000
26	HET1_NL_LNY_0	9	-1159.27	0.09	2336.54	2384.14	0.000	0.000
27	HET1_RPL_LNY_0	14	-1034.42	0.18	2096.83	2170.88	0.000	0.000
28	HET1_NRPL_LNY_0	15	-1034.73	0.18	2099.46	2178.79	0.000	0.000
29	HET1_CL_LNY_1	14	-1129.09	0.11	2286.18	2360.23	0.000	0.000
30	HET1_NL_LNY_1	15	-1129.08	0.11	2288.16	2367.49	0.000	0.000
31	HET1_RPL_LNY_1	20	-1017.64	0.20	2075.29	2181.07	0.000	0.000
32	HET1_NRPL_LNY_1	21	-1017.16	0.20	2076.31	2187.38	0.000	0.000

Note: "_0" denotes a no-covariates specification and "_1" denotes a with-covariates specification. AIC is an abbreviation for Akaike's Information Criterion and BIC is an abbreviation for the Bayesian Information Criterion. This table is continued on the next page.

Table 2.3 Model Fits and Weights for All Estimated Models (Continued)

#	Model Names	# para.	Log-Likelihood	Pseudo R^2	AIC	BIC	AICW (%)	BICW (%)
33	HET2_CL_Y_0	9	-1115.72	0.12	2249.45	2297.05	0.000	0.00
34	HET2_NL_Y_0	10	-1114.71	0.12	2249.43	2302.32	0.000	0.00
35	HET2_RPL_Y_0	15	-994.24	0.22	2018.48	2097.81	0.003	26.04
36	HET2_NRPL_Y_0	16	-991.14	0.22	2014.28	2098.90	0.023	8.78
37	HET2_CL_Y_1	17	-1083.27	0.15	2200.53	2290.44	0.000	0.00
38	HET2_NL_Y_1	18	-1083.23	0.15	2202.46	2297.66	0.000	0.00
39	HET2_RPL_Y_1	23	-976.39	0.23	**1998.78**	2120.43	52.635	0.00
40	HET2_NRPL_Y_1	24	**-975.68**	**0.23**	1999.36	2126.29	39.502	0.00
41	HET2_ CL_LNY_0	9	-1154.93	0.09	2327.86	2375.46	0.000	0.00
42	HET2_NL_LNY_0	10	-1154.55	0.09	2329.10	2381.99	0.000	0.00
43	HET2_RPL_LNY_0	15	-1028.28	0.19	2086.55	2165.89	0.000	0.00
44	HET2_NRPL_LNY_0	16	-1028.36	0.19	2088.72	2173.34	0.000	0.00
45	HET2_CL_LNY_1	15	-1124.74	0.11	2279.48	2358.82	0.000	0.00
46	HET2_NL_LNY_1	16	-1124.74	0.11	2281.48	2366.10	0.000	0.00
47	HET2_RPL_LNY_1	21	-1012.61	0.20	2067.22	2178.29	0.000	0.00
48	HET2_NRPL_LNY_1	22	-1011.9	0.20	2067.80	2184.15	0.000	0.00
49	HET3_CL_Y_0	9	-1115.52	0.12	2249.03	2296.63	0.000	0.00
50	HET3_NL_Y_0	10	-1114.47	0.12	2248.94	2301.83	0.000	0.00
51	HET3_RPL_Y_0	15	-993.93	0.22	2017.86	**2097.19**	0.004	48.30
52	HET3_NRPL_Y_0	16	-991.03	0.22	2014.07	2098.69	0.025	10.84
53	HET3_CL_Y_1	17	-1106.09	0.13	2246.17	2336.08	0.000	0.00
54	HET3_NL_Y_1	18	-1106.09	0.13	2248.17	2343.37	0.000	0.00
55	HET3_RPL_Y_1	23	-986.12	0.22	2018.25	2139.89	0.003	0.00
56	HET3_NRPL_Y_1	24	-985.85	0.22	2019.70	2146.63	0.002	0.00
57	HET3_ CL_LNY_0	9	-1154.70	0.09	2327.41	2375.01	0.00	0.00
58	HET3_NL_LNY_0	10	-1154.28	0.09	2328.56	2381.45	0.00	0.00
59	HET3_RPL_LNY_0	15	-1027.82	0.19	2085.63	2164.97	0.00	0.00
60	HET3_NRPL_LNY_0	16	-1027.37	0.19	2086.74	2171.36	0.00	0.00
61	HET3_CL_LNY_1	15	-1139.07	0.10	2308.13	2387.47	0.00	0.00
62	HET3_NL_LNY_1	16	-1138.87	0.10	2309.74	2394.36	0.00	0.00
63	HET3_RPL_LNY_1	21	-1026.79	0.19	2095.58	2206.65	0.00	0.00
64	HET3_NRPL_LNY_1	22	-1026.87	0.19	2097.75	2214.10	0.00	0.00

Note: "_0" denotes a no-covariates specification and "_1" denotes a with-covariates specification. AIC is an abbreviation for Akaike's Information Criterion and BIC is an abbreviation for the Bayesian Information Criterion.

Table 2.4 Sensitivity Analysis of Model Specifications and Model Fits

Specification Variable	Log-Likelihood	Akaike's Information Criterion (AIC)	The Bayesian Information Criterion (BIC)
Intercept	-1134.334** (-286.526)	2268.668** (286.526)	2268.668 (286.526)
Number of parameters	1.862** (3.749)	-1.723* (-1.735)	3.566** (3.590)
HET	0.644 (0.443)	-1.289 (-0.443)	-1.289 (-0.443)
RPL	106.017** (31.265)	-212.034** (-31.265)	-212.034** (-31.265)
NL	-1.142 (-0.674)	2.284 (0.674)	2.284 (0.674)
Nonlinear	-37.500** (-32.203)	75.001** (32.203)	75.001** (32.203)
CovariateP	18.148** (4.961)	-36.296** (-4.961)	-36.296** (-4.961)
CovariateS	-3.146 (-0.815)	6.292 (0.815)	6.292 (0.815)
Order	1.963* (1.864)	-3.926* (-1.864)	-3.926* (-1.864)
*RPL*HET*	3.668** (2.583)	-7.336** (-2.583)	-7.336** (-2.583)
*RPL*CovariateP*	-13.212** (-10.400)	26.423** (10.400)	26.423** (10.400)
*RPL*Nonlinear*	2.072* (1.704)	-4.145* (-1.704)	-4.145* (-1.704)
*NL*HET*	0.130 (0.092)	-0.261 (-0.092)	-0.261 (-0.092)
*NL*CovariateP*	-0.278 (-0.219)	0.557 (0.219)	0.557 (0.219)
*NL*Nonlinear*	-0.897 (-0.738)	1.795 (0.738)	1.795 (0.738)
*NL*RPL*	0.744 (0.612)	-1.489 (-0.612)	-1.489 (-0.612)
Number of observations	64	64	64

Notes: *CovariateP* and *CovariateS* are the same set of demographic information (they are eight variables for linear specifications and six for non-linear specifications). *CovariateP* enters the preference function, *CovariateS* enters the scale function. *Order* is 1 for the first choice task each respondent answered. t-ratios are in parentheses. ** denotes the 5% level, and * denotes the 10% level.

Table 2.5 Discriminatory Analysis for Model Selection – Likelihood-Ratio Tests

Block	#	Model Names	# para.	Log-Likelihood	Nested by Models within Block	Nested by Models across Block	LR Test Results
1	1	HLM_CL_Y_0	7	-1121.28	2,3,4,5, 6,7, 8	17, 33,49	Reject
1	2	HLM_NL_Y_0	8	-1121.27	4,6,8	18,34,50	Reject
1	3	HLM_RPL_Y_0	13	-1003.45	4,7,8	19,35,51	Reject
1	4	HLM_NRPL_Y_0	14	-1000.95	8	20,36,52	Reject
1	5	HLM_CL_Y_1	15	-1087.09	6,7,8	21,37	Reject
1	6	HLM_NL_Y_1	16	-1086.91	8	22,38	Reject
1	7	HLM_RPL_Y_1	21	-983.48	8	23,39	Reject
1	8	HLM_NRPL_Y_1	22	-982.69	n.a.	24,40	Reject
2	9	HLM_ CL_LNY_0	7	-1160.19	10,11,12,13,14,15,16	25,41,57	Reject
2	10	HLM_NL_LNY_0	8	-1160.19	12,14,16	26,42,58	Reject
2	11	HLM_RPL_LNY_0	13	-1040.12	12,15,16	27,43,59	Reject
2	12	HLM_NRPL_LNY_0	14	-1040.73	16	28,44	Reject
2	13	HLM_CL_LNY_1	13	-1129.12	14,15,16	29,45	Reject
2	14	HLM_NL_LNY_1	14	-1129.12	16	30,46	Reject
2	15	HLM_RPL_LNY_1	19	-1024.51	16	31,47	Reject
2	16	HLM_NRPL_LNY_1	20	-1023.34	n.a.	32,48	Reject
3	17	HET1_CL_Y_0	8	-1118.71	18,19,20,21,22,23,24	33	Reject
3	18	HET1_NL_Y_0	9	-1118.04	20,22,14	34	Reject
3	19	HET1_RPL_Y_0	14	-998.75	20,23,24	35	Reject
3	20	HET1_NRPL_Y_0	15	-995.94	24	36	Reject
3	21	HET1_CL_Y_1	16	-1086.06	22,23,24	37	Reject
3	22	HET1_NL_Y_1	17	-1086.06	24	38	Reject
3	23	HET1_RPL_Y_1	22	-980.24	24	39	Reject
3	24	HET1_NRPL_Y_1	23	-978.93	n.a.	40	Reject
4	25	HET1_ CL_LNY_0	8	-1159.54	26,27,28,29,30,31,32	41	Reject
4	26	HET1_NL_LNY_0	9	-1159.27	28,30,32	42	Reject
4	27	HET1_RPL_LNY_0	14	-1034.42	28,31,32	43	Reject
4	28	HET1_NRPL_LNY_0	15	-1034.73	32	44	Reject
4	29	HET1_CL_LNY_1	14	-1129.09	30,31,32	45	Reject
4	30	HET1_NL_LNY_1	15	-1129.08	32	46	Reject
4	31	HET1_RPL_LNY_1	20	-1017.64	32	47	Reject
4	32	HET1_NRPL_LNY_1	21	-1017.16	n.a.	48	Reject

Notes: "n.a." denotes a model is not nested by other models within a block. This table is continued on the next page.

Table 2.5 Discriminatory Analysis for Model Selection – Likelihood-Ratio Tests (Continued)

Block	#	Model Names	# para.	Log-Likelihood	Nested by Models within Block	Nested by Models across Block	LR Test Results
5	33	HET2_ CL_Y_0	9	-1115.72	34,35,36,37,38,39,40	53	Reject
5	34	HET2_NL_Y_0	10	-1114.71	36,38,40	54	Reject
5	35	HET2_RPL_Y_0	15	-994.24	36,39,40	55	Reject
5	36	HET2_NRPL_Y_0	16	-991.14	40	56	Reject
5	37	HET2_CL_Y_1	17	-1083.27	38,39,40		Reject
5	38	HET2_NL_Y_1	18	-1083.23	40		Reject
5	39	HET2_RPL_Y_1	23	-976.39	40		Not Reject
5	40	HET2_NRPL_Y_1	24	-975.68	n.a.		-
6	41	HET2_ CL_LNY_0	9	-1154.93	42,43,44,45,46,47,48	61	Reject
6	42	HET2_NL_LNY_0	10	-1154.55	44,46,48	62	Reject
6	43	HET2_RPL_LNY_0	15	-1028.28	44,47,48	63	Reject
6	44	HET2_NRPL_LNY_0	16	-1028.36	48	64	Reject
6	45	HET2_CL_LNY_1	15	-1124.74	46,47,48		Reject
6	46	HET2_NL_LNY_1	16	-1124.74	48		Reject
6	47	HET2_RPL_LNY_1	21	-1012.61	48		Not Reject
6	48	HET2_NRPL_LNY_1	22	-1011.90	n.a.		-
7	49	HET3_CL_Y_0	9	-1115.52	50,51,52		Reject
7	50	HET3_NL_Y_0	10	-1114.47	56		Reject
7	51	HET3_RPL_Y_0	15	-993.93	56		Reject
7	52	HET3_NRPL_Y_0	16	-991.03	n.a.		-
7	53	HET3_CL_Y_1	17	-1106.09	54,55,56		Reject
7	54	HET3_NL_Y_1	18	-1106.09	56		Reject
7	55	HET3_RPL_Y_1	23	-986.13	56		Not Reject
7	56	HET3_NRPL_Y_1	24	-985.85	n.a.		-
8	57	HET3_ CL_LNY_0	9	-1154.70	58,59,60		Reject
8	58	HET3_NL_LNY_0	10	-1154.28	60		Reject
8	59	HET3_RPL_LNY_0	15	-1027.82	60		Not Reject
8	60	HET3_NRPL_LNY_0	16	-1027.37	n.a.		-
8	61	HET3_CL_LNY_1	15	-1139.07	62,63,64		Reject
8	62	HET3_NL_LNY_1	16	-1138.87	64		Reject
8	63	HET3_RPL_LNY_1	21	-1026.79	64		Not Reject
8	64	HET3_NRPL_LNY_1	22	-1026.87	n.a.		-

Notes: "n.a." denotes a model is not nested by other models within a block. "-" denotes an LR test cannot be performed.

Table 2.6 Model Fits and Weights for Remaining Models after Likelihood-Ratio Tests

#	Model Names	AICW (%)	BICW (%)	AIC Ranking		BIC Ranking	
				Remaining Models	All Models	Remaining Models	All Models
39	HET2_RPL_Y_1	57.108	0.000	**1**	**1**	2	10
40	HET2_NRPL_Y_1	42.859	0.000	**2**	**2**	3	14
47	HET2_RPL_ LNY_1	0.000	0.000	6	17	8	23
48	HET2_NRPL_LNY_1	0.000	0.000	7	18	9	27
52	HET3_NRPL_Y_0	0.027	100	3	7	1	3
55	HET3_RPL_Y_1	0.003	0.000	4	10	4	15
56	HET3_NRPL_Y_1	0.002	0.000	5	12	5	16
59	HET3_RPL_ LNY_0	0.000	0.000	8	21	6	17
60	HET3_NRPL_LNY_0	0.000	0.000	9	24	7	20
63	HET3_RPL_LNY_1	0.000	0.000	10	27	10	31
64	HET3_NRPL_LNY_1	0.000	0.000	11	29	11	32

Notes: AICW and BICW are model weights that are calculated based on the AIC and BIC criteria, respectively. Ranks are calculated based on the AIC or BIC value. The top model (i.e., ranked as the first) has the lowest AIC or BIC.

Table 2.7 Three Preferred Models based on the AIC or BIC Criterion

Variable	AIC Preferred Model HET2_ RPL_Y_1	AIC Preferred Model HET2_NRPL_Y_1	BIC Preferred Model HET3_NRPL_Y_0
SQ	2.798** (2.956)	2.570** (2.847)	2.338** (5.201)
SQ_SD	2.063** (4.796)	1.955** (3.898)	2.211** (5.524)
MICI	-1.60E-04** (-5.382)	-1.51E-04** (-5.358)	-1.82E-04** (-5.891)
MICI_SD	1.27E-04** (4.103)	1.23E-04** (4.039)	1.31E-04** (4.037)
MICD	-0.150** (-4.967)	-0.143** (-4.932)	-0.163** (-5.085)
MICD_SD	0.118** (2.779)	0.117** (2.886)	0.144** (2.876)
CANI	-0.020** (-3.959)	-0.018** (-3.844)	-0.020** (-3.766)
CANI_SD	0.031** (4.155)	0.030** (4.264)	0.029** (3.833)
CAND	-0.100** (-4.097)	-0.094** (-4.117)	-0.093** (-3.898)
CAND_SD	0.122** (3.406)	0.120** (3.561)	0.121** (3.204)
SQCE3	-1.609** (-3.757)	-1.636** (-3.954)	-1.419** (-3.029)
SQCE3_SD	2.158** (3.482)	1.678** (2.871)	3.258** (4.073)
BILL	-0.009** (-5.614)	-0.009** (-5.647)	-0.010** (-6.089)
NEST_NONSQ_SD	-	0.838 (1.250)	-0.115 (-0.144)

Notes: t-ratios are in parentheses. ** denotes the 5% level, and * denotes the 10% level. AIC is an abbreviation for Akaike's Information Criterion and BIC is an abbreviation for the Bayesian Information Criterion. This table is continued on the next page.

Table 2.7 Three Preferred Models based on the AIC or BIC Criterion (Continued)

Variable	AIC Preferred Model HET2_ RPL_Y_1	AIC Preferred Model HET2_NRPL_Y_1	BIC Preferred Model HET3_NRPL_Y_0
*AGE65*SQ*	-1.075* (-1.952)	-0.988* (-1.831)	-
*INCOME*SQ*	1.56E-05 (0.842)	1.59E-05 (0.876)	-
*INCOME2*SQ*	-1.79E-10 (-1.543)	-1.76E-10 (-1.536)	-
*ENGLISH*SQ*	-0.112 (-0.265)	-0.108 (-0.265)	-
*CITYSIZE*SQ*	-0.492** (-3.102)	-0.447** (-2.996)	-
*ILLNESS*SQ*	-0.465 (-0.432)	-0.533 (-0.509)	-
*MALE*SQ*	1.036** (2.592)	1.024** (2.676)	-
*MARRY*SQ*	0.776* (1.795)	0.801* (1.900)	-
*ORDER*SQ*	-	-	-0.9047** (-2.759)
SCALE_CE3	0.047 (0.240)	0.160 (0.811)	-0.320 (-1.665)
SCALE_ORDER	-0.564** (-2.714)	-0.555** (-2.591)	-
Number of Obs.	1464	1464	1464
Log-likelihood	-976.39	-975.68	-991.03

Notes: t-ratios are in parentheses. ** denotes the 5% level, and * denotes the 10% level. AIC is an abbreviation for Akaike's Information Criterion and BIC is an abbreviation for the Bayesian Information Criterion.

Table 2.8 Marginal Willingness-to-Pay Estimates from Models with Non-zero AIC Weight

#	Model Names	AICW (%)	MICI	MICD	CANI	CAND
			Weights based on 64 models			
7	HLM_RPL_ Y_1	0.324	0.016	13.887	2.354	9.542
8	HLM_NRPL_Y_1	0.264	0.016	13.860	2.360	9.434
23	HET1_RPL_Y_1	3.053	0.016	14.473	2.205	10.375
24	HET1_NRPL_Y_1	4.163	0.016	14.576	2.133	10.347
35	HET2_RPL_Y_0	0.003	0.017	15.468	2.142	10.135
36	HET2_NRPL_Y_0	0.023	0.017	16.017	2.152	10.767
39	HET2_RPL_Y_1	52.635	0.017	15.897	2.126	10.606
40	HET2_NRPL_Y_1	39.502	0.017	16.078	2.074	10.624
51	HET3_RPL_Y_0	0.004	0.017	14.819	1.870	8.373
52	HET3_NRPL_Y_0	0.025	0.018	15.773	1.898	9.066
55	HET3_RPL_Y_1	0.003	0.017	15.588	2.061	9.951
56	HET3_NRPL_Y_1	0.002	0.017	15.588	2.061	9.951
	Mean		0.017	15.169	2.120	9.931
	Standard deviation		0.000	0.809	0.148	0.715
	AIC weighted		**0.017**	**15.858**	**2.109**	**10.588**
			Weights base on 11 remaining models			
39	HET2_RPL_Y_1	57.108	0.017	15.897	2.126	10.606
40	HET2_NRPL_Y_1	42.859	0.017	16.078	2.074	10.624
52	HET3_NRPL_Y_0	0.027	0.018	15.773	1.898	9.066
55	HET3_RPL_Y_1	0.003	0.017	15.588	2.061	9.951
56	HET3_NRPL_Y_1	0.002	0.017	15.588	2.061	9.951
	Mean		0.017	15.785	2.044	10.039
	Standard deviation		0.000	0.210	0.086	0.638
	AIC weighted		**0.017**	**15.975**	**2.104**	**10.613**

Notes: AIC is an abbreviation for Akaike's Information Criterion. AICW denotes model weights that are calculated based on the AIC criterion.

Table 2.9 Marginal Willingness-to-Pay Estimates from Models with Non-zero BIC Weight

#	Model Names	BICW (%)	MICI	MICD	CANI	CAND
			Weights based on 64 models			
3	HLM_RPL_Y_0	0.552	0.017	13.719	2.324	9.586
4	HLM_NRPL_Y_0	0.057	0.016	13.878	2.344	9.471
19	HET1_RPL_Y_0	4.566	0.016	13.862	2.236	9.807
20	HET1_NRPL_Y_0	0.867	0.017	14.480	2.235	10.432
35	HET2_RPL_Y_0	26.041	0.017	15.468	2.142	10.135
36	HET2_NRPL_Y_0	8.778	0.017	16.017	2.152	10.767
51	HET3_RPL_Y_0	48.302	0.017	14.819	1.870	8.373
52	HET3_NRPL_Y_0	10.838	0.018	15.773	1.898	9.066
	Mean		0.017	14.752	2.150	9.705
	Standard deviation		0.000	0.915	0.179	0.766
	BIC weighted		**0.017**	**15.143**	**1.991**	**9.207**
			Weights based on 11 remaining models			
52	HET3_NRPL_Y_0	100	0.018	15.773	1.898	9.066
	Mean		0.018	15.773	1.898	9.066
	Standard deviation		0.000	0.000	0.000	0.000
	BIC weighted		**0.018**	**15.773**	**1.898**	**9.066**

Notes: BIC is an abbreviation for the Bayesian Information Criterion. BICW denotes model weights that are calculated based on the BIC criterion.

Appendix 2.1 Variants of Random Utility Models

Table A2.1 Variants of Random Utility Models

Random Utility Models	Algebraic Forms of Models	Assumptions about the Error Term Structure	Behavioural Implications
	(a) Error structure specification		
Conditional Logit Model (CL) /Multinomial Logit Model (MNL)	$P_{nj} = \exp(\mu V_{nj}) \Big/ \sum_{j \in C} \exp(\mu V_{nk}), \mu = 1$	Type one extreme value distribution, $Var(\varepsilon_{nj}) = Var(\varepsilon_{nk})$ $\operatorname{cov}(\varepsilon_{nj}, \varepsilon_{nk}) = 0, j \neq k$	The IIA holds across all alternatives. No correlation between alternatives.
Nested Logit Model (NL)	$P_{nj} = P_{nj\|B_k} P_{B_k}$ $P_{nB_k} = \exp(W_{nk} + \lambda_k I_{nk}) \Big/ \sum_{l=1}^{K} \exp(W_{nl} + \lambda_l I_{nl})$ $P_{nj\|B_k} = \exp(Y_{nj} / \lambda_k) \Big/ \sum_{q \in B_k} \exp(Y_{nq} / \lambda_k)$ $Where,\ I_{nk} = \ln \sum_{j \in B_k} \exp(Y_{nj} / \lambda_k)$	Generalized extreme value distribution, $Var(\varepsilon_{nj}) = Var(\varepsilon_{nk}), j, k \in B_k$ $\operatorname{cov}(\varepsilon_{nj}, \varepsilon_{nk}) \neq 0, j, k \in B_q$ $Var(\varepsilon_{nj}) \neq Var(\varepsilon_{nl}), j \in B_{m,} k \in B_q, m \neq q$ $\operatorname{cov}(\varepsilon_{nj}, \varepsilon_{nl}) = 0, j \in B_{m,} k \in B_{q,} m \neq q$	The IIA assumption is partially relaxed. Error terms are correlated between alternatives within a nest. Error terms are uncorrelated for alternatives in different nests.
Heteroscedastic Extreme Values (HEV)	$P_{nj} = \exp(\mu_j V_{nj}) \Big/ \sum_{j \in C} \exp(\mu_k V_{nk}), \mu_i \neq \mu_j$	Type one extreme value distribution, $Var(\varepsilon_{nj}) \neq Var(\varepsilon_{nk})$, $\operatorname{cov}(\varepsilon_{nj}, \varepsilon_{nk}) = 0, j \neq k$	The IIA assumption is relaxed. Error terms are independent but non-identically distributed

Notes: n indexes decision makers; j and k index alternatives; i indexes choice formats; C and B denote choice sets; V indicates indirect utility function; ε denotes error term; μ is the scalar parameter. For the NL model, I denotes the inclusive value. $P_{nj|B}$ is the probability of choosing j conditional on j belongs to choice set B (for more details see Train, 2003). This table is continued on the next page.

Table A2.1 Variants of Random Utility Models (Continued)

Random Utility Models	Algebraic Forms of Models	Assumptions about the Error Term Structure	Behavioural Implications
	(b) Preference heterogeneity		
Random Parameter Logit Model (RPL) or Mixed Logit (ML)	$P_{nj} = \exp(\mu V_{nj}) \Big/ \sum_{j \in C} \exp(\mu V_{nk}),\ \mu = 1$ $\overline{P}_{nj} = \int P_{nj} f(\beta \mid \theta) d\beta$	A mix of type I extreme value distribution and a standard normal distribution or logistic distribution.	Preferences are different across individuals.
Models with covariates	$V_{nj} = \beta X_j + \theta' Z_n X_j + \varepsilon_{nj}$ (compared to a model without covariates: $V_j = \beta X_j + \varepsilon_j$)		Preferences are different across individuals.
	(c) Scale heterogneity		
Heteroscedastic CL –alternative specific scale parameter	$P^i_{nj} = \exp(\mu_j V_{nj}) \Big/ \sum_{j \in C} \exp(\mu_k V_{nk})$	Type I extreme value distribution, $Var(\varepsilon_{nj}) \neq Var(\varepsilon_{nk}), j \neq k;$	The scale parameter differs across alternatives.
Heteroscedastic CL – choice format specific scale parameter	$P^i_{nj} = \exp(\mu^i V_{nj}) \Big/ \sum_{j \in C} \exp(\mu^i V_{nk})$	Type I extreme value distribution, $Var(\varepsilon^i) \neq Var(\varepsilon^{i'}), i \neq i';$	The scale parameter differs across different datasets, but constant across alternatives.
Heteroscedastic CL-individual specific	$P^i_{nj} = \exp(\mu^i_n V_{nj}) \Big/ \sum_{j \in C} \exp(\mu^i_n V_{nk})$ $\mu^i_n = \exp(\alpha SQ + \theta' S_n + \gamma Z)$	Type I extreme value distribution, $Var({\varepsilon_{nj}}^i) \neq Var({\varepsilon_{nj}}^{i'}), i \neq i';$ $Var(\varepsilon_{nj}) \neq Var(\varepsilon_{mj}), n \neq m;$	The scale parameter differs across different datasets and individuals, but constant across alternatives.

Notes: n indexes decision makers; j and k index alternatives; i indexes choice formats; C denotes choice set; V indicates indirect utility function; ε denotes error term; $\boldsymbol{X}$ is a vector of attributes, $\boldsymbol{S}$ is a vector of social-demographic variables; $\boldsymbol{Z}$ is a vector of contextual variables; $f(.)$ is probability density function (pdf); $\phi(.)$ is normal distribution pdf. μ is the scalar parameter.

Appendix 2.2 Likelihood-Ratio Tests for All Estimated Models (64 Models)
Table A2.2.1 Likelihood-Ratio Tests for Models 1 to 16

Block	#	Models	# para.	Log-Likelihood	Likelihood-Ratio Tests within Block				Likelihood-Ratio Tests across Block			LR Test Results
					Against CL [a]	Against NL [a]	Against RPL [a]	Against with-Covariates [b]	HET1 [c]	HET2 [d]	HET3 [e]	
1	1	HLM_ CL_Y_0	7	-1121.28	-	-	-	**68.38**	**5.14**	**11.11**	**11.52**	Reject
1	2	HLM_ NL_Y_0	8	-1121.27	0.01	-	-	**68.72**	**6.47**	**13.12**	**13.61**	Reject
1	3	HLM_ RPL_Y_0	13	-994.48	**235.65**	-	-	**39.95**	**9.40**	**18.43**	**19.05**	Reject
1	4	HLM_ NRPL_Y_0	14	-993.52	**240.66**	**240.65**	**5.01**	**36.53**	**10.02**	**19.62**	**19.83**	Reject
1	5	HLM_ CL_Y_1	15	-1087.00	-	-	-	-	2.06	**7.65**	-	Reject
1	6	HLM_ NL_Y_1	16	-1084.76	0.35	-	-	-	1.72	**7.37**	-	Reject
1	7	HLM_ RPL_Y_1	21	-978.08	**207.21**	-	-	-	**6.49**	**14.18**	-	Reject
1	8	HLM_ NRPL_Y_1	22	-977.52	**208.81**	**208.46**	1.59	-	**7.51**	**14.01**	-	Reject
2	9	HLM_ CL_LNY_0	7	-1160.19	-	-	-	**62.12**	1.29	**10.51**	**10.96**	Reject
2	10	HLM_ NL_LNY_0	8	-1160.19	0			**62.14**	1.85	**11.29**	**11.82**	Reject
2	11	HLM_ RPL_LNY_0	13	-1040.12	**240.14**	-		**31.22**	**11.40**	**23.68**	**24.60**	Reject
2	12	HLM_ NRPL_LNY_0	14	-1040.73	**238.92**	**238.93**	-1.22	**34.77**	**12.00**	**24.74**	**26.72**	Reject
2	13	HLM_ CL_LNY_1	13	-1129.12	-	-	-	-	0.06	**8.76**	-	Reject
2	14	HLM_ NL_LNY_1	14	-1129.12	0.00	-	-	-	0.08	**8.76**	-	Reject
2	15	HLM_ RPL_LNY_1	19	-1024.51	**209.23**	-	-	-	**13.73**	**23.80**	-	Reject
2	16	HLM_ NRPL_LNY_1	20	-1023.34	**211.57**	**211.56**	2.34	-	**12.37**	**22.88**	-	Reject

Notes: [a] Likelihood-Ratio (LR) values are calculated based on Model *i* against either a conditional logit (CL) model, a nested logit (NL) model, or a random parameters logit (RPL) model within each four-model sub-block; [b] LR values for Model *i* are calculated based on Model *i* against Model (*i*+4); [c] LR values for Model *i* are calculated based on Model *i* against Model (*i*+16); [d] LR values for Model *i* are calculated based on Model *i* against Model (*i*+32); [e] Model *i* against Model (*i* +52).

Table A2.2.2 Likelihood-Ratio Tests for Models 17 to 32

Block	#	Models	# para.	Log-Likelihood	Likelihood-Ratio Tests within Block				Likelihood-Ratio Tests across Block			Likelihood - Ratio Test Results
					Against CL [a]	Against NL [a]	Against RPL [a]	Against with-Covariates [b]	HET1	HET2 [c]	HET3	
3	17	HET1_CL_Y_0	8	-1118.71	-	-	-	**65.29**	-	**5.97**	-	Reject
3	18	HET1_NL_Y_0	9	-1118.04	1.34	-	-	**63.97**	-	**6.65**	-	Reject
3	19	HET1_RPL_Y_0	14	-998.75	**239.91**	-	-	**37.03**	-	**9.03**	-	Reject
3	20	HET1_NRPL_Y_0	15	-995.94	**245.54**	**244.20**	**5.63**	**34.02**	-	**9.60**	-	Reject
3	21	HET1_CL_Y_1	16	-1086.06	-	-	-	-	-	**5.59**	-	Reject
3	22	HET1_NL_Y_1	17	-1086.06	0.01	-	-	-	-	**5.65**	-	Reject
3	23	HET1_RPL_Y_1	22	-980.24	**211.64**	-	-	-	-	**7.69**	-	Reject
3	24	HET1_NRPL_Y_1	23	-978.93	**214.27**	**214.25**	2.62	-	-	**6.50**	-	Reject
4	25	HET1_CL_LNY_0	8	-1159.54	-	-	-	**60.90**	-	**9.23**	-	Reject
4	26	HET1_NL_LNY_0	9	-1159.27	0.54	-	-	**60.38**	-	**9.44**	-	Reject
4	27	HET1_RPL_LNY_0	14	-1034.42	**250.25**	-	-	**33.55**	-	**12.28**	-	Reject
4	28	HET1_NRPL_LNY_0	15	-1034.73	**249.62**	**249.08**	-0.62	**35.14**	-	**12.74**	-	Reject
4	29	HET1_CL_LNY_1	14	-1129.09	-	-	-	-	-	**8.70**	-	Reject
4	30	HET1_NL_LNY_1	15	-1129.08	0.02	-	-	-	-	**8.68**	-	Reject
4	31	HET1_RPL_LNY_1	20	-1017.64	**222.89**	-	-	-	-	**10.07**	-	Reject
4	32	HET1_NRPL_LNY_1	21	-1017.16	**223.87**	**223.85**	0.98	-	-	**10.51**	-	Reject

Notes: [a] Likelihood-Ratio (LR) values are calculated based on Model *i* against either a conditional logit (CL) model, a nested logit (NL) model, or a random parameters logit (RPL) model within each four-model sub-block; [b]LR values for Model *i* are calculated based on Model *i* against Model (*i*+4); [c] LR values for Model *i* are calculated based on Model *i* against Model (*i*+16).

Table A2.2.3 Likelihood-Ratio Tests for Models 33 to 48

Block	#	Models	# para.	Log-Likelihood	Likelihood-Ratio Tests within Block				Likelihood-Ratio Test Results
					Against CL[a]	Against NL[a]	Against RPL[a]	Against with-Covariates[b]	
5	33	HET2_CL_Y_0	9	-1115.72	-	-	-	**64.92**	Reject
5	34	HET2_NL_Y_0	10	-1114.71	2.02	-	-	**62.97**	Reject
5	35	HET2_RPL_Y_0	15	-994.24	**242.97**	-	-	**35.70**	Reject
5	36	HET2_NRPL_Y_0	16	-991.14	**249.17**	**247.15**	**6.20**	**30.92**	Reject
5	37	HET2_CL_Y_1	17	-1083.27	-	-	-	-	Reject
5	38	HET2_NL_Y_1	18	-1083.23	0.08	-	-	-	Reject
5	39	HET2_RPL_Y_1	23	-976.39	**213.75**	-	-	-	**Not Reject**
5	40	HET2_NRPL_Y_1	24	-975.68	**215.18**	**215.10**	1.43	-	-
6	41	HET2_CL_LNY_0	9	-1154.93	-	-	-	**60.37**	Reject
6	42	HET2_NL_LNY_0	10	-1154.55	0.76	-	-	**59.62**	Reject
6	43	HET2_RPL_LNY_0	15	-1028.28	**253.30**	-	-	**31.33**	Reject
6	44	HET2_NRPL_LNY_0	16	-1028.36	**253.14**	**252.38**	-0.16	**32.92**	Reject
6	45	HET2_CL_LNY_1	15	-1124.74	-	-	-	-	Reject
6	46	HET2_NL_LNY_1	16	-1124.74	0.00	-	-	-	Reject
6	47	HET2_RPL_LNY_1	21	-1012.61	**224.26**	-	-	-	**Not Reject**
6	48	HET2_NRPL_LNY_1	22	-1011.90	**225.68**	**225.68**	1.42	-	-

Notes: [a] Likelihood-Ratio (LR) values are calculated based on Model *i* against either a conditional logit (CL) model, a nested logit (NL) model or a random parameters logit (RPL) model within each four-model sub-block; [b]LR values for Model *i* are calculated based on Model *i* against Model (*i*+4).

Table A2.2.4 Likelihood-Ratio Tests for Models 49 to 64

Block	#	Models	# para.	Log-Likelihood	Likelihood-Ratio Tests within Block: Against CL[a]	Against NL[a]	Against RPL[a]	Against with-Covariates[b]	Likelihood-Ratio Test Results
7	49	HET3_CL_Y_0	9	-1115.52	-	-	-	-	Reject
7	50	HET3_NL_Y_0	10	-1114.47	2.10	-	-	-	Reject
7	51	HET3_RPL_Y_0	15	-993.93	**243.17**	-	-	-	Reject
7	52	HET3_NRPL_Y_0	16	-991.03	**248.97**	**246.87**	**5.79**	-	-
7	53	HET3_CL_Y_1	17	-1106.09	-	-	-	-	Reject
7	54	HET3_NL_Y_1	18	-1106.09	0.00	-	-	-	Reject
7	55	HET3_RPL_Y_1	23	-986.12	**239.92**	-	-	-	**Not Reject**
7	56	HET3_NRPL_Y_1	24	-985.85	**240.47**	**240.48**	0.55	-	-
8	57	HET3_CL_LNY_0	9	-1154.70	-	-	-	-	Reject
8	58	HET3_NL_LNY_0	10	-1154.28	0.84	-	-	-	Reject
8	59	HET3_RPL_LNY_0	15	-1027.82	**253.77**	-	-	-	**Not Reject**
8	60	HET3_NRPL_LNY_0	16	-1027.37	**254.67**	**253.83**	0.90	-	-
8	61	HET3_CL_LNY_1	15	-1139.07	-	-	-	-	Reject
8	62	HET3_NL_LNY_1	16	-1138.87	0.40	-	-	-	Reject
8	63	HET3_RPL_LNY_1	21	-1026.79	**224.55**	-	-	-	**Not Reject**
8	64	HET3_NRPL_LNY_1	22	-1026.87	**224.39**	**223.99**	-0.17	-	-

Notes: [a] Likelihood-Ratio (LR) values are calculated based on Model *i* against either a conditional logit (CL) model, a nested logit (NL) model or a random parameters logit (RPL) model within each four-model sub-block; [b] LR values for Model *i* are calculated based on Model *i* against Model (*i*+4).

Appendix 2.3 An Explanation of Nesting Relationships among All Estimated Models

In Table 2.5 Column 2, models are named 1 to 64 based on the order of presentation to facilitate explanation for nesting relationships among all estimated models. Table 2.5 Column 6 indicates ordered names of the models (i.e., 1 - 64) that have more general specifications than the model for a given row. For example, for the first eight models (i.e., Block 1), Model 1 (HLM_CL_Y_0) is nested by all other models within the block. Model 2 (HLM_NL_Y_0) is nested by three models in the block: HLM_NRPL_Y_0, HLM_NL_Y_1 and HLM_NRPL_Y_1. Both Model 6 (HLM_NL_Y_1) and Model 7 (HLM_RPL_Y_1) are nested by Model 8 (HLM_NRPL_Y_1), but they are non-nested by each other. Model 8 is the most complex model in the block, so it is not nested by any other model in the block. The nesting structure is exactly the same for the next five blocks. Within each block, the first model is nested by all other models, the second model is nested by the 4^{th}, 6^{th} and 8^{th} models, the third model is nested by the 4^{th}, 7^{th} and 8^{th} models, the fourth model is nested by the 8^{th} model only, so are the 6^{th} and 7^{th} models. The fifth model is nested by the 6^{th}, 7^{th} and 8^{th} models, and the 8^{th} is not nested by any model within the block. For the 7^{th} and 8^{th} blocks (i.e., models with a HET3 specification), the nesting structure is slightly different. Models with a no-covariates specification are not nested by their with-covariates counterparts (Table 2.2), so the first model in the two blocks (e.g., HET3_CL_Y_0 in the 7^{th} block) is only nested by the 2^{nd}, 3^{rd} and 4^{th} (HET3_NL_Y_0, HET3_RPL_Y_0, and HET3_NRPL_Y_0, respectively). The 2^{nd} and 3^{rd} models are nested by the 4^{th} model and the 4^{th} model is not nested by any other model within the block. Similarly, the 5^{th} model is nested by the 6^{th}, 7^{th} and 8^{th}, and the 6^{th} and 7^{th} are nested by 8^{th}, while the 8^{th} model is not nested by any model in the block. Results of pair-wise likelihood-ratio (LR) tests within each block are presented in Appendix 2.2.

Nesting relationships also widely exist across blocks (Table 2.5 Column 7). Across blocks, each linear-in-income specification HLM model is nested by its HET1 or HET2 counterpart. So are the non-linear specifications. Each HET1 model is also nested by its HET2 counterpart. Furthermore, each no-covariates HLM model is nested by its HET3 counterpart.. For example, Model 1 in Block 1 is nested by Models 17, 33, and 49, which are the first model in Blocks 3, 5, and 7 respectively. In fact, since Model 1 is nested by Model 17, and Model 17 is nested by the other seven models in Block 3 (Models 18 to 24), Model 1 is necessarily nested by these seven models. For brevity, we do not list all nesting relationships across blocks. LR tests are conducted to determine if the null hypothesis about a simpler model in one block against a model in a different block can be rejected or not. Like the within-block LR test, "Reject" is reported if a model is rejected once by pair-wise LR tests across blocks. The across-block LR test results are presented in the second last columns of Tables 2.2.1-2.2.4 (Appendix 2.2).

The last column of Table 2.5 therefore is a final result about whether a null hypothesis is ever rejected from either a within-block LR test or an across-block LR test. For example, Model 8 is on the top of a nest tree within block 1. A within-block LR test cannot be performed. However, across blocks, Model 8 is nested by either Model 24 or Model 40. Across-block LR tests indicate that the null that Model 8 is a true model is rejected at least once. So the final LR test result for Model 8 is "Reject".

Chapter 3 The Role of Altruism in the Valuation of Municipal Drinking Water Risk Reductions

3.1 Introduction

Municipal water quality improvement programs can be considered to be public goods. Once a preferred disinfection method is chosen and provided, everyone in the community will be able to access drinking water of a given quality. The level of preferred disinfection method could be chosen using Samuelson's decision rule for provision of a public good. According to Samuelson's rule, the derived total benefits of a public good are equal to the summation of individual marginal benefits only if these benefits are purely motivated by self-interest. However, there is increasing evidence to show that the demand for public goods is driven by both self-interest and altruistic motives (Holmes 1990; Andreoni 1990; Johansson 1994; Flores 2002). Therefore, the presence of altruism in individuals' willingness-to-pay (WTP) for a public good is potentially a serious problem for decision makers. This paper examines the role of altruism in the demand for a public good that reduces multiple health risks. The effect of altruism on the demand for the public good is revealed from WTP estimates for different public risk reductions. In addition the paper provides value of statistic life estimates (VSL) derived from these WTP estimates.

Research indicates that chlorine, the dominant disinfection method used to reduce risks of microbial illnesses in water, can potentially produce carcinogenic disinfection by-products (Mills et al. 1998). Alternative disinfection methods may create fewer carcinogenic disinfection by-products but are not only generally more expensive and may also be less effective. This suggests the existence of a trade-off between types of risks and risk reductions and expenditures in water quality management. To identify public preferences for risk reductions in cancer and microbial illnesses from drinking-water quality improvements, an internet-based survey was conducted across Canada during the summer of 2004. In the survey, people were asked a) to vote for or against a proposed water treatment program which reduces either one or both types of drinking-water health risks but at the cost of a higher water bill; or b) to choose between alternative programs describing different levels of health risks and water bills and a status quo situation. We incorporate altruism into individuals' utility functions when modeling choice decisions involving risk-dollar tradeoffs as well as risk-risk tradeoffs so that the magnitude of altruism can be estimated explicitly.

A major difficulty in accounting for altruism empirically is to distinguish altruistic motives from self-interested motives in a consumer's choice decision (Jones-Lee 1992). In this paper data on self-protection measures against health risks from drinking water are used to distinguish the demand for the public good by individuals who are driven solely by altruistic motives from those who might have both self-interest and altruistic motives. The paper is organized as follows. Section 3.2 reviews the current literature on accounting for altruism in valuing public goods. Section 3.3 introduces a theoretical model incorporating altruism into a utilitarian framework and describes two approaches for implementing the model. Section 3.4 describes the survey, data and model specifications. Section 3.5 reports model estimation results, WTP estimates and values of statistical life and values of statistical illness estimates. The last section concludes the paper.

3.2 The Role of Altruism in the Demand for Public Goods

Several studies suggest that the demand for public goods is driven both by self-interest and altruistic motives. For example, Johansson (1994) suggests that people are concerned with not only their own health, but also the health of others. Studies on voting behaviour provide evidence of the existence of altruism regarding environmental quality (Holmes 1990). Flores (2002) indicates that there can be legitimate altruistic values resulting from increases in public goods. In fact, the existence of altruism in human behaviour is widely acknowledged in various literatures, including social psychology, sociology, economics, environmental economics, political behaviour and sociobiology (Piliavin and Charng 1990). In the environmental valuation literature, Krutilla (1967) was one of the earliest researchers to propose accounting for non-use value in valuing environments. He suggests that people value particular natural environments not only because they obtain value from their use or planned use but also from other motivations. Kopp (1992) argues that some of these motivations are due to individuals' "ethical concerns", which can be interpreted to include altruism.

In order to provide a full accounting of the benefits of environmental preservation, environmental economists have started to account for altruism in benefit-cost analysis (Krutilla 1967; Holmes 1990). According to the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) of 1980 and the Oil Pollution Act of 1990, non-use values, including altruistic value, have to be taken into account in natural resource damage assessment. Johansson (1994) suggests that a willingness-to-pay approach to valuing health risk reductions is problematic in general if altruistic motives are ignored because there is empirical evidence on the magnitude of the willingness-to-pay for a reduction of the health risks faced by others (for example, Jones-Lee 1992). From a welfare economics perspective, Flores (2002) indicates that a selfish benefit-cost test, which ignores altruism in the demand for public goods, could potentially lead to a rejection of welfare-improving changes in public goods. This finding simply suggests that Samuelson's rule for optimal provision of public goods does not hold in the presence of altruism. The theoretical framework based on self-interested values is not consistent with cases of preference interdependence over public goods. Consequently, McConnell (1997) concludes that the role of altruism in benefit-cost analysis is sufficiently troubling to warrant empirical research on motives for non-use values.

The literature typically examines two types of altruism[22]: non-paternalistic and paternalistic altruism. Non-paternalistic altruism implies that an altruist derives utility from his or her beneficiaries' general wellbeing and respects their preferences: "if the beneficiaries are happy, then I am happy". Paternalistic altruism, on the other hand, implies that an altruist derives utility from his or her beneficiaries' consumption of a particular good (McConnell 1997; Lazo, McClelland and Schulze 1997; Flores 2002). "It is much like literal paternalism, for example, when parents insist that their children eat their carrots."(McConnell 1997, p. 32) Depending on which type is present, altruism may or may not lead to double counting in benefit-cost analysis (McConnell 1997; Lazo, McClelland and Schulze 1997; Flores 2002) and have an impact upon welfare measures

[22] Jones-Lee (1991) also suggests two types of altruism but uses different terms: safety-focused altruism and pure altruism. His safety-focused altruism is similar to paternalistic altruism and pure altruism is similar to non-paternalistic altruism.

and the calculation of optimal Pigovian taxes (Flores 2002; Johansson 1997).[23] McConnell (1997) suggests that individuals' willingness-to-pay for non-paternalistic altruistic reasons should not be included in benefit-cost analysis since benefits received by the altruist and the beneficiaries imply double counting.[24] In the case of paternalistic altruism, however, the altruist actually derives self-interested utility by acting altruistically, so this type of altruism is self-motivated and should be counted as a part of the altruist's self-interested value. Empirically, however, it is difficult to determine whether the observed altruism is paternalistic or non-paternalistic since motives are not observable. McConnell (1997) suggests that paternalistic motives are more plausible in the case of a project involving natural resources.

The role of altruism in decision-making has been mainly investigated empirically in several areas. Andreoni (1990, 1995) provides an analytical model that can explain impure altruism, which means individuals derive utility from the acting of giving or donating to a public good whose provision can be supplied both publicly and privately. He found weak evidence of existence of altruism regardless of public investments in the provision of a public good. Voting behaviour is another popular area since "voting is a means for expressing individual preference for alternative social states" (Holmes 1990, p. 140). Holmes (1990) examines the effects of altruism on political choices regarding environmental health risks using a residualization method that involves the construction of a proxy measure for altruism by controlling for "other" factors that cannot be clearly categorized as altruistic or self-interest variables. He then recovers altruistic pro-environmental preferences by comparing votes across cases. He finds that altruism plays a role in voting decisions. A similar study by Deacon and Shapiro (1975) reaches the opposite conclusion. To rectify the discrepancy between these two studies, Popp (2001) examines the role of altruism in the demand for environmental quality using a referendum-style survey. The micro-level data he uses enables him to construct a more direct measure of altruism when compared to the aggregate data used in the previous two studies. Specifically, he examines how an individual's life expectancy affects his/her

[23] Johansson (1997) finds optimal externality-correcting taxes are different if the types of altruistic behaviour are different.

[24] However, Flores (2002) suggests that non-paternalistic altruism may play a role in a benefit-cost analysis of generic changes in public goods when "preference interdependence between public goods and the distribution of income" is present (p. 294). This paper does not assume such preferences.

willingness to pay for the benefits of an environmental project. He hypothesizes that, if people are motivated only by self-interest, the amount individuals are willing to pay for environmental protection should fall as life expectancy decreases since older people will not be alive to enjoy the benefits of preserving resources for later years. He finds the existence of weak altruism, i.e., people are concerned with both self-interest and the interest of future generations regarding provision of environmental quality. However, Popp does not include the number of children as an explanatory variable, which could be very important in understanding altruism among household members.

Few empirical studies have tried to identify the magnitude of altruistic values. One reason might be that whether altruistic value should be incorporated into economic decision making is still a controversial issue given that the nature and factors affecting altruism are not fully understood.[25] This study provides an empirical example of how to estimate the demand for a public good taking altruism into account. The focus of this paper is to decompose total WTP into a self-motivated component and an altruistically motivated component. However, like previous studies, we are unable to distinguish between the two types of altruism due to a lack of data on motivations for altruism.

3.3 Accounting for Altruism – Theoretical and Empirical Framework

3.3.1 Theoretical Framework

In this study the goal is to identify and isolate self-interested WTP from total WTP for a public good. One way to do this is to assume the existence of impure altruism as suggested by Andreoni (1989, 1990). We define individual n's utility to be a function of his or her private consumption of a public good g_n, the total provision of the public good G and a Hicksian composite good Z. Hence, $u_n = u_n(g_n, G, Z)$, where $\partial u_n / \partial g_n > 0$, $\partial u_n / \partial G \geq 0$. Let $\mathbf{q}_n$ denote a vector of attributes associated with g_n. Then, the indirect utility function can be specified as $V_n = V(\mathbf{q}_n, G, Y_n)$, where Y_n is individual n's income and the price for Z is normalized to 1. Let $V_n^0 = V(\mathbf{q}_n^0, G^0, Y_n)$ be the utility level before

[25] For example, pure altruism, to some extent, could be independent of one's income level which makes the prediction of altruistic behaviour highly unlikely (Andreoni 1995).

an improvement in the provision of G and let the utility level after the change be V_n^1 = $V(\mathbf{q}_n^1, G^1, Y_n)$. Then, the compensating variation (CV)[26] or total WTP for an improvement in the quality of the public good is the solution to the equation $V(\mathbf{q}_n^0, G^0, Y_n) = V(\mathbf{q}_n^1, G^1, Y_n - CV_n)$. Alternatively, total WTP can be expressed explicitly as a change in expenditure functions:

(3.1) $WTP_{Total_n} = e(\mathbf{q}_n^0, G^0, U^0) - e(\mathbf{q}_n^1, G^1, U^0)$

where U^0 denotes the utility level at the status quo. Adding and subtracting the term $e(\mathbf{q}_n^1, G^0, U^0)$ on the right hand side of the equation, we get:

(3.2) $WTP_{Total_n} = e(\mathbf{q}_n^0, G^0, U^0) - e(\mathbf{q}_n^1, G^0, U^0) + e(\mathbf{q}_n^1, G^0, U^0) - e(\mathbf{q}_n^1, G^1, U^0)$

Altruistic WTP is defined as the difference between the last two items:

(3.3) $WTP_{Altrm_n} = e(\mathbf{q}_n^1, G^0, U^0) - e(\mathbf{q}_n^1, G^1, U^0)$

This is the amount that an individual is willing to pay for an improvement in the quality of the public good G despite the fact that the individual does not derive private benefits from such a change. This is the WTP for other people's benefit. The self-interested WTP, WTP_{Self}, is the difference between the first two terms of Equation 3.2:

(3.4) $WTP_{Self_n} = e(\mathbf{q}_n^0, G^0, U^0) - e(\mathbf{q}_n^1, G^0, U^0)$

This is considered as self-interested WTP because it is the amount an individual would be willing to forgo to secure private benefits from a quality change of a public good holding the quality level of the public good for other people constant. Therefore, total WTP can be decomposed into two parts: altruistic WTP and self-interested WTP:

(3.5) $WTP_{Total_n} = WTP_{Altrm_n} + WTP_{Self_n}$

For our particular case study, we assume that individuals who have already engaged in self-protection against drinking water related health risks obtain no private gains if the level of attributes of the public good improves from G^0 to G^1. Thus, their indirect utility functions before and after the improvement in the quality of the public good are $V^0(\mathbf{q}^1, G^0, Y)$ and $V^1(\mathbf{q}^1, G^1, Y)$. The WTP for an improvement in the level of attributes of a public good is considered to be purely altruistic, i.e., WTP_{Altrm}. From

[26] CV is one type of measurement of WTP. It is the amount an individual is willing to pay for an improvement of a good in question and remained as well-off as before the improvement.

Equation 3.5, the self-interested WTP_{Self} can be obtained by subtracting WTP_{Altrm} from WTP_{Total}.

There are a number of advantages associated with the application of this theoretical framework. First, each individual's utility function is hypothesized to be both self-interest and altruistically motivated. Depending on the magnitude of these two components of WTP an individual can be purely selfish, purely altruistic or in-between. Thus, compared to the model used by Flores (2002), in which a utility function has to be defined differently for the altruist and for his or her beneficiaries, our approach is easier to implement empirically. We do not have to differentiate an altruist from his or her beneficiaries before applying our model. Second, since individuals' preferences are defined over levels of attributes associated with a public good, the WTP for an improvement in the levels of attributes can be directly estimated, which is exactly what this study aims to do. Third, one does not have to identify the type of altruism before actually estimating it. If altruism is paternalistic, WTP_{Total} is an individual's self-interested WTP. If altruism is non-paternalistic, WTP_{Self} is an individual's self-interested WTP, and WTP_{Altrm} has to be excluded from benefit cost analysis to avoid double counting. At the aggregate level, to conform to Samuelson's rule, if altruism is paternalistic, the total benefits (TB) of a public good are,

$$(3.6) \quad TB|_{\text{paternalistic altruism}} = \sum_{n=1}^{N} WTP_{Total_n}$$

Given non-paternalistic altruism, however, TB is

$$(3.7) \quad TB|_{\text{non-paternalistic altruism}} = \sum_{n=1}^{N} WTP_{Self_n}$$

However, in our specific case we must also make several assumptions about the nature of self protection and the benefits of the public good. We discuss these assumptions below.

3.3.2 Empirical Framework

Johansson (1994) suggests that it is possible to extract information on the magnitude of WTP associated with different forms of altruism by letting different sub-samples of respondents respond to different valuation questions. A key to successful decomposition of total WTP is to identify those people who derive utility solely from private gains from their own consumption of the public good.

In this study, for people who did not engage in self-protection against drinking water health risks, the proposed water treatment program could provide them with private benefits in addition to benefits for other people, so their elicited WTP is considered to be WTP_{Total}. For people who did engage in self-protection, it is hypothesized that there are no private gains from the proposed program, so their elicited WTP, if is greater than zero, will be WTP_{Altrm}. The self-interested WTP, WTP_{Self}, is the difference between WTP_{Total} and WTP_{Altrm}. This is a rather strong assumption in that it excludes private benefits derived from the provision of water treatment outside a household, such as in a park, or in a public library. In fact, even for those who have installed water filter systems at home, individuals might be willing to pay for the public good for private reasons, i.e., to save future costs of installing and/or maintaining their water filter systems. Our analysis assumes that such private benefits are zero. Therefore, the altruistic value estimated in this study might be biased upward.

Once different motives for the demand for the public good are identified based on actual self-protection behaviour, the size of benefits for other people and total benefits can be estimated directly, and private benefits can be derived subsequently. In terms of demand modelling exercises, we adopt two approaches to account for the differences in motives. One is a sample segmentation method and the other is an interaction method. The sample segmentation method uses a dummy variable to indicate whether a respondent engaged in self-protection to segment the entire sample into two sub-groups. We then estimate different demand models for the sub-groups. The interaction method, on the other hand, assumes one's total willingness to pay for risk reductions decreases as one spends more on self-protection. It thus approximates a dose-response function between one's willingness to pay decision and the level of risks one faces (resulting from changes in the levels of self-protection). This method estimates one model using the entire sample based on an augmented utility function that includes interactions between attributes of the public good and expenditure on self-protection to control for different levels of altruism. Using these two different approaches also provides a sensitivity analysis on the decomposition of WTP into altruistic and self-interested components.

3.4 Survey, Data and Model Specifications

This paper uses the pooled dataset CE23 introduced in Chapter 1. A summary of the sample descriptive statistics can be found in Chapter 1 Appendix 1.3. We find that about 45% of the sample engaged in averting behaviour[27] against health risks from drinking water suggesting that many Canadians are concerned with water quality. Our sample excludes those observations defined as "yea-saying" data. "Yea-saying" data are those respondents who stated that they were willing to pay any amount to reduce the health risks in the surveys. It is possible that these individuals did not make tradeoffs between attributes of a good or between attributes and money, and therefore, inclusion of their responses in the analysis might lead to erroneous inference. In this study, about 10% of the survey responses are identified as the yea-saying data. Since each respondent answered four choice tasks, we have a total of 1464 observations from 366 respondents

In this study, each option is characterized as a bundle of health risk attributes associated with different water treatment programs and the costs that such a program would add to the annual household water bill. A status quo option is included as a baseline program that does not involve any increase in the water bill. The alternatives are characterized with a reduction in at least one type of health risks, as well different greater-than-zero increases in the water bill. According to random utility theory, individual *n*'s utility associated with alternative *j* has two components: deterministic utility (V_{nj}) and stochastic utility (ε_{nj}). The deterministic utility in this study is specified as, in the most basic form,

$$(3.8) \quad V_{nj} = \beta_1 SQ_j + \beta_2 MICI_j + \beta_3 MICD_j + \beta_3 CANI_j + \beta_5 CAND_j + \beta_6 BILL_j + \beta_7 SQCE3_j$$

[27] Averting behaviour in this study is specifically defined as whether one installed water filter systems (or water container systems that filter water) at home.

where *SQ* is the alternative specific constant (ASC) for the status quo option. It is included to capture unobserved utility associated with staying at the status quo (Adamowicz et al. 1998; Scarpa, Ferrini and Willis 2005). An interaction term *SQCE3* between the status quo (*SQ*) and a version dummy variable for the 3-alternative conjoint design (*CE3*) is included to account for the choice format effect on preferences.[28]

Equation 3.8 is specified with no covariates. A with-covariates specification includes additional socio-demographic variables in the indirect utility function. These covariates can be used to explain preference heterogeneity. The number of interaction terms can be large since it is the product of a number of socio-demographic variables and a number of attributes. For a small or moderate sample size, too many interaction terms might mask some important relationships; therefore, only interaction terms between socio-demographic variables and the status quo ASC (SQ) enter the model specification. A with-covariates specification is,

$$(3.9) \quad V_{nj} = \beta_k \sum_{k=1}^{6} MAIN_{kj} + \beta_7 SQCE3_j + \gamma SQ_j * \mathbf{S}_n$$

where MAIN_k defines the k^{th} main attributes, $k = 1$ to 6, indicating *SQ, MICI, MICD, CANI, CAND* and *BILL* respectively, $\boldsymbol{\gamma}$ is a vector of parameters, and $\boldsymbol{S_n}$ is a vector of individual n's socio-demographic variables. These interaction terms are defined in Table 3.1, along with the status quo ASC, a version dummy variable, attributes and expenditure variables on self-protection against drinking water related health risks.

3.4.1 The Sample Segmentation Method

To isolate self-interested and altruistic values, the sample segmentation method splits the entire sample according to a variable that indicates whether one engaged in self-protection behaviour. This indicator variable is created from information on a respondent's annual household expenditure on installing and maintaining water filtration systems (*FEXP*) since drinking water related health risks are considered to be negligible if such systems are installed at home (Adamowicz, Dupont and Krupnick 2005). For

[28] Analysis is also conducted by version, i.e., estimating models using either the CE2 or CE3 dataset. Generally speaking, results based on different datasets are similar. However, results from CE23 are more robust. We only report the CE23 result due to space limitations.

individuals who installed water filter systems, i.e., *FEXP* > 0, it is hypothesized that they have sufficient protection against various tap-water related health risks and will not obtain private benefits if they agree to a proposed public program that aims to reduce the risks. For individuals who did not install water filter systems, i.e., *FEXP* = 0, it is hypothesized that they were exposed to drinking water health risks. Therefore, they might derive both private benefits and altruistic benefits if they agree to the proposed program that aims to reduce the risks.

We did find some individuals with zero expenditure on water filter systems who spent a substantial amount on purchased bottled water to drink at home (Table 3.2). As purchased bottled water is considered to be free from the two types of drinking-water related health risks (i.e., microbial risks and cancer risks), these people might be able to avoid tap-water related health risks[29] so that they might not necessarily derive private benefits from the purchase of the public good.[30] Therefore, a "cleaner" sample which includes only those who did not have any protection against the drinking-water health risks is also defined. This sample includes observations satisfying two conditions (*FEXP* = 0 and *WPHS = 0*), where WPHS is annual household expenditure on purchasing bottled water consumed at home. Similarly, for the sample satisfying *FEXP* > 0, we further define a subset of the sample that has a greater-than-zero expenditure on purchased water. This sub-sample thus consists of observations that satisfy both *FEXP* > 0 and *WPHS* > 0 conditions. The stratification is summarized in Figure 3.1.

Observations satisfying both *FEXP* = 0 and *WPHS* = 0 conditions are called sub-sample 1, which is a subset of sub-sample 2 that only has to satisfy the condition *FEXP* = 0. For both sub-samples, individuals' WTP for the public program is considered to be driven by both self interest and altruistic reasons. However, WTP for risk reductions revealed by sub-sample 1 is expected to be higher than that in sub-sample 2 because individuals in sub-sample 1 probably face higher health risks in general. Compared to sub-sample 1, sub-sample 2 contains individuals that have non-zero expenditure on

[29] It is assumed that purchased water can avoid various drinking water related health risks to a large degree if it meets industry product standards.

[30] On the other hand, the reasons for purchasing water to drink at home can include convenience, odour, taste, and/or health concerns, so tap-water might still be an important source for drinking water at home for these people. These people might also derive private benefits as well as other benefits from purchasing the public good.

purchased bottle water, which might have provided some protection against the health risks. Table 3.2 reports the mean of *FEXP* and *WPHS* for each sub-sample.

Sub-sample 3 consists of respondents whose expenditures on water filter systems are greater than zero (*FEXP* > 0). Its subset satisfying both *FEXP* > 0 and *WPHS* > 0 conditions is called sub-sample 4. It is hypothesized that individuals in both sub-samples are no longer exposed to the health risks from tap drinking water, and therefore, they derive only altruistic benefits from the purchase of the public good. However, individuals in sub-sample 4 may have more protection against the risks compared to those in sub-sample 3. Therefore, on average, the WTP for risk reduction revealed by individuals in sub-sample 4 is expected to be less than WTP obtained for sub-sample 3.

Individual models will be estimated based on sub-samples 1 to 4 (Models 1 to 4). The derived WTP estimates from each model are called WTP1, WTP2, WTP3 and WTP4 accordingly (Figure 3.1). Assuming individuals are willing to pay more for larger risk reductions, WTP estimates derived from Model 1 to Model 4 are expected to be decreasing, i.e., WTP1 > WTP2 > WTP3 > WTP4. However, since motives for purchasing bottled water are quite varied, our key hypothesis is WTP2 > WTP3.

3.4.2 The Interaction Method

The interaction method accounts for differences in motives for risk reductions through interactions between deterministic utility V_{nj} and an expenditure variable on self-protection against the risks. The primary variable *FEXP* used in the sample segmentation method is again used as an indictor variable. However, it is now used as a continuous variable in order to capture continuous changes in levels of altruism due to changes in the levels of self-protection. Thus, this method allows us to explore two interesting issues. One is an examination of the need to control for endogeneity between choice decisions regarding risk reduction and risk preferences (Louviere et al. 2005) and the other is to investigate whether there is a dose-response relationship between the willingness to pay for risk reductions and levels of self-protection.

Algebraically, individual n's utility associated with an alternative water treatment program j, based on Equation 3.9 (a with-covariates specification) is specified as, (3.10) $V_{nj} = \beta_k \sum_{k=1}^{6} MAIN_{kj} + \beta_7 SQCE3_j + \gamma SQ_j * \mathbf{S}_n + \delta_k FEXP_n * \sum_{k=1}^{6} MAIN_{kj}$

where MAIN_k, *SQCE3* and $\boldsymbol{S_n}$ are defined earlier for Equation 3.9; $FEXP_n$ is invididual n's annual total expenditure on installing and maintaining water filter systems at home, and δ_k is the coefficient on the k^{th} main variable to be estimated.

However, this specification is correct only when *FEXP* is exogenous. If *FEXP* is correlated with the error term ε, the estimator will be biased. A risk-averse individual may not only take more self-protection measures than a risk-neutral individual, but may also be willing to pay higher amounts for a program to reduce the risks. One way to handle this endogeneity is to create an instrumental variable that is correlated with *FEXP*, but uncorrelated with the error term ε (Greene 2003; Louviere 2005). One natural instrumental variable is predicted *FEXP*, which can be created based on a regression relationship between the actual *FEXP* and a vector of exogenous variables, such as age, income and so forth. The predicted *FEXP* should be correlated with the actual *FEXP* but uncorrelated with its error term. However, since *FEXP* is non-negative in our study, a tobit model that handles censored data is appropriate (Green 2003). In the tobit model, the actual dependent variable is a latent variable for *FEXP*. Denote Z_n^* as latent filter expenditures for individual n,.

(3.11) $Z_n^* = \mathbf{x}_n \beta + e.$

$FEXP_n = 0$, if $Z_n^* \leq 0$;

$FEXP_n = Z_n^*$, if $Z_n^* > 0$.

where Z_n^* can be negative, zero or positive, and $\boldsymbol{x_n}$ is a vector of socio-demographic variables for individual n.

Once a tobit regression for *FEXP* is estimated, the predicted latent expenditure $\hat{Z}_n^*$ can be used as an instrumental variable for $FEXP_n$. Equation 3.10 thus becomes:

(3.12) $V_{nj} = \beta_k \sum_{k=1}^{6} MAIN_{kj} + \beta_7 SQCE3_j + \gamma SQ_j * \mathbf{S}_n + \delta_k \hat{Z}_n^* \sum_{k=1}^{6} MAIN_{kj}$

Using this specification, a total WTP (WTP_{Total}) for a certain type of risk reduction is the estimated WTP given $FEXP$ = 0, whilst an altruistic WTP (WTP_{Altrm}) is the estimated WTP given $FEXP$ > 0.

(3.13) $WTP_{Total} = WTP\,|_{FEXP=0}$ and,

$$WTP_{Altrm} = WTP\,|_{FEXP>0}$$

However, the estimated WTP is a function of estimated parameters and the latent expenditure $\hat{Z}^*$, yet $\hat{Z}^*$ differs from $FEXP_n$. Therefore, the mean WTP_{Total} of a sample is estimated at the mean of $\hat{Z}^*$ satisfying $FEXP = 0$ and the mean WTP_{Altrm} of a sample is estimated at $\hat{Z}^*$ for the sample satisfying $FEXP > 0$.[31] For comparison, we call this model Model 5 and its derived willingness-to-pay estimates WTP5. Our hypothesis is $WTP5_{Total} > WTP5_{Altrm}$.

Note that this hypothesis is parallel to the hypothesis WTP2 > WTP3 to be tested in the sample segmentation method. The left hand side of both hypotheses is the WTP estimated given $FEXP$ = 0 (or total WTP) and the right hand side is the WTP estimated given $FEXP$ > 0 (or altruistic WTP).

The estimation of Models 1 to 5 starts with a conditional logit (CL) specification that is widely used in modelling choice decisions involving the choice of one preferred option out of a finite number of alternatives. WTP estimates are subsequently derived from the estimated models and hypotheses are then tested. We also investigate the effect of preference heterogeneity on WTP estimates and decomposition by estimating a random parameters logit (RPL) model.[32] To facilitate estimation, only the main attributes, the status-quo alternative specific constant *SQ* and the four risk attributes, are estimated as random parameters. These preference parameters are individual specific, although in aggregate they are assumed to be subject to a statistical distribution that is characterized

---

[31] Alternatively, one could estimate individual WTP using $\hat{Z}_n^*$ and $FEXP_n$ and take the mean of these individual WTPs.

[32] The RPL model is widely used to capture both unobserved preference heterogeneity and to account for the panel data structure of choice experiments. Since each respondent was asked to complete four choice tasks in the water survey, our data are a type of "panel".

with an estimated mean and standard deviation. To avoid large variances in welfare estimates, the price effect (*BILL*) is specified to be non-random.[33]

3.5 Results and Discussions

3.5.1 Model Estimation

In this section, the estimated CL models are presented first, followed by the RPL models. Models 1 to 4 are estimated using sub-samples of the dataset based on the sample segmentation method while Model 5 is estimated with the entire dataset but employing the interaction method. Table 3.3 reports estimated CL models specified with covariates (Table 3.3).[34] A total of 13 covariates are included, and they are *AGE65, INCOME, INCOME2, ENGLISH, CITYSIZE, ILLNESS, MALE, MARRY, KID06, KID012, KID137, INDEXA* and *INDEXB* (see their definitions in Table 3.1).

One common finding from these results is that the estimated coefficients on all risk attributes (*MICI, MICD, CANI* and *CAND*) and the price attribute (*BILL*) are negative as expected, and all of them are significant at the 1% level (Table 3.3). The estimated coefficients for the status-quo alternative specific constant *SQ*, are positive and significant. Since there are many interaction terms between *SQ* and demographic variables, it is easier to interpret an overall status quo effect from the no-covariates specification. The estimated coefficients for *SQ* from these models are found to be positive and significant which implies that, on average, people derive utility from staying at the status quo. This could be due to an endowment effect or for other reasons that have been documented (Adamowicz et al. 1998; Dhar 1997). The absolute value of *BILL* increases from Model 1 to Model 4, indicating that people become more price sensitive when private benefits derived from the consumption of the public good decrease as the level of exposure to drinking water related health risks decreases.

Since socio-demographic variables enter as interaction terms with *SQ*, the demographic effects can only be used to explain how people with different characteristics

---

[33] To facilitate welfare calculation, a normal distribution is assumed for random parameters and price effect is assumed to be fixed. Refer to pp. 32-33 and footnote 15 for explanations.

[34] Since no-covariates models are nested within their corresponding with-covariates models, likelihood-ratio (LR) tests were conducted to determine a preferred specification. The test results indicate that adding covariates significantly improves model fit except for Model 4.

differ in choosing the status quo option. Some variables are found to be important in explaining choice decisions in at least one model, although none of them has a consistent pattern (in terms of signs and significance level) across all models. Annual household income seems to have a positive quadratic effect in Models 1 and 2 so, at lower income levels, respondents are more likely to choose the status quo, but when one's income goes beyond a certain level, he or she is more likely to choose the program. Such income effects do not appear in Model 3 or Model 4, which are estimated based on responses from households that installed water filter systems at home. It is probably because these people are more similar in terms of income. Older people (aged over 65, *AGE_65*) are more likely to choose the program when they did not install water filter systems at home. Households in smaller cities or communities (*CITYSZ_SQ*) are more likely to choose the program with lower risks but higher bills. A male respondent (*MALE_SQ*) is more likely to choose the status quo option that does not involve additional costs. Health problems of one's family members (*INDEXB_SQ*) are found to have no significant effect on choosing the status quo option except in Model 4. Marital status (*MARRY_SQ*) also is found to be significant in Model 3 only. Other demographic variables are found to be not statistically significant. The demographic effects are in general more consistent within the group which did not install water filter systems (Model 1 and Model 2) or within the group which installed water filter systems (Model 3 and Model 4), than across the groups. Model 2 and Model 3, as based on more encompassing datasets than Model 1 and Model 4 respectively, have more significantly estimated coefficients.

Table 3.4 presents the Tobit model estimated by regressing *FEXP* on a vector of exogenous variables. Since the marginal effects of Tobit models differ from the estimated coefficients by a factor that accounts for the proportion of data falling short of the lower bound (zero in this case), estimated marginal effects are also reported. Most variables are found to be insignificant except for *ENGLISH* and *KID612*. Respondents whose language is English spent, on average, $40 more on installing and maintaining water filter systems at home. This relationship perhaps explains why coefficients on *ENG_SQ* are not significant in Models 1 to 4, which are estimated using sub-samples segmented by *FEXP* as the key variable. In addition, households with kids aged between 6 and 12 years old spent on average $33.5 more on the filter expenditures than other households. An interesting finding is that *WPHS* is found to be negatively related to *FEXP*. Bottled water seems to be considered as a substitute for water filter systems regardless of the motivations for purchasing bottled water (in fact, fewer than 11% of respondents stated that they purchased bottled water out of health concerns). The correlation between *FEXP* and the predicted latent expenditure variable Z* is about 0.23.

Two sets of results for Model 5 are presented in Table 3.5. The first uses predicted Z* to interact with the main attributes while the other uses *FEXP*. Overall, these models have more significant coefficients than Models 1 to 4, which may arise since the sample segmentation method employs small sample sizes. The estimated coefficients for risk attributes and *BILL* are significant with expected signs. Some filter expenditure interaction terms are found to be significant, indicating preferences for risk reductions might change as household expenditures on water filter systems increase. Although it is difficult to compare the fit of the two models, their log-likelihood values are very close in value. However, the estimated coefficients from the two models differ both in sign and degree of significance. For example, *FMICD* is positive and significant in the predicted Z* model but negative and significant in the *FEXP*. Some estimated coefficients are found to be significant in one of the models, but not in the other. If the predicted Z* model is a preferred model that controls for endogeneity, then a large bias in the estimates might have been resulted if this problem is ignored. For models estimated with the predicted Z*, significant demographic effects of choosing the status quo option are household income (*INCM_SQ, INCM2_SQ*), age (*AGE65_SQ*), city size (*CITYSZ_SQ*) and gender (*MALE_SQ*). Signs of these estimates are the same as the ones estimated in previous models. Marital Status (*MARRY_SQ*) is found to have significant positive effects on choosing the status quo in Model 5.

Table 3.6 reports estimated models specified with random parameters. Due to limited sample sizes, Models 1 and 4 cannot be estimated.[35] Most estimated standard deviations are significant at the 5% level. Heterogeneity in the SQ and the risk attributes

[35] Note that these are no-covariates RPL models. Often, covariates enter a RPL model as shifters in the mean of individual level parameters (Hu 2004). With-covariates RPL models are estimated, but most estimated covariates are not significantly different from zero. Therefore, only no-covariates RPL models are presented.

is highly significant and consistent across models. The estimated mean and standard deviation of these variables are of same order, implying a large variability in preferences.

Comparing Model 2 and Model 3, respondents in Model 3 are in general more risk averse than those in Model 2 (more negative coefficients in risk attributes). They are also more likely to choose the program (smaller value of the estimated mean of *SQ*), and they are more price sensitive (larger absolute value of the coefficient of *BILL*). Recall that respondents in Model 2 are those who did not install water filter systems and respondents in Model 3 are those who installed them. This suggests that more risk averse respondents were more likely to take self-protection measures against the risks and they tend to choose the program to reduce the risks. However, since they already protected themselves from the health risks, they are more price sensitive in their willingness-to-pay for other people's safety. The estimated mean and standard deviation of the interaction term *SQCE3* is found to be significant in Model 3 but not in Model 2, indicating the framing effect of the choice format (two alternatives versus three) might differ across respondents with different risk preferences.

The RPL Model 5 incorporates the predicted latent filter expenditure variable ($\hat{Z}^*$) as a covariate in the means of random parameters. Significant effects are found in *MICI*, *MICD* and *CANI*. These shifters have positive signs, opposite to the estimated coefficients for these attributes. The more spent on water filter systems, the less risk averse against those risks one becomes because he or she is less prone to these risks. This inference is the opposite of our earlier finding based on the estimated RPL Models 2 and 3. The magnitudes of these shifters are small though. This difference may arise because Model 5 controls for endogeneity while Models 2 and 3 do not.

So far, we have discussed Models 1 to 5 for both CL and RPL specifications. These models can be summarized as follows. First, preferences for risk reductions from drinking water are heterogeneous. While some heterogeneity can be explained using socio-demographic information, a significant portion remains unobservable. Second, when the endogeneity between risk preferences and the self-protection decision is controlled, the substitution effect of purchased bottled water for the public good program (that improves drinking water quality) is supported.

3.5.2 *Welfare Estimates*

The marginal WTP for a risk reduction is an individual's willingness to pay for a unit reduction in the level of a specific health risk. Marginal WTP estimates for reductions in four different types of risk (*MICI, MICD, CANI* and *CAND*) are derived from the estimated CL Models 1 to 5 and RPL Models 2, 3 and 5. They are reported in Table 3.7. Standard deviations of the WTP estimates are based on Krinsky-Robb simulations (Krinsky and Robb 1986). The WTP estimates derived from the CL specification are positive and significant at the 1% level. So are the WTP estimates derived from the RPL. In fact, the mean WTP estimates derived from CL models are close to their RPL model counterparts, although the former are slightly larger than the latter with two exceptions (WTP3 and $WTP5_{Altrm}$ for *MICD*).

Table 3.7 indicates that the WTP estimates differ by health risk outcomes (mortality versus morbidity) and cause (microbial versus cancer). If we take WTP1 as an example, then the marginal WTP for a reduction in the risk of death is much higher than that for a reduction in the risk of an illness. Moreover, the WTP to avoid a cancer illness is substantially higher than that for microbial illness.

Turning to the main hypothesis for this paper we first compare WTP estimates derived from Models 1 - 4. The results are consistent with our hypothesis. That is, in general, WTP1 > WTP2 > WTP3 > WTP4. This order holds across all types of health risk reductions, which gives us more confidence in decomposing the WTP estimates according to motive. WTP1 and WTP2 are both considered to be total WTP that includes self-interested WTP and altruistic WTP. The fact that WTP1 is generally higher than WTP2 supports our hypothesis that purchased water is a substitute for safe drinking water and individuals' WTP for the program decreases with a reduction in the level of risks. WTP3 and WTP4 are considered to be only altruistically motivated if we assume away the existence of private benefits on cost savings from a reduced need to self protect in the future. Note that in terms of magnitude, WTP4 estimates are lower than their WTP3 counterparts, suggesting that altruistic WTP for the public program decreases as one's self-protection level is higher.

Since WTP1 > WTP2 > WTP3 > WTP4, our subtraction method can be used to decompose the total WTP values. Two versions of self-interested values are derived by subtracting WTP4 from WTP1 and by subtracting WTP3 from WTP2. The results are shown in Table 3.8. The difference between WTP4 and WTP1, representing a measure of self-interested WTP, is larger than the difference between WTP3 and WTP2. We can consider them as upper and lower bounds for self-interested WTP. The WTP estimates derived from Model 5 also support our hypothesis that $WTP5_{Total}$ is greater than $WTP5_{Altrm}$ for all types of health risks. Self-interested WTP ($WTP5_{Self}$) is the difference between $WTP5_{Total}$ and $WTP5_{Altrm}$ (Table 3.8). However, these differences in WTP estimates are not statistically significant (standard deviations of self-interested WTP estimates are derived using Krinsky-Robb simulations and are presented in Table 3.8), probably due to limited sample size. However, given the fact that both total and altruistic WTP estimates are found to be positive and significant in this study, our finding of the decreasing order of WTP estimates from total WTP estimates to altruistic WTP estimates supports our hypotheses.. Therefore, we still apply the subtraction method to derive a crude measure of self-interested WTP and compare them with other published estimates. Nonetheless, the magnitude of these self-interested WTP estimates themselves should not be used for policy making purposes.

Examination of WTP estimates derived from the RPL Models 2 , 3 and 5 indicates that estimated total WTP is greater than altruistic WTP with one exception (i.e., for MICI, WTP2 < WTP3). Therefore, in general, both RPL models and CL models support our hypothesis. Consequently, self-interested WTP estimates can be calculated from the pairs of WTP estimates since our hypothesis is supported.

Table 3.8 presents three pairs of self-interested WTP and altruistic WTP based on, respectively, WTP1 and WTP 4, WTP2 and WTP3 and $WTP5_{Total}$ and $WTP5_{Altrm}$. For each pair, the percentage of self-interested WTP relative to total WTP is also presented. Based on "cleaner or stricter samples", either completely no self-protection (WTP1) or definitely no exposure to the health risks (WTP4), self-interested WTP ranges from 35% to 63% of one's total WTP. Based on loosely segmented samples (depending on whether *FEXP* = 0 or not), self-interested WTP ranges from almost 1% to 36% of one's total WTP. Based on levels of expenditure on water filter systems, self-interested WTP ranges from 12% to 25% of one's total WTP. The third interval is much narrower and the upper bound is less than half of the first two intervals as a result of using instrumental variables in the interaction method. The $WTP5_{Total}$ is calculated at the mean of the latent expenditure variable $\hat{Z}^*$ of the sub-sample satisfying $FEXP = 0$ ($\overline{\hat{Z}^*}|_{FEXP=0}$) while $WTP5_{Altrm}$ is calculated at the mean of $\hat{Z}^*$ of the sub-sample satisfying $FEXP > 0$ ($\overline{\hat{Z}^*}|_{FEXP>0}$). The mean difference in $\overline{\hat{Z}^*}$ is smaller than the mean difference in *FEXP* for the $FEXP = 0$ group and the $FEXP > 0$ group. Consequently, $WTP5_{Self}$ derived from Model 5 accounts for a much smaller proportion of total WTP. Therefore, we are not as confident in WTP_{Self} derived from the interaction method as from the sample segmentation method. Based on the latter, altruistic WTP accounts for about 37% to 99% of total WTP. Table 3.8 also reports self-interested WTP and altruistic WTP values derived from RPL models. Self-interested WTP estimates are only calculated for pairs of WTP estimates supporting our hypothesis. General speaking, the self-interested WTP estimates account for a smaller proportion of total WTP based on the RPL specification as compared to the CL specification. The self-interested WTP5 estimates given the RPL specification seem to be quite different from the rest. Incorporating heterogeneity seems to affect the proportion of the self-interested WTP of total WTP for risk reductions in microbial deaths the most: 36% of total WTP indicated by the CL models versus 22% by the RPL models based on WTP2 and WTP3 estimates.

In summary, WTP estimates from both the sample segmentation method and the interaction method using the CL specification support our hypothesis that WTP for a public good consists of two parts: a self-interest motivated part and an altruistically motivated part.[36] However, the relative magnitude of the two parts differs substantially

[36] After each choice question, respondents were asked to indicate whether they would increase their future tap water consumption if the program they voted for was implemented. Positive responses indicate respondents' choice decisions are motivated by self-interest even if they installed filter system and had non-zero expenditures on purchased bottled water at home. To assess the robustness of our results, these responses were used to calibrate the sub-samples (Appendix Table A3.1.1) and the sample segmentation method was reapplied using the future consumption calibrated sub-samples. About 10% of the responses are re-categorized as a result. Based on the calibrated sub-samples, the hypothesis of the decreasing order of WTP values from sub-sample 1 to sub-samples 4 (WTP1 to WTP4) cannot be rejected in most cases. The self-interested WTP1 for mortality risk reductions are about 20% less than their pre-calibration counterparts and most of other estimates are about 0-10% less than their original counterparts (except for WTP4 for CANI). The interaction method cannot be easily applied using the calibrated sample because of

depending on how we distinguish motives and how we account for the difference in motivations. Overall, WTP estimates from CL models seem more robust, and will be used for further analysis. The sample segmentation method, compared to the interaction method seems to provide more plausible results in deriving the decomposition of WTP estimates.

3.5.3 Values of Statistical Life and Values of Statistical Illness

Willingness-to-Pay (WTP) estimates for health risk reductions are widely used to derive values of statistical life (VSL) or values of statistical illness (VSI). In this paper, the estimated WTP is considered to be a household WTP. We need to convert the household WTP estimates into individual estimates before we calculate VSLs or VSIs. The average number of people in a household in Canada is 2.6 (Statistics Canada 2005). Recall that the risk level is derived based on a 35 year-period within a community of 100,000 people. Therefore, a VSL due to microbial death, is CAN\$26.8 million ($19.875 \times 100{,}000 \times 35 / 2.6$), according to WTP1, or CAN\$ 21 million according to WTP2 or CAN\$19 million, according to $WTP5_{Total}$. These and subsequent values are in \$2004 constant dollars.

Depending on the type of altruism, the VSLs may differ substantially. If we assume paternalistic altruism, WTP_{Total} should be used to derive VSL estimates. We call this type of VSL (VSL (P)). If we assume non-paternalistic altruism, then according to our assumption in Equation 5, the WTP_{Self} should be used for the calculation of VSL. This type of VSL is called VSL(NP). Table 3.9 reports the paternalistic and non-paternalistic VSL and VSI estimates based on the three pairs of WTP estimates from the CL specifications in Table 3.8. These values differ substantially depending on the cause of death or illness and the nature of altruism. Paternalistic VSLs and VSIs are greater than their non-paternalistic counterparts by the altruistic component of WTP. For microbial death, the paternalistic VSL estimates vary from CAN\$19 million to CAN\$26.8 million, while the non-paternalistic VSLs vary from CAN\$4.8 million to CAN\$13.4 million. For cancer death, the paternalistic VSL estimates vary from

---

collinearity problems (only those who voted for the program can answer this question). Based on this sensitivity analysis we feel that the results presented here are relatively robust.

CAN$14.4 million to CAN$22.5 million, while the non-paternalistic VSLs vary from CAN$1.6 million to CAN$13.1 million.[37]

In the health risk valuation literature, extensive efforts have been undertaken to estimate VSLs (Viscusi and Aldy 2003); however, most estimates are based on WTP for reducing private risks rather than public risks. As Strand (2004) points out in a policy context, VSLs apply to public goods. Our "good" or risk change certainly has public good characteristics. Compared to these studies, our paternalistic results seem to be on the upper bound of these published estimates. Costa and Kahn (2003) suggest that it is likely that the value of risk reductions has increased over time as per capita income increases. Our estimates reflect that increase. On the other hand, the non-paternalistic VSLs are very similar to these published estimates derived in a private risk reduction context. For example, based on the WTP2 and WTP3 estimates, the estimated VSLs are between $4.9 and $7.6 million depending on the causes of deaths (microbial death versus cancer death) and the estimated VSIs for risk reductions in microbial illnesses are $135 and for risk reductions in cancer illnesses are $1.2 million (Table 3.9). If paternalistic VSLs are considered to be public VSLs, and non-paternalistic VSLs are considered to be private VSLs, our estimates of private VSLs are consistent with current VSL estimates based on market data. However, as we discussed earlier, non-paternalistic VSLs may also be one type of public VSL. It is a matter of how to account for altruism empirically. In fact, a society is more likely to be composed of citizens with heterogeneity over the degree and type of altruism. The public VSLs are likely to fall in between the interval of our paternalistic VSLs and non-paternalistic VSLs.

As mentioned earlier, non-paternalistic VSL estimates are equal to self-interested VSL estimates, which are calculated based on self-interested WTP. Self-interested WTP is derived by subtracting altruistic WTP from total WTP. However, it has to be pointed out that our altruistic WTP might be overestimated because there also may be joint benefits associated with consuming purchased water, such as the taste, convenience and the lack of odour. Further analysis is needed to account for the jointness in benefits. Therefore, our non-paternalistic VSLs might be underestimated. The non-paternalistic

[37] Based on the future consumption calibrated sample (see footnote 36), the paternalistic VSLs vary from $15 million to $21 million and the non-paternalistic VSLs vary from $3.4 million to $7 million regardless of the type of mortality risk (Appendix Table A3.1.3).

VSL for microbial death are about 25% - 50% that of the paternalistic VSL, while the non-paternalistic VSL for cancer death is 11% to 58% of that of the paternalistic VSL. Jones-Lee (1992) argues that the VSL for a "caring" society will be some 10% to 40% larger than the value that would be appropriate for a society of purely self-interested individuals. Our finding shows that a society of individuals who are paternalistically altruistic value people's lives about 40% to 90% more than a society of purely self-interested individuals or a society of individuals who are non-paternalistically altruistic.

3.6 Conclusions

This paper is an empirical study of how VSLs and VSIs are affected by altruism. Variables indicating whether an individual has taken self-protection measures against risks are used to distinguish the demand for the public good by individuals who are solely altruistically driven from those who might have both self-interested and altruistic motivations. We decompose total WTP by different motives and estimate the magnitude of altruistic values in the demand for a public good.

Two different approaches (sample segmentation and interaction) are adopted to test the hypothesis that total WTP is greater than its altruistic component WTP. Results from both methods, in general, support our hypothesis. Self-protection against health risk decreases one's willingness-to-pay for a public program to reduce the risk. The behavioural information on self-protection can be used to differentiate demand for the program with different motives. There is also significant heterogeneity in preferences for risk reductions. Although the results derived from models based on the interaction approach might be subject to more measurement errors, they indicate that endogeneity might be an issue in this study. The endogeneity issue in choice modelling has started to receive more attention (Louviere et al. 2005) and our examination indicates there is a need to control for endogeneity in this study. For example, comparing Model 2 and Model 3 without taking into account risk endogeneity suggests that more risk averse respondents were more likely to take self-protection measures against the risks and they are willing to pay more to reduce the risks. However, they are more price sensitive in their willingness-to-pay for other people's safety. Taking endogeneity into account

(Model 5), it is found that the more spent on water filter systems, the less risk averse against those risks one becomes because he or she is less prone to these risks. The contradictory inference with and without accounting for endogeneity indicates there is a need to control for endogeneity in this study.

Our WTP and VSL estimates suggest that altruism plays a significant role in people's valuation of health risk reductions: 40% to 90% of an individual's total WTP for a marginal risk reduction appears to arise from feelings of altruism. However, our results are derived based on the assumption that perceived health risks from tap drinking water were reduced to zero if water filter systems were installed and we also assume reducing health risks was the main reason for installing the filter systems at home. While one may incline to challenge our assumption, this is an empirical question that has not been adequately addressed in the literature when using expenditure data. That is, as researchers, we can never be sure about the reasons behind real expenditures. As such, it is possible that our altruistic values are over-estimated. Moreover, the effect of altruism on WTP or VSL estimates depends on the nature of assumptions made about the form of altruism: paternalistic or non-paternalistic. If paternalistic altruism is assumed, the VSL estimates of an altruistic society is about 40% to 90% more than that of a society comprised of purely self-interested individuals. In contrast, if non-paternalistic altruism is assumed, the VSLs of an altruistic society are the same as that of a society of purely self-interested individuals. This finding indicates that a society with paternalistic altruism necessarily values life higher than a society with non-paternalistic altruism. As non-paternalistic and paternalistic VSLs differ substantially, which value should we use in public policy decisions? According to McConnell (1997), paternalistic motives are more likely to be present in the valuation of natural resources. In this study, we find that people are willing to pay for other people's safe drinking water, which supports McConnell's view about the context in which paternalistic altruism is more plausible. However, since this study does not distinguish between the types of altruism at the conceptual level, our evidence could be *ad hoc*. Further research on understanding and testing for the nature of altruism is thus needed.

References

Adamowicz, W. L., P. Boxall, M. Williams and J. Louviere. 1998. Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics* 80(1): 64-75.

Adamowicz, W. L., D. Dupont and A. Krupnick. 2005. *Willingness to Pay to Reduce Community Health Risks from Municipal Drinking Water: A Stated Preference Study*, Report for the U.S. Environmental Protection Agency, 62pp.

Alberini, A., M. Cropper, A. Krupnick and N. B. Simon. 2004. Does the value of a statistical life vary with age and health status? Evidence from the US and Canada. *Journal of Environmental Economics and Management* 48(1): 769-792.

Andreoni, J. 1989. Giving with impure altruism: applications to charity and Ricardian equivalence. *The Journal of Political Economy* 97(6): 1447-1458.

Andreoni, J. 1990. Impure altruism and donations to public goods: a theory of warming glow giving. *The Economic Journal* 100: 464-477.

Costa, D. L. and M. E. Kahn. 2003. The rising price of nonmarket goods. *American Economic Review* 93(2): 227-232.

Deacon, R. and P. Shapiro. 1975. Private preference for collective goods revealed through voting on referenda. *American Economic Review* 65(5): 943-955.

Dhar, R. 1997. Context and task effects on choice deferral. *Marketing Letters* 8(1): 119-130.

Flores, N. E. 2002. Non-paternalistic altruism and welfare economics. *Journal of Public Economics* 83: 293-305.

Greene, W. 2003. *Econometric Analysis*. 5 ed., Pearson Education Inc.

Holmes, T. 1990. Self-interest, altruism, and health-risk reduction: an economic analysis of voting behavior. *Land Economics* 66(2): 140-149.

Holmes, T. and W. L. Adamowicz. 2002. Attributes based methods. In: *A Primer on the Nonmarket Valuation.* (eds.) P. Champ, T. Brown and K. Boyle. 171-174. Kluwer: Amsterdam.

Hu, W. 2004. *Three Essays on Genetically Modified Food Labelling and Consumer Behaviour,* PhD Dissertation. Department of Rural Economy, University of Alberta.

Johansson, O. 1997. Optimal Pigovian taxes under altruism. *Land Economics* 73(3): 297-308.

Johansson, Per-Olov. 1994. Altruism and the value of statistical life: empirical implications. *Journal of Health Economics* 13: 111-118.

Jones-Lee, M. W. 1991. Altruism and the value of other people's safety. *Journal of Risk and Uncertainty* 91(4): 213-219.

Jones-Lee, M. W. 1992. Paternalistic altruism and the value of statistical life. *Economical Journal* 102: 80-90.

Kahneman, D. and J. L. Knetsch. 1992. Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22(1): 57-70.

Kopp, R. J. 1992. *Ethical Motivations and Nonuse Value*, Resources for the Future, Washington.

Koop, G. and L. Tole. 2004. Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? *Journal of Environmental Economics and Management* 47: 30-54.

Krinsky, I. and A. L. Robb. 1986. On approximating the statistical properties of elasticities. *The Review of Economics and Statistics* 68(4): 715-719.

Krutilla, J. V. 1967. Conservation reconsidered. *American Economic Review* 57(4): 777-786.

Lazo, J. K., G. H. McClelland and W. D. Schulze. 1997. Economic theory and psychology of non-use values. *Land Economics* 73(3): 358-371.

Louviere, J., K. Train, M. Ben-Akiva, C. Bhat, D. Brownstone, T. A. Cameron, R. Carson, J. DeShazo, D. Fiebig, W. Greene, D. Hensher and D. Waldman. 2005. Recent progress on endogeneity in choice modelling. *Marketing Letters* 16(3): 255-265.

McConnell, K. E. 1997. Does altruism undermine existence value? *Journal of Environmental Economics and Management* 32: 22-37.

Mills, C. J., R. J. Bull, K. P. Cantor, J. Reif, S. E. Hrudey and P. Huston. 1998. Workshop report. health risks of drinking water chlorination by-products: report of an expert working group. *Chronic Diseases in Canada* 19(3): 91-102.

Mitchell, R.C. and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Resources for the Future, Washington.

Piliavin, J. A. and Hong-Wen Charng. 1990. Altruism: a review of recent theory and research. *Annual Review of Sociology* 16: 27-65.

Popp, D. 2001. Altruism and the demand for environmental quality. *Land Economics* 77(3): 339-349.

Scarpa, R., S. Ferrini and K. G. Willis. 2005. Performance of error component models for status-quo effects in choice experiments. In: *Applications of Simulation Methods in Environmental and Resource Economics.* 247-274. Springer Publisher: Dordrecht, The Netherlands.

Statistics Canada. 2005. *Household Size, by Province and Territory (2001 Census).* Statistics Canada.

Strand, J. 2004. Public and private-good values of statistical lives: results from a combined choice-experiment and contingent valuation survey. [accessed 20 May 2006]. Available at http://folk.uio.no/jostrand/lifepaper02.pdf.

Viscusi, W. K. and J. Aldy. 2003. The value of a statistical life: a critical review of market estimates throughout the world. *The Journal of Risk and Uncertainty* 27(1): 5-76.

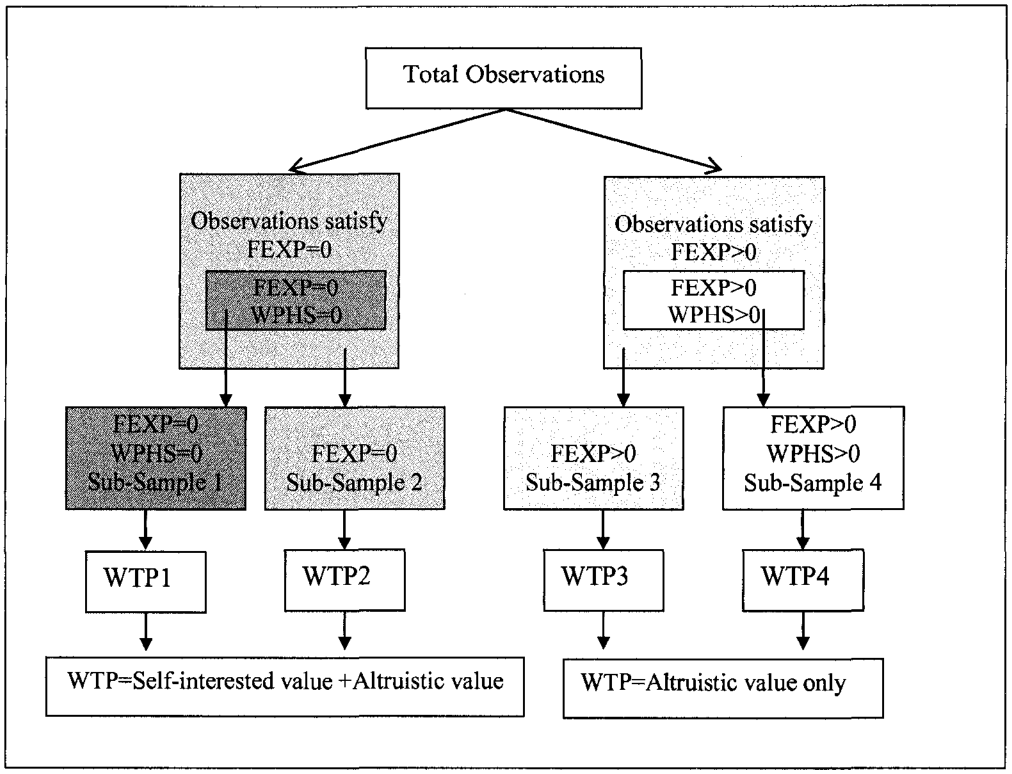


Figure 3.1 Sample Segmentation Based on Actual Household Expenditures on Self-protection against Health Risks from Tap Drinking Water

Notes: *FEXP* is annual household expenditure on installing and maintaining water filter systems at home and WPHS is annual household expenditure on purchasing bottled water consumed at home. Sub-sample 1 is a subset of Sub-sample 2, and Sub-sample 4 is a subset of Sub-sample 3. The health risks from drinking water are considered to decrease from Sub-sample 1 to Sub-sample 4. It is hypothesized that WTP1 > WTP2 > WTP3 > WTP4.

Table 3.1 Definition of Variables

Variable	Definition	Interaction Term
SQ	Dummy variable, equals 1 if an alternative is the status quo option and 0 otherwise.	*F_SQ = FEXP*SQ*
MICI	Number of microbial illness cases over a 35-year period from drinking tap water in the community.	*FMICI = FEXP*MICI*
MICD	Number of deaths due to microbial illnesses over a 35-year period from drinking tap water in the community.	*FMICD = FEXP*MICD*
CANI	Number of cancer cases over a 35-year period from drinking tap water in the community.	*FCANI = FEXP*CANI*
CAND	Number of cancer deaths over a 35-year period from drinking tap water in the community.	*FCAND = FEXP*CAND*
BILL	The increase in the current water bill.	*FBILL = FEXP*BILL*
CE3	1 if an individual is faced with a choice set of 3 alternatives, and 0 if faced with a choice set of 2 alternatives.	
SQCE3	Interaction between SQ and CE3.	
	Socio-demographic variables	
AGE65	Dummy variable, equals 1 if an individual is equal to or over 65 years old and 0 otherwise.	*AGE65_SQ =AGE65*SQ*
ILLNESS	1 if an individual has ever being ill due to drinking tap water, 0 otherwise	*ILL_SQ = ILLNESS*SQ*
MARRY	1 if an individual is married and 0 otherwise.	*MARRY_SQ = MARRY*SQ*
MALE	1 if male, and 0 otherwise.	*MALE_SQ = MALE*SQ*
INCOME	Annual after-tax income of a household	*INCM_SQ = INCOME*SQ*
INCOME2	Squared household income.	*INCM2_SQ =INCOME2*SQ*
KID06	1 if a household has kid(s) under 6, and 0 otherwise.	*KID06_SQ = KID06*SQ*
KID612	1 if a household has kid(s) aged between 6 and 12, and 0 otherwise.	*KID612_SQ = KID612*SQ*
KID137	1 if a household has kid(s) aged between 13 and 17, and 0 otherwise.	*KID137_SQ = KID137*SQ*
CITYSIZE	Categorical variables from 1 to 6, ranging from 1 denoting 1000,000 plus to 6 denoting 1499 and under.	*CITYSZ_SQ = CITYSIZE*SQ*
ENGLISH	1 if English is the corresponding language for a respondent and 0 otherwise.	*ENG_SQ = ENGLISH*SQ*
INDXA	Index of number of health problems an individual has experienced, such as food allergies and heart diseases.	*INDXA_SQ = INDXA*SQ*
INDXB	Index of number of health problems of the household members of an individual.	*INDXB_SQ = INDXB*SQ*
	Expenditure on Self-protection	
FEXP	Annual household expenditure on installing and maintaining water filter system at home.	
WPHS	Annual expenditure on purchasing bottled water consumed at home.	

Table 3.2 A Summary of Filter Expenditure Variable (FEXP) and Expenditure on Bottled Water Purchases (WPHS) in each Sub-sample

Samples	Number of Observations	*FEXP* Mean	*FEXP* Standard Deviation	*WPHS* Mean	*WPHS* Standard Deviation
Sub-Sample 1 (*FEXP* = 0 and *WPHS* =0)	432	0	0	0	0
Sub-Sample 2 (*FEXP* = 0)	828	0	0	106.84	172.87
Sub-Sample 3 (*FEXP* > 0)	636	89.49	146.32	86.94	149.06
Sub-Sample 4 (*FEXP* > 0 and *WPHS* > 0)	292	79.34	130.26	189.37	170.35
All observations	1464	38.88	106.12	98.20	163.20

Notes: *FEXP* is annual household expenditure on installing and maintaining water filter systems at home and *WPHS* is annual household expenditure on purchasing bottled water consumed at home. All expenditures are in 2004 Canadian dollars.

Table 3.3 Estimated Conditional Logit Models for each Sub-Sample

	Model 1 (Sub-Sample 1)	Model 2 (Sub-Sample 2)	Model 3 (Sub-Sample 3)	Model 4 (Sub-Sample 4)
SQ	1.308**	1.325**	0.959**	1.548*
	(2.100)	(3.501)	(1.971)	(1.787)
MICI	-6.81E-05**	-7.14E-05**	-9.54E-05**	-1.12E-04**
	(-5.347)	(-7.990)	(-8.836)	(-6.734)
MICD	-0.059**	-0.068**	-0.059**	-0.072**
	(-3.214)	(-5.271)	(-3.977)	(-3.183)
CANI	-0.014**	-0.013**	-0.012**	-0.013**
	(-3.943)	(-4.914)	(-4.063)	(-2.721)
CAND	-0.049**	-0.053**	-0.050**	-0.050**
	(-2.832)	(-4.283)	(-3.568)	(-2.350)
BILL	-2.96E-03**	-0.004**	-0.006**	-0.007**
	(-3.145)	(-6.415)	(-6.946)	(-5.612)
ILL_SQ	-2.378**	-1.357**	2.19E-01	7.63E-01
	(-2.019)	(-2.212)	(0.466)	(1.076)
INCM_SQ	3.31E-05**	1.72E-05**	-2.62E-06	-1.57E-05
	(2.963)	(2.065)	(-0.257)	(-0.917)
INCM2_SQ	-2.42E-10**	-1.61E-10**	1.65E-11	4.47E-11
	(-3.256)	(-3.047)	(0.244)	(0.419)
ENG_SQ	0.021	0.180	-0.180	0.511
	(0.085)	(1.009)	(-0.673)	(1.059)
AGE65_SQ	-0.588*	-0.809**	-0.246	0.227
	(-1.843)	(-3.298)	(-0.817)	(0.387)
CITYSZ_SQ	-0.283**	-0.264**	-0.171**	-0.105
	(-3.209)	(-4.415)	(-2.358)	(-0.835)
K06_SQ	-0.395	-0.346	-0.065	-0.130
	(-0.504)	(-0.911)	(-0.159)	(-0.242)
K612_SQ	0.408	0.198	-0.199	-0.345
	(0.863)	(0.609)	(-0.561)	(-0.749)
K137_SQ	0.125	0.327	0.714	0.312
	(0.298)	(0.968)	(1.656)	(0.427)
MALE_SQ	0.364	0.579**	0.515**	-0.032
	(1.507)	(3.515)	(2.512)	(-0.098)
MARRY_SQ	0.277	0.131	0.648**	0.551
	(1.065)	(0.686)	(3.067)	(1.625)
INDXA_SQ	0.109	0.010	0.101	-0.064
	(1.661)	(0.214)	(1.482)	(-0.521)
INDXB _SQ	-0.083	-0.053	-0.027	-0.210**
	(-1.054)	(-0.963)	(-0.468)	(-1.962)
SQCE3	-0.448*	-0.520**	-0.721**	-0.539*
	(-1.842)	(-3.104)	(-3.644)	(-1.69)
Number of Obs.	432	828	636	292
Log-likelihood	-307.982	-601.383	-466.437	-202.438

Notes: t-ratios are in parentheses. ** denotes the 5% level, and * denotes the 10% level.

Table 3.4 Results of Tobit Regressions on the Filter Expenditure Variable (*FEXP*)

Variable	Coefficient	Marginal effect
CONSTANT	-140.984** (-2.856)	-51.966** (-2.982)
ILLNESS	33.759 (0.571)	12.443 (0.571)
AGE65	19.217 (0.577)	7.083 (0.576)
WPHS	-0.141* (-1.925)	-0.052* (-1.926)
MARRY	-11.019 (-0.447)	-4.061 (-0.448)
INCOME	5.03E-04 (0.430)	1.85E-04 (0.430)
INCOME2	-3.73E-09 (-0.492)	-1.38E-09 (-0.492)
ENGLISH	110.962** (3.921)	40.900** (3.968)
CITYSIZE	-5.467 (-0.660)	-2.015 (-0.66)
KID06	12.256 (0.250)	4.518 (0.25)
KID612	90.818** (2.189)	33.475** (2.178)
KID137	-23.663 (-0.501)	-8.722 (-0.502)
INDXA	-4.532 (-0.654)	-1.671 (-0.654)
INDXB	11.190 (1.651)	4.124 (1.652)
SIGMA	179.019 (16.611)	-
Scale factor for marginal effect		0.368
Number of observations	366	
Log-likelihood	-1161.34	

Note: *FEXP* is annual household expenditure on installing and maintaining water filter systems at home.

Table 3.5 Estimated Conditional Logit Model – the Interaction Method

Variable	Interacting with Z*	Interacting with *FEXP*	Variable (Cont'd)	Interacting with Z*	Interacting with *FEXP*
SQ	1.685** (4.255)	1.175** (4.054)	*MALE_SQ*	0.573** (4.553)	0.582** (4.633)
MICI	-9.74E-05** (-5.363)	-7.91E-05** (-10.799)	*MARRY_SQ*	0.295** (2.161)	0.289** (2.122)
MICD	-0.115** (-4.414)	-0.056** (-5.441)	*INDXA_SQ*	0.027 (0.714)	0.030 (0.806)
CANI	-0.014** (-2.709)	-0.012** (-5.933)	*INDXB_SQ*	-0.045 (-0.983)	-0.047 (-1.216)
CAND	-0.0541** (-2.192)	-0.048** (-4.801)	*F_SQ*	-0.011 (-1.333)	-0.001 (-1.219)
BILL	-2.73E-03* (-1.919)	-4.49E-03** (-7.928)	*FMICI*	3.52E-07 (1.018)	-2.81E-08 (-0.327)
SQCE3	-0587** (-4.686)	-0.588** (-4.688)	*FMICD*	1.07E-03** (2.152)	-1.93E-04* (-1.704)
ILL_SQ	-0.342 (-1.025)	-0.316 (-0.905)	*FCANI*	3.25E-05 (0.333)	-2.12E-06 (-0.102)
INCM_SQ	1.22E-05* (1.938)	1.24E-05** (1.979)	*FCAND*	6.30E-05 (0.135)	-9.59E-05 (-0.905)
INCM2_SQ	-1.14E-10** (-2.778)	-1.14E-10** (-2.814)	*FBILL*	-4.69E-05* (-1.665)	-1.70E-05** (-2.083)
ENG_SQ	0.014 (0.052)	0.006 (0.041)			
AGE65_SQ	-0.590** (-3.160)	-0.596** (-3.191)			
CITYSZ_SQ	-0.231** (-4.915)	-0.226** (-5.005)			
K06_SQ	-0.297 (-1.095)	-0.310 (-1.152)			
K612_SQ	0.017 (0.051)	-0.030 (-0.129)			
K137_SQ	0.415 (1.586)	0.412 (1.590)			
Number of Obs.	1464	1464	Log-likelihood	-1079.19	-1079.99

Note: Z* is an instrumental variable for *FEXP*, annual household expenditure on installing and maintaining water filter systems at home.

Table 3.6 Estimated Random Parameters Logit Models

Variable	Model 2	Model 3	Model 5
SQ	2.064**	1.953**	2.639**
	(4.783)	(4.043)	(6.293)
SQ_SD	2.183**	1.442**	2.641**
	(4.948)	(3.112)	(8.728)
MICI	-1.23E-04**	-1.96E-04**	-1.67E-04**
	(-5.963)	(-5.588)	(-6.769)
MICI_SD	1.16E-04**	1.32E-04**	1.29E-04**
	(3.645)	(3.800)	(4.851)
MICD	-0.118**	-0.147**	-0.186**
	(-4.730)	(-3.925)	(-5.762)
MICD_SD	0.093	0.198**	0.124**
	(1.699)	(3.220)	(3.397)
CANI	-0.087**	-0.097**	-0.033**
	(-3.911)	(-3.221)	(-5.076)
CANI_SD	0.102**	0.148**	0.030**
	(2.442)	(3.250)	(4.541)
CAND	-0.021**	-0.021**	-0.093**
	(-3.807)	(-3.195)	(-3.217)
CAND_SD	0.028**	0.033**	0.139**
	(3.214)	(3.479)	(4.243)
SQCE3	-1.290	-1.660**	-1.288*
	(-2.406)	(-2.866)	(-1.723)
SQCE3_SD	2.271**	2.950**	-
	(3.671)	(3.711)	
BILL	-0.008**	-0.012**	-0.009**
	(-5.503)	(-5.536)	(-7.709)
F_SQ	-	-	-0.017
			(-1.303)
FMICI	-	-	4.18E-07**
			(0.822)
FMICD	-	-	1.43E-03**
			(2.286)
FCANI	-	-	3.62E-04**
			(2.763)
FCAND	-	-	1.55E-04
			(0.244)
Number of Obs.	828	636	1464
Log-likelihood	-555.88	-429.65	-992.13

Notes: t-ratios are in parentheses. ** denotes the 5% level, and * denotes the 10% level. A model estimating *SQCE3* as a random parameter was attempted, but the model was failed to converge. The expenditure interaction terms are between the predicted latent filter expenditure variable Z* and attributes. "_SD" denotes the standard deviation of a variable, e.g., *SQ_SD* denotes the estimated standard deviation for *SQ*.

Table 3.7 Marginal Willingness-to-Pay Estimates

Attributes	Sample Segmentation Method				Interaction Method	
	(FEXP = 0 & *WPHS* =0)	*(FEXP* = 0)	*(FEXP* > 0)	*(FEXP* > 0 & *WPHS* >0)	Full sample	
	WTP1	WTP2	WTP3	WTP4	$WTP5_{Total}$	$WTP5_{Altrm}$
			The conditional logit specification			
MICI	0.023**	0.01634**	0.01629 **	0.015**	0.017**	0.015**
	(0.008)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
MICD	19.875**	15.670**	10.005**	9.753**	14.157**	10.680**
	(8.836)	(3.759)	(2.833)	(3.517)	(2.572)	(2.183)
CANI	4.875**	2.983**	2.079**	1.756**	2.639**	2.325**
	(1.879)	(0.729)	(0.569)	(0.681)	(0.493)	(0.443)
CAND	16.731**	12.238**	8.549**	6.768**	10.714**	9.578**
	(6.844)	(3.045)	(2.391)	(2.783)	(2.062)	(1.858)
			The random parameters logit specification			
MICI	-	0.016**	0.017**	-	0.016**	0.015**
		(0.003)	(0.003)		(0.005)	(0.006)
MICD	-	15.596**	12.238**	-	13.144	11.515
		(3.297)	(3.168)		(10.205)	(7.119)
CANI	-	2.751	1.821	-	1.819	1.406
		(2.929)	(2.566)		(2.405)	(1.463)
CAND	-	11.450**	8.260**	-	9.233**	9.055
		(0.723)	(0.570)		(3.832)	(6.832)

Notes: Standard deviations are in parentheses. ** denotes the 5% level, and * denotes the 10% level. The RPL results for WTP1 and WTP4 are not available because of small sample sizes. *FEXP* is annual household expenditure on installing and maintaining water filter systems at home and *WPHS* is annual household expenditure on purchasing bottled water consumed at home. Different WTP estimates are derived using different sub-samples satisfying conditions that are defined by either *FEXP* alone (greater than zero or not) or both variables when the sample segmentation method is used.

Table 3.8 Self-interested Willingness-to-Pay and Altruistic Willingness-to-Pay Estimates

Health Risks	Based on WTP1 and WTP4			Based on WTP2 and WTP3			Based on $WTP5_{Total}$ and $WTP5_{Altrm}$		
	Self-interested WTP		Altruistic WTP	Self-interested WTP		Altruistic WTP	Self-interested WTP		Altruistic WTP
	$WTP1_{Self}$	% of total WTP (WTP1)	$WTP1_{Altrm}$	$WTP2_{Self}$	% of total WTP (WTP2)	$WTP2_{Altrm}$	$WTP5_{Self}$	% of total WTP ($WTP5_{Total}$)	$WTP5_{Altrm}$
				The conditional logit specification					
MICI	0.008 (0.008)	35%	0.015	0.0001 (0.004)	1%	0.01629	0.002 (0.003)	12%	0.015
MICD	9.951 (9.494)	50%	9.753	5.642 (4.881)	36%	10.005	3.561 (3.341)	25%	10.680
CANI	3.090 (1.988)	63%	1.756	0.924 (0.925)	31%	2.079	0.299 (0.662)	11%	2.325
CAND	9.759 (7.314)	58%	6.768	3.639 (3.748)	30%	8.549	1.173 (2.756)	11%	9.578
				The random parameters logit specification					
MICI	--	--	--	--	--	0.017	0.0009 (0.008)	6%	0.015
MICD	--	--	--	3.370 (4.531)	22%	12.238	1.692 (12.283)	13%	11.515
CANI	--	--	--	0.871 (3.829)	32%	1.821	0.357 (2.786)	20%	1.406
CAND	--	--	--	3.202 (0.889)	28%	8.260	0.278 (7.677)	3%	9.055

Notes: 1) Standard deviations for self-interested WTP estimates are in parentheses. They are derived based on Krinsky-Robb simulation using 2000 draws. 2) $WTP1_{Self}$=WTP1-WTP4, where WTP1 is total WTP and WTP4 is altruistic WTP ($WTP1_{Altrm}$); $WTP2_{Self}$=WTP2-WTP3, where WTP2 is total WTP and WTP3 is altruistic WTP ($WTP2_{Altrm}$).

Table 3.9 Comparison of Values of Statistical Life and Values of Statistical Illness

Health risks	Values of Statistical Life and Values of Statistical Illness Estimates		
	Assuming paternalistic altruism		
	VSL(P) and VSI(P) (based on $WTP1_{Total}$)	VSL(P) and VSI(P) (based on $WTP2_{Total}$)	VSL(P) and VSI(P) (based on $WTP5_{Total}$)
MICI	30,962	21,996	22,885
MICD	26,754,808	21,094,231	19,057,500
CANI	6,562,500	4,015,577	3,552,500
CAND	22,522,500	16,474,231	14,422,692
	Assuming non-paternalistic altruism		
	VSL(NP) and VSI(NP) (based on $WTP1_{Self}$)	VSL(NP) and VSI(NP) (based on $WTP2_{Self}$)	VSL(NP) and VSI(NP) (based on $WTP5_{Self}$)
MICI	10,769	135	2,692
MICD	13,395,577	7,595,000	4,793,654
CANI	4,159,615	1,243,846	402,500
CAND	13,137,115	4,898,654	1,579,031

Notes: VSL(P) and VSI(P) denote values of statistic life and values of statistical illness cases when paternalistic altruism is assumed, and VSL(NP) and VSI(NP) denote values of statistic life and values of statistical illness cases when non-paternalistic altruism is assumed. $WTP1_{Total}$, $WTP2_{Total}$ and $WTP5_{Total}$ are WTP1, WTP2 and $WTP5_{Total}$ from Table 3.7. $WTP1_{Self}$ and $WTP2_{Self}$ and $WTP5_{Self}$ are from Table 3.8.

Appendix 3.1

Table A3.1.1 Comparison of Sub-samples and Future Consumption Calibrated Sub-samples

Samples	Number of Observations	Future Consumption Calibrated Samples	Number of Observations
Sub-Sample 1 (*FEXP*=0 and *WPHS*=0)	432	Sub-Sample 1 (*FEXP*=0 and *WPHS*=0, or *FutureConsume*=1)	535
Sub-Sample 2 (*FEXP*=0)	828	Sub-Sample 2 (*FEXP*=0 or *FutureConsume*=1)	908
Sub-Sample 3 (*FEXP*>0)	636	Sub-Sample 3 (*FEXP*>0 and *FutureConsume*=0)	556
Sub-Sample 4 (*FEXP*>0 and *WPHS*>0)	292	Sub-Sample 4 (*FEXP*>0 and *WPHS*>0, and *FutureConsume*=0)	248
All observations	1464	All observations	1464

Notes: *FEXP* is annual household expenditure on installing and maintaining water filter systems at home and *WPHS* is annual household expenditure on purchasing bottled water consumed at home. *FutureConsume* is a dummy variable indicates whether respondents stated that they would increase their future consumption of tap water if the program they voted for was implemented.

Table A3.1.2 Comparison of Willingness-to-Pay Estimates between the Original Sub-samples and Future Consumption Calibrated Sub-samples

Attribute	WTP1	WTP2	WTP3	WTP4
		Original sub-samples		
MICI	0.023	0.01634	0.01629	0.015
MICD	19.875	15.670	10.005	9.753
CANI	4.875	2.983	2.079	1.756
CAND	16.731	12.238	8.549	6.768
		Future consumption calibrated sub-samples		
MICI	0.021	0.0158	0.0159	0.015
MICD	15.342	13.916	9.974	10.552
CANI	4.219	2.744	1.948	1.407
CAND	12.698	11.137	8.624	7.467

Notes: All estimates are significant at the 1% level. Original sub-samples are defined in Table 3.2. Future consumption calibrated sub-samples are defined in Table A3.1.1.

Table A3.1.3 Comparison of Values of Statistical Life and Values of Statistical Illness Using Future Consumption Calibrated Sub-samples based on the Sample Segmentation Method

Health Risk	Values of Statistical Life and Values of Statistical Illness Estimates	
	Assuming paternalistic altruism	
	VSL(P) and VSI(P) (based on $WTP1_{Total}$)	VSL(P) and VSI(P) (based on $WTP2_{Total}$)
MICI	28,083	21,255
MICD	20,652,146	18,733,267
CANI	5,679,651	3,693,478
CAND	17,093,147	14,992,620
	Assuming non-paternalistic altruism	
	VSL(NP) and VSI(NP) (based on $WTP1_{Self}$)	VSL(NP) and VSI(NP) (based on $WTP2_{Self}$)
MICI	7,891	-156
MICD	6,448,042	5,307,294
CANI	3,786,079	1,071,424
CAND	7,040,890	3,383,530

Notes: VSL(P) and VSI(P) denote values of statistic life and values of statistical illness cases when paternalistic altruism is assumed, and VSL(NP) and VSI(NP) denote values of statistic life and values of statistical illness cases when non-paternalistic altruism is assumed. They are calculated based on WTP estimates presented in Table A3.1.2. Refer to Table 3.9 for the calculation of these estimates. Refer to Table A3.1.1 for the definition of future consumption calibrated sub-samples.

Chapter 4 Why Does Choice Format Affect Preference Elicitation?

4.1 Background

Stated choice methods have been widely used to obtain values of goods that are not traded in the market. In the case of choice experiments, multiple elicitation formats have been used, including "optimal" designs based on the criteria of statistical efficiency and stated preference survey design (Mitchell and Carson 1989; DeShazo 2002). The elicitation format, however, could influence the process of decision-making, causing the elicited preferences to be context-dependent. For example, DeShazo (2002) found that behavioural responses depended on the sequences of willingness to pay (WTP) questions in a double-bounded survey format. A study done by Cameron et al. (2002) found datasets from some elicitation methods can be pooled using proper econometric models while others cannot. When datasets from different elicitation methods cannot be pooled, the notion of a common underlying preference structure must be rejected. The possibility of context-dependent preferences casts a shadow on the validity of stated choice methods (Tversky and Kahneman 1991; Tversky and Simonson 1993; Swait et al. 2002).

In a preliminary analysis of the datasets collected using the 2004 Canadian drinking water survey (hereafter the water survey, which has been introduced in Chapter 1), it was found that the underlying preferences for health risk reductions revealed in two sub-samples administered with different survey formats were different. One adopted a 2-alternative choice format (CE2) and the other adopted a 3-alternative choice format (CE3). Since the water survey employed a random split sample approach, one would expect that preferences inferred from one sub-sample should be no different from another. However, the results show that both the likelihood-ratio test for pooling the datasets using a basic conditional logit specification[38] and a test for common parameters on alternative specific attributes between the two sub-samples are rejected (Appendix 4.1 Tables A4.1.1 and A4.1.2). Statistically significant differences in the estimated marginal

[38] The model is specified with alternative specific attributes only, i.e., it includes *SQ, MICI, MICD, CANI* and *CAND* and *BILL* only. Definitions of these variables are introduced in Chapter 1 (Table 1.1).

utilities for cancer morbidity risk reduction and for income are found in the two choice formats. The inferred willingness-to-pay estimates for cancer morbidity risk reductions from the two datasets are also significantly different. In addition, a χ^2-test for equal frequency of choosing the status quo option in the two sub-samples is also rejected[39] (Appendix 4.1 Tables A4.1.3 and A4.1.4). It implies that, relative to a status quo option, a proposed alternative that is likely to be rejected using a binary choice format could be accepted if a trinary choice format was used instead. Adding a third alternative to a choice set seems to make consumers more likely to move away from a status quo. This two-versus-three choice format effect has also been found in a few other studies (Adamowicz, Dupont and Krupnick 2005; Alevy, List and Adamowicz 2006). This chapter explores why consumers' stated preferences for risk reductions in drinking water are influenced by choice formats differing in the number of alternatives. If the choice format systematically affects consumers' decisions, it is important to control for its impact on preference elicitation. Identifying factors underlying the phenomenon enables us to predict changes in responses caused by a change of choice format, and we are more likely to reveal "stable and innate" preferences (McFadden 2001).

In the literature on choice modelling, there has been extensive research on the influence of choice formats, or more generally, choice environment on willingness to pay (WTP) responses (see for example, Cameron et al. 2002; DeShazo and Fermo 2002; Breffle and Rowe 2002; Caussade et al. 2005). But these studies seem to offer insufficient explanations for the choice format effect found in this study. For example, studies on the effects of elicitation methods on WTP responses mainly focus on comparisons across more dissimilar methods, such as across an open-ended contingent valuation method (CVM), a discrete choice CVM and/or CVM with a payment card (Halvorsen and Sælensminde 1998; Cameron et al. 2002); between revealed preference data and stated preference data (for example, Ben-Akiva et al. 1994; Cameron et al. 2002), or across rating data, ranking data, attitudinal data and choice data (Louviere et al. 1999; Hensher, Louviere and Swait 1999; Layton and Lee 2006).

[39] The status quo option characterizes a baseline condition, and it is kept the same across all choice tasks in all survey formats. The χ^2 test is conducted based on a contingence table.

Studies comparing WTP estimates within an elicitation method, such as the stated choice approach, have been fewer, mostly focusing on addressing more general issues, such as how design dimensions of a choice experiment affect decision-making. The number of alternatives is often one of dimensions of the experimental design. Other dimensions include the number of attributes, correlation structure between attributes and the number of choice tasks (DeShazo and Fermo 2002; Swait and Adamowicz 2001b; Caussade et al. 2005). For example, Caussade et al. (2005) find that variance in responses differs significantly only when the number of alternatives in a choice set increases from three to four (the choice sets in their study vary from two to nine alternatives) [40]. Alternatively, researchers are concerned about how overall choice environment in terms of task complexity affects choice decisions. For example, measuring complexity as the number of single and multiple attribute changes in the alternatives differing from a status quo option, Moon et al. (2004) find that increased complexity increases the probability the status quo option being chosen in a choice experiment. Swait and Adamowicz (2001a), measuring complexity as entropy based on information theory, also suggest that increased complexity may also make a status quo option more attractive. This is the opposite of what we find in this study. In the same study, they find that as a choice task gets more complex, the variance of responses would first decrease and then increase, which is often referred to as an inverted U shape relationship between response variance and number of alternatives. Mazzotta and Opaluch (1995) also report a similar finding. Results from a few recent studies, e.g., Hensher (2006) and Carlsson and Martinsson (2007), however, suggest that the number of choice sets does not have a significant effect on WTP responses.

Overall, these studies have mainly relied on one-dimensional quantification of the effect of the number of alternatives on decision-making. They offer limited insights for the two-versus-three choice format phenomenon, in which only a discrete change in the number of alternatives is involved. Other researchers have started to examine qualitative aspects of the issue with the insights from behavioural and psychological research. For example, respondents may adopt different decision rules under different choice formats at

[40] In their experiment, choice sets do not contain a status quo option.

different stages of decision-making. Different choice formats are likely to be associated with different degrees of process heterogeneity (Johnson et al. 2006).

Since the early 1980s, researchers in marketing, after observing many violations of "rationality" as defined in traditional choice theory, have started to bring the psychology of consumer choice decision making into economic choice modelling (Huber, Payne and Puto 1982). Various context effects on choice decisions have been examined. Studies concerned with the effect of introducing a third product on the changes in market share of two incumbent products have found strong evidence of the existence of a higher-order rule in choice decision making, which says that the value of an alternative depends on the choice set (e.g., Huber, Payne and Puto 1982; Johnson and Meyer 1984; Simonson 1989; Simonson and Tversky 1992). The documented context effects include: similarity effects (Simonson and Tversky 1992), asymmetric dominance effects (Huber, Payne and Puto 1982); tradeoff contrast and extremeness aversion (Simonson and Tversky 1992), attraction effects, and compromise effects (Simonson 1989). These studies offer many explanations for why these context effects occur. For example, it is found that different contexts are associated with different perceptual framing of the level of the attributes, thus different relative attractiveness of each alternative, different evaluation processes, or different decision making strategies (Huber, Payne and Puto 1982; Johnson and Meyer 1984; Dhar 1997). Other explanations for these behavioural "anomalies" include reasoning based choice, loss aversion, and information cue effects (Simonson and Tversky 1992; Dhar 1997; Dhar, Nowlis and Sherman 2001). After a series of experimental studies, Simonson and Tversky (1992) concluded that "context effects are both common and robust, representing the rule rather than the exception"(p. 293).

However, these marketing studies mainly focus on providing empirical evidence of the existence of "anomalies" in decision making introduced by a change in context. Little effort is devoted to model or control for the context effect since most of these studies are not as concerned with studying the underlying preference for attributes. The choice format effect on preference elicitation for drinking water risk reduction differs from the above documented context effects in a number of aspects. First of all, we are concerned about the provision of a public good, not private goods that are normally used in marketing studies. Context effects might affect public choice decisions differently than private choice decisions. Secondly, since it is a public good decision, the "state-of-the-world" format is used in the choice experiment. Except for the status quo option, there is no experienced utility associated with the proposed alternatives. There might be preference uncertainty in the attributes of the alternatives as well as in their outcomes. Therefore, it is likely that there is an anchoring effect at the status quo in the water survey data. It is important for us to understand the interaction between the status quo effect and the choice format effect. Thirdly, most context effects examined in marketing studies are limited to choices involving only two attributes. The choice tasks in our study are clearly more complex. While choices in the marketing literature are largely brand specific, the proposed alternatives in this study are generic and unlabelled.

As can be seen, despite being seemingly simple, the current literature does not offer sufficient explanation to understand the two-versus-three choice format effect. Nevertheless, this issue is important for two reasons. First of all, both choice formats are widely used for valuing non-market goods and services (e.g., Adamowicz et al. 1998; Breffle and Rowe 2002; Holmes and Boyle 2005). Carson et al. (2000) suggest that in the studies concerned with consumer preferences on public goods, a single binary choice is less susceptible to strategic bias than a single mutinomial choice over more than two alternatives, from the perspective of incentive compatibility theory.[41] In fact, however, a vast majority of stated choice surveys adopt the three-alternative format, probably for the sake of offering a better contrast as an alternative format to the contingent valuation method (CVM) that essentially contains two alternatives. It is therefore important to understand the choice format effect. Alternatively, if one considers the 2-alternative choice format as a close proxy to the referendum contingent valuation design, answers to the discrepancies in inferences derived from the two choice formats could help to understand the discrepancy between the CVM and the choice experiment (CE) method.

If inferred preferences for risk reductions differ significantly depending which survey choice format is used, the reliability of the estimates is questionable. Depending on the size of the variation in these estimates, they may no longer be informative for policy making in allocating resources (Garber-Yonts 2000). On the other hand, if we can

[41] In their study, Carson et al. (2000) suggest that a single binary choice is preferred to a sequence of binary choices, and it is also preferred to either a single or a sequence of multinomial choice(s). We assume decision independence across choices in this study.

find unified preferences under multiple choice formats so that the inferred preferences are context free or averaged over different contexts, the derived inferences could be more robust and valuable since their values are more transferable across different contexts. Therefore, it is crucial to test for construct validity by comparing inferences derived from different survey formats to ensure the quality and validity of a survey design. Secondly, as status quo bias is often found in valuing environmental goods (Adamowicz et al. 1998; Garber-Yonts 2000), the effect of choice format on status quo bias could have important empirical implications. If anchoring to a status quo option is a bias, in the case of our water survey, adopting a trinary choice format seems to mitigate this bias. The focus of this paper is to search for a model specification that recovers unified preferences for risk reductions from drinking water from the two different choice formats.

This paper is organized as follows. First, the paper explains how choice behaviour implied by the choice format effect deviates from the one implied by a standard RUM model. Then the paper explores two different ways to augment the standard RUM model so that the choice format effect on preference elicitation can be controlled within a random utility framework. One is to augment the model with contextual variables that directly characterize choice formats. Another is to develop an extended RUM model based on behavioural decision theory assuming reference-dependent preferences. The ultimate goal is to find model specifications for which common or unified preferences can be derived from the two choice formats.

4.2 Deviations from the Standard Random Utility Models Implied by the Choice Format Effect

The choice format effect, or context effects in general, has been an area of intensive interest in marketing research since the 1980s, partly due to many observed violations of "rationality" defined in traditional economic choice theory. Results from these studies have challenged traditional economic theory, for which "rationality" is the fundamental principle. One of the principles is the existence of independent preferences, i.e., preference between options does not depend on the presence or absence of other options (Tversky and Simonson 1993; Sen 1982). This property, called independence of irrelevant alternatives, is found to be often violated in the real world. A simple conditional logit (CL) model, normally specified with a linear preference function (McFadden 1974), assumes that the independence of irrelevant alternatives (IIA) assumption is satisfied:

$$(4.1)\qquad V_{nj} = \alpha(Y_n - PRICE_j) + \boldsymbol{\beta}\mathbf{X}_j + \varepsilon_{nj}$$

$$\varepsilon_{nj} \sim i.i.d(0, \pi^2/6)$$

where Y_n is income for individual n, $PRICE_j$ is the price of alternative j, $\mathbf{X}_j$ is a vector of attributes of alternative j, V_{nj} is the indirect utility function, α and β are parameters to be estimated. Since error terms are assumed to be independently and identically (IID) distributed, a closed form solution is available:

$$(4.2)\qquad P_{nj} = \exp(\mu V_{nj}) \Big/ \sum_{j \in C} \exp(\mu V_{nk})$$

where μ is the scale parameter, which is inversely related to the variance of error terms. It is fixed at 1 in a CL model to facilitate parameter estimation. Based on Equation 4.2, the logarithm of the odds ratio between choosing alternative j and k is a linear sum of the net utilities of attributes of the two alternatives ($\Delta V_{n,jk}$) (Equation 4.3).

$$(4.3)\qquad \ln(\frac{P_{nj}}{P_{nk}}) = \mu(V_{nj} - V_{nk}) = \mu\Delta V_{n,jk}$$

Equation 4.3 implies that the odds ratio is independent of the attributes of other alternatives in the choice set. Clearly, the choice format effect implies a violation of the IIA assumption: the odds ratio between a status quo option and its alternative being chosen changes as a third alternative is added into the choice set.

One explanation for context effects that is widely documented is choice complexity (Mazzotta and Opaluch 1995; Swait and Adamowicz 2001a, 2001b; Swait et al. 2002; Moon, Adamowicz and Boxall 2004). The two-versus-three choice format effect can be considered a type of choice complexity effect. Swait et al. (2002) suggest that context effects may enter into four elements in the structure of decision making: the preference, the error term, the decision strategy and choice set formation. The first two elements are directly related to decision outcome while the last two elements are related to decision process through which decision outcome is then affected. Therefore, algebraically, incorporating both direct and indirect impacts of a choice format effect on a random utility model can be expressed as, in a general sense,

(4.4) $$V_{nj} = \alpha(Y_n - PRICE_j) + \boldsymbol{\beta}\mathbf{X}_{nj} + f(\Omega)$$ [42], or

(4.5) $$\mu = \mu(\Omega)$$

where Ω denotes choice set or choice context. It thus affects both the preference function (Equation 4.4) and the scale function (Equation 4.5).

Equations 4.4 and 4.5 can be used to assess how decision outcome in terms of preference and variance in responses are directly affected by choice format. However, the choice format might also affect the preference function and scale function indirectly through its impact on decision process. It is difficult to disentangle these two types of effects. The examination of the choice format effect on the decision process is particularly difficult because most standard economic surveys do not collect process data. We, as economists, are in general more concerned with end points of decision making (outcome) not mid-points (process), unlike behaviourists or psychologists (Louviere et al. 1999). However, we can specify behavioural models underlying decision processes and examine whether the choice format effect on preference elicitation can be controlled and predicted in such models. Nonetheless, due to a lack of process data, it has to be acknowledged that the specified behavioural model may at best approximate the decision process.

4.3 A Modelling Framework for Controlling for the Choice Format Effect on Choice Decisions

We refer to the approach proposed by Swait and Adamowicz (2001b) in analyzing choice complexity to assess the two-versus-three choice format effect. According to Swait et al. (2002), context affects different components of a choice decision simultaneously. Although it is desirable to estimate a choice model that allows choice complexity to affect all the components of a decision structure, identification problems might arise. Thus, our modelling strategy is to allow choice complexity to affect one component of a

[42] Equation 4.4 implicitly assumes linear additive context effects on the preference function. In fact, context can affect preferences in a non-linear or non-additive fashion; a more general expression could be $V_{nj,\Omega} = f(V_{nj}, \Omega)$, where V_{nj} is defined in Equation 4.1.

choice decision at a time. For example, it may affect error components (i.e., the scale function): responses to complex choice tasks may have larger variances. It may affect taste (i.e., the preference function): respondents may attempt to make tradeoffs between maximizing product utility and minimizing cognitive effort at the same time. It may affect choice set formation: respondents may selectively assess a subset of information when information load increases. It may also affect decision rule: decision heuristics might be employed to make a decision easier. In both cases (redefining the choice set or using a heuristic decision rule), respondents tend to reconstruct a choice scenario by simplifying it when a choice task gets more complex.

While capturing the effect of choice complexity on the scale function or the preference function in a random utility model is straightforward, capturing its effect on choice set formation or decision rule is more difficult for a number of reasons. First of all, it is difficult to map the exact relationship between decision process and decision outcome. Due to individual differences in knowledge, experience, and other characteristics, different decision processes may lead to the same outcome, yet the same decision rule used by different individuals might results in different outcomes. So, changes in choice set formation or decision rule might affect both preference functions and scale functions in many different ways. Therefore, it is difficult to identify a causal relationship between decision processes and outcomes. Secondly, the impact of choice complexity on choice set formation or decision rule is difficult to identify since they affect decision outcome indirectly. Changes in choice set formation or decision rule are difficult to measure or quantify unless specific survey questions on decision process (for example, through mouse tracing or verbal protocol) are included to collect such data. Thirdly, implicitly, both decision rule and choice set formation are external to a random utility model in terms of model specification. A random utility model, once specified, already has a well-defined choice set and a compensatory decision rule is assumed. It is much more difficult to specify a model that explains both preference heterogeneity and process heterogeneity (Johnson et al. 2006). However, we can analyze decision strategy change in a random utility framework. For example, Swait and Adamowicz (2001b) identify two types of decision strategies using a latent class model within a random utility framework. For another example, non-compensatory tradeoffs may be captured using a linear specification by allowing for decision kinks (Elrod, Johnson and White 2005). Therefore, we first develop models that control for the choice format effect on decision outcome and then explore ways to hypothesize its effect on decision process.

4.3.1 Modelling the Impact of Choice Format Using Contextual Variables

Our test for the two-versus-three choice format effect using contextual variables is conducted as follows. We test for pooling of the CE2 and CE3 datasets under models that are augmented with contextual variables. If pooling cannot be rejected, it is suggested that common parameters on risk preferences are obtained from the two sub-samples.

First of all, we search for contextual variables that characterize different choice formats or choice environments. Entropy, a single summary measure of choice complexity of a choice task, proposed by Swait and Adamowicz (2001b), is considered a good candidate. Entropy can capture the multidimensional aspects of the choice format effect. It is defined as the cross-product between the probability of each alternative being chosen in a choice set and the logarithm of the probability (denote as H). The probability (of alternative j, j = 1 to J) is characterized by an attribute vector (denoted as $\pi(x_j)$) (Swait and Adamowicz 2001a, 2001b),

$$(4.6) \quad H(\pi_x) = -\sum_{j=1}^{J} \pi(x_j)\log \pi(x_j)$$

where $\pi(x_j)$ can be calculated using a conditional logit (CL) formula. Overall, it measures how similar alternatives are in a choice set. The more similar are the alternatives, the more difficult for one to choose. It is directly affected by the number of alternatives in a choice set. The larger choice set size, the higher entropy value; hence, the more complex a choice task. Adding a non-dominated alternative always increase the level of complexity of a choice decision. Other components of a choice set, like number of attributes or the degree of attribute correlation also affect entropy because the probability of an alternative being chosen is a function of its attributes (Swait and Adamowicz 2001b).

Alternatively, multiple variables can be used to explicitly capture different aspects of a choice environment (DeShazo and Fermo 2002; Caussade et al. 2005). For example, the number of alternatives and levels of attributes are used to capture quantitative information, and the number of tradeoffs between attributes, dispersion of attribute levels within each alternative and the dispersion of the standard deviation across alternatives are used to capture the structure of information (DeShazo and Fermo 2002). Another commonly used contextual variable is the order of a choice task. Based on Equations 4.4 and 4.5, we will show how to incorporate these contextual variables into a random utility model shortly.

In this study, individual n's indirect utility associated with alternative j is specified as,

(4.7) $V_{nj} = \beta_1 SQ_j + \beta_2 MICI_j + \beta_3 MICD_j + \beta_4 CANI_j + \beta_5 CAND_j + \beta_6 BILL_j$

where SQ is the alternative specific constant (ASC) for the status quo option.[43] In a more concise form the specification is, $V_{ij} = \beta_k \sum_{k=1}^{6} MAIN_k$ where $MAIN_k$ defines the k[th] main attribute, k = 1 to 6, indicating *SQ, MICI, MICD, CANI, CAND* and *BILL* respectively.[44] Let $\mathbf{Z}$ be a vector of choice task specific contextual variables. They enter the preference function as,

(4.8) $V_{njt}\_P = \sum_{k=1}^{6} \beta_k MAIN_k + \sum_{k=1}^{6} \gamma_k MAIN_k * \mathbf{Z}_t$

where t indicates the order of choice tasks one faces. If they enter the scale function,

(4.9) $V_{njt}\_S = \mu^{nit} V_{nj}, \quad \mu^{nit} = f(\mathbf{Z}_t)$

If they enter both the preference function and the scale function,

(4.10) $V_{njt}\_PS = \mu^{nit} V_{njt}\_P, \quad \mu^{nit} = f(\mathbf{Z}_t)$

Therefore, we search for a model that is augmented with contextual variables, by which pooling of the CE2 and CE3 datasets cannot be rejected statistically.

Equation 4.8 is a simple CL model specified without covariates. A with-covariates specification when contextual variables enter the preference function is,

(4.11) $V_{njt}\_PW = \sum_{k=1}^{6} \beta_k MAIN_k + \sum_{k=1}^{6} \gamma_k MAIN_k * \mathbf{Z}_t + \varphi SQ_j * \mathbf{S}_n$

[43] It is customary to include the status-quo alternative specific constant in the specification of an indirect utility function due to the widely documented status quo effect in similar studies in which the status-quo option is a baseline situation and the utility of an alternative is state-dependent (Scarpa, Ferrini and Willis 2005).

[44] See Chapter 2 Table 2.1 for definitions.

where **S** is a vector of socio-demographic variables.[45] A model when contextual variables enter the scale function can be similarly specified. To estimate a more general model that better captures preference heterogeneity, a random parameters logit (RPL) specification can be used.

In the next section, we first introduce behavioural theories that explain how choice environment affects reference point adoption processes and then hypothesize behavioural models that might offer some insights for the choice format effect found in the water survey.

4.3.2 Explaining the Choice Format Effect Using a Reference-Dependent Model

While traditional economic analysis focuses on mapping information input into output and treating the decision process as a black box by invoking normative "rationality" assumptions, some economists have found that behavoural models that bring the elements affecting decision process into economic analysis can be used to predict these behavioural "anomalies" (McFadden 1999). As a result, gradually more attention has been given to understand human bounded rationality by conducting inter-disciplinary research on human behavioural decision making at both theoretical and empirical levels. For example, Huber at al. (1982) and Bateman et al. (2005) found that choice decisions can be affected even by adding a dominated alternative. Research on cognitive processes of decision making also indicate that contexts can affect the allocation of attention across information within a choice set. A failure to control for individuals' propensity to attend to only a subset of the information in the choice set might lead to downward bias in welfare estimates (DeShazo and Fermo 2004). However, due to a lack of data to measure attention, it is difficult to formulate a model based on attention theory.

In fact, many behavioural "anomalies" defined according to neoclassical expected utility theory (EUT) can be explained or predicted by Prospect Theory (Kahneman and Tversky 1979). Prospect Theory differs from EUT by incorporating the psychological

[45] These covariates are socio-demographic variables, and they enter as interactions terms of the status quo alternative specific constant. . They are *AGE65, INCOME, INCOME2, ENGLISH, CITYSIZE, ILLNESS, MALE* and *MARRY* (see their definitions in Chapter 2 Table 2.1).

aspects of human decision making into economic analysis (Plous 1993). From a psychological perspective,

> *"[b]ehaviour is local, adaptive, learned, dependent on context, mutable, and influenced by complex interactions of perceptions, motives, attitudes, and affect." (McFadden 1999, p. 75)*

Based on Prospect theory, Tversky and Kahneman (1991) propose a reference-dependent preference theory of consumer choice, in which preferences are defined over certain reference points. Due to loss aversion, consumers value their losses relative to a reference point more than an equivalent size of gains. An individual's real-valued function U (i.e., reference-dependent utility function rather than utility function as defined in EUT) is decomposed into different reference functions R of multiple dimensions of preference (or attributes) x_k with respect to a reference state r. For a real-valued function with multiple dimensions of preference and additive in reference functions evaluated at reference state r, and assuming a constant loss version holds, the model is

$U_r(x_1, x_2, \ldots, x_k) = \sum_k R_k(x_k)$, where

$$R_k(x_k) = \begin{cases} u_k(x_k) - u_k(r_k), & \text{if } x_k \geq r_k, \\ \lambda_k[u_k(x_k) - u_k(r_k)], & \text{if } x_k < r_k \end{cases} \tag{4.12}$$

λ_k is a positive scalar that gives a higher weight to a loss relative to a reference level r_k than the gain. Reference functions can be constructed separately around different dimensions of underlying preferences. Weighting of losses or gains can also vary across different reference states.

Using a reference-dependent model, Hu et al. (2006) find evidence of reference-dependent preferences in the demand for genetically modified canola oil. Schweitzer (1995) indicates that different contexts might lead to different reference point adoption processes. In examining how an outcome is evaluated when two reference points, the status quo and its alternative, provide conflicting information about the "goodness" of the outcome, Boles and Messick (1995) found a complete reversal of preferences resulting from a shift in a reference point. Adding a third alternative may induce a change in perceptual framework of relative attractiveness of each alternative in a choice set.

Developing a reference-dependent (RD) model within the RUM framework is straightforward. It involves augmenting a standard model with loss and gain variables (Hardie, Johnson and Fader 1993; Hu 2004). As Equation 4.11 indicates, the specification of a reference-dependent random utility model (RDRUM) varies as a reference state changes; and real-valued functions can be constructed separately for each attribute of an alternative. As a result, we try to search for a RDRUM that pools the datasets statistically.

For the binary choice format, it is very likely that the status quo option serves as a reference state. Real-valued functions (or reference functions) of attributes are constructed individually relative to the status quo levels. For the trinary choice format, the status quo may or may not be the reference level. A discussion of a few possible reference states is thus in order. Two types of reference state can be specified: alternative specific reference states and attribute-based reference states.[46] For alternative specific reference states, three reference-dependent models can be estimated depending on which alternative is chosen as a reference state, the status quo (SQ) option, the first alternative or the second alternative.[47] Research on compromise effects and reasoning-based choices implies a compromise alternative is likely to be chosen as a reference state (Simonson 1989; Simonson and Tversky 1992). A fourth alternative specific RD model is therefore based on a compromise alternative.[48] Compromise effects may also be achieved at the attribute level, i.e., an alternative with balanced attributes is more likely to be chosen (Moon, Adamowicz and Boxall 2004). An attribute-based reference state means that a reference point for each attribute is always set at the average level of the attribute of all alternatives. So, five different reference-dependent models can be specified for the trinary

---

[46] It is likely that there are other possible reference states. Fox example, respondents may anchor their responses at the first choice task that they observe (Carlsson et al. 2007). We assume decision independence across choice tasks in this study.

[47] The alternative program(s) are unlabelled or generic by design in the water survey, therefore, it is less likely for a respondent to anchor at one alternative program versus the other except for the position effect (middle or far right), if there is any.

[48] According to Simonson (1989), a compromise alternative is the one lying in between two other alternatives in a two-dimensional space (two attributes). For a choice set with more attributes, a compromise alternative can be defined as the one with the two shortest Euclidean distances with the other two alternatives. For example, for alternative A, B and C, with attribute X1, X2, X3, and the Euclidean distance between AB = $|X1_A - X1_B| + |X2_A - X2_B| + |X3_A - X3_B|$, If AB is larger than BC and AC, then C is the compromise alternative. Note that X1, X2 and X3 should be normalized between 0 and 1 to avoid the distance being dominated by attributes with large scale.

choice dataset. For the binary choice dataset, we can estimate two reference-dependent models, namely the status-quo-dependent model and the alternative-dependent model.

In the water survey, the price level is fixed at zero for the status quo and is greater than zero for proposed alternatives. As a result, the value function of the price dimension of an option relative to the status quo is always negative. Therefore, only reference point effects of non-price attributes are examined. The four risk attributes used to describe water quality are *MICD, MICI, CAND* and *CANI* as introduced in Chapter 1. Gains are defined when a risk attribute level is lower than a reference level *r*; while Losses are defined conversely; no gain or loss is defined when a risk attribute level of an alternative is equal to a reference level.

Algebraically, to specify a basic status-quo-reference-dependent (SQRD) model, the indirect utility function (V_{nj_SQRD}) for individual *n* derived from purchasing a public good *j* that reduces drinking water health risks (before socio-demographic effects are taken into account) is specified as,

$$\begin{aligned} V_{nj\_SQRD} &= \beta_1 MICI + \beta_2 MICD + \beta_3 CANI + \beta_4 CAND + \beta_5 BILL \\ &\quad + \beta_{11} MICIG + \beta_{12} MICIL + \beta_{21} MICDG + \beta_{22} MICDL \\ &\quad + \beta_{31} CANIG + \beta_{32} CANIL + \beta_{41} CANIG + \beta_{42} CANIL \end{aligned} \tag{4.13}$$

where *j* is not the status quo. For *j* at the status quo, the indirect utility function is

$$V_{nj\_SQRD} = \beta_0 SQ + \beta_1 MICI + \beta_2 MICD + \beta_3 CANI + \beta_4 CAND + \beta_5 BILL \tag{4.14}$$

For other types of reference states, the models can be similarly defined and are not listed here. The only difference is that all the gain and loss dummy variables have to be recalculated relative to reference points implied by a new reference state.

It is noteworthy that there might be heterogeneity in the reference point effects. Different individuals may have different weights for each attribute. Some attributes are more important in decision making for some people than for others. Individuals with different health conditions may have different levels of loss aversion to different health risks. Preference heterogeneity in the reference point effects can be captured by using a random parameters logit model (e.g., Hu 2004). That is, we assume effects of the gain and loss dummy variables are randomly distributed across people rather than fixed as implied in Equation 4.13.

In summary, this section attempts to explore model specifications that generate common preferences for risk reductions by controlling for the choice format effect. Two types of strategies are adopted. One is to augment the standard random utility models with contextual variables that characterize the choice format and the other is to use a behavioural model that allows for reference-dependent preferences to explain the phenomenon.

4.4 Controlling for Impacts of Choice Format on Choice Decisions

In this section, context-variable augmented random utility models (RUMs) and reference-dependent RUMs are estimated. To estimate the former, contextual variables that characterize the two-versus-three choice format are first selected. Then we examine whether common risk preferences can be obtained in the augmented RUMs. After that, we estimate various RUMs that allow for reference-dependence preferences and see whether they are able to provide some insights into the choice format effect.

A range of contextual variables that might have impacts on choice decisions are listed in Table 4.1. Table 4.1 reports their definitions and basic sample statistics for CE2, CE3 and for the pooled dataset CE23. Appendix 4.2 explains how entropy values are calculated for each choice occasion (Table A4.2.1) and provides information on correlation relationships between some major contextual variables (Table A4.2.2). Note that the standard deviation of Number of Attributes whose levels Differ across Alternatives (*NADA*) for CE23 is more than triple that of either CE2 or CE3. The standard deviation of *NADA* for CE23 mainly arises from differences in the group mean between the two datasets. Two high correlations (higher than 0.5) are found between cumulative entropy (*CUMENTRO*) and *NADA* in CE2 and the correlation between cumulative entropy (*CUMENTRO*) and cumulative dispersion of standard deviation across alternatives (*CUMDISSD*) in CE3.

4.4.1 Models Controlling for Effects of Choice Format on Decision Outcome

Since contextual variables enter the preference function by interacting with the main variables (MAIN), the number of parameters multiplies as the number of contextual variables increases. Thus, we start by estimating a model with each individual contextual variable listed in Table 4.1, use hypothesis tests to determine whether it is preferred to the basic model (Equation 4.7), and then estimate a more general model by including more contextual variables. In the end, it is found that *NADA, ENTROPY* and *ORDER* are the factors that are generally found to have significant effects in various model specifications. In a more general model, contextual variables are allowed to affect both the preference and scale functions. Table 4.2 reports the likelihood-ratio (LR) test results for pooling the datasets using the three types of models that allow choice format to influence a choice decision. For each type of model, both a CL specification and a RPL specification are estimated (Table 4.2).[49]

Table 4.2 indicates that strict pooling of CE2 and CE3 is always rejected when the contextual variables enter the scale function alone. Estimated as simple CL models, the datasets can be pooled if the contextual variables are allowed to enter the preference function regardless how they enter the preference function, i.e., through full interaction or partial interaction, and regardless of whether they also enter the scale function as well. Thus, there is strong evidence that the choice format affects preference elicitation. However, the effect can be controlled by incorporating these contextual variables in the preference function and common preferences underlying different sub-samples can be recovered. When estimated as RPL models, however, a pooled model cannot be rejected only when the contextual variables enter the preference function through a full

[49] All models are estimated with covariates, i.e., they include interactions between the eight socio-demographic variables and the status-quo alternative specific constant (SQASC). However, when contextual variables enter both preference and scale functions, the estimated model does not converge for both a CL and RPL specification. Therefore, rather than letting the three contextual variables interact with all six main variables, we estimate models in which the contextual variables interact with the SQASC only when they are included in both functions (even so, estimated RPL models do not converge). For comparison, two models are estimated with the two different types of interactions between the contextual variables and alternative specific attributes when they enter the preference function only. One allows for full interaction between the contextual variables and the main attributes (full interaction) and the other only allows for the interactions with the SQASC (partial interaction).

interaction. RPL models specified with contextual variables interacting with SQASC LR tests indicate that RPL models are preferred to their CL counterparts. Thus, the RPL model with the preference function augmented with *NADA*, *ENTROPY* and *ORDER* fully interacting with alternative specific attributes is chosen as the preferred model to explain the choice format effect.

Table 4.3 reports the estimated augmented RPL model for the pooled dataset CE23 (see Appendix 4.4 for other estimated models, Tables A4.4.1-A4.4.6). Half of the interactions between the main attributes and NADA are significant. The coefficient on the interaction between *SQ* and *NADA* is negative and coefficients on the interactions with risk attributes are estimated to be positive. As the number of tradeoffs among risk attributes increases in a choice task, the SQ option is less likely to be chosen, and respondents are less risk averse to microbial deaths and cancer illnesses. It is likely that increases in the number of tradeoffs provides better preference matching so that individuals' risk preferences are less anchored at the status quo level and individuals are more willing to make tradeoffs among different risk reductions. The coefficients on the interactions between *SQ* and *BILL* with *ENTROPY* are found to be significant and positive. As entropy increases, individuals are more likely to stay at the status quo. As a choice task becomes more difficult (with increased similarity between alternatives), respondents were more likely to choose not to purchase (Dhar 1997; Adamowicz et al. 1998). This finding is widely documented by many other studies (Swait and Adamowicz 2001a; Scarpa, Ferrini and Willis 2005). Also, increased complexity makes respondents less sensitive to price. Other things being equal, it implies, respondents are likely to pay more to reduce per unit health risk reduction. It is found that there is an order effect (*ORDER*) on the preference for the status quo (*SQ*) and on the price effect (*BILL*) albeit the effect is marginally significant (i.e., at the 10% level). Respondents are less likely to choose SQ and are less price sensitive when they answered the first two tasks than the last two.

It is also found that *NADA* and *ENTROPY* are correlated (the correlation is 0.54). However, they have opposite effects on the preference for the SQ. The overall effect thus depends on relative size of the two different effects. It has to be noted that when the model is allowed to be augmented with one contextual variable at a time in the preference function, *NADA* is the only variable with which pooling the dataset is not rejected. So *NADA,* number of attributes whose levels differ across alternatives, is more important in explaining the choice format effect than *ENTROPY* in this study. The choice format effect in this study is more likely to be specifically related to the changes in number of tradeoffs in attribute levels than the increased information load as indicated by entropy levels. *ENTROPY* might be more suitable to capture continuous relationships between the level of choice complexity and the number of alternatives in a choice set rather than a discrete change from two alternatives to three alternatives. Increases in the level of choice complexity from a 2-alternative choice task to a three-alternative task may not cause substantial increases in cognitive burden on respondents (Caussade et al. 2005). *NADA* seems to characterize the major difference between the 2-alt and 3-alt choice format datasets in our study.

4.4.2 A Behavioural Model Incorporating Reference-Dependent Preferences

The direct modelling of the impacts of choice format described above is outcome-oriented rather than process-oriented. To assess deeper behavioural reasons underlying the choice complexity effect[50], different behavioural models might be developed. For example, as a choice decision gets more complex, individuals may redefine choice sets or use decision heuristics to simplify the choice task. In this section, a behavioural model is hypothesized to test whether choice complexity induces changes in the decision rule. It is hypothesized that more complex choice tasks induce changes in reference points upon which respondents construct value functions. We test for the choice format effect by pooling the two datasets (CE2 and CE3) using reference-dependent models. If the datasets can be pooled under a specification in which reference dependent preferences are allowed, the choice format effect might result from ignoring reference point effects in model specification. We specify two types of reference-dependent models differing in the specification about reference points upon which value functions are constructed.[51] One is

[50] The finding that increased choice complexity increases preference for the status quo option can be considered a behavioural reason as well.
[51] Other reference-dependent models proposed in Section 4.3 were estimated, but statistical tests show these augmented models do not perform better than the models without loss and gain variables.

a status-quo-reference-dependent (SQRD) model and the other is an attribute-compromise (AC) model.

Table 4.4 lists variables indicating gains and losses in the four types of risk reductions of an alternative relative to their reference levels. We include interactions between the gain and loss variables and *BILL* to examine whether value functions of each risk attribute are constructed differently at different price levels. Table 4.4 lists both linear and quadratic terms of gain and loss variables as well as the interaction terms between linear gain and loss variables and *BILL*.

For the SQRD model, to conduct the pooling test, models are estimated for CE2, CE3 and CE23 separately. An LR test is then used to determine whether the datasets can be pooled under either specification. If the datasets can be pooled, the choice format effect might be due to reference-dependent preferences.[52]

Table 4.5 reports LR test results of pooling the CE2 and CE3 datasets based on various estimated SQRD models using both a CL specification and a RPL specification. Two types of specifications are used: Model 1 includes gain and loss variables linearly only while Model 2 includes both the linear terms and the interactions between linear gain and loss variables and *BILL*.[53] For the purpose of comparison, model fits of models specified without incorporating gain and loss variables (Model 0) are also reported. In addition to log-likelihood values, the AIC and BIC values are also provided to facilitate model selection. Table 4.6 indicates that LR tests for pooling the datasets are rejected across all specifications (Table 4.6 last column). Therefore, we cannot provide evidence that the choice format effect is due to status-quo-reference-dependent preferences.

[52] For the attribute-compromise model, rather than testing for pooling the datasets the same way as we do to the SQRD model, we test for equality of risk preferences by imposing restrictions on model common parameters on risk attributes in a pooled model specification (using the pooled dataset CE23) while allowing extra model parameters on gains and loss variables that are only relevant to the CE3 dataset in the utility function (Louviere et al. 2000). In this case, it is equivalent to the LR test of a nested model with equality constraints imposed on the common model parameters against a more general model that does not have such restrictions. We are interested in the parameters on *SQ, MICI, MICD, CANI, CAND* and *BILL*.

[53] Models included quadratic gain and loss variables were initially specified, but they cannot be estimated due to multi-collinearity problems.

Tables 4.6 and 4.7 report model fits of various AC models based on a CL specification and a RPL specification[54], respectively. In each table, log-likelihood values, AIC and BIC values are reported for the CE3 dataset and the pooled CE23 dataset. Five pairs of models differ in how the gain and loss variables are included in the indirect utility function. Each pair of models is a contrast between models estimated without and with interactions between a version dummy variable CE2 (equal to 1 if the CE2 dataset and 0 otherwise) and main attributes: *SQ, MICI, MICD, CANI, CAND* and *BILL*. For CE23, LR tests conducted based on the difference in the pair of LL values without and with these interactions. For CE3, only the simpler model is estimated, i.e., the one without the interaction terms.[55] The five pairs of models, named as Model 0 to 4, are specified with an increased number of variables indicating gains and losses except for Model 3.[56] Table 4.6 indicates that LR tests for the common model parameters between the two datasets are rejected across all CL specifications.[57] For example, the LR test value for pooling the model specified with quadratic gain and loss variables (Model 2) is 15.74, which is greater than the χ^2 table value at the 5% level with 6 degree of freedom (12.59). Table 4.7 reports the same results when the RPL specification is used. In the marketing literature, when the compromise effects are of concern, the existence of a compromise level of attributes in a choice set is often built into the experimental design. Therefore, the rejection of common parameters may be a result of the data in this case—too few choice sets have three levels of attributes in which the compromise effect is likely to occur. However, we do not have concrete evidence that reference dependence is the cause of the choice format effect.

---

[54] For the RPL specification, only *SQ, MICI, MICD, CANI* and *CAND* are estimated as random parameters (normal distributed). It would be desirable to estimate random parameters on gain or loss variables. However the number of gain or loss variables is too large to estimate.

[55] For each pair, the models are named as Model #.0 versus Model #.1, where #.0 indicates the simpler version, and #.1 indicates the complex version or the pooled version

[56] The Model 0 pair (i.e., Model 0.0 and Model 0.1) does not contain any gain and loss variable; the Model 1 pair uses a linear expression of the gain and loss variables; the Model 2 pair specifies a quadratic effect of the gain and loss variables; the Model 3 pair, contains the linear terms and the interaction terms between the linear gain and loss variables and *BILL*; and the Model 4 pair contains a quadratic term and the interaction terms. Therefore, Models 2 and 3 are each nested within Model 4, and themselves are non-nested models.

[57] Nonetheless, LR test values decrease as an AC model becomes more general. The LR test value based on Model 4 is very close to the critical value at the 5% level (13.08 versus 12.59).

We have attempted to derive common preference inferences from the two datasets of different choice formats by estimating augmented random utility models that allow for reference-dependent preferences. Respondents' preferences for risk reductions are allowed to be either anchored at the status-quo level and or at intermediate levels of each attribute. We find that the choice format effect on preferences still persists in these reference-dependent models. There is weak evidence of the impact of choice format on reference point adoption processes since for CE2 a status-quo-reference-dependent model provides a better fit while for CE3 an attribute-compromise model fits the data slightly better based on AIC and BIC criteria.

4.5 Conclusions

This chapter examines the choice format effect that was found in a preliminary analysis of the stated preference data that investigated Canadians' preferences for health risk reductions. The choice format effect found in this study, in particular, is that preferences for health risk reductions from drinking water inferred from survey questions differing in the number of alternatives included in a choice set were different. With the increased popularity of stated preference surveys to investigate consumers' preferences in non-market valuation studies, it is important to understand this phenomenon so that we can strive to control for the effects and derive context-free preferences. More importantly, it has become standard practice to use multiple choice formats in survey design to ensure construct validity (Mitchell and Carson 1989; Cameron et al. 2002). Both the two-alternative and three-alternative formats are very popular choice formats (Adamowicz et al. 1998; Breffle and Rowe 2002). Therefore, the analysis was conducted to unveil the choice format effect in this study could be of general interest to stated preference researchers.

In this study, effort was made to search for a model specification that reconciles preference discrepancies derived from the two difference choice formats. In other words, we strive to derive common parameters on preferences for health risk reductions from 2-alternative choice questions and 3-alterantive choice questions. This is achieved when contextual variables are explicitly included in the preference function. Two variables are found to be important in explaining the choice format effect. One is *NADA*, the number of attributes whose levels differ across alternatives, and the other is *ENTROPY*, a summary measure of information load that a choice set contains. It is found that, on one hand, the more complex is a choice task, the more likely is an individual to choose a status quo option; on the other hand, the more tradeoffs offered in a choice set, the more likely one chooses a non-status quo option. Results from estimating RUMs expanded with gain and loss variables, however, indicate that the choice format effect can be not explained using a reference-dependent model. Nonetheless, for the 2-alternative choice question, a status-quo-reference-dependent model fits the data better while for the 3-alternative choice question, compromise effects might have existed in making attribute tradeoffs.

Results from this paper suggest that choice formats can affect preference elicitation by imposing cognitive costs on decision making (negative utility). However, it is possible to control for the choice format effect by explicitly including it in the preference function for decision making. It is therefore important to use multiple choice formats in survey design so that the effect of choice format can be systematically analyzed and controlled. In addition, we might also want to build some variation into the entropy level of choice sets at the design stage so that we are able to capture and control it in the estimation stage.

References

Adamowicz, W. L., P. Boxall, M. Williams and J. Louviere. 1998. Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics* 80(1): 64-75.

Adamowicz, W. L., D. Dupont and A. Krupnick. 2005. *Willingness to Pay to Reduce Community Health Risks from Municipal Drinking Water: A Stated Preference Study*, Report for the U.S. Environmental Protection Agency, 62pp.

Alevy, J. E., J. A. List and W. L. Adamowicz. 2006. "More is less" preference reversals and non-market valuation. *Working Paper*. Department of Agricultural & Resource Economics, University of Maryland.

Bateman, I. J., A. Munro and G. L. Poe. 2005. Asymmetric dominance effects in choice experiments and contingent valuation. [accessed 25 October 2006]. Available at: http://www.uea.ac.uk/env/cserge/pub/wp/edm/edm_2005_06.pdf.

Ben-Akiva, M., M. Bradley, T. Morikawa, J. Benjamin, T. Novak, H. Oppewal and V. Rao. 1994. Combining revealed and stated preferences data. *Marketing Letters* 5(4): 335-350.

Boles, T. L. and D. M. Messick. 1995. A reverse outcome bias: The influence of multiple reference points on the evaluation of outcomes and decisions. *Organizational Behavior and Human Decision Process* 61(3): 262-275.

Breffle, William S. and Robert D. Rowe. 2002. Comparing choice question formats for evaluating natural resource tradeoffs. *Land Economics* 78(2): 298-314.

Cameron, T. A., G. L. Poe, R. G. Ethier and W. D. Schulze. 2002. Alternative non-market value-elicitation methods: are the underlying preferences the same? *Journal of Environmental Economics and Management* 44: 391-425.

Carlsson, F. and Martinsson. P. 2007. How much is too much? – An investigation of the effect of the number of choice sets, starting point and the choice of bid vectors in choice experiments, *Environmental and Resource Economics*, accepted for publication.

Carson, R., T. Groves and M. Machina. 2000. Incentive and informational properties of preference questions. Department of Economics, University of California, San Diego. [accessed 17 May 2007]. Available at: http://econ.ucsd.edu/~rcarson/cgm.pdf.

Caussade, S., J. Ortúzar, L. I. Rizzi and D. A. Hensher. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B* 39: 621-640.

DeShazo, J. R. 2002. Designing transactions without framing effects in iterative question formats. *Journal of Environmental Economics and Management* 43: 360-385.

DeShazo, J. R. and G. Fermo. 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44: 123-143.

———. 2004. Implications of rationally-adaptive pre-choice behavior for the design and estimation of choice models. [accessed 15 November 2006]. Available at: http://faculty.spa.ucla.edu/deshazo/pdf/16/Microsoft%20Word%20%20deshazo_fermo_081004.pdf.

Dhar, R. 1997. Context and task effects on choice deferral. *Marketing Letters* 8(1): 119-130.

Dhar, R., S. M. Nowlis and S. J. Sherman. 2001. Trying hard or hardly trying: an analysis of context effects in choice. *Journal of Consumer Psychology* 9(4): 189-200.

Elrod, T., R. D. Johnson and J. White. 2005. A new integrated model of noncompensatory and compensatory decision strategies. *Organizational Behavior and Human Decision Processes* 95(1): 1-19.

Garber-Yonts, Brian. 2000. *A Choice Experiment Analysis of Public Preferences for Conservation of Biological Diversity in the Oregon Coast Range*. PhD Dissertation. Oregon State University.

Halvorsen, B. and K. Sælensminde. 1998. Difference between willingness-to-pay estimates from open-ended and discrete-choice contingent valuation methods: the effects of heteroscedasticity. *Land Economics* 74(2): 262-282.

Hardie, B. G. S., E. J. Johnson and P. S. Fader. 1993. Modeling loss aversion and reference dependence effects on brand choice. *Marketing Science* 12(4): 378-394.

Hensher, D., 2006. Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environmental and Resource Economics* 34(1): 7-44.

Hensher, D., J. Louviere and J. Swait. 1999. Combining sources of preference data. *Journal of Econometrics* 89: 197-221.

Holmes, T. and K. Boyle. 2005. Dynamic learning and context-dependence in sequential, attribute-based, stated preference valuation questions. *Land Economics* 81(1): 114-126.

Hu, W. 2004. *Three Essays on Genetically Modified Food Labelling and Consumer Behaviour*, PhD Dissertation. Department of Rural Economy, University of Alberta.

Hu, W., W. L. Adamowicz and M. M. Veeman. 2006. Labeling context and reference point effects in models of food attribute demand. *American Journal of Agricultural Economics* 88(4): 1034-1049.

Huber, J., J. W. Payne and C. Puto. 1982. Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *The Journal of Consumer Research* 9(1): 90-98.

Johnson, E. J., B. G. S. Hardie, R. J. Meyer and J. Walsh. 2006. Observing Unobserved Heterogeneity: Using Process Data to Enhance Choice Models. [accessed 15 September 2007]. Available at: http://marketing.wharton.upenn.edu/Marketing-_Content_Management/Marketing_files/Publication_Files/Observing%20%20Unobserved%20Heterogeneity.pdf.

Johnson, E. J. and R. Meyer. 1984. Compensatory choice models of noncompensatory processes: The effect of varying context. *Journal of Consumer Research* 11: 528-541.

Kahneman, D. and A. Tversky. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263-291.

Layton, D. F. and S. T. Lee. 2006. Embracing model uncertainty: strategies for response pooling and model averaging. *Environmental and Resource Economics* 34(1): 51-85.

Louviere, J., R. J. Meyer, D. Bunch, R. Carson, B. Dellaert, W. M. Hanemann, D. Hensher and J. Irwin. 1999. Combining sources of preference data for modeling complex decision process. *Marketing Letters* 10(3): 205-217.

Mazzotta, M. and J. Opaluch. 1995. Decision making when choices are complex: a test of Heiner's hypothesis. *Land Economics* 71: 500-515.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behaviour. *Frontiers in Econometrics*. (ed.) P. Zarembka, 105-142. New York: Academic Press.

McFadden, D. 1999. Rationality for economists? *Journal of Risk and Uncertainty* 19(1-3): 73-105. Special Triple issue on: Preference Elicitation.

———. 2001. Economic choices. *The American Economic Review* 91(3): 351-378.

Mitchell, R. C. and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Resources for the Future. USA: The Johns Hophkins University Press.

Moon, A., W. L. Adamowicz and P. Boxall, 2004. Complexity in decision making: The effect of task structure and context on participant response. Paper presented at the Northeastern Agricultural and Resource Economics Association – Canadian

Agricultural Economics Society Conference, 20-23 June 2004, Halifax, Nova Scotia.

Plous, S. 1993. *The Psychology of Judgment and Decision Making.* The United States of America: McGraw-Hill.

Scarpa, R., S. Ferrini and K. G. Willis. 2005. Performance of error component models for status-quo effects in choice experiments. In: *Applications of Simulation Methods in Environmental and Resource Economics.* (eds.) R. Scarpa and A. Alberini, 247-274. Springer Publisher: Dordrecht, The Netherlands.

Schweitzer, M. 1995. Multiple reference points, framing, and the status quo bias in health care financing decisions. *Organizational Behavior and Human Decision Processes* 63(1): 69-72.

Sen, A. 1982. *Choice, Welfare and Measurement.* B. Blackwell: Oxford.

Simonson, I. 1989. Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research* 16: 158-174.

Simonson, I. and A. Tversky. 1992. Choice in context: tradeoff contrast and extremeness aversion. *Journal of Marketing Research* 29(3): 281-295.

Swait, J. and W. L. Adamowicz. 2001a. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Process* 86(2): 141-167.

———. 2001b. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. *Journal of Consumer Research* 28: 135-148.

Swait, J., W. L. Adamowicz, M. Hanemann, A. Diederich, J. Krosnick, D. Layton, W. Provencher, D. Schkade and R. Tourangeau. 2002. Context dependence and aggregation in disaggregate choice analysis. *Marketing Letters* 13(3): 195-205.

Tversky, A. and D. Kahneman. 1991. Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics* 106(4): 1039-1061.

Tversky, A. and I. Simonson. 1993. Context-dependent preferences. *Management Science* 39(10): 1179-1189.

Table 4.1 Definition of Contextual Variables

Variable	Definition	Mean (Standard Deviation) CE2	CE3	CE23
ENTROPY	Unit entropy value of a choice set assuming equal weight for each attribute	0.506 (0.160)	0.823 (0. 234)	0.666 (0.256)
CUMENTRO	Cumulative entropy value	0.712 (0.557)	1.200 (0.945)	0.958 (0.814)
NADA	The number of attributes whose levels differ across alternatives	3.830 (0.90)	12.230 (1.47)	8.076 (4.378)
SD_j	The mean standard deviation among the normalized attribute levels of alternative *j* in a choice set	0.655 (0.124)	0.674 (0.149)	0.667 (0.140)
DISSD	The dispersion of the SD of attributes levels across alternatives in a choice set	0.141 (0.070)	0.119 (0.073)	0.130 (0.073)
CUMDISSD	Cumulative DISSD	0.326 (0.327)	0.502 (0.305)	0.415 (0.328)
ORDER	1 if a choice task faced by an individual is the first two choice tasks out of a total of four choice tasks, and 0 for the last two choice tasks	0.5 (-)	0.5 (-)	0.5 (-)

Notes: According to DeShazo and Fermo (2002), $SD_j = \sqrt{(\sum_{i=1}^{K}(x_{ij} - \bar{x}_j)^2)/J}$, where x_{ij} is the normalized i^{th} attribute level of alternative *j*; *K* is the total number of attributes in alternative *j*. $DISSD = \sqrt{(\sum_{j=1}^{J}(SD_j - Average\,SD)^2 / J)}$, where Average $SD = (\sum_{j=1}^{J} SD_j)/J$.

Table 4.2 Likelihood-Ratio Tests for Pooling Datasets Using Context-Variable Augmented Models (Italicized LR Values Indicate Pooling Cannot be Rejected)

Model	# para.	Log-likelihood CE2	Log-likelihood CE3	Log-likelihood CE23	Likeli-hood-Ratio Test	χ^2 table value
In the preference function only						
Contextual variables interact with all main variables (full interaction)						
CL	26	-372.82	-675.76	-1065.11	*33.05*	46.19
RPL	31	-345.78	-593.42	-959.52	*40.65*	52.19
Contextual variables interact with the SQ ASC only (partial interaction)						
CL	16	-385.75	-689.90	-1085.77	*20.22*	27.59
RPL	21	-361.13	-603.13	-981.93	35.36	33.92
In the scale function only						
CL	16	-380.34	-689.84	-1091.86	43.38	27.59
RPL	21	-359.78	-603.19	-990.57	55.19	33.92
In both preference and scale functions						
Contextual variables interact with the SQ ASC only (partial interaction)						
CL	18	-379.18	-689.16	-1082.38	*28.07*	31.41
RPL	23	-	-	-	-	-

Notes: 1) For the RPL specification, effects of all main variables except for *BILL* are assumed to follow a normal distribution in the population. 2) All models are estimated with eight covariates (interactions between eight socio-demographic variables and the status-quo alternative specific constant). 3) RPL models cannot be estimated when contextual variables enter both preference and scale function. 4) χ^2 table values are at the 5% level.

Table 4.3 Estimated Pooled Random Parameters Logit Model CE23 with the Contextual Variables in the Preference Function

Variable	CE23	Variable (Cont'd)	CE23
SQ	2.757**	*SQ * NADA*	-0.320**
	(2.423)		(-4.335)
SQ_SD	2.396**	*MICI * NADA*	5.01E-06
	(8.520)		(1.022)
MICI	-2.18E-04**	*MICD * NADA*	0.013**
	(-4.102)		(2.002)
MICI_SD	1.13E-04**	*CANI * NADA*	0.005**
	(4.850)		(4.321)
MICD	-0.225**	*CAND * NADA*	0.002
	(-2.817)		(0.330)
MICD_SD	0.119**	*BILL * NADA*	-1.65E-04
	(3.281)		(-0.559)
CANI	-0.054**	*SQ * ENTROPY*	3.068**
	(-3.843)		(2.960)
CANI_SD	0.026**	*MICI * ENTROPY*	1.06E-05
	(3.302)		(0.118)
CAND	-0.042	*MICD * ENTROPY*	-0.094
	(-0.642)		(-0.776)
CAND_SD	0.111**	*CANI * ENTROPY*	-0.027
	(3.414)		(-1.501)
BILL	-0.016**	*CAND * ENTROPY*	-0.090
	(-4.705)		(-0.873)
*AGE65*SQ*	-1.215	*BILL * ENTROPY*	0.014**
	(-2.444)		(2.855)
*INCOME*SQ*	1.80E-05	*SQ * ORDER*	-0.826*
	(1.000)		(-1.809)
*INCOME2*SQ*	-1.85E-10	*MICI * ORDER*	-4.86E-06
	(-1.648)		(-0.174)
*ENGLISH*SQ*	-0.040	*MICD * ORDER*	0.019
	(-0.095)		(0.431)
*CITYSIZE*SQ*	-0.414**	*CANI * ORDER*	-0.004
	(-3.201)		(-0.416)
*ILLNESS*SQ*	-0.886	*CAND * ORDER*	0.038
	(-0.934)		(0.966)
*MALE*SQ*	1.055**	*BILL * ORDER*	-0.004*
	(2.883)		(-1.700)
*MARRY*SQ*	0.545		
	(1.299)		
Number of Obs.	1464	Log-likelihood	-959.52

Notes: t-ratios are in parentheses. ** denotes the 5% level, and * denotes the 10% level. See definitions of socio-demographic variables in Chapter 2 Table 2.1.

Table 4.4 Description of Gain and Loss Variables

Variable	Description
MICIG	= 1 if the alternative involves lower microbial illness risk; = 0 otherwise
MICDG	= 1 if the alternative involves lower microbial death risk; = 0 otherwise
CANIG	= 1 if the alternative involves lower cancer illness risk; = 0 otherwise
CANDG	= 1 if the alternative involves lower cancer death; = 0 otherwise
MICIL	= 1 if the alternative involves higher microbial illness risk; = 0 otherwise
MICDL	= 1 if the alternative involves higher microbial death risk; = 0 otherwise
CANIL	= 1 if the alternative involves higher cancer illness risk; = 0 otherwise
CANDL	= 1 if the alternative involves higher cancer death; = 0 otherwise
MIG	= MICI*MICIG
MDG	= MICD*MICDG
CIG	= CANI*CANDG
CDG	= CAND*CANDG
MIL	= MICI*MICIL
MDL	= MICD*MICDL
CIL	= CANI*CANDL
CDL	= CAND*CANDL
MIG2	= MIG*MIG
MDG2	= MDG*MDG
CIG2	= CIG*CIG
CDG2	= CDG*CDG
MIL2	= MIL*MIL
MDL2	= MDL*MDL
CIL2	= CIL*CIL
CDL2	= CDL*CDL
MIGBL	= MIG*BILL
MDGBL	= MDG*BILL
CIGBL	= CIG*BILL
CDGBL	= CDG*BILL
MILBL	= MIL*BILL
MDLBL	= MD*BILL
CILBL	= CIL*BILL
CDLBL	= CDL*BILL

Notes: 1) Other alternative attribute variables and socio-demographic variables are defined in Chapter 2 Table 2.1. 2) Gain and loss variable names are created by appending either a "G" (indicating Gain) or an "L" (indicating Loss) to the names of risk attributes.

Table 4.5 Log-Likelihood, AIC and BIC Values of Various Status-Quo-Reference-Dependent Models (Likelihood-Ratio Values in Bold Indicate Pooling is Rejected)

Model Specification		# para.	Log-likelihood			AIC		BIC		LR Test for Pooling
			CE2	CE3	CE23	CE2	CE3	CE2	CE3	
						The conditional logit specification				
Model 1.0	**MAIN** + SQ*__Z__	14	-398.46	-716.16	-1130.95	824.92	1460.33	889.10	1524.82	**32.66**
Model 1.1	Model 0 + **RISKG** + **RISKL**	22	-381.14 [a]	-687.03	-1094.75	806.28	1418.06	907.15	1519.41	**53.16**
Model 1.2	Model 1+**RISKG***BILL + **RISKL***BILL	30	-375.91	-678.10 [a]	-1084.59	811.81	1416.20	949.36	1554.40	**61.16**
						The random parameters logit specification				
Model 2.0	**MAIN** + SQ*__Z__	19	-363.20	-607.47	-992.54	764.40	1252.94	851.51	1340.47	**43.74**
Model 2.1	Model 0 + **RISKG** + **RISKL**	27	-353.34 [a]	-598.78 [a]	-985.85	760.68	1251.57	884.47	1375.95	**67.44**
Model 2.2	Model 3+**RISKG***BILL + **RISKL***BILL	35	-347.65	-593.15	-977.23	765.30	1256.31	925.77	1417.54	**72.86**

Notes: 1) **MAIN** is a vector of attribute variables that describe each alternative, i.e., *SQ, MICI, MICD, CANI, CAND* and *BILL*. **Z** is a vector of socio-demographic variables: *AGE65*, *INCOME, INCOME2, ENGLISH, CITYSIZE, ILLNESS, MALE* and *MARRY*. **RISKG** is a vector of variables that indicates gains in health risk reductions: MIG, *MDG, CIG* and *CDG*; **RISKL** is a vector of variables indicates losses in health risk reductions: *MIL, MDL, CIL* and *CDL*. 2) Models with squared gain and loss variables included cannot be estimated due to collinearity problems. 3) For RPL models, coefficients on *SQ, MICI* and *MICD, CANI* and *CAND* are assumed to be normally distributed. Other variables are estimated to have fixed effects only. [a] denotes preferred models across columns based on likelihood-ratio tests.

Table 4.6 Log-Likelihood, AIC and BIC Values of Various Attribute-Compromise Conditional Logit Models (Likelihood-Ratio Values in Bold Indicate Pooling is Rejected)

Model Name	Model Specification	# para.	CE3			CE23			LR Test for Common Parameters
			Log-Likelihood	AIC	BIC	Log-Likelihood	AIC	BIC	
Model 0.0	**MAIN** + SQ*Z	14	-716.16	1432.32	1534.36	-1099.09	2226.18	2300.22	
Model 0.1	Model 0.0 + **MAIN***CE2	20	-	-	-	-1079.94 [a]	2199.88	2305.66	**38.30**
Model 1.0	Model 0.0 +**RISKG** + **RISKL**	22	-684.17 [a]	1368.34	1528.70	-1084.57	2213.14	2329.50	
Model 1.1	Model 1.0 +**MAIN***CE2	28	-	-	-	-1073.53	2203.06	2351.15	**22.08**
Model 2.0	Model 1.0 + **RISKG2** + **RISKL2**	30	-681.68	1363.36	1582.03	-1078.93	2217.86	2376.53	
Model 2.1	Model 2.0+**MAIN***CE2	36	-	-	-	-1071.06	2214.12	2404.52	**15.74**
Model 3.0	Model 1.0 +**RISKG***BILL + **RISKL***BILL	30	-678.81	1357.62	1576.29	-1078.72	2217.44	2376.11	
Model 3.1	Model 3.0 + **MAIN***CE2	36	-	-	-	-1068.14	2208.28	2398.68	**21.16**
Model 4.0	Model 2.0 +**RISKG***BILL + **RISKL***BILL	38	-677.07	1354.14	1631.12	-1072.97	2221.94	2422.92	
Model 4.1	Model 4.0 + **MAIN***CE2	44	-	-	-	-1066.43	2220.86	2453.57	**13.08**

Notes: 1) **MAIN** is a vector of attribute variables that describe each alternative, i.e., *SQ, MICI, MICD, CANI, CAND* and *BILL*. **Z** is a vector of socio-demographic variables: *AGE65, INCOME, INCOME2, ENGLISH, CITYSIZE, ILLNESS, MALE* and *MARRY*. **RISKG** is a vector of variables that indicates gains in health risk reductions: *MIG, MDG, CIG* and *CDG*, and **RISKG2** is squared **RISKG; RISKL** is a vector of variables indicates losses in health risk reductions: *MIL, MDL, CIL* and *CDL*, and **RISKL2** is squared **RISKL;** 2) Models for CE2 cannot be estimated since there is no middle level of attributes for a 2-alternative choice set. 3) Likelihood-Ratio (LR) tests are conducted for common parameters on *SQ, MICI, MICD, CANI, CAND* and *BILL* between two models. [a] denotes preferred models across columns based on LR tests.

Table 4.7 Log-Likelihood, AIC and BIC Values of Various Attribute-Compromise Random Parameters Logit Models (Likelihood-Ratio Values in Bold Indicate Pooling is Rejected)

Model Name	Model Specification	# para.	CE3 Log-Likelihood	CE3 AIC	CE3 BIC	CE23 Log-Likelihood	CE23 AIC	CE23 BIC	LR Test for Common Parameters
Model 0.0	**MAIN** + SQ***Z**	19	-603.78[a]	1245.56	1346.05	-992.54	2023.08	2123.57	
Model 0.1	Model 0.0 + **MAIN***CE2	25	-	-	-	-972.73[a]	1995.46	2127.68	**39.62**
Model 1.0	Model 0.0 +**RISKG** + **RISKL**	27	-598.17	1196.34	1393.14	-987.57	2029.14	2171.94	
Model 1.1	Model 6.0 +**MAIN***CE2	33	-	-	-	-966.54	1999.08	2173.62	**42.05**
Model 2.0	Model 1.0 + **RISKG2** + **RISKL2**	35	-594.97	1189.94	1445.05	-979.56	2029.11	2214.23	
Model 2.1	Model 2.0+**MAIN***CE2	41	-	-	-	-964.34	2010.67	2227.52	**30.44**
Model 3.0	Model 1.0 +**RISKG***BILL + **RISKL***BILL	35	-591.22	1182.44	1437.55	-980.11	2030.23	2215.34	
Model 3.1	Model 3.0 + **MAIN***CE2	41	-	-	-	-960.73	2003.47	2220.31	**38.76**
Model 4.0	Model 2.0 +**RISKG***BILL + **RISKL***BILL	43	-589.65	1179.30	1492.72	-973.69	2033.38	2260.81	
Model 4.1	Model 4.0 + **MAIN***CE2	49	-	-	-	-959.12	2016.24	2275.40	**29.14**

Notes: 1) **MAIN** is a vector of attribute variables that describe each alternative in this study, i.e., *SQ, MICI, MICD, CANI, CAND* and *BILL*. **Z** is a vector of socio-demographic variables: *AGE65, INCOME, INCOME2, ENGLISH, CITYSIZE, ILLNESS, MALE* and *MARRY*. **RISKG** is a vector of variables that indicates gains in health risk reductions: *MIG, MDG, CIG* and *CDG*, and **RISKG2** is squared **RISKG; RISKL** is a vector of variables indicates losses in health risk reductions, includes *MIL, MDL, CIL* and *CDL*, and **RISKL2** is squared **RISKL**;. 2) Models for CE2 cannot be estimated since there is no middle level of attributes for a 2-alternative choice set. 3) For RPL models, coefficients on SQ, MICI and MICD, CANI and CAND are assumed to be normally distributed. Other variables are estimated to have fixed effects only. [a] denotes preferred models across columns based on LR tests.

Appendix 4.1

Table A4.1.1 Likelihood-Ratio Tests for Pooling the CE2 and CE3 Datasets

Model	Log-likelihood	Number of Observations
Full sample		
CE2	-458.29	812
CE3	-786.50	812
CE23 (Pooled CE2 and CE3)	-1262.62	1624
Likelihood-Ratio Test	**35.62****	
Yea-sayers removed sample		
CE2	-398.46	724
CE3	-716.16	740
CE23 (Pooled CE2 and CE3)	-1130.95	1464
Likelihood-Ratio Test	**32.66****	

Notes: 1) CE2 stands for the sample who were asked to answer binary choice questions and CE3 stands for the sample who were asked to answer trinary choice questions 2) χ^2 table values at the 5% level: $\chi^2(6) = 12.59$ (p value = 0.05); $\chi^2(12) = 21.03$. 3)Yea-sayers are identified in the samples who stated that they were willing to pay anything for health risk reductions.

Table A4.1.2 Test for Equal Preferences between the CE2 Model and the CE3 Model

Variable	Estimated Models				Test for Equal Coefficients	
	CE2		CE3			
	Coefficient	Standard Deviation	Coefficient	Standard Deviation	Difference	Standard Deviation
<u>Full sample</u>						
SQ	1.067	0.172	0.523	0.123	**0.540****	0.201
MICI	-8.66E-05	1.16E-05	-7.62E-05	7.40E-06	-9.34E-06	1.30E-05
MICD	-0.074	0.017	-0.054	0.011	-0.021	0.020
CANI	-0.022	0.003	-0.008	0.002	**-0.015****	0.004
CAND	-0.046	0.015	-0.058	0.011	0.012	0.018
BILL	-0.006	0.001	-0.004	0.001	**-0.002****	0.001
<u>Yea-sayers removed sample</u>						
SQ	1.132	0.183	0.528	0.127	**0.617****	0.222
MICI	-8.77E-05	1.25E-05	-7.55E-05	7.83E-06	-1.24E-05	1.54E-05
MICD	-0.081	0.018	-0.053	0.011	-0.029	0.021
CANI	-0.020	0.004	-0.008	0.002	**-0.013****	0.004
CAND	-0.041	0.016	-0.048	0.011	0.008	0.019
BILL	-0.006	0.001	-0.004	0.001	**-0.002****	0.001

Notes: 1) Models are estimated as basic condition logit models with only main attributes included. The main attributes include *SQ, MICI, MICD, CANI, CAND* and *BILL*. See definitions of these variables in Chapter 1 Table 1.1. 2) Standard deviations are calculated based on Krinsky-Robb simulation using 2000 draws. 3) Test for equal coefficients is conducted by testing whether the differences in coefficient estimates are significantly different from zero. 4) Yea-sayers are identified in the samples who stated that they were willing to pay anything for health risk reductions.

Table A4.1.3 lists counts and percentage of "yes" response to the status-quo (SQ) option and the Non-SQ options under a binary choice format (CE2) and trinary choice format (CE3). The percentage of choosing SQ is substantially higher when a dichotomous choice format is used than a multiple choice format. It seems that the sequence of choice decision does not affect the proportions of choosing SQ versus non-SQ options.

Table A4.1.3 Counts of "Yes" Responses in CE2 and CE3

Choice	CE2		CE3	
	Count	%	Count	%
Full sample				
SQ	514	63.30	323	39.77
Non-SQ	298	36.69	489	60.22
Total	812		812	
Yea-sayers removed sample				
SQ	479	66.16	306	41.35
Non-SQ	245	33.84	434	58.65
Total	724		740	

Notes: Yea-sayers are identified in the samples who stated that they were willing to pay anything for health risk reductions.

Results of a χ^2 square test of independence between the number of choosing the SQ option and choice format is shown in Table A4.1.4. The test statistics is based on the deviance between the actual frequency and the expected frequency assuming independence across the two categories: SQ versus Non-SQ options,

$$\chi^2 \text{ square} = \sum \frac{(x_{ij} - E_{ij})^2}{E_{ij}}$$

where the expected frequency assuming independence E_{ij} is,

$$E_{ij} = \frac{N_{i.} N_{.j}}{N_{..}}$$

and $N_{i.}$ is the sum across columns and $N_{.j}$ is the sum across rows.

Table A4.1.4 Comparison of Actual and Expected Frequency of Choosing SQ versus Non-SQ

	Actual		Expected		Number of Observations
	CE2	CE3	CE2	CE3	
<u>Full sample</u>					
SQ	514	323	418.5	418.5	837
Non-SQ	298	489	393.5	393.5	787
$N_{.j}$	812	812	812	812	1624
χ^2	**89.93****				
<u>Yea-sayers removed sample</u>					
SQ	479	306	388.2	396.8	785
Non-SQ	245	434	335.8	343.2	679
$N_{.j}$	724	740	724	740	1464
χ^2	**90.57****				

Notes: $\chi^2(6)$ =12.59 (p value = 0.05). ** denotes the 5% level.

Appendix 4.2

Calculation of Entropy

To calculate entropy, we need to calculate $\pi(x_j)$ first. It is the probability of each alternative being chosen in a choice set. It is standard practice to invoke the equal weights assumption on attributes in order to calculate $\pi(x_j)$. The expected utility of each alternative is then the sum of marginal utilities provided by each attribute. The probability of alternative *j* being chosen is,

$$\pi(x_j) = \frac{e^{\sum_{k=1}^{K} x_k}}{\sum_{j=1}^{J} e^{\sum_{k=1}^{K} x_k}} \tag{A4.2.1}$$

where x_k is the level of attribute *k* ($k = 1$ to 5 in this study, and indicates, in an order, *MICI, MICD, CANI, CAND* and *BILL*). To avoid expected utility being dominated by attributes of large magnitude, attribute values are recoded into levels that are orthogonal to each other and vary from -1 to 1 (they are orthogonal polynomials). Table A4.2.1 lists levels of attributes in terms of their actual values and their corresponding recoded levels.

Table A4.2.1 Unit Codes for Calculating Entropy for Each Choice Set

Attribute	Level					
	1	2	3	4	5	6
Actual values						
MICI	7500	15000	23000	30000	-	-
MICD	5	10	15	20	-	-
CANI	50	75	100	125	-	-
CAND	10	15	20	25	-	-
BILL	0	25	125	150	250	350
Coded values						
MICI	1	0.333333	-0.33333	-1	-	-
MICD	1	0.333333	-0.33333	-1	-	-
CANI	1	0.333333	-0.33333	-1	-	-
CAND	1	0.333333	-0.33333	-1	-	-
BILL	1	0.6	0.2	-0.2	-0.6	-1

Table A4.2.2 Correlation Matrix between Contextual Variables by Dataset

Variable	Variable				
			CE2		
	ENTROPY	*CUMENTRO*	*NADA*	*DISSD*	*CUMDISSD*
ENTROPY	1				
CUMENTRO	0.25	1			
NADA	-0.18	0.57	1		
DISSD	0.21	0.35	0.45	1	
CUMDISSD	0.27	0.43	0.33	0.12	1.00
			CE3		
	ENTROPY	*CUMENTRO*	*NADA*	*DISSD*	*CUMDISSD*
ENTROPY	1				
CUMENTRO	0.12	1			
NADA	-0.29	0.13	1		
DISSD	-0.31	0.15	0.49	1	
CUMDISSD	0.07	0.66	0.25	0.18	1.00

Appendix 4.3

Table A4.3.1 Estimated Conditional Logit Models for CE2, CE3 and CE23 (Contextual Variables in the Preference Function Only - Full Interaction)

Variable	CE2	CE3	CE23
SQ	-7.311	1.455	1.516**
MICI	6.98E-04	-2.22E-04	-1.12E-04**
MICD	0.182	-0.230	-0.121**
CANI	-0.010	-0.059	-0.029**
CAND	0.473	0.178	-0.010
BILL	-0.016	-0.003	-0.008**
*SQ * NADA*	0.988	-0.110	-0.161**
*MICI * NADA*	-1.06E-04	8.19E-06	1.48E-06
*MICD * NADA*	-0.058	0.015	0.007*
*CANI * NADA*	0.004	0.004	0.002**
*CAND * NADA*	-0.080	-0.013	-0.001
*BILL * NADA*	1.45E-03	-4.62E-04	-1.54E-04
*SQ * ENTROPY*	7.580	1.622*	1.566**
*MICI * ENTROPY*	-6.17E-04	2.82E-05	9.16E-06
*MICD * ENTROPY*	-0.214	-0.027	-0.052
*CANI * ENTROPY*	-0.052	-0.002	-0.012
*CAND * ENTROPY*	-0.326	-0.074	-0.046
*BILL * ENTROPY*	0.003	0.007	0.007**
*SQ * ORDER*	0.394	-0.662	-0.473*
*MICI * ORDER*	-6.89E-05	2.26E-05	2.49E-06
*MICD * ORDER*	0.034	0.028	0.030
*CANI * ORDER*	0.013	-0.001	-0.004
*CAND * ORDER*	-0.049	0.001	0.021
*BILL * ORDER*	0.006	-0.004*	-0.002
*AGE65*SQ*	-0.613**	-0.637**	-0.621**
*INCOME*SQ*	1.90E-05**	6.04E-06	1.32E-05**
*INCOME2*SQ*	-1.54E-10**	-8.08E-11	-1.21E-10**
*ENGLISH*SQ*	0.170	-0.117	0.013
*CITYSIZE*SQ*	-0.135**	-0.308**	-0.228**
*ILLNESS*SQ*	-0.476	-0:544	-0.360
*MALE*SQ*	0.538**	0.618**	0.579**
*MARRY*SQ*	0.061	0.466**	0.227**
Number of Observations	724	740	1464
Log-likelihood	-385.86	-689.94	-1086.83

Notes: ** denotes the 5% level, and * denotes the 10% level.

Table A4.3.2 Estimated Conditional Logit Models for CE2, CE3 and CE23 (Contextual Variables in the Preference Function Only - Partial Interaction)

Variable	CE2	CE3	CE23
SQ	0.111	0.866	1.361**
MICI	-8.40E-05**	-7.46E-05**	-7.82E-05**
MICD	-0.073**	-0.053**	-0.061**
CANI	-0.018**	-0.008**	-0.012**
CAND	-0.027	-0.047**	-0.045**
BILL	-0.007**	-0.004**	-0.005**
*SQ * NADA*	0.035	-0.028	-0.076**
*SQ * ENTROPY*	0.950	0.361	0.242
*SQ * ORDER*	-0.135	-0.045	-0.183
*AGE65*SQ*	-0.616**	-0.635**	-0.596**
*INCOME*SQ*	1.73E-05**	9.10E-06	1.17E-05**
*INCOME2*SQ*	-1.36E-10**	-1.01E-10**	-1.10E-10**
*ENGLISH*SQ*	0.133	-0.110	0.008
*CITYSIZE*SQ*	-0.140**	-0.296**	-0.223**
*ILLNESS*SQ*	-0.286	-0.568	-0.356
*MALE*SQ*	0.530**	0.642**	0.571**
*MARRY*SQ*	0.062	0.442**	0.262**
Number of Observations	724	740	1464
Log-likelihood	-385.75	-689.90	-1085.77

Notes: ** denotes the 5% level, and * denotes the 10% level.

Table A4.3.3 Estimated Conditional Logit Models for CE2, CE3 and CE23 (Contextual Variables in the Scale Function)

Variable	CE2	CE3	CE23
SQ	0.107	0.912	1.213**
MICI	-2.24E-05	-8.42E-05	-1.22E-04**
MICD	-0.021	-0.066	-0.105**
CANI	-0.005	-0.011	-0.018**
CAND	-0.011	-0.057	-0.072**
BILL	-0.002	-0.005	-0.007**
*AGE65*SQ*	-0.159	-0.661	-0.763**
*INCOME*SQ*	5.27E-06	1.10E-05	1.73E-05**
*INCOME2*SQ*	-4.06E-11	-1.15E-10	-1.54E-10**
*ENGLISH*SQ*	0.012	-0.141	-0.047
*CITYSIZE*SQ*	-0.020	-0.314	-0.264**
*ILLNESS*SQ*	-0.067	-0.701	-0.366
*MALE*SQ*	0.136	0.699	0.825**
*MARRY*SQ*	0.013	0.467	0.311
NADA	0.1832	-0.0189	-0.0493**
ENTROPY	1.7183**	0.2592	0.2308
ORDER	-0.3602	-0.1227	-0.2553**
Number of Observations	724	740	1464
Log-likelihood	-372.82	-689.84	-1091.86

Notes: ** denotes the 5% level, and * denotes the 10% level.

Table A4.3.4 Estimated Conditional Logit Models for CE2, CE3 and CE23 (Contextual Variables in both Preference and Scale Functions)

Variable	CE2	CE3	CE23
SQ	0.121	0.484	1.329**
MICI	-1.25E-05	-7.90E-05	-7.92E-05**
MICD	-0.011	-0.061	-0.071**
CANI	-0.002	-0.009	-0.013**
CAND	-0.005	-0.054	-0.048**
BILL	-0.001	-0.004	-0.005**
*SQ * NADA*	-0.027	-0.004	-0.074**
*SQ * ENTROPY*	0.071	0.519	0.227
*SQ * ORDER*	-0.018	-0.012	-0.130
*AGE65*SQ*	-0.099	-0.651	-0.571**
*INCOME*SQ*	3.40E-06	1.02E-05	1.20E-05**
*INCOME2*SQ*	-2.63E-11	-1.09E-10	-1.10E-10**
*ENGLISH*SQ*	0.009	-0.141	-0.032
*CITYSIZE*SQ*	-0.011	-0.305	-0.205**
*ILLNESS*SQ*	-0.044	-0.665	-0.370
*MALE*SQ*	0.086	0.685	0.558**
*MARRY*SQ*	0.007	0.470	0.245**
NADA	0.2861*	-0.0153	-0.0287*
ENTROPY	1.8257**	0.2537	0.5181
ORDER	-0.3323	-0.1415	-0.153
Number of Observations	724	740	1464
Log-likelihood	-380.86	-689.56	-1084.56

Notes: ** denotes the 5% level, and * denotes the 10% level.

Table A4.3.5 Estimated Random Parameters Logit Models for CE2 and CE3 (Contextual Variables in the Preference Function Only - Full Interaction)

Variable	CE2	CE3	Variable (Cont'd)	CE2	CE3
SQ	-10.921	0.558	*SQ * ORDER*	-1.138	-0.850
SQ_SD	1.984**	2.892**	*MICI * ORDER*	-8.94E-05	1.41E-05
MICI	1.31E-03	-2.97E-04	*MICD * ORDER*	-0.096	0.005
MICI_SD	5.02E-05	1.22E-04**	*CANI * ORDER*	0.053	0.005
MICD	0.706	-0.205	*CAND * ORDER*	-0.086	0.016
MICD_SD	-0.185	0.134**	*BILL * ORDER*	0.002	-0.005
CANI	-0.048	-0.066	*AGE65*SQ*	-1.180	-1.134
CANI_SD	0.018	0.030**	*INCOME*SQ*	3.69E-05	1.79E-06
CAND	0.933	0.294	*INCOME2*SQ*	-2.92E-10**	-1.13E-10
CAND_SD	0.050	-0.122**	*ENGLISH*SQ*	0.379	-0.421
BILL	-0.006	-0.010	*CITYSIZE*SQ*	-0.234	-0.588
*SQ * NADA*	1.403	-0.095	*ILLNESS*SQ*	-1.122	-0.743
*MICI * NADA*	-2.17E-04	1.03E-05	*MALE*SQ*	0.926	1.185
*MICD * NADA*	-0.154	0.009	*MARRY*SQ*	0.187	1.184
*CANI * NADA*	0.007	0.006			
*CAND * NADA*	-0.153	-0.020			
*BILL * NADA*	-2.02E-03	-5.06E-04			
*SQ * ENTROPY*	12.876	3.526			
*MICI * ENTROPY*	-1.05E-03	1.16E-05			
*MICD * ENTROPY*	-0.663	-0.039			
*CANI * ENTROPY*	-0.044	-0.022			
*CAND * ENTROPY*	-0.715	-0.182			
*BILL * ENTROPY*	0.005	0.012			
Number of Obs.	724	740	Log-likelihood	-345.78	-593.42

Notes: ** denotes the 5% level, and * denotes the 10% level.

Table A4.3.6 Estimated Random Parameters Logit Models for CE2, CE3 and CE23 (Contextual in the Preference Function Only - Partial Interaction)

Variable	CE2	CE3	CE23
SQ	0.352	0.450	2.581**
SQ_SD	1.865**	2.973**	2.278**
MICI	-1.46E-04**	-1.42E-04**	-1.39E-04**
MICI_SD	1.06E-04**	1.26E-04**	1.19E-04**
MICD	-0.146**	-0.131**	-0.132**
MICD_SD	0.145	0.136**	0.126**
CANI	-0.030**	-0.010**	-0.018**
CANI_SD	0.026**	0.028**	0.024**
CAND	-0.053	-0.080**	-0.071**
CAND_SD	-0.027	0.162**	0.140**
BILL	-0.010**	-0.008**	-0.009**
*SQ * NADA*	0.041	0.005	-0.160**
*SQ * ENTROPY*	1.876	1.162	0.703
*SQ * ORDER*	-0.432	-0.113	-0.500**
*AGE65*SQ*	-1.094*	-1.122	-1.143**
*INCOME*SQ*	2.83E-05	1.33E-05	1.57E-05
*INCOME2*SQ*	-2.17E-10*	-1.88E-10	-1.72E-10
*ENGLISH*SQ*	0.385	-0.439	0.031
*CITYSIZE*SQ*	-0.244	-0.629**	-0.426**
*ILLNESS*SQ*	-0.557	-0.707	-0.544
*MALE*SQ*	0.865*	1.427**	1.034**
*MARRY*SQ*	0.093	1.216*	0.623
Number of Observations	724	740	1464
Log-likelihood	-361.13	-603.13	-981.93

Notes: ** denotes the 5% level, and * denotes the 10% level.

Table A4.3.7 Estimated Random Parameters Logit Models for CE2, CE3 and CE23 (Contextual Variables in the Scale Function Only)

Variable	CE2	CE3	CE23
SQ	0.047	3.332	1.442
SQ_SD	0.095	7.756	2.348**
MICI	-7.69E-06	-3.05E-04	-1.46E-04**
MICI_SD	-5.66E-06	1.88E-03	1.24E-04**
MICD	-0.009	-0.294	-0.138**
MICD_SD	-0.008	0.060	0.128**
CANI	-0.001	-0.029	-0.020**
CANI_SD	0.001	0.324	0.026**
CAND	-0.004	-0.188	-0.088**
CAND_SD	-0.004	-0.046	0.136**
BILL	-0.001	0.013	-0.009**
*AGE65*SQ*	-0.060	-2.377	-0.984
*INCOME*SQ*	1.79E-06	3.33E-05	1.84E-05
*INCOME2*SQ*	-1.26E-11	-4.46E-10	-1.79E-10
*ENGLISH*SQ*	0.006	-1.017	-0.043
*CITYSIZE*SQ*	-0.010	-1.415	-0.396**
*ILLNESS*SQ*	0.000	-2.139	-0.772
*MALE*SQ*	0.034	3.252	1.197**
*MARRY*SQ*	0.003	2.506	0.554
NADA	1.304	-0.073	-0.0251
ENTROPY	-1.672	0.330	0.5825
ORDER	-0.645	-0.294	-0.3966
Number of Observations	724	740	1464
Log-likelihood	-359.78	-603.20	-990.57

Notes: ** denotes the 5% level, and * denotes the 10% level.

Chapter 5 Conclusions and Future Research

5.1 Conclusions

In Canada, provincial governments are largely responsible for the safety and security of drinking water. Each government develops its own water strategy to address various water management issues within the region. For example, as part of its provincial water strategy, *Water for Life: Alberta's Strategy for Sustainability*, periodic assessment of waterworks facilities have been conducted across Alberta. Results from the most recent summary report on the current facility assessment in Alberta indicates that about $290 million in capital investment is needed to upgrade current facilities to address immediate concerns about the functionality of these facilities (i.e., water quality, source, treatment and operation issues) (Alberta Environment 2004). Over 70% of these concerns are related to disinfection issue in water treatment or ongoing operation and monitoring (Alberta Environment 2004).[58] In Saskatchewan, shortly after the North Battleford incident, regulation for water treatment has been strengthened and about $87 million is going to be invested to upgrade current waterworks facilities within the next 20-30 years (Government of Saskatchewan 2004). Apart from large capital investment, the water strategy in both provinces emphasizes a collaborative approach between the government and citizens. As such, information on the public's opinion about drinking water quality and disinfection preferences in particular, is highly needed.

This thesis investigates Canadians' preferences for different municipal water treatment technologies that differ in their effectiveness in reducing microbial risk versus cancer risk based on their responses to a series of hypothetical choice questions. It is found that Canadians are willing to pay higher amounts to reduce microbial mortality risk from their drinking water than to reduce cancer mortality risk. In the new and updated 2007 guideline for Canadian drinking water quality[59], regulation for microbiological parameter is strengthened: the microbiological parameter for E. coli becomes 0 per 100ml

[58] The percentage is calculated using the information presented in Figure 3-4 Predominant Issues in Waterworks Facility Assessment: Summary Report (Alberta Environment 2004).
[59] http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/sum_guide-res_recom/revised-revisees_e.html#t1.

while no parameter is provided in a previous guideline. In contrast, the chemical parameter for Total Trihalomethanes (TTHMs) remains unchanged from the previous guideline (100 μg/L). Our result suggests that the direction of the regulatory changes seems to be appropriate.

Apart from identifying Canadians' preferred water treatment technology, this thesis also addresses some important empirical and methodological issues in the valuation of health risk reductions and some of these issues are relevant to non-market environmental valuation as well. Therefore, contributions of this thesis are both empirical and methodological to the health risk valuation literature as well as to the environmental valuation literature in general.

One of the major empirical contributions is that we provide value of statistical life (VSL) estimates in a new yet important policy context with some of the "best practice" assumptions invoked. In the environmental valuation literature, values of risk reduction are of the most important for policy analysis, especially the value of mortality risk reduction. According to Krupnick (2002) and Kochi et al. (2006), the benefits of mortality risk reduction capture about 75 to 90 percent of total policy benefits in regulatory analysis for pollution control. However, reported VSL estimates are found to have a considerably wide range, varying from $0.1 million up to $87.6 million US dollars (Mrozek and Taylor 2002; Viscusi and Aldy 2003; Kochi, Hubbell and Kramer 2006). Results from several review papers on published VSL estimates suggest that VSL estimates are affected by choice of medium (like air or water), valuation method, policy context, study area, population characteristics, the nature of risk and many other factors. Thus, it is recommended that the use of VSL should be matched with the case or similar policy context (Krupnick 2004). To provide a better match, there is a continued need to estimate VSLs based on "best-practices" across different policy contexts (USEPA Science Advisory Board 2007). In Canada, VSLs are needed when in the economic evaluation of health benefits (or health outcomes) of alternative polices or developing Canadian wide standards for air and water quality. However, the recommended official use of VSLs ($ 6.1 million in 2004 dollars) for health benefit assessment of public projects, according to the Canadian Cost-Benefit Analysis Guide 2007, is mostly based on U.S. wage-risk tradeoff studies (Treasury Board of Canada Secretariat 2007). Our

VSLs derived from a municipal drinking water risk reduction context based on stated preference survey are likely to be more appropriate for assessing health benefits of alternative public projects or environmental policies in Canada.

First of all, the nature of risk in our study is different from that in labour studies. Decisions to reduce health risks from drinking water involve both risk-risk and risk-dollar tradeoffs rather than risk-dollar tradeoffs only as in many wage-risk tradeoff studies. According to the Scientific Advisory Board (SAB) report on the U.S. Environmental Protection Agency's (USEPA) white paper *Valuing the Benefits of Fatal Cancer Risk Reduction*, VSLs based on wage-risks should not be used for fatal cancer risk reduction (USEPA 2000). Cases involving risk-risk tradeoffs are likely to be more challenging in regulatory decision makings.

In addition, there have been relatively few VSL estimates derived in a drinking water context. Many VSL estimates derived outside of wage-risk tradeoff contexts are derived in the air quality context (Deck and Chestnut 2006). For example, in an up-to-date database, the Environmental Valuation Reference Inventory (EVRI), maintained by Environment Canada, about 145 human health impact studies are based on air quality, and 95 are based on water quality, of which 46 studies are on drinking water quality (Appendix Table A1.2). However, only 4 studies are conducted in Canada (versus 20 in the U.S.A.), and none of them are concerned about tradeoffs between microbial risks and cancer risks in treating drinking water. Moreover, our VSL estimates are derived in a context where both mortality risk and morbidity risk are involved in contracting microbial disease or cancer. Controlling for the collinear relationship between mortality and morbidity risk is found to be important to avoid overestimation of value of mortality risk reductions (Bosworth, Cameron and DeShazo 2005). Hence, our VSL estimates are derived from a realistic context, which is characterized by multiple risk tradeoffs and collinear relationship between mortality risk and morbidity risk in each type of risk.

The model averaging approach outlined in the first paper shows that our derived willingness-to-pay (WTP) estimates are found to be robust across a wide range of model specifications. Furthermore, our survey adopted two types of stated preference techniques: both CVM and CE methods.[60] In a paper using the same dataset, the derived WTP estimates are shown to be reliable after a number of tests including scope tests and convergent validity tests (Adamowicz et al. 2007). Krupnick (2002) and Deck and Chestnut (2006), after reviewing current studies on VSLs, recommend the use of a stated preference approach. Goldberg and Roosen (2007), in their study examining consumers' WTP for health risk reduction from food using both CVM and CE methods, suggest that WTP estimates derived from the CE are more robust than those derived from the CVM.[61] Their results suggest that it might be beneficial to use the CE method when complex tradeoffs are involved. In sum, our VSL estimates can be used to make informed resource allocation decisions in cases with similar public policy contexts.

Methodologically, the thesis addresses how to handle uncertainty in WTP estimates using a model averaging approach, and how some of the major uncertainties in choosing model specification come from within a random utility framework (Chapter 2). The thesis also shows how to calibrate WTP estimates when willingness-to-pay responses for public risk reductions contain element of altruism (Chapter 3). In addition, the thesis also provides suggestions on survey design and how to control for the effect of choice format on preference elicitation (Chapter 4). Both methods should help improve the transferability of WTP estimates or VSL estimates by controlling for variations in model specification and survey design.

In the first paper (Chapter 2), we show how to derive WTP estimates that are not subject to a particular model specification by using model specification weighted averages in a random utility framework. In considering various specification possibilities of a random utility model, an experimental design approach is used, i.e., hierarchical levels are built into various model specification choices and their impacts on model fit are systematically assessed. It is found that our WTP estimates are similar among the best fitting models although a model averaging approach certainly improves the robustness of our estimates by explicitly communicating the process by which the estimates are

[60]A search of the Environmental Valuation Reference Inventory database (Appendix 1.1) indicates that about half of the health impact studies use the contingent valuation method (CVM) (a total of 183 studies) and only 2 studies use the choice experiment (CE) method.
[61] They find that WTP estimates based on CE is scope sensitive and convex in risk level in all scenarios, however, WTP based on CVM is scope sensitive only for the single health risk reductions scenario and embedding is observed for multiple risk reductions.

derived. Among a variety of estimated models, capturing unobserved heterogeneity in preferences improves model fit the most. Heterogeneity in scale or variance, in contrast, does not matter much in this study. Our study shows that decision complexity seems to affect preferences, but not scale. Although these findings are specific to the data and case used in the study, our results suggest that it is important to capture the way heterogeneity enters a model (preferences versus scale), and to control the way complexity affects preferences or scale through experimental design. We also reveal the relative efficiency of various model specifications and their interaction effects, which can be very useful for practioners who are interested in estimating RUMs. For example, it appears to be important to control for heteroscedasticity when a mixed logit specification is used. Recently, there has been a growing interest in applying meta analysis to synthesize different VSL estimates to reduce variability in published VSL estimates (for example, Kochi, Hubbell and Kramer 2006). However, in a recent report to the EPA, the SAB suggests that meta analysis should be used to determine criteria to select appropriate VSL studies to be matched with the case under study in terms of policy context and population characteristics (USEPA Science Advisory Board 2007). In contrast, the model averaging approach is used to improve robustness of estimates from a single study by taking the weighted average of individual estimates derived from various models with competing statistical performance.

In the second paper (Chapter 3), we distinguish an individual's willingness-to-pay by motivations based on actual self-protection expenditure data and provide our VSL estimates in a public good provision context (i.e., the altruism paper, Chapter 3). Our results confirm that individuals are willing to pay for other people's health risk reductions. In other words, the public VSL is greater than the private VSL, which is consistent with the finding reported by Strand (2004) and yet opposite to the finding reported by Johannesson et al. (1996). We report different VSL estimates conditional on the assumptions about the nature of altruism. For instance, if we assume non-paternalistic altruism, the VSL is estimated to be $7.6 million (in 2004 Canadian dollars) for one microbial death reduction and $4.9 million for one cancer death reduction. When paternalistic altruism is assumed, the estimated VSLs are more than doubled to their non-paternalistic counterparts. Although our estimated VSLs differ significantly by risk type (microbial or cancer risk), the estimated VSL to avoid one microbial death is much higher than the VSL to avoid one cancer death. This is probably due to the acute nature of microbial death caused by waterborne diseases.

In the third paper (Chapter 4), based on extended RUMs, we are able to reconcile preference differences inferred from two different survey formats: a two-alternative choice format and a three-alternative choice format. It appears that choice context affects preference elicitation, but the effect can be controlled and predicted if contextual variables are included in the preference function. It is thus suggested to design variation in the choice environment at the survey design stage and subsequently control it at the estimation stage. It is also suggested that multiple choice formats should be used in survey development so that the choice format effect can be captured and the revealed preferences are free or independent from context effects or are averaged over various contexts. The choice format paper provides an example of how to take into such contextual effects.

5.2 Future Research

There are several other interesting issues in the valuation of health risk reductions that remain unaddressed.

One issue is about the understanding of risks and human ability to process very small risk changes. We are dealing with a very small risk change – are people understanding these risk changes? Responses to debriefing questions on the understanding of the level of risks presented in the water survey indicate that over 90% participants stated that they understood that number of deaths and illness cases were occurred over a 35-year period within a community of 100,000 people. Our risk communication seemed to be effective, however, it is not certain whether respondents truly understood small risk tradeoffs given the fact that humans tend to make poor judgments under uncertainty (Kahneman and Tversky 1982). It is also found that a majority of respondents said they did not know using chlorine to disinfect water increases the risk of contracting bladder cancer although most of them believed the scientific information we presented about the cancer risk. It has been reported that estimated VSLs seem to vary in the level of risk that is presented in a survey (Viscusi and Aldy 2003; Alberini et al. 2004). For example, Alberini et al. (2004) found VSL estimates based on WTP for the 1-in-1000 risk reduction are 4 times greater than those derived based on WTP for the 5-in-1000 risk reduction. Viscusi and Aldy (2003), in their review of published VSLs, report that the mean risk level is between 0.00001-0.00025 and the estimated VSLs are between $3.9 and $21.7 million (2000 US dollars). The risk level in this study at the margin is 2.9E-07 (1 in 100,000 during a 35 year period), which is about 35 - 862 times smaller than previous studies. While our VSL estimates fall in the ballpark of the range of published VSL estimates, it is not clear how our VSL estimates are affected by the level of risk versus other factors like policy contexts (e.g., the public good property of program and nature of risk-risk tradeoffs), characteristics of respondents or other factors.

Another vein of issues in the valuation of health risk reductions that we did not pursue in this thesis are the latency effects of the cancer risk and characteristics of individuals and affected population on risk valuation. Recent research suggests that individuals are willing to pay higher amounts to reduce an immediate risk than a risk with a long latency period (Alberini et al. 2006; Alberini et al. 2007). Other research examines how individuals' age and health status affects their WTP to reduce health risks and whether VSLs should be age-adjusted (Alberini et al. 2004; DeShazo and Cameron 2005). Although a recent SAB report recommends the use of age independent VSL (USEPA Science Advisory Board 2007), there is increasing concern about using a mean VSL to measure the values of risk reductions that might be meaningful at the individual level (National Transportation Safety Board 2006). Raucher (2003) reports that VSLs adjusted for age, income growth, latency and discounting to avoid a bladder cancer is less than a quarter of the mean EPA-derived VSL estimates. In our study, we sidestep the latency issue associated with contracting bladder cancer by presenting both microbial and cancer risks over a 35-year period. Our finding that VSLs to avoid a microbial death is larger than that to avoid a cancer death indicates that respondents might have accounted for the acute nature of microbial disease and are willing to pay more to reduce an immediate risk. However, it is not clear that the risk premium to reduce microbial mortality risk versus the cancer risk comes from discounted rate of time preferences or preference to reduce one type of risk versus another.

On the issue of accounting for altruism in the demand for safe drinking water – have we really captured all the elements? Due to the joint production of benefits associated with better drinking water quality and imperfect substitution of drinking water for bottled water, it is possible that our altruistic values are overestimated. For example, some respondents who have installed filter systems at home and who voted yes for a new public program also stated that they were likely to consume more water if the opted new program was in place. So they might not be as altruistic as we assume. Although sensitivity analysis of willingness-to-pay estimates indicates the effect is relatively small. Another issue is the assumption about the nature of altruism — paternalistic versus non-paternalistic. Since public VSLs differ significantly depending on the type of altruism, it is important to invoke appropriate assumptions. If there is heterogeneity in the nature of altruism among altruists, how should VSLs be calibrated at the aggregate level?

Further research can also be directed to pursue other related issues in health risk valuation. For example, in the model averaging paper, it is found that including unobserved heterogeneity in preferences is important, but heterogeneity in scale or variance seems not as important. This also includes the fact that complexity seems to affect preferences, but not scale. These findings are specific to the data and case used in the study, but it would be interesting if they are found in other data sets and cases. As a public risk valuation study, we haven't taken into account characteristics of performance of a public program (such as the term of a project, immediate benefits or delayed benefits, public funds managerial methods, etc.) or the public's trust level in government, which might be important in citizens' voting decisions (Alberini et al. 2007). In the choice format paper, our finding that extended RUM models can be used to control for the choice format effect may only apply to the two-versus-three choice format phenomenon. To examine whether our approach can be used to explain other choice format effects, we could include choice formats of more alternatives, and compare whether there are changes in the factors that explain the choice format effect as we found in this study. Finally, we employed a stated preference method - it is always interesting to conduct experiments or revealed preference techniques to test whether there is hypothetical bias in WTP estimates.

In summary, this thesis reveals Canadians' preference for health risk reductions from drinking water. It is found that Canadians prefer a water treatment technique reducing both microbial risk and cancer risk although effectiveness in reducing microbial risk is more important in their choice decisions for different treatments. Currently, large capital investment in waterworks facilities is needed in many provinces, our results on Canadian's WTP for risk reductions can be used to make informed resource allocations decision to improve drinking water quality or to draw a line for choosing between different public projects involving human health benefits. This thesis has also addressed some important methodological issues in health risk valuation and environmental valuation, such as variation in welfare estimates caused by uncertainty in model selection, and the magnitude of altruistic value in individuals' WTP for risk reductions and how to calibrate the impact of choice format on preference elicitation. Although some other issues remain unaddressed, we have started to understand choices and preferences in the valuation of public health risk reductions. For instance, we develop testable hypothesis to show that WTP estimates can be decomposed by motives and we further show how to derive or estimate each component of WTP using both stated choice data and actual expenditure data within a random utility framework. We also found that people are willing to pay for other people's safety, and the magnitude of public VSL estimates depends on the assumption invoked about the nature of altruism. Additionally, we have also provided suggestions for both survey design and model estimation. It is found that while survey design can affect individuals' preference, its impact can be and should be controlled through careful survey design and estimation. Therefore, we suggest that it is important to build variation in complexity in survey development and it is also important to include unobserved heterogeneity and to extend a model to allow complexity affect both preference and scale. These suggestions, although derived in the context of Canadian drinking water risk reductions, are likely to be useful to other environmental valuation studies.

References

Adamowicz, W. L., D. Dupont, A. Krupnick and J. Zhang. 2007. *Valuation of Cancer and Microbial Disease Risk Reductions in Municipal Drinking Water*. RFF-DP-07-39. Resources for the Future.

Alberini, A., M. Cropper, A. Krupnick and N. B. Simon. 2004. Does the value of a statistical life vary with age and health status? Evidence from the US and Canada. *Journal of Environmental Economics and Management* 48(1): 769-792.

Alberini, A., M. Cropper, A. Krupnick and N. Simon. 2006. Willingness-to-pay mortality risk reductions: Does latency matters? *Journal of Risk and Uncertainty* 32(3): 231-245.

Alberini, A., S. Tonin, M. Turvani and A. Chiabai. 2007. Paying for permanence: Public preferences for contaminated site cleanup. *Journal of Risk and Uncertainty* 34(2): 155-178.

Alberta Environment. 2004. *Waterworks Facility Assessment Summary Report*. Alberta Environment. [accessed 4 December 2007]. Available at http://www.waterforlife.gov.ab.ca/docs/wwSummary_Report.pdf.

Bosworth, R., T. A. Cameron and J. R. DeShazo. 2005. Advances in evaluating the demand for risk prevention policies. *Working Paper*. [accessed 12 October 2006]. Available at http://www2.bren.ucsb.edu/~kolstad/events/OccWkshp/Bosworth-_Cameron_DeShazo.pdf.

Deck, L. and L. Chestnut. 2006. *Recommended Health Benefit Assessment Methods for the 2007 AQMP Socioeconomic Assessment.* Denver, Colorado: Stratus Consulting.

DeShazo, J. R. and Cameron, T. A. 2005. The effect of health status on willingness to pay for morbidity and mortality risk reductions. *California Center for Population Research Papers.* CCPR-050-05. [accessed 25 May 2006]. Available at http://repositories.cdlib.org/ccpr/olwp/CCPR-050-05.

Goldberg, I. and J. Roosen. 2007. Scope insensitivity in health risk reduction studies: a comparison of choice experiments and the contingent valuation method for valuing safer food. *Journal of Risk and Uncertainty* 34: 123-144.

Government of Saskatchewan. 2004. *Significant Progress in Saskatchewan's Water Strategy*. Webpage. [accessed 7 December 2007]. Available at http://www.gov.sk.ca/news?newsId=437c9a31-469f-49b6-b399-2aa68a175d3a.

Johannesson, M., P. Johansson and O'Conor R. M. 1996. The value of private safety versus the value of public safety. *Journal of Risk and Uncertainty* 13: 263-275.

Kahneman, D. and A. Tversky. 1982. Judgment under uncertainty: heuristics and biases. *Judgment under Uncertainty: Heuristics and Biases*, 555pp. Cambridge University Press: New York.

Kochi, I., B. Hubbell and R. Kramer. 2006. An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis. *Environmental & Resource Economics* 34: 385-406.

Krupnick, A. 2002. The value of reducing risk of death: a policy perspective. *The Journal of Policy and Management* 21(2): 275-282.

Krupnick, A. 2004. *Valuing Health Outcomes: Policy Choices and Technical Issues.* RFF Reports, Resources for the Future.

Mrozek, J. R. and L. O. Taylor. 2002. What determines the value of a life: a metal analysis. *Journal of Policy Analysis and Management* 21(2): 253-270.

National Transportation Safety Board. 2006. *Morbidity and Mortality: How Do We Value the Risk of Illness and Death?* EE-0497-02. Washington, DC 20594.

Raucher, R. S. 2003. Benefit-cost analysis and drinking water regulation. In: *Drinking Water Regulation and Health,* (ed.) P. Frederick. 225-249. John Willey & Sons.

Strand, J. 2004. Public and private-good values of statistical lives: results from a combined choice-experiment and contingent valuation survey. [accessed 20 May 2006]. Available at http://folk.uio.no/jostrand/lifepaper02.pdf.

Treasury Board of Canada Secretariat. 2007. Canadian Cost-Benefit Analysis Guide: Regulatory Proposals. Government of Canada. [accessed 20 November 2007]. Available at http://www.tbs-sct.gc.ca.

USEPA. 2000. *An SAB Report on EPA's White Paper Valuing the Benefits of Fatal Cancer Risk Reduction.* EPA-SAB-EEAC-00-013. U.S. Environmental Protection Agency. [accessed November 20 2007]. Available at http://www.epa.gov/sab/pdf/eeacf013.pdf.

USEPA Science Advisory Board. 2007. *SAB's Advisory on EPA's Issues in Valuing of Mortality Risk Reduction.* EPA-SAB-08-00. U.S. Environmental Protection Agency. [accessed 20 November 2007]. Available at http://epa.gov/sab/pdf/sab-08-001.pdf.

Viscusi, W. K. and J. Aldy. 2003. The value of a statistical life: a critical review of market estimates throughout the world. *The Journal of Risk and Uncertainty* 27(1): 5-76.

Appendix A

Public Policies for Water Management Discrete Choice Questionnaire

INTRODUCTION AT SITE

Welcome!

Thanks for agreeing to take part in our survey today!

As you go through the survey, please take the time to answer each question, as you can only move from one screen to the next in the survey after answering all the questions.

Preferences for Public Policies in Canada

Health Canada Santé Canada

UNIVERSITY OF ALBERTA

Brock University

A research project to support policy making and decision making. Sponsored by Health Canada. Conducted by researchers from the University of Alberta and Brock University.

We are conducting a survey with Canadians that presents options for the provision of cleaner and safer tap water. This survey asks for your opinions about public programs that may provide a benefit to you and your household but may also have some impact on your public utility bills. The survey will give you information about these programs and ask you a series of questions. This research is being conducted by the Canadian government through Health Canada and the Canadian Water Network – your opinions are important.

The research is being sponsored by Brock University and has been approved by the Brock University Research Ethics Board (File No. 02.330). If you have questions about research survey rights, please contact Michelle Roy at michelle.roy@ipsos-reid.com.

[SECTION 1: PRE-DISCRETE CHOICE QUESTIONNAIRE]

1. From which of the following sources do you get water in your household?

CHECK ALL THAT APPLY

Tap water from a municipal utility
Purchased bottled water
Water delivery service of spring, pure or distilled water (e.g. Culligan, Canadian Springs, Sparkling Springs, etc.)
Well water / natural well
Other (please specify)
Don't know

[THANK & TERMINATE IF <u>DO NOT</u> SELECT TAP WATER FROM A MUNICIPAL UTILITY IN Q1. PLEASE INCLUDE DQS AT Q1 IN THE DATAFILE AND APPEND PANEL DEMOGRAPHIC VARIABLES.]

[ASK Q2 IF MORE THAN ONE ITEM SELECTED IN Q1]

2. From which source do you get **most** of the water you and members of your household **drink** at home?

CHECK ONE ONLY

[INSERT LIST OF ITEMS SELECTED IN Q1

THE WATER YOU DRINK

This survey will focus on options for the provision of cleaner and safer tap water to your home. First we would like some information about the water you drink at home.

[NEW SCREEN]

3. Which, if any, of the following have you experienced with the tap water in your home over the past year?

CHECK ALL THAT APPLY

Rusty colour
Sediment (particles at the bottom of a glass)
Unpleasant smell (e.g., musty, chlorine)
Unpleasant taste (e.g., musty, chlorine)
Hard water / mineral deposits
Pollutants or other contamination
Other (please specify)
None of the above

4. Looking forward two years, do you expect the quality of your tap water at home to be…?

CHECK ONE ONLY

Worse than today
Same as today
Better than today
Don't know

[NEW SCREEN]
There are three sources of water to use in the home that will be discussed in this survey:

(i) Untreated tap water
(ii) Treated tap water (filtered or boiled in the home)
(iii) Purchased water (bottled or from home delivery)

5. For the three water sources, please indicate the percentage of water you **personally** consume **at home** that comes from each source – both now and one to two years ago.

IF YOUR ANSWER IS ZERO, YOU MUST SELECT 0% IN THE DROP-DOWN BOX

[PROGRAMMER NOTE: FORMAT SHOULD BE AS BELOW. TOTAL SHOULD ADD AUTOMATICALLY AND MUST ADD UP TO 100%]

	% Consumed Now	**% Consumed 1 to 2 Years Ago**
Untreated tap water	10%	25%
Treated tap water	80%	65%
Purchased water	10%	10%
Total	**100%**	**100%**

6. Thinking about **your own** personal water consumption at home from **all** sources, would you say you are drinking…?

CHECK ONE ONLY

More than the amount consumed 1 to 2 years ago
About the same amount of water as 1 to 2 years ago
Less than the amount consumed 1 to 2 years ago
Don't know

7. For the three water sources, please indicate the percentage of water members of your **household** other than yourself consume **at home** that comes from each source – both now and one to two years ago.

IF YOUR ANSWER IS ZERO, YOU MUST SELECT 0% IN THE DROP-DOWN BOX

[PROGRAMMER NOTE: FORMAT SHOULD BE AS BELOW. TOTAL SHOULD ADD AUTOMATICALLY AND MUST ADD UP TO 100%]

	% Consumed Now	% Consumed 1 to 2 Years Ago
Untreated tap water	10%	25%
Treated tap water	80%	65%
Purchased water	10%	10%
Total	**100%**	**100%**

8. Thinking about your household's water consumption at home from all sources, would you say members of your household, other than yourself, are drinking …?

CHECK ONE ONLY

More than the amount consumed 1 to 2 years ago
About the same amount of water as 1 to 2 years ago
Less than the amount consumed 1 to 2 years ago
Don't know

[ASK Q9 AND Q10 IF PURCHASED BOTTLED WATER OR WATER DELIVERY SERVICE SELECTED IN Q1, ELSE SKIP TO Q11]

9. What is the **primary** reason your household uses **purchased** water?

CHECK ONE ONLY

Convenience
Taste
Health concerns about tap water
Other (please specify)
Don't know

10. In an average month, how much money do you estimate that your **household** spends on **purchased** water to drink at home?

PLEASE ENTER YOUR BEST ESTIMATE

Nothing ($0)
$ _________ **[PROGRAMMER NOTE: MAY BE A RANGE]**
Don't know

11. Which, if any, of the following types of water filtration or treatment systems do you use at home?

CHECK ALL THAT APPLY

Container style water filter (e.g. Brita type systems)
Water filtration system that is attached to a tap
Water filtration system attached to a refrigerator
Water softener system
Fluoridation not already in your municipal water
None
Other [SPECIFY] ___________________
Don't Know

[ASK Q12 ONLY IF CONTAINER STYLE WATER FILTER OR WATER FILTRATION SYSTEM ATTACHED TO A TAP SELECTED IN Q11. ASK FOR EACH ITEM SELECTED IN Q11.]

[PROGRAMMER NOTE: SET UP Q12 TO 14 AND Q15 & Q16 AS A LOOP]

12. Do you own or rent your…[**INSERT ITEM**: CONTAINER STYLE WATER FILTER/WATER FILTRATION SYSTEM ATTACHED TO TAP FROM Q11]?

Own
Rent
Don't know

[**ASK Q13 FOR EACH ITEM <u>OWNED</u> IN Q12**]

13. Approximately, how much did you spend to buy your…[**INSERT ITEM**]? **[PROGRAMMER NOTE: MAY BE A RANGE]**

PLEASE ENTER YOUR BEST ESTIMATE

Nothing (we did not purchase it)
$ _________
Don't know

[ASK Q14 FOR EACH ITEM RENTED IN Q12]

14. Approximately, how much do you spend per month to rent…**[INSERT ITEM]**? **[PROGRAMMER NOTE: MAY BE A RANGE]**

PLEASE ENTER YOUR BEST ESTIMATE

$ __________
Don't know

[ASK Q15 and 16 ONLY IF CONTAINER STYLE WATER FILTER OR WATER FILTRATION SYSTEM ATTACHED TO A TAP SELECTED IN Q11]

15. How much do you spend for **each** replacement filter for your…**[INSERT ITEM FROM Q11]**?

$ __________
Don't know

16. And, how frequently do you replace the filters for this home system?

Weekly
Once a month
Once every two to three months
Once every four months
Twice a year
Once a year
Less than once a year
Don't know

17. How often, if ever, do you boil your tap water at home before drinking it (i.e., to make it safer or taste better, not for making a hot beverage such as tea)?

CHECK ONE ONLY

Always
Weekly
Monthly
Never
Don't know

YOUR VIEWS ON THE SAFETY OF DRINKING YOUR TAP WATER AT HOME

The following questions are specific to the quality of your household tap water (i.e., **not** water that you treat or purchase).

[NEW SCREEN]

18. Which of the following statements **best** reflects your **personal** opinion about health concerns you might have with the tap water in your home?

CHECK ONE ONLY

Drinking tap water **does not** pose a problem for my health or my family's health.
Drinking tap water poses a **minor** problem for my health or my family's health.
Drinking tap water poses a **moderate** problem for my health or my family's health.
Drinking tap water poses a **serious** problem for my health or my family's health.

19. For each of the following items that may be present in a household's tap water, please indicate if you have **heard** about it as a concern with drinking tap water and if any of these items have been a special concern in your **community**.

CHECK ALL THAT APPLY IN EACH COLUMN

Microbe -- E. coli
Microbe -- Cryptosporidium
Microbe -- Giardia
Chemical -- Fluoride
Chemical -- Trihalomethanes
Chemical -- Pesticides
Metals -- Iron, Lead, Mercury
None of the above

Heard about as a drinking water concern **[COLUMN 1]**
Specific concern in my community **[COLUMN 2]**

20. Considering each of these, how much of a health concern do you personally believe each poses in your home's tap water?

Microbe -- E. Coli
Microbe -- Cryptosporidium
Microbe -- Giardia
Chemical -- Fluoride
Chemical -- Trihalomethanes
Chemical -- Pesticides
Metals -- Iron, Lead, Mercury

No health concern
Minor health concern
Moderate health concern
Serious health concern
Don't know/not sure

21. Does anyone in your household have any health conditions that require them to take special care with the water they drink?

CHECK ONE ONLY

Yes
No
Don't know

22. To the best of your knowledge, have you or has anyone in your household ever become sick from drinking the tap water in your home?

CHECK ONE ONLY

Yes
No
Don't know

23. In your opinion, what is the **primary** way safer tap water for your home should be paid for?

CHECK ONE ONLY

Increase Federal, Provincial or Municipal taxes
Increase prices to tap water users
Charge businesses that worsen water quality
Other (please specify)
There is no need for safer tap water
Don't know

[SECTION 2: CVM AND DISCRETE CHOICE INFORMATION SCREENS]

[NOTE: FOR ALL INFORMATION SCREENS, FONTS SHOULD BE 12-POINT. THE WIDTH SHOULD BE OPTIMIZED FOR THE ONLINE SCREEN – I.E., NO HORIZONTAL SCROLLING AND NO, OR A MINIMUM OF, VERTICAL SCROLLING]

[INFORMATION SCREEN 1]

A CASE FOR STUDY

We would like your opinions about the management of tap water quality to your home and your community.

Please read the information below before moving to the next section of the survey.

Health Effects of Microbes in Tap Water

Water utilities are concerned with providing tap water that is as free as possible from microbes. While many people are familiar with the harm caused by the bacteria, E. coli, in Walkerton, Ontario, this is not the only microbe of concern. Over the last 10 years, several communities across Canada have experienced problems with other microbes such as cryptosporidium and giardia. All of these microbes cause similar problems:

- Symptoms of microbial illness:
  - Stomach pain or cramps, nausea or vomiting, diarrhea, blood in stools, and low-grade fevers.
  - Symptoms appear soon after infection.
  - Typically, a microbial infection lasts for about two weeks.
  - Death can result if a sensitive person gets the disease, but death is rare. Death would occur **soon** after infection.
- Treatment of Illness:
  - Over the counter and prescription drugs.
  - Rest.
  - In more severe cases, fluid loss can lead to hospitalisation.
- Sensitive Groups:
  - People with weak immune systems including the very young, the very old, those who have had chemotherapy and those who have HIV-AIDS.
- Tap Water Treatment:
  - Providers of tap water typically disinfect the water supply with chlorine. Chlorine is used around the world because it is cheap and fairly effective against microbes. More expensive technologies are available to reduce further the effects of microbes.

[NEW SCREEN]

24. Before this survey, did you know that chlorine is used to kill microbes in drinking water?

Yes
No
Don't know/not sure

[INFORMATION SCREEN 2]

Health Effects of Chlorine

When tap water is disinfected with chlorine, various by-products including Trihalomethanes (THMs) are produced. Scientists believe that THMs are an indicator for substances in the tap water that are linked to increased cases of bladder cancer when water is consumed over long periods of time.

- Symptoms of bladder cancer:
  - Urgent and frequent need to urinate, blood in your urine, pain during urination, and pain from the tumour.
  - Symptoms for this cancer do not occur immediately after drinking tap water, rather they take years to show since it takes years for this cancer to develop.
  - For about one in five cases, death occurs within five years from diagnosis.
- Medical Treatment of Illness:
  - Surgery, radiation, and chemotherapy are used to treat bladder cancer.
  - Side effects from surgery may include a long recuperation period and the need for colostomy (bag for body wastes).
  - Side effects of chemotherapy include loss of hair, change in taste or smell, mouth sores, possible loss of fertility, fatigue and less ability to deal with infections.
- Sensitive Groups:
  - Occurs most frequently in male smokers over the age of 70, but other older people can also get this cancer.
- Tap Water Treatment:
  - Providers of tap water can lower the chlorine levels in the water supply.
  - Less chlorine lowers cancer risks but raises microbial risks.
  - More expensive water treatment technologies are available to reduce both cancer risks and microbial risks.

[NEW SCREEN]

25. Before this survey, did you know that using chlorine to disinfect drinking water can increase one's chances of getting bladder cancer?

Yes
No
Don't know/not sure

[INFORMATION SCREEN 3]

Health Effects of Microbes and THMs in Tap Water

You won't need to remember these numbers. We just want to give you some idea of the risks people face.

First we list effects from all causes, then we list effects from drinking tap water only.

Microbial Health Effects in Numbers	Cancer Health Effects in Numbers
From all causes of microbial disease	**From all causes of cancer**
♦ Scientists estimate that for every 100,000 people: - Over a 35-year period, microbes from all sources (food, tap water and direct contact such as swimming), lead to 2.5 million cases of microbial infection. This means that a person may likely suffer multiple episodes of microbial illness over this period. - Over a 35-year period, about 100 deaths occur from microbes from all sources.	♦ Scientists estimate that for every 100,000 people: - Over a 35-year period, 27,000 people will contract cancer of all types. - Of these 27,000 people, 7,000 deaths are due to cancer of all types.
From drinking tap water	**From drinking tap water**
♦ Scientists estimate that for every 100,000 people drinking tap water: - Over a 35-year period, 23,000 people will get some sort of microbial infection. - Of those infected, 15 will die over the 35-year period. Death often occurs soon after infection.	♦ Scientists estimate that for every 100,000 people drinking tap water: - Over a 35-year period, 100 people will contract bladder cancer. - Of these, approximately 20 persons will die within 5 years as a direct consequence of the cancer. - Out of the 80 who do not die, some will be fully cured; others will experience cancer symptoms, and require medical interventions and drugs over their remaining lifetime.

This information is summarized in the following screen.

Sources for Health Effects Estimates **[THIS SHOULD BE A HYPERLINK]**

[INFORMATION IN POP-UP: Pierre Payment and Merry S. Riley (2002), Resolving the Global Burden of Gastrointestinal Illness: A Call To Action, *American Academy of Microbiology*; Donald Wigle. (2000), "Safe Drinking Water: A Public Health Challenge," *Chronic Diseases in Canada,* Volume 19; and Canadian Cancer Statistics (2002), Statistics Canada.]

[INFORMATION SCREEN 4]

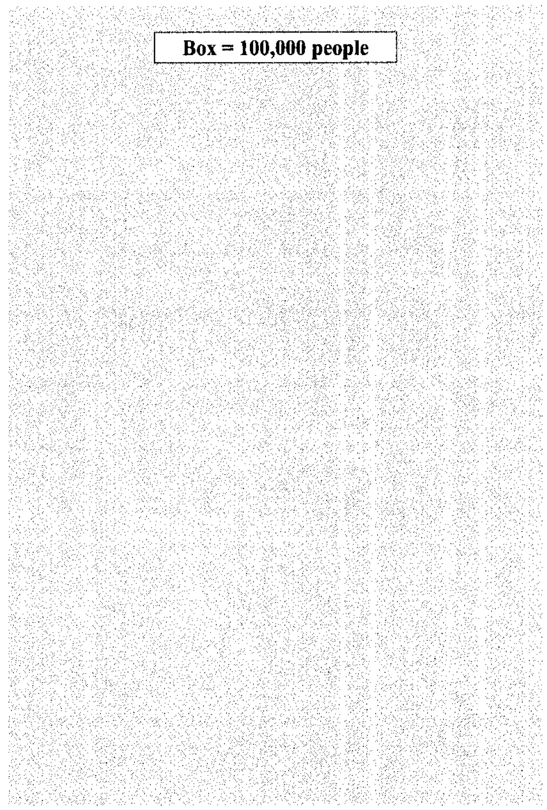
For a community of 100,000 people, over a 35-year period, illnesses and deaths from microbial disease and cancer will be approximately...

	MICROBIAL DISEASE		*CANCER*	
	Illnesses	**Deaths**	**Illnesses**	**Deaths**
From all Causes	2,500,000	100	27,000	7,000
	⬇	⬇	⬇	⬇
From Drinking Tap Water	23,000	15	100	20

On the next two screens, this situation is shown with pictures.

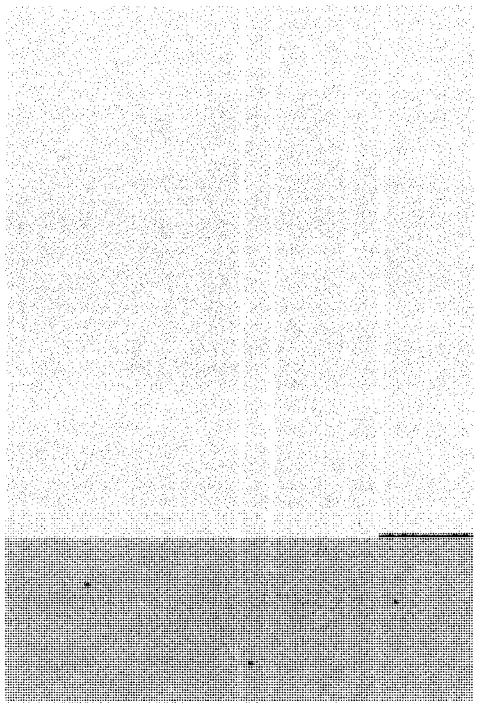
[INFORMATION SCREEN 5]

- The light blue box represents a 100,000 person community.



[INFORMATION SCREEN 6]

- ▒ → The orange section of the picture shows cases of microbial illness. The orange covers 23% of the box, representing 23,000 microbial illnesses per 100,000 people.
- ▓ → The red section of the picture shows cases of bladder cancer. The red section represents 100 cases of cancer per 100,000 people.
- The small black squares [■] show deaths, from microbial illness if they're inside an orange section, or from bladder cancer if they're inside a red section.
  - The three black squares inside the orange section represent 15 deaths from microbial illness per 100,000 people.
  - The four black squares inside the red section represent 20 deaths from bladder cancer per 100,000 people.



In the following screens, we will be using pictures like this one to help you understand the choices you will be voting on.

Although the boxes may be smaller, they will be proportionately correct.

[INFORMATION SCREEN 7]

MANAGEMENT OPTIONS FOR TREATING DRINKING WATER

Chlorination isn't the only way your municipal water utility can disinfect your tap water. Other ways to purify tap water include ozonation and ultraviolet techniques. These water purification techniques are safe methods to disinfect drinking water, and are currently used in other countries to successfully reduce adverse health effects from microbes and Trihalomethanes.

If any of the methods are adopted for use by your water utility, the costs of your tap water will increase. Different methods have different costs and different effects on health.

Currently, the average household in Canada pays between $250 to $400 per year for its tap water.

Here's Where You Can Have a Say

In this survey, we will ask you to consider management programs to reduce health effects from drinking water produced by your water utility.

Compare the alternative management programs to the current situation. Choose your preferred option as if you were voting in a referendum. You will vote several times.

Please vote each time independently from the other votes.

Assume that the drinking water from your tap at your home will taste the same, smell the same and have the same colour no matter what choice you make.

[INFORMATION SCREEN 8]

Please note

We know that how people vote in surveys is often not a reliable indication of how people will actually vote. In surveys some people ignore the sacrifices they would need to make if their vote actually meant they would have less money to spend. In a recent survey like this one, 55% of the people in a community voted for a new program. When the program was put to a vote for real, only 40% actually voted for the program. Therefore, we'd like you to vote in this survey as if your vote was real -- imagine that you actually will have to dig into your pocket and pay the additional charges on your household's water bill if the majority agreed to go ahead with a program.

Some people might choose to vote to keep the current situation because they think:

- It is too much money for the type and number of health improvements.
- The community's tap water is safe enough.
- There are other places, including other health prevention options, where my money would be better spent.
- No one in my household drinks tap water, so this doesn't concern me.

Other people might choose one of the management options because they think:

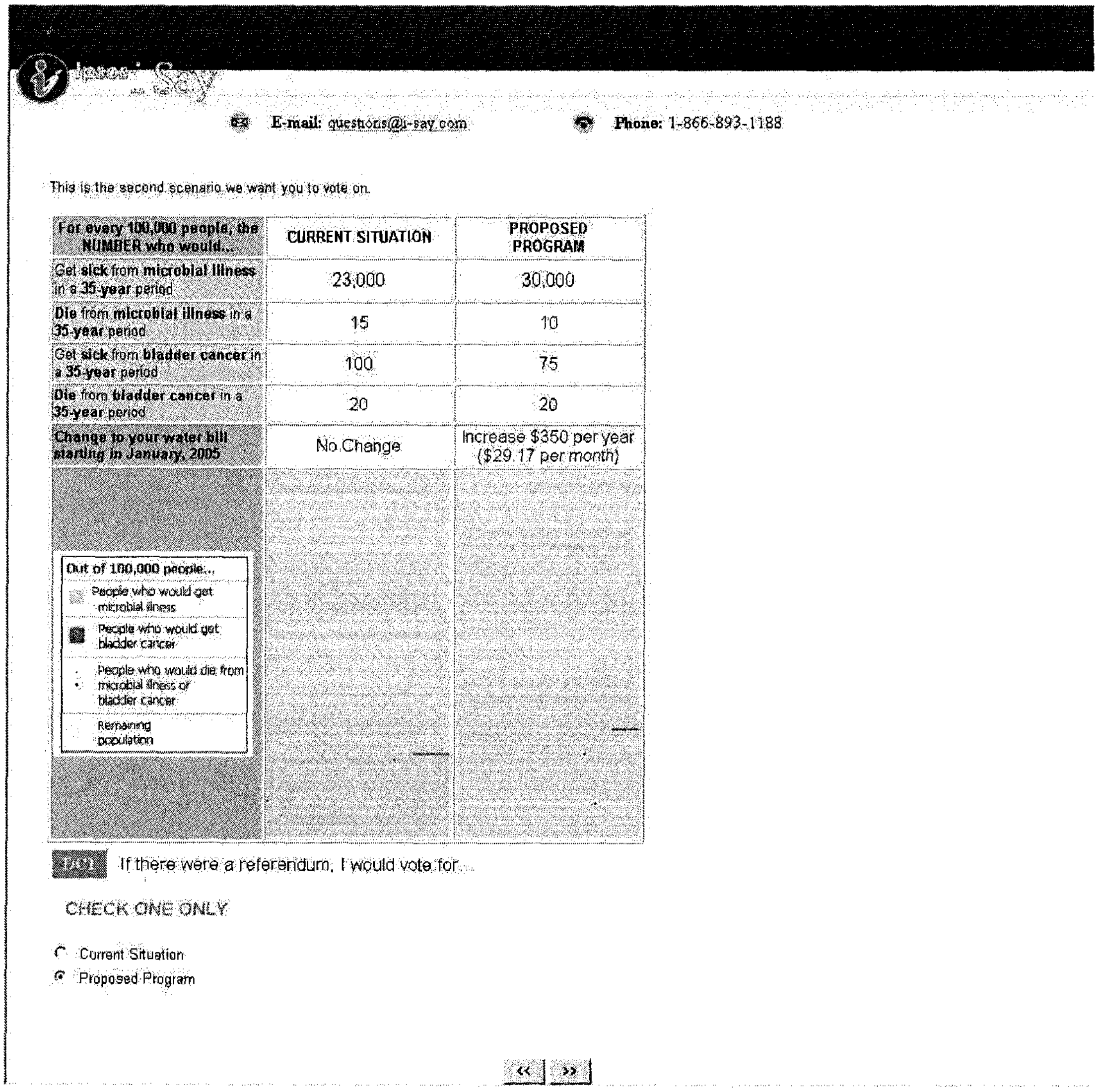
- The reduction in health effects is worth the money.
- The community's water is not safe enough.
- This is a good use of money compared to other things I can spend my money on.

[SECTION 3: CVM SECTION]

(Not included in this thesis)

[SECTION 4: DC SECTION]

Here is an example for a two-alterative choice question.



E-mail: questions@i-say.com Phone: 1-866-893-1188

This is the second scenario we want you to vote on.

For every 100,000 people, the NUMBER who would...	CURRENT SITUATION	PROPOSED PROGRAM
Get sick from microbial illness in a 35-year period	23,000	30,000
Die from microbial illness in a 35-year period	15	10
Get sick from bladder cancer in a 35-year period	100	75
Die from bladder cancer in a 35-year period	20	20
Change to your water bill starting in January, 2005	No Change	Increase $350 per year ($29.17 per month)

Out of 100,000 people...

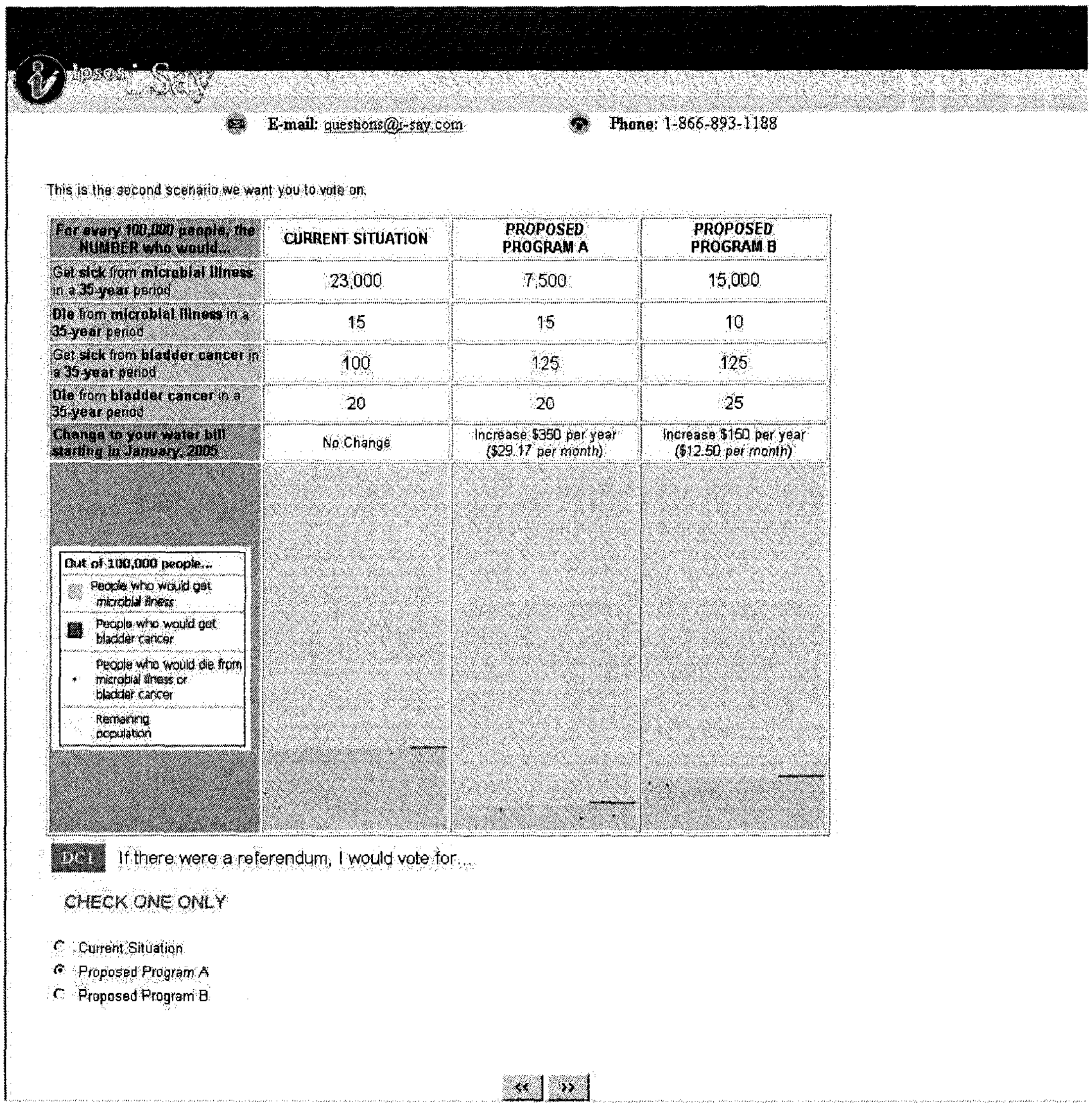
- People who would get microbial illness
- People who would get bladder cancer
- People who would die from microbial illness or bladder cancer
- Remaining population

DC1 If there were a referendum, I would vote for...

CHECK ONE ONLY

- Current Situation
- Proposed Program

Here is an example for a three-alterative choice question.



Ipsos i-Say

E-mail: questions@i-say.com Phone: 1-866-893-1188

This is the second scenario we want you to vote on.

For every 100,000 people, the NUMBER who would...	CURRENT SITUATION	PROPOSED PROGRAM A	PROPOSED PROGRAM B
Get sick from microbial illness in a 35-year period	23,000	7,500	15,000
Die from microbial illness in a 35-year period	15	15	10
Get sick from bladder cancer in a 35-year period	100	125	125
Die from bladder cancer in a 35-year period	20	20	25
Change to your water bill starting in January, 2005	No Change	Increase $350 per year ($29.17 per month)	Increase $150 per year ($12.50 per month)

DC1 If there were a referendum, I would vote for...

CHECK ONE ONLY

- Current Situation
- Proposed Program A
- Proposed Program B

<< >>

[SECTION 5: POST-DISCRETE CHOICE QUESTIONNAIRE]

PLEASE TELL US ABOUT YOURSELF

For a variety of reasons, people of different age, gender, and background may face different health effects from drinking tap water. In order to best understand and utilize survey results, it will be important for us to know some of these details about you.

Please be assured all information provided will be kept **strictly confidential**.

[NEW SCREEN]

26. Compared to others your age would you say your health is...?
CHECK ONE ONLY

Much better
Somewhat better
About the same
Somewhat worse
Much worse
Don't know

27. Compared to your general health now, do you expect your health ten years from now to be...?
CHECK ONE ONLY

Much better
Somewhat better
About the same
Somewhat worse
Much worse
Don't know

28. In the past 12 months, have you ever been a patient overnight in a hospital, nursing home or convalescent home?

Yes
No
Decline to respond

29. Which, if any, of the following long-term health conditions do you or members of your household have? [PLEASE SET UP AS GRID]

CHECK ALL THAT APPLY. PLEASE CHECK AT LEAST ONE RESPONSE IN EACH COLUMN.

[ACROSS THE TOP]]

Myself
A member(s) of my household

[ALONG THE SIDE]

Food allergies
Any other allergies
Asthma
Arthritis or rheumatism
Back problems, excluding arthritis
High blood pressure
Migraine headaches
Chronic bronchitis or emphysema
Sinusitis
Diabetes
Epilepsy
Heart disease
Cancer (Please specify type)
Stomach or intestinal ulcers
Effects of a stroke
Any other long-term condition that has been diagnosed by a health professional (Please specify)
None of the above

30. Are you…?
Male
Female
Decline to respond

31. What is your age in years?

ENTER NUMBER. PLEASE DO NOT ENTER DECIMALS **[RANGE 18 TO 120]**

[RECORD RESPONSE]
Decline to respond

32. Which of the following is the highest level of education you have completed?

CHECK ONE ONLY

Grade school or some high school
Completed high school
Post-secondary technical school
Some university or college
Completed college diploma
Completed university undergraduate degree
Completed post-graduate degree (masters or Ph.D.)
Decline to respond

33. What is your current employment status?

CHECK ONE ONLY

Working full time outside the home or self employed
Working part time outside the home or self employed
Student
Homemaker
Retired
Unemployed
Decline to respond

We would like your general opinion about the level of income taxes you pay. **[ON SAME SCREEN AS Q34]**

34. Do you consider that the amount of income tax you pay is…?

CHECK ONE ONLY

Too high
About right
Too low
Don't know

35. Do you consider that the amount you pay for your water bill is...?

CHECK ONE ONLY

Too high
About right
Too low
Don't know
I do not pay a separate bill for my home's tap water

36. We would also like your opinions about spending on public services. For each of the publicly-provided services listed below, please indicate if you personally think funding for these services should be reduced substantially, reduced somewhat, not changed, increased somewhat, or increased substantially. **[RANDOMIZE ORDER OF PRESENTATION]**

Education services in elementary and secondary schools
Support for colleges and universities
Policing services
Health care services
Providing and maintaining natural areas and wildlife refuges
Providing clean tap water
Providing roads and highways

Reduced substantially
Reduced somewhat
Not changed
Increased somewhat
Increased substantially
Not sure

QUESTIONS ABOUT THE SURVEY YOU JUST COMPLETED

37. Please indicate your agreement or disagreement with each of the following statements regarding the survey you just completed. **[RANDOMIZE ORDER OF PRESENTATION, EXCEPT STATEMENT 1 SHOULD ALWAYS BE FIRST]**

The survey was clear and easy to understand
In making my decision about which programs to choose I thought about the size of my current water bill
In making decisions about the which treatment options I preferred I considered whether I could afford to pay a higher water bill
Reducing health effects among sensitive people was important to my vote
There was too much information about health effects
The information provided about health effects helped me decide how to vote

Strongly agree
Somewhat agree
Neither agree nor disagree
Somewhat disagree
Strongly disagree
Don't know

38. Please indicate your agreement or disagreement with each of the following statements regarding water treatment options. **[RANDOMIZE ORDER OF PRESENTATION]**

I was concerned that there will be side effects associated with the proposed water treatment options
The water treatment options presented here do not substantially improve health effects
The decision about water treatment options should be left to experts
The public should not have to pay for new water treatment options
I am willing to see my household water bill increase by as much as it takes to reduce deaths and illnesses from drinking tap water

Strongly agree
Somewhat agree
Neither agree nor disagree
Somewhat disagree
Strongly disagree
Don't know

39. Do you believe the information presented in this survey about the microbial effects associated with drinking tap water in your community?

Yes
No
Don't know/not sure

[ASK Q40 IF NO OR DK/NS IN Q39]
40. Do you think that these microbial effects are larger or smaller than those presented in the survey for your community?

CHECK ONE ONLY

Much Larger
Somewhat Larger
Somewhat Smaller
Much Smaller
Don't know/not sure

41. Do you believe the information presented in this survey about the cancer effects associated with drinking household tap water in your community?

Yes
No
Don't know/not sure

[ASK Q42 IF NO OR DK/NS IN Q41]
42. Do you think that these cancer effects are larger or smaller than those presented in the survey for your community?

CHECK ONE ONLY

Much Larger
Somewhat Larger
Somewhat Smaller
Much Smaller
Don't know/not sure

43. Comparing health effects from drinking bottled water to health effects from drinking your home's tap water, do you think that bottled water is:

CHECK ONE ONLY

Much more safe than tap water
A little safer than tap water
About as safe as tap water
A little less safe than tap water
Much less safe than tap water
Don't know/not sure

44. Did you understand that your water bill would increase for the foreseeable future if any of these programs were put in place?

Yes
No
Don't know/not sure

45. When you looked at the numbers of health effects from drinking your home's tap water, did you understand that these numbers were for a 35-year period?

Yes
No
Don't know/not sure

46. When you voted, did you understand that these numbers related to a community population of 100,000?

Yes
No
Don't know/not sure

47. When you were making your choices between alternative programs where cancer and microbial illness and mortality were being reduced, how important were each of the characteristics below to your decision?

Numbers of microbial illnesses
Numbers of deaths from microbial illness
Numbers of cancer illnesses
Numbers of deaths from cancer illnesses
Total number of illnesses (microbial plus cancer)
Total number of deaths (microbial plus cancer)
Costs to my household

Extremely important
Very important
Somewhat important
Not very important
Not at all important
Don't know

48. Considering yourself, your family, and your community please assign a total of 10 points among the three groups according to their influence on your program choices. For example, if you thought only about your family, but not yourself or the community in making your choices, you would allocate all 10 points to your family. **[PROGRAMMER NOTE: TOTAL SHOULD SUM AUTOMATICALLY AND MUST SUM TO 10]**

Myself
My family (not including myself)
Others in the community

49. How certain do you believe scientists are about microbial illness arising from drinking tap water?

CHECK ONE ONLY

Very certain
Somewhat certain
Somewhat uncertain
Very uncertain
Don't know/not sure

50. How certain do you believe scientists are about bladder cancer arising from drinking tap water?

CHECK ONE ONLY

Very certain
Somewhat certain
Somewhat uncertain
Very uncertain
Don't know/not sure

[PROGRAMMER NOTE: PLEASE DISABLE THE BACK BUTTON. WE DO NOT WANT RESPONDENTS TO BE ABLE TO GO BACK AND CHANGE THEIR ANSWERS]

51. When you chose among the programs and status quo, what did you notice about the relationship between cancer cases and deaths from cancer in any of the programs?

I did not compare cancer deaths to cases
The relationship between cancer deaths and cases was always the same
The relationship between cancer deaths and cases was always different
I do not remember what I noticed

THANK YOU VERY MUCH FOR YOUR COOPERATION!

POST-SURVEY DEBRIEF

Thank you for your participation in our survey about the views that Canadians have about public policies for water management that has been approved by the Brock University Research Ethics Board (File No. 02.330). We hope that you enjoyed your participation. Your help has been very valuable to us.

If you wish to have further information about this research entitled "Valuing Drinking Water Quality", please contact Tech Support at panel@i-say.com or 1-888-618-2056.

At this point we would also like to explain more about this research. The study is being conducted by researchers at the Department of Economics, Brock University in St. Catharines, ON, the Department of Rural Economy at the University of Alberta in Edmonton, AB, and Resources for the Future, Washington, DC. The purpose is to better understand how people view potential risks and benefits of different water treatment techniques and their likely choices in a voting context. The results of this study will be reported in terms of averages only and individual responses will be confidential. The report of these results will be publicly available and may help to improve Canadian water management policies.

To access the quoted information on possible health effects of different water treatment programs, please click on the following links: **[LINKS SHOULD OPEN IN A SEPARATE WINDOW]**

Information about guidance procedures for safe drinking water in Canada:
http://www.hc-sc.gc.ca/hecs-sesc/water/publications/tap_water_guidance/toc.htm

Information on water quality in Canada from Environment Canada with links to provincial information: http://www.ec.gc.ca/water/en/manage/qual/e_qual.htm

Information about microbial risks from the American Academy of Microbiology
http://www.asmusa.org/acasrc/Colloquia/GIDiseasesReport.pdf

Information about chlorinated by-products (THMs):
http://www.hc-sc.gc.ca/hecs-sesc/water/chlorinated_disinfection.htm and
http://www.hc-sc.gc.ca/pphb-dgspsp/publicat/cdic-mcc/19-3/c_e.html

Information about bladder cancer from the National Cancer Institute of Canada:
http://www.cancer.ca/vgn/images/portal/cit_86751114/14/33/195986411niw_stats2004_en.pdf**[ENGLISH]**
http://www.cancer.ca/vgn/images/portal/cit_86755361/27/54/195991114CCS_stats2004_fr.pdf **[FRENCH]**

We expect to have a report summarizing the results of our survey available in December. It can be accessed via the following web site:
http://spartan.ac.brocku.ca/~ddupont/tap_water_quality.html

Again, please be assured that all information from your personal responses will be treated with strict confidentiality and will not be made available to anyone other than the researchers. Participant's responses will not be individually identified. If you have any questions about the interview or the study in general and its results, please contact Tech Support at panel@i-say.com or 1-888-618-2056. Please click on the following link to access our Privacy Policy http://www.i-say.ca/legal/privacy.cfm.