Unclassified **ENV/EPOC/WPNEP(2010)10/FINAL**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **03-Dec-2010**

English - Or. English

ENVIRONMENT DIRECTORATE
ENVIRONMENT POLICY COMMITTEE

ENV/EPOC/WPNEP(2010)10/FINAL
Unclassified

Working Party on National Environmental Policies

Meta-analysis of stated preference VSL studies: Further model sensitivity and benefit transfer issues

This paper was prepared by Henrik Lindhjem, Vista Analyse, Norway, and Ståle Navrud, Department of Economics and Resource Management, Norwegian University of Life Sciences, with input from Vincent Biausque, Ecole Nationale de la Statistique et de l'Administration Economique, Paris, and Nils Axel Braathen of the OECD Secretariat.

The paper gives a technical description of meta-analyses that have been made of estimates of the value of a statistical life in stated preferences surveys, and is an input a user's guide for policy makers on the use of VSL values in policy assessments that is being prepared.

For more information, please contact Nils Axel Braathen; tel: + 33 (0) 1 45 24 76 97;
email: Nils-Axel.Braathen@oecd.org

JT03293914

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

FOREWORD

This paper was prepared by Henrik Lindhjem, Vista Analyse, Norway, and Ståle Navrud, Department of Economics and Resource Management, Norwegian University of Life Sciences, with input from Vincent Biausque, Ecole Nationale de la Statistique et de l'Administration Economique, Paris, and Nils Axel Braathen of the OECD Secretariat.

The paper gives a technical description of meta-analyses that have been made of estimates of the value of a statistical life in stated preferences surveys, and is an input a user's guide for policy makers on the use of VSL values in policy assessments that is being prepared.

The paper is issued under the responsibility of the Secretary-General of the OECD.

TABLE OF CONTENTS

META-ANALYSIS OF STATED PREFERENCE VSL STUDIES: FURTHER MODEL SENSITIVITY AND BENEFIT TRANSFER ISSUES

1. Introduction

1.1 Background and objectives

1. This paper is a deliverable under the project "PIMAVE – Policy implications of meta-analysis of value of statistical life (VSL) estimates" funded by the European Commission and coordinated through the Environment Directorate of the OECD. It builds on and continues the work documented in Braathen *et al.* (2009). Since that, study the meta data have been updated with more recent studies (as of March 2010) and more information has been added from studies and collected from authors to arrive at a more complete and comprehensive database.

2. Based on this updated dataset, the main objectives of this paper are to:

- Consider screening procedures for the VSL estimates based on quality of studies and other factors that will make the dataset more amenable and appropriate for statistical analysis and for use in benefit transfer (BT) applications for policy use.
- Conduct some further meta-analysis (MA) regression models on subsets of the data generated by the above screening criteria to investigate robustness and sensitivity of results.
- Consider some MA models for BT and policy use and the reliability of those models.

3. One screening criterion under the first bullet that has been paid particular attention to is to contact the authors themselves, to ask their view on whether they think estimates should be included in the analysis or not. A range of other objective and subjective criteria for managing the large number of available VSL estimates in our meta-dataset are also discussed. Many of these criteria screen based on indicators of quality of the underlying study or the estimate in question.

4. The full dataset for this phase of the project still consists of all studies globally that could be found in English that use stated preference (SP) methods (*i.e.* contingent valuation, contingent ranking or choice experiments) to estimate *adult* willingness to pay (WTP) for mortality risk reductions from environmental, health and transport policies. Hence, revealed preference studies (e.g. hedonic wage risk studies) are not included as they are considered too methodologically and conceptually different from SP studies[1]. Further, the small, but growing number of SP studies asking parents to value mortality risk changes for their children specifically, have not been included.

5. The paper is primarily written by Henrik Lindhjem[2], with comments and inputs from Ståle Navrud[3], Nils Axel Braathen[4] and Vincent Biausque[5]. This paper builds on and expands on some of the work documented in Biausque (2010). It has also benefitted from comments from Alan Krupnick of Resources for the Future (as a reviewer of Braathen *et al.*, 2009) and other participants at the conference

1 As also recommended by US EPA (2006).
2 Norwegian Institute of Nature Research (NINA) and Vista.
3 Department of Economics and Resource Management, Norwegian University of Life Sciences
4 OECD, Environment Directorate.
5 Ecole Nationale de la Statistique et de l'Administration Economique, Paris.

"Valuation of Environment-Related Health Impacts", held 17-18 September 2009 in Prague,[6] and from Stephen White of the European Commission.

6. The paper is closely linked with the final output of the PIMAVE project, the policy report, where Ståle Navrud is the primary author. The aim of these two papers and Biausque (2010) is to demonstrate how the meta-dataset can be analysed and how results can be used for research and for BT and policy purposes. However, the papers do not intend to be exhaustive in this regard, as that would be impossible given the share size and complexity of the database and the limited time and resources available. As the meta-dataset is made available online, the intention would be that interested researchers and policy-makers can use the data for further analysis and adapt parts of the dataset for their particular purposes (see www.oecd.org/env/policies/VSL for the full database). Being open source, the information in the database can potentially receive a greater quality control for errors and misinterpretations and new studies can be added to the database as they become available. This paper, Biausque (2010) and the policy paper aim to point the way for further use of the database by researchers and policy-makers. Ultimately, greater transparency and consciousness about people's WTP for risk reductions should lead to more efficient allocation of public funds to those mortality risk reductions that matter most.

1.2 Outline of paper

7. The next chapter briefly discusses characteristics of the database and issues related to screening out VSL estimates based on assessment of quality of the underlying studies or other factors. As the literature on MA is fairly silent about this and typically do not make explicit why and how studies have been excluded from the dataset, it is worth considering this here, as from the start studies were excluded from our raw database.

8. Chapter 3 explains and discusses further MA regressions done on the dataset compared to Braathen *et al.* (2009). This work includes extensive preliminary and explorative analysis to narrow down the dataset, test screening criteria and to form basis for choosing the potentially most important explanatory variables from the wide range of theoretically and empirically relevant variables coded in the database. Four main ways of screening the data are chosen, for which several meta-regression models with combinations of explanatory variables are run. The results demonstrate sensitivities and prepare the ground for choosing MA functions that may be used for BT. The chapter concludes with some further considerations of sensitivity that have not been analysed in detail here. The analysis in Chapter 3 expands and refines the analysis in Braathen *et al.* (2009).

9. Chapter 4 discusses some issues to consider when choosing and applying the estimated meta-functions to generate VSL estimates for BT purposes. Further, issues of reliability and sensitivity of meta-analytic BT are briefly discussed in relation to simple BT techniques more commonly used in practice. Chapter 5 sums up and concludes.

10. The paper can be read independently, but references are made to Braathen *et al.* (2009) and Biausque (2010) for more detailed descriptions of the database and further discussion of econometric modelling.

2. Meta-data and screening considerations

11. This chapter first gives a brief update on the dataset, before discussing screening procedures for the VSL estimates based on quality of studies and other factors that will make the dataset more amenable and appropriate for statistical analysis and for use in benefit transfer (BT) applications for policy use.

---

6 www.oecd.org/document/22/0,3343,en_21571361_36146795_37920982_1_1_1_1,00.html.

2.1 Meta-data

12. The database has been collected and updated over a period of 3 years, with final entries made during early 2010. The database contains the most extensive set of information for any meta-database in environmental economics to the knowledge of the authors. The total dataset contains more than 1000 observations of mean[7] VSL estimates from SP studies globally.

13. Further details and description of the database are given in Chapters 3 and 4 of Braathen *et al.* (2009) and in the annexes to that report. The database itself is posted on the OECD website[8].

14. The summary statistics for the full sample and trimmed samples of mean VSL are given in Table 1. The mean VSL of the database is ca 6 million in US 2005 dollars. Weighing this estimate so that each *study* gets equal weight (rather than each *estimate*, as there are more estimates from each study), the mean increases to ca 7 million. Screening out the highest and lowest 2.5% of the sample, scale these two values down to almost the same value of USD 4.5 million.

15. There is wide variation in the estimates, from a minimum of USD 4450 to a maximum of USD 197 million. The distribution is heavily right-skewed, as can seen from the medians being much lower than the means. The variation is not just due to the fact that many countries (poor and rich) are included. For USA, the minimum and maximum VSL values reported are USD 37 222 and USD 138 000 000, respectively.

Table 1. Summary of the estimates of value of statistical life (VSL).
US dollars 2005.

	Full sample	**Trimmed sample****
Mean VSL (standard deviation)	6 256 797 (453 874)	4 577 084 (181 355)
Weighted mean VSL* (standard deviation)	7 360 216 (824 671)	4 592 717 (301 182)
Median	2 554 708	2 554 708
Minimum value	4 450	57 626
Maximum value	197 000 000	35 700 000
Number of observations	937	891

*Weighted by the inverse of the number of observations from each SP survey
**Highest and lowest 2.5% of the values taken out of the sample

16. The summary statistics for the three categories of risk that are used for classification are also reported; environment, health and traffic. See Table 2. The mean (unweighted and untrimmed) is 8.7, 4.7 and 6.9 million US dollars. These differences will be investigated further in the regressions in the next chapter.

7 We consider *mean* VSL estimates from survey samples reported in studies, *i.e.* not medians (which fewer studies report) or individually stated VSL estimates from full SP datasets.

8 www.oecd.org/env/policies/VSL.

Table 2. VSL by risk category
Unweighted and untrimmed, US dollars 2005.

	Environment	Health	Traffic
Mean (st.dev)	8 690 887 (1 459 615)	4 698 226 (394 035)	6 9396 34 (812 836)
Median	3 027 388	1 462 561	3 041 077
Minimum value	24 421	4450	21 086
Maximum value	197 000 000	75 400 000	112 000 000
Number of obs.	217	455	265

2.2 MA trade-offs and sensitivity of choices

17. As is well-known to meta-analysts, there is a trade-off between the number of possible and interesting explanatory variables that can be included to explain variation in VSL, and the information actually available about these variables in the studies collected. Choosing fewer variables will give a dataset with fewer holes, as it is more likely that the information is found across more studies. This choice of balancing number of studies and variables to arrive at a final dataset for analysis and BT is to some extent more art than science. There is little guidance in the MA literature, although some newer studies have begun to explore such questions and the sensitivity of results to this choice (Lindhjem and Navrud 2008; Nelson and Kennedy 2009; Rosenberger and Johnston 2009; Johnston and Rosenberger 2010).

18. A related issue is that even if the variable we are trying to understand and explain, VSL, is consistently defined across studies, the VSL estimates may vary due to many heterogeneous factors, such as differences in methodologies, country-variation, risk types valued, etc. And there is a limit to how much variation (or heterogeneity) in a meta-dataset that can be meaningfully modelled in meta-regressions with a range of explanatory variables.[9] There is no agreement in the literature on where this limit is. USEPA (2006) represents perhaps the most conservative view, while there are several examples of published studies where the value analysed is very ambiguously defined and heterogeneity in possible explanatory factors great. Such examples in the environmental economics literature include coral reefs, wetlands (Brander, Florax *et al.*, 2006; Brander, van Beukering *et al.*, 2007), biodiversity (Jacobsen and Hanley 2009; Richardson and Loomis 2009) among others.

19. Many MA studies are not explicit about their protocol for collecting, coding, including and analysing studies in final meta-analyses. There is also little in the way of sensitivity analyses of results to such protocols and choices during the process of collecting and coding data. Hence, our approach here has been to include as many studies as could be found and code a rich set of variables from each study (inevitably creating some holes in the dataset). Further, data from studies have been supplemented with information from official statistics, from authors and to some extent our own calculations based on information in studies. This makes the database very detailed and rich, but also makes it necessary to decide on some protocols for screening data when conducting analysis. This will depend on the objective of the analysis. Further, the need for sensitivity analysis following such choices is emphasised here.

20. Some initial sensitivity considerations were made in Braathen *et al.* (2009), as different models incorporating different variables lead to different observations being dropped from the regressions, depending on whether or not information regarding a given variable is available for a given estimate. In the following, several potential criteria for excluding certain studies or observations from analysis based on subjective and more objective factors, are discussed. In MA generally, it is a controversial issue to screen studies based on quality, as there is not always agreement about what constitutes quality in general and

9 In addition, there is the usual consideration in econometrics not to include too many explanatory variables compared to the number of observations, which would lead to an over-specified model.

required quality for a certain purpose, specifically. Still, there are good reasons to explore ways to do this, since good studies provide better information that is closer to the "truth" in some sense. It is, however, important to be transparent about how such choices are made and their potential implications.

2.3 Screening studies based on quality and other criteria

21. Some screening criteria and the ones used as basis for the meta-regressions are discussed in the next chapter.

Screening based on quality criteria

22. Many characteristics of a study may indicate lower quality in SP research in particular and survey research more generally. If a study has a high share of respondents seemingly protesting to the valuation question, you may say that there are aspects of the scenario description or other weaknesses of the questionnaire that has resulted in this. However, it is difficult to judge what would be an acceptable level and whether protesting varies between cultures. This is therefore a type of quality criterion which perhaps is too ambiguous to use, in addition to the fact that not all studies report this information. No estimate has therefore been excluded based on this criterion.

23. Another similar criterion is to exclude studies based on whether an external "scope test" has been performed and passed. This means that two independent samples have received different levels of risk change to check if respondents' WTP vary positively with the risk change (Hammitt and Graham 1999). This quality criterion originates from the NOAA panel's recommendations for SP research (Arrow, Solow *et al.*, 1993). It was applied by Krupnick (2007) to screen studies in a literature review of the relationship between VSL and age. However, the importance of significant scope effects for quality of studies have been downplayed in recent research on theoretical and empirical grounds (Amiran and Hagen, 2010).

24. 24. Another criterion, of a more general sort, is whether a study is published or not. In the MA literature, it is generally not recommended to exclude studies on the basis of this. Published studies may not always provide the most suitable information needed for the purpose of the MA (e.g. since the aim often is to provide methodological twists, not to report value estimates per se) and working papers and reports may often be better than papers published in bad quality journals. Further, published studies may be systematically different in some way than unpublished studies. To reduce this potential publication bias, it is better to "err on the side of inclusion" (Stanley and Jarrel, 2005).

25. Two potentially more objective criteria are related to the quality of survey research more generally. If the response rate of the SP survey is too low, it indicates that too few of the less interested respondents have been included potentially giving higher value estimates. This can be supported in principle, but in practice, few studies are thorough in providing their net response rates and the ways response rates are calculated and reported vary (*e.g.* for web-based surveys from pre-recruited panels).

26. More recent studies may be considered to be of higher quality, due to gradual methodological innovation and refinement over time. However, instead of choosing an arbitrary year (*i.e.* years after the NOAA panel recommendations) and exclude older studies, survey year is typically included in regressions to control for such effects.[10] Some of the regressions do this.

27. Another relatively objective criterion is sample size. Larger samples give statistically more precise estimates and are generally associated with higher budget and (one would hope) quality of studies.

10 There may of course be other time trends captured in this variable, *e.g.* effects of wealth increases not reflected in GDP numbers or changes in relative importance/scarcity of the good valued over time.

This criterion is chosen in our screening. Further, (admittedly somewhat arbitrary) a sample size of 100 for a subsample behind any particular VSL estimate as a minimum, is applied. Further, a requirement of 200 for a full sample of a survey is applied. Sample size was also used by Krupnick (2007).

28. There are also other characteristics of SP research that could be considered, but are difficult or controversial to use in practice, for example WTP question format (dichotomous choice recommended by the NOAA panel vs. other formats), whether a study used debriefing protocols, and found results consistent with economic theory in regression analysis (SEPA 2006; Krupnick 2007).

29. An additional option that has been applied in this project is to ask authors themselves to assess whether an estimate should be included in the MA or not. Specifically, the question was:

> "It would be excellent if you could indicate if you think that a given VSL estimate ought to be included in our analysis. We would like you to distinguish between four 'options': "Only", "Yes", "Perhaps" and "No". Please use "Only" to indicate *the* preferred estimate from a given survey (if any), "Yes" to indicate that the estimate is one among several estimates that ought to be included, "Perhaps" to indicate that you are in doubt and "No" if you think that a given estimate definitively should not be included in the meta-analysis. "

30. This process yielded opinions from the authors regarding 627 out of a total of 1 095 observations, or slightly under 60%. It was decided to remove from the initial sample all those estimates for which the authors had answered "No". There were too few of the authors who wanted to recommend one specific VSL estimate (*i.e.* reply "only" to our question), for us to be able to use that information. Hence, their opinion of exclusion is used instead.

Heterogeneity considerations and screening based on other criteria

31. To maintain a higher degree of homogeneity, only studies that had calculated mean VSL directly, rather than *e.g.* based on a calculated median (or indeed studies reporting only median VSL), are included. Further, studies that asked people's willingness to accept (WTA) compensation for risk increases are excluded, as this is seen as conceptually different from WTP for a risk reduction (or less commonly to avoid an increase). WTA is not bounded by income and there are also problems relating to protest from respondents.

32. Some studies were of particular occupational or other groups (*e.g.* scientists living near a nuclear power plant, commuters of a certain type or similar). It was attempted to include those studies which had more representative samples of a broader population. There is, however, some variation in which age groups are targeted with different surveys. It was not tried to separate between age groups in terms of representativeness of a broader population, but attempted instead to control for age for respondents in preliminary regressions. Some estimates that were split by subgroups, *e.g.* according to age, income or similar, where also included.

33. In preliminary analysis the effect of limiting the dataset to western or OECD vs. other countries or other types of geographical splits was considered and checked, with the idea that risk preferences may be different between countries. No significant such differences were found, beyond which can be controlled with differences in income.

34. Further, in our sensitivity analysis in Chapters 3 and 4, trimming the data for very small or large VSL estimates was considered. It was also found in preliminary analyses a close relationship between the risk change proposed to respondents and the income level. It was therefore decided to include both these variables in most of the regressions, even if some observations are lost, since this is not reported in all

studies (se Chapter 3). In the econometric analysis, there is also consideration of what to do with the issue that some studies contribute (many) more estimates than other studies into the meta-data.

35. In the next chapter, the screening criteria are introduced step by step for different meta-regressions to investigate robustness of results. As it is an important issue to better control the heterogeneity of the data – something which not always possible to do well enough through a broad range of explanatory variables in regressions – it was expanded on an idea from Braathen *et al.* (2009). In one of the model runs the methodological heterogeneity is limited by analysing only estimates from studies that use variations of the same "good practice" questionnaire. The idea is that if some of the methodological variation can be reduced, more of the variation in VSL estimates can be explained by risk characteristics and other policy-relevant factors.

36. For each of the four subsets of meta-data generated by the screening criteria, several meta regressions were run, including explanatory variables that have been found to be important from theory and from extensive preliminary statistical analysis on the dataset. The (almost) full range of variables is defined in Braathen *et al.* (2009), a subset of which was used in this paper.

3. Further meta-regressions and sensitivity analysis

37. This chapter first discusses the extensive preliminary analysis done to narrow down the number of explanatory variables to be included in the meta-regressions. These variables are listed and defined and it is discussed how they can be expected to relate to VSL, either from theory or from previous empirical studies. Then the econometric approach chosen is explained.

38. Second, the mentioned data-screening criteria are introduced step by step and sensitivity of results to these criteria and to the inclusion of different types of explanatory variables in the regression models is demonstrated. The reason for doing this is to better understand how risk context, methodological, socio-economic and other variables determine the observed VSL estimates, and to search for models that can potentially be used for benefit transfer (BT) and policy purposes. If there, for example, are few or no policy-relevant variables that show robust relationships with VSL, there is no basis in the data to argue that mortality risks should be valued differently based on these variables.

3.1 Preliminary analysis, variable definitions and expectations

39. The explanatory variables are of three main types: (1) characteristics of the risk change and context in which it is valued (type of risk, controllability of risk, size of risk, etc.); (2) characteristics of the methods applied in the different studies (ways the WTP question is asked, survey mode, econometric estimation procedures, etc.), and (3) characteristics of the population asked to value the risk change (socio-economics, such as income and age). In addition, meta-analysts sometimes include variables that cover quality dimensions of the studies or other types of variables. For many variables there are *a priori* expectations of relationship with VSL from theory or empirical studies, while others are typically more explorative.

40. Braathen *et al.* (2009) in their Table 2 lists 50 variables in the four categories. The variables were chosen on the basis of theory and empirical expectations. Many of them were preliminarily tested in that study. Further explorative analysis was carried out for this project, based on that work. In particular, more combinations and recoding of variables were tested on different subsamples. All those results are not reported here, but a few important points are summarised as a motivation for the short-list of variables presented in Table 3 below and used for the meta-regressions to follow.

Table 3. Explanatory meta analysis variables and expected relationships with VSL.

Variable	Description	Sign
Dependent variable		
lnvsl	Natural logarithm of mean VSL estimates in PPP-adjusted USD 2005 (mean, annual WTP divided by annual risk change, PPP adjusted based on AIC*)	
Risk valuation context variables:		
lnrchrisk	Continuous: Log of change in mortality risk on an annual basis per 1000 (normalised per year from study info).	0/-
public	Binary: 1 if public good; 0 if private (risk affects only the individual asked or her household).	+/-
environ	Binary: 1 if environment-related risk change; 0 if health-related	?
traffic	Binary: 1 if traffic-related risk change; 0 if health-related	+/?
latent	Binary: 1 if risk change occurs after a certain time; 0 if the risk change is immediate	-
cancer	Binary: 1 if reference to cancer risk in survey; 0 if not	+
household	Binary: 1 if WTP is stated on behalf the household; 0 if WTP is only for the individual asked	+
Methodological variables:		
noexplan	Binary: 1 if no visual tool or specific explanation of the risk change was used in survey; 0 if otherwise	+/?
turnbull	Binary: 1 if WTP was estimated using Turnbull, non-parametric method; 0 parametric method	-
Income and survey year:		
lngdp	Continuous: Log of mean GDP, USD 2005, AIC-adjusted	+
lnyear	Continuous: Log of year of data collection, adjusted to start at ln2.	+/-

*AIC – Actual Individual Consumption. A measure of the individual goods and services that households actually consume as opposed to what they actually purchase.

41. Of the risk context variables, no consistent relationships in the data between VSL and the duration of the risk change was found, whether the risk was acute or chronic, whether degree of suffering was mentioned in the survey, degree of individual control over the risk, or special ways the risk change was displayed to respondents in the survey (ladders, grids etc). Some significance was found related to whether or not the risk change was latent or immediate, whether it affected private individuals or their household members as opposed to the public at large, whether the risk change was related to cancer and the size of the risk change itself. These four variables were therefore included in the main meta-regressions reported here. There were some indications that baseline risks may affect VSL in some regressions in Biausque (2010) (Appendix). Few studies report the baseline risk, and theoretically it is not expected to influence WTP and VSL very much, at least for small levels of risk (see *e.g.* Eeckhoudt and Hammitt, 2001). This variable was therefore not included in further regressions.

42. Cancer risk is associated with some dread and can be expected to influence WTP and VSL for such risks positively (van Houtven, Sullivan *et al.*, 2008). Latency of risks should normally give lower WTP and VSL, as the risk change will occur some time in the future. Respondents are known to discount the future at a positive rate and should value an immediate risk change higher. The relationship between the risk change and WTP should, from standard economic theory, be positive and approximately proportional. This implies that VSL should be unaffected by the change in risk, at least for small changes in risk and for low baseline risks (Hammitt and Graham 1999). However, what is typically found in practical SP studies is that peoples' WTP is generally very insensitive to the size of the risk change; hence, lower risk changes tends to result in higher VSL estimates. This result has not to our knowledge been documented across many studies in a meta-analysis before.

43. The effect of whether the risk change affects the individual or her household members versus the public at large is trickier to determine. One effect, that of altruism, would pull in the direction of higher WTP and VSL for public risk changes. On the other hand, private risk changes are typically something the family or individual controls through buying a helmet or a product that reduces risk. In other words, the risk change is more concrete and direct when it is private compared to a public risk programme that may be associated with diffuse, abstract, uncertain and delayed effects. This factor would pull in the other direction, so the net effect is an empirical question. The "household" variable is also included, to indicate if the respondent is asked to state WTP for himself or behalf of his household. This inclusion reflects an emerging literature emphasising the importance of family resource allocation models in determining expenditures for household and other goods (see *e.g.* Lindhjem and Navrud, 2009).

44. As part of the risk characteristics variables, the type or category of risk was also included; *i.e.* environment, health[11] or traffic. There is some evidence in the literature that characteristics of typical risks under each of these categories may give different WTP and corresponding VSL estimates. However, the categories themselves may be too general to give clear indications in the data. In preliminary analysis some different results were found, and it was decided to include them in the main meta-regressions, as they are highly policy-relevant variables.

45. Of the methodological variables, a number of variables typically included in MA studies were tested in preliminary analysis, such as survey mode, type of WTP elicitation method (dichotomous choice, open-ended, etc.), type of payment vehicle, etc. No clear relationships with VSL were found for these variables. However, some patterns related to the way the risk change was displayed to respondents was found. Especially if there was no clear explanation orally or through visual tools of the risk change, VSL tended to be higher. In other words, respondents seem to overrate risks that are not carefully explained and displayed. Hence, this variable is included in the main regressions. The variable "turnbull" is also included, which indicates if the authors used a non-parametric approach,[12] typically giving a conservative or lower-bound estimate of WTP and therefore VSL.

46. Of socio-economic and other variables, it was decided to retain only GDP per capita and the year of the survey (for a subset of the regressions). Most studies report mean income from the total sample, but not for subsamples from which many of our estimates typically are derived. Not to lose these observations and those where no sample income has been reported at all, it was therefore decided to use GDP per capita instead as a proxy for individual wealth. The correlation between log of GDP and log of reported income was found to be very high (more than 0.9), so this is a good proxy. The relationship with VSL can be expected to be positive.

47. The relationship between year and VSL is empirically undetermined. New studies may use more stringent methodologies, yielding more conservative estimates, an argument sometimes found in the MA literature. Increased wealth or scarcity of the good valued, on the other hand, may in increase values over time, as mentioned in Chapter 2.

---

11 The distinction between the environment and health categories is not always obvious, in part because some health risks are caused by environmental problem - *e.g.* air or water pollution. In the classifications made here, the focus has been on whether or not an *explicit* reference to an environmental problem was made in the valuation question posed to the sample. If that was not the case, the survey was classified as being "health-related". This is, for example, the case with some well-known surveys using a questionnaire developed by Krupnick, Alberini and Cropper *et al.*, which in several cases refer to environmental problems in the titles of the papers presenting the surveys (see Chapter 3.5 below).

12 Except in the cases of simple mean from an open-ended WTP question.

48. Some investigations were conducted of the relationship between different characterisations of the age of the samples from which VSL estimates were drawn, but no clear relationships were found in the data. For a subset of the data, Biausque (2010) found indications of an inverted U-shaped relationship between VSL and mean age of the sample. The relationship is known to be both theoretically and empirically undermined, and different studies get different results (Krupnick, 2007). Hence, this variable is left out from further analysis.

49. Braathen *et al.*, 2009) checked for differences in VSL between countries and groups of countries (such as OECD vs. non-OECD) other than due to income, but found no clear patterns. Further analysis for this paper did not reveal anything new. Hence, it seems that the GDP variable is the best variable to differentiate between countries. Some negative correlation between the degree of risk reduction and level of GDP per capita was found (see Biausque ,2010, page 17). This means that studies in lower income countries have used larger risk reductions in the surveys, perhaps to reflect more realistic changes given the relatively higher baseline risks.

50. Finally, it was decided not to include additional variables from Braathen *et al.* (2009) related to study quality and other factors, for example related to whether studies are published or not (see discussion in Chapter 2).

3.2 Meta-regression approach

51. A number of meta-regression models were considered and tested. The following model, based on fairly standard practice in the MA literature, was used:

$$(1)\ \mathrm{lnvsl}_{si} = \beta_0 + \beta_1 \mathrm{lngdp}_{si} + \textstyle\sum_k \beta_k X_{si}(k) + \varepsilon_{si}$$

where lngdp_{si} is the natural logarithm of per capita GDP for observation *i* in survey group *s* and X_{si} is a vector of binary dummy variables (except for one model where log of survey year is included). This model is estimated using ordinary least squares (OLS). However, since the number of observations varies widely across survey groups *s,* the ordinary least squares are weighted by the reciprocal of the number of observations in each group, so as to weight each survey group equally (as opposed to giving equal weight to each individual VSL estimate). There seems to be no general agreement on what is the best strategy in the case where there are many estimates from the same surveys. Mrozek and Taylor (2002), a much-cited MA of VSL estimates from labour market studies, apply this weighting scheme as do other MA studies in the environmental economics field (see *e.g.* van Houtven *et al.*, 2007).

52. Moreover, the "cluster" option is used for estimating robust standard errors in order to account for the correlation between different observations within the same survey group. This is also a common strategy in the MA literature (Nelson and Kennedy, 2009). A random effects model was used in Braathen *et al.* (2009), but a simpler approach for interpretation and use in BT, was chosen here. A technical reason for this choice was the concern that the random effects model involves stronger assumptions than a clustered OLS and may introduce bias into the estimations. Clustered OLS (with or without some kind of weighting) is still the most common approach in the MA literature.

53. A log-log model is also applied, since this provides the best fit to the data. As mentioned in Chapter 2, the VSL distribution is highly skewed with a long right tail. Using double-log has the additional advantage that the estimated coefficients have a natural interpretation as elasticities.

54. USEPA (2006) recommends weighting estimates with their precision, preferably with the standard deviation (or alternatively sample sizes). Information about standard deviation was only available for 254 observations, either from the studies or calculated based on information in the studies. Biausque

(2010) provides some preliminary regressions of this weighting strategy. We do not pursue that approach further in this paper.

55. In the following the meta-regression results are presented for four different samples, starting with the full, unscreened, dataset.

3.3 Full dataset – no screening

56. For sake of comparison, the section starts by reporting results for the full dataset where no screening criteria have been applied. Five regression models are run gradually increasing the number of explanatory variables, see results in Table 4. Note that the risk change variable is not included here, since many studies do not report this information. Compared to Table 1 (937 estimates), 15 estimates have been excluded that were based on willingness to accept (WTA) compensation for a risk increase, rather than WTP for a reduction (or to avoid an increase) (see also discussion in Chapter 2.3).

Table 4. Regression results, full sample.

	Model I	Model II	Model III	Model IV	Model V
lngdp	1.507***	1.479***	1.544***	1.498***	1.331***
	(0.204)	(0.203)	(0.194)	(0.196)	(0.212)
envir		-0.00717	0.278	0.124	0.159
		(0.431)	(0.444)	(0.353)	(0.330)
traffic		0.550**	0.590**	0.651**	0.459*
		(0.265)	(0.269)	(0.304)	(0.271)
public			-0.387	-0.399	-0.390
			(0.362)	(0.360)	(0.332)
household			-0.266	-0.215	-0.177
			(0.307)	(0.296)	(0.275)
cancerrisk				0.733**	0.908**
				(0.360)	(0.345)
latent				-0.548	-0.489
				(0.384)	(0.337)
noexplan					1.057***
					(0.292)
turnbull					-0.172
					(0.486)
Constant	-0.551	-0.495	-1.022	-0.634	0.790
	(2.053)	(2.025)	(1.948)	(1.974)	(2.135)
Observations	922	922	922	922	922
R-squared	0.421	0.445	0.460	0.488	0.544

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

57. Starting with Model I, it can be seen that including only log GDP per capita explains more than 40% of the variation in the VSL estimates (R-squared equals 42.1%). Despite the fact that it is a full and unscreened dataset, the R-squared compare favorably with many MA studies in the literature. The number of observations is the same for all five models, so they are internally comparable.

58. Increasing the number of variables from Model I to V increases the explained variation to 54% in model 5. The GDP per capita is highly significant for all five models, yielding an elasticity of GDP to VSL by between 1.3 and 1.5. This is high compared to other studies (typically based on individual surveys rather than meta-analysis). Traffic risks show significantly higher VSL across all five models compared to the "hidden" category of health-related risks (the coefficient on "traffic" is significant and positive). There is a significant cancer premium in Models IV and V where this variable is included. Model V also shows that surveys where respondents have not been carefully explained the risk change by the use of visual tools or proper oral explanation, the VSL tends to be higher.

59. Latent risks seem to be valued in the same way as immediate risks, and the turnbull, household and public are also not significant.

3.4 First level screening procedure

60. For the next three subsets of the data, the risk reduction reported in the survey is included as an explanatory variable. Some observations are lost as this information is not always reported, but something is gained as the model is more appropriate. The model used from here on is therefore:

$$\ln vsl_{si} = \beta_0 + \beta_1 \ln gdp_{si} + \beta_2 \ln chrisk_{si} + \sum_k \beta_k X_{si}(k) + \varepsilon_{si}$$

where lnchrisk_{si} is the natural logarithm of the risk change, and the other variables are as explained in Section 3.2 and defined in Table 3 above.

61. The screening criteria used are (see discussion in Chapter 2):

- If no value for the risk change has been reported, the study is excluded (243 observations dropped).
- Subsamples smaller than 100 observations and main survey samples less than 200 observations are left out (118 observations are dropped).
- Samples that are not representative of a broad population are left out (140 observations dropped).

62. Compared to the full dataset above, this dataset is likely to be of higher quality. Results of five regression models using the same explanatory variables as in Section 3.3, except for the risk change variable, are reported in Table 5.

Table 5. Regression results, first level screening

	Model I	Model II	Model III	Model IV	Model V
lngdp	1.042***	1.060***	1.069***	1.025***	0.936***
	(0.204)	(0.167)	(0.146)	(0.155)	(0.197)
lnchrisk	-0.434***	-0.524***	-0.547***	-0.567***	-0.543***
	(0.0882)	(0.0854)	(0.0793)	(0.0596)	(0.0613)
envir		-1.225***	-0.521*	-0.695*	-0.584*
		(0.362)	(0.294)	(0.348)	(0.337)
traffic		-0.436*	-0.214	-0.282	-0.333
		(0.254)	(0.288)	(0.256)	(0.206)
public			-1.053***	-0.956***	-0.913***
			(0.250)	(0.248)	(0.243)
household			0.0968	0.132	0.0921
			(0.266)	(0.221)	(0.223)
cancerrisk				0.351	0.443
				(0.311)	(0.304)
latent				-0.467	-0.369
				(0.379)	(0.367)
noexplan					0.667***
					(0.195)
turnbull					-0.237
					(0.571)
Constant	-0.0290	-0.640	-0.919	-0.644	0.319
	(2.086)	(1.766)	(1.558)	(1.703)	(2.262)
Observations	421	421	421	421	421
R-squared	0.720	0.784	0.823	0.835	0.853

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

63. The number of observations has now been reduced from 922 to 421. As before, the GDP per capita is highly significant, though the elasticity has dropped to around unity. Interestingly, the risk change is also highly significant, but negative. This means that respondents do not state WTP in accordance with theory, but in accordance with what is commonly observed in individual SP studies. WTP does not increase in proportion with the risk reduction, therefore yielding lower VSL. It is a robust result also found for other specific datasets, combinations of variables and meta-regression approaches in Braathen et al. (2009). This finding is a potential problem for policy and research, as using lower risk change levels in the surveys would ensure higher VSL estimates. The reasons for this phenomenon is complex and may have to do with people's difficulty in understanding small probabilities (Kahneman and Tversky 2000; Gilovich, Griffin *et al.*, 2002).

64. It can also be observed that the traffic variable is no longer significant, though the environment-related risk is (negative coefficient for the variable "envir"). The public variable is now significant and negative, meaning perhaps that the effect of altruism is outweighed by our previously suggested effect of directness of a private risk reduction compared to a more indirect effect of a public program. Finally, the "noexplan" variable is still positive and significant.

65. The R-squared is high for all models. It is interesting to observe that the combinations of the risk change and income explains almost 75% of the variation in VSL estimates. Adding the other explanatory variables increases the R-squared to 85%, which is high by any measure in the MA literature.

3.5 Estimates from studies using same "best-practice" questionnaire

66. This section proceeds by limiting the dataset to studies using variations of the same type of questionnaire. The idea is that if more of the methodological variation can be eliminated, the effects of other and more relevant variables for benefit transfer will come out more clearly. The questionnaire in questions is the one designed by Maureen Cropper, Anna Alberini and Alan Krupnick, see *e.g.* Alberini *et al.* (2004). It values health risk reductions, *e.g.* using grids for displaying risk changes and training respondents to understand risk changes, etc. In some ways, the surveys using this approach can be regarded as good practice compared to many other approaches.

67. In addition, the questionnaire has been used in several countries, ensuring variation in some of the policy relevant variables (such as income). The screening criteria used were the same as in section 3.4, except the limitation that observations should come from only the mentioned type of questionnaire survey was added.

68. The variables "household", "envir", "traffic", "cancerrisk" and "public" drop out as the values for these are the same for all observations. The survey year was added to the two regression models presented in Table 6.

Table 6. Regression results – "best practice" questionnaire sample

	Model I	Model II
lngdp	0.591***	0.493***
	(0.102)	(0.0902)
lnchrisk	-0.554**	-0.508**
	(0.157)	(0.164)
latent		-0.227***
		(0.0525)
turnbull		-0.591*
		(0.249)
lnyear	0.486**	0.485***
	(0.152)	(0.131)
Constant	3.029*	4.476**
	(1.505)	(1.335)
Observations	169	169
R-squared	0.762	0.814
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

69. The number of observations drops to 169. The two models explain around 80% of the variation in VSL estimates. It can be seen that both risk change and income are highly significant with the expected signs. The income elasticity has dropped to around 0.5 compared to 1 in the previous models. This is more in line with the rest of the literature, though levels up to and above 1 also have been observed in some studies. In model 2, which included "turnbull" and "latent" variables, both of which have the expected signs and are significant. In other words, VSL tends to be lower for risk reductions that are latent and for

estimation procedures using the lower bound, conservative non-parametric turnbull estimator. It can also be noted that newer studies tend to give higher VSL estimates, unclear for which reasons. One possibility may be that it picks up differences between countries not explained by the GDP.

70. These results are quite encouraging for benefit transfer, as coefficients have the expected signs and most of the variation in VSL can be explained by other factors than methodological differences between studies (which is a common problem in MA and meta-analytic benefit transfer, MA-BT).

3.6 Estimates recommended by authors

71. The final sample uses author recommendations to exclude certain estimates, as explained in Chapter 2.3. Details of these specific results are given in Biausque (2010), but briefly reiterated here for sake of comparison and completeness. Hence, in addition to the screening criteria in Section 3.4, screening was done based on author recommendation to exclude a particular estimate from further analysis the answers to (which drops an additional 55 observations compared to models in Table 5). It is worth noting that many of the estimates authors recommend for exclusion are screened out anyway based on our other criteria given in Section 3.4. Results are displayed in Table 7.

Table 7. Regression results for sample where author recommendations are used

	Model I	Model II	Model III	Model IV	Model V
lngdp	1.022***	1.029***	1.075***	1.006***	0.894***
	(0.206)	(0.166)	(0.150)	(0.157)	(0.200)
lnchrisk	-0.445***	-0.576***	-0.552***	-0.581***	-0.568***
	(0.0942)	(0.0951)	(0.0846)	(0.0625)	(0.0606)
envir		-1.395***	-0.634**	-0.882**	-0.790**
		(0.365)	(0.308)	(0.357)	(0.330)
traffic		-0.635**	-0.317	-0.369	-0.460**
		(0.283)	(0.296)	(0.266)	(0.211)
public			-0.949***	-0.800***	-0.713***
			(0.255)	(0.262)	(0.228)
household			-0.0405	-0.0279	-0.127
			(0.279)	(0.234)	(0.242)
cancerrisk				0.460	0.587*
				(0.330)	(0.319)
latent				-0.419	-0.304
				(0.381)	(0.367)
noexplan					0.715***
					(0.204)
turnbull					-0.277
					(0.571)
Constant	0.0433	-0.726	-0.978	-0.575	0.509
	(2.098)	(1.727)	(1.575)	(1.703)	(2.294)
Observations	366	366	366	366	366
R-squared	0.719	0.798	0.832	0.845	0.865
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

72. The same variables as in Section 3.4 were used again for five different regressions. As can be seen, the R-squared is again very high at between 72 and 87%. The "envir "variable is again negative, reflecting lower VSL values for such environment-related risk changes compared to health-related risks. "Traffic" is this time negative and significant in Models II and V. The variable "public" is highly significant and negative in all three models where it is included. Risk change and income are as always significant, the income elasticity again increasing to around 1. There is a cancer premium in model 5, but not in Model IV. Latency and turnbull are not significant.

3.7 Summary of results and further sensitivity issues

73. Overall, through our four different screening procedures, the following results were found:

- Gradually higher explanatory power of the models as stricter screening criteria are applied. Heterogeneity is gradually reduced in the data. The full sample also has fairly high R-squared
- Fairly robust results. Effects of income (GDP) and risk change are strong positive and negative, respectively. Income elasticity of VSL seems to be around 1 for most regressions (although lower for the "best practice" studies in Section 3.5).
- Strong indication that public risks are valued *lower* than private risks.
- Some indication that environment and traffic-related risks are valued lower than health risks.
- Some indication that latency of risk changes reduces VSL, as expected.
- No proper explanation of the risk change in the survey tends to increase WTP and corresponding VSL estimates.
- Indications that Turnbull-estimates are lower than other estimates, as expected since this non-parametric procedure gives a lower bound on WTP and VSL.

74. Further sensitivity analysis of some of the results from Sections 3.3-3.6 is conducted in the Annexes, related to the weighting strategy applied and to trimming high and low values from the samples. The results are generally found to be robust to these model changes. R-square drops somewhat for all models, but significance for the majority of variables is retained. The income elasticity seems to drop a bit, while the latency variable comes out more robustly as significantly negative than in the models in Chapter 3.

4. Considerations for benefit transfer and policy use

75. This chapter first discusses some considerations when choosing which estimated meta regression models to use for benefit transfer purposes. The second part of the chapter makes a brief assessment of how accurate a subset of the most suitable models are for predicting values. The last section demonstrates the use of the meta regression models and other, more simple BT techniques to estimate the VSL value for a hypothetical policy situation.

4.1 Introduction – Some theoretical and practical considerations

76. The previous chapter has gone through three ways to screen the dataset and run meta regression models on subsets of the data. The next question is which of these models are the most suitable for predicting values that could be used for policy purposes. By "predicting" it is meant running the regressions which estimate the coefficients determining the influence of each variable (as done in Chapter 3) and then inserting variable values corresponding to a policy situation of interest (e.g. a public risk program giving a risk change of 1/10 000 for a country with certain GDP) and adding up to a VSL estimate. In a particular BT situation, the values for methodological variables will have to be chosen based on some "best practice" consideration or set equal to the mean of the variable in the dataset or similar. This

procedure of using the estimated meta-function to predict or estimate a value for policy purposes, is sometimes called meta-analytical benefit transfer (MA-BT).

77. The more explanatory power (the higher R-squared) the meta models have, the more accurate they generally are in predicting values. The more significant variables influencing VSL, the higher generally is the R-squared and the explanatory power of the model. The next section therefore assesses this accuracy for a selection of the meta regression models from Chapter 3.

78. There is generally no one single, most appropriate or correct meta model for policy use. There is no such agreement in the literature or among practitioners. As has been shown, the results vary between model specifications and subsets of the data. An even if some results are fairly robust, coefficient values will not be identical. These differences in coefficients may have large impacts on the estimated VSL in a particular context. However, based on the analysis in Chapter 3, more confidence can be had in the models where estimates have been screened out than in the models run on the full, unscreened dataset.

79. The final section of this chapter illustrates the use of MA-BT compared to other BT techniques (such as choosing a value from a similar study, making simple adjustment based on GDP differences, taking a raw average from studies in the same country or the whole sample etc).

4.2 Accuracy of benefit transfer: Out-of-sample transfers

80. This section compares the accuracy of the different meta-regression models. A measure frequently used to assess the accuracy of benefit transfers is transfer error, (TE), defined as:

$$TE = \frac{|VSL_T - VSL_B|}{VSL_B} * 100\% \quad ,$$

where T = Transferred (predicted) value from study site(s), B = Estimated true value ("benchmark") at policy site. TE is a measure of how many% the estimated and transferred value "missed" the true value for a particular policy context, assuming that we could know what this "true" value is. When a VSL estimate is needed for assessing value of mortality risks of a certain policy proposal, the true VSL value is of course not known in practice. Studies testing transfer errors often use a "benchmark" value for this true value, often the VSL estimate from a good study, and then test how different BT techniques perform when predicting this value.

81. Validity has traditionally required "that the values, or the value functions generated from the study site, be statistically identical to those estimated at the policy site" (Navrud and Ready, 2007b), *i.e.* that TE is statistically indistinguishable from zero. More recently, BT validity assessment has shifted focus somewhat to the concept of reliability for policy use, which requires that TE is relatively small (but not necessarily zero). This shift comes from the realisation that BT can be considered valid even if the standard hypothesis of TE=0 is rejected — in fact the most appropriate null hypothesis is that TE is TE > 0 since environmental and other benefits from theory should be assumed to vary between contexts (Kristofersson and Navrud, 2005). However, there is no agreement on maximum TE levels for BT to be reliable for different policy applications, though 20 and 40% have been suggested (Kristofersson and Navrud, 2007).

82. To utilize the measure of TE to assess BT accuracy of our meta regression models, a data splitting technique, or BT simulation, is used. N different MA-BT functions are estimated using N-1 of the data for each run, since the VSL estimate predicted is taken out. Then the overall mean and median TE for all the N models taken together, sometimes termed the mean and median Absolute Percentage Error is calculated (Brander, Florax *et al*., 2006).

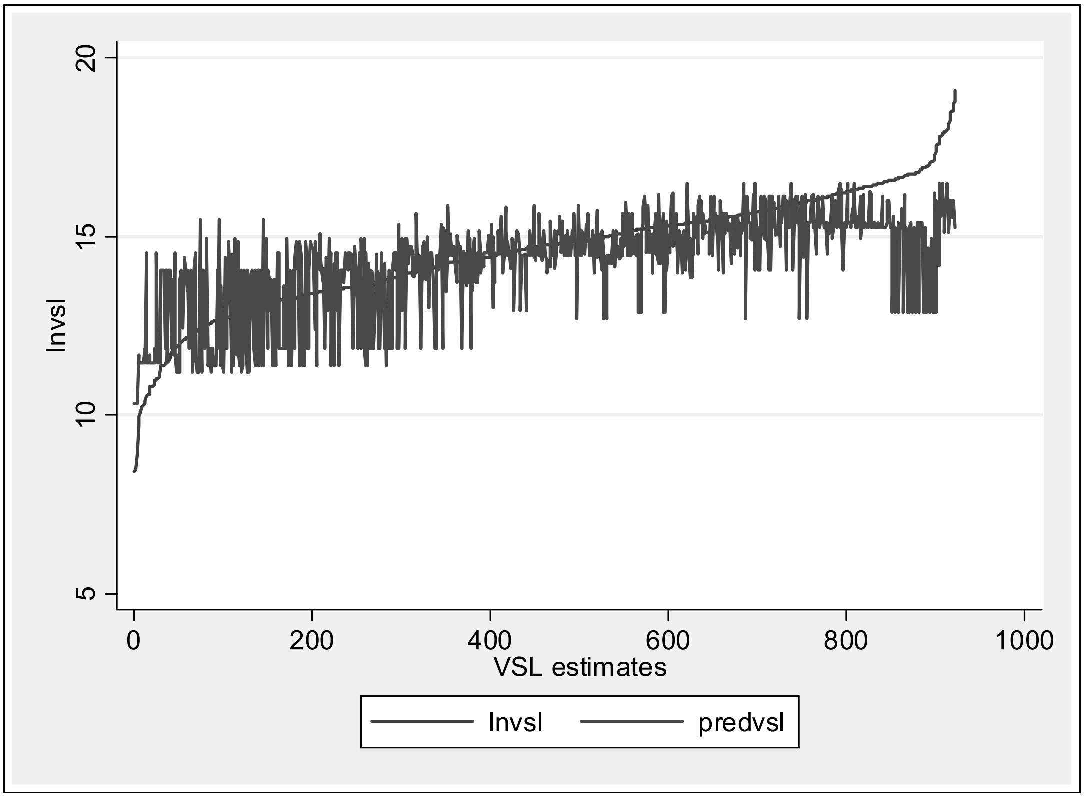
83. In the following this procedure is conducted for the simplest (Model I) and fullest models (Model V) from Sections 3.3, 3.4 (no screening and first level screening) and 3.6 (author-recommended estimates) and the fullest model from 3.5 ("best practice" questionnaire). These results are also displayed graphically, *i.e.* the predicted values (zig-zag line) and the VSL values in the dataset are compared in ascending order from the lowest to the highest VSL estimates in the dataset. The difference represents the TE for BT for each VSL value.

Full dataset - no screening, Model V

84. Figure 1 shows the results for Model V from the unscreened sample. Mean and median TE are 121 and 63%, respectively. That means that on average the values transferred miss the "true" benchmark value, the value to be predicted, by 121%. That result is quite high and as expected with a full model of the unscreened sample. As can be seen from the figure, the predictions particularly miss at the high and the low end of the values, *i.e.* the further out in the tales of the distribution. This is as expected.

85. The simpler models would have even higher TE. The mean TE for Model I for example, where GDP is the only variable included, is 236%. The mean TE was also estimated for a trimmed Model V where the 2.5% highest and lowest VSL values were taken out. This version of the model reduced the TE somewhat to 99%.

Figure 1 Plot of lnVSL and predicted or transferred lnVSL from Model V of the unscreened sample

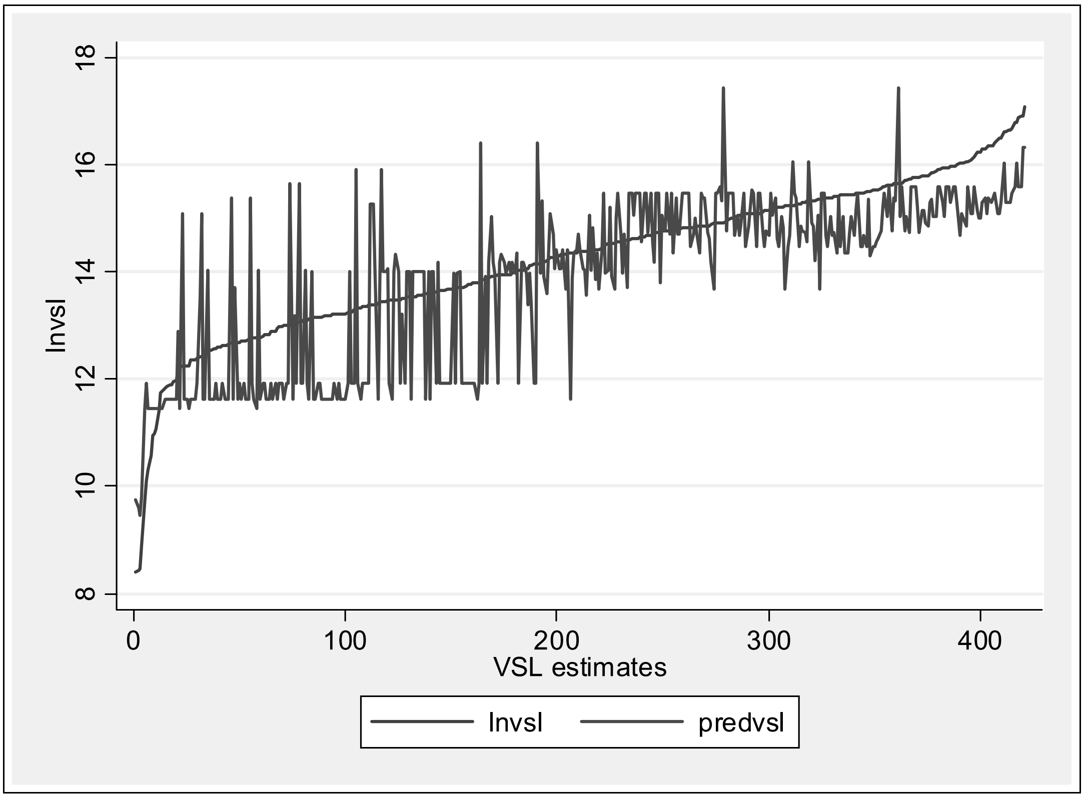


First level screening procedure – simple Model I

86. Two accuracy simulations were conducted for Models I and V of the sample that underwent first level screening (the first shown in Figure 2). The overall mean TE for Model I (only variables risk change and GDP included) was found to be 100% and the median 60%. The trimmed version reduced mean TE to

79.5%. Hence, screening reduces the TE somewhat compared to the full sample. However, the TE level is still quite high.

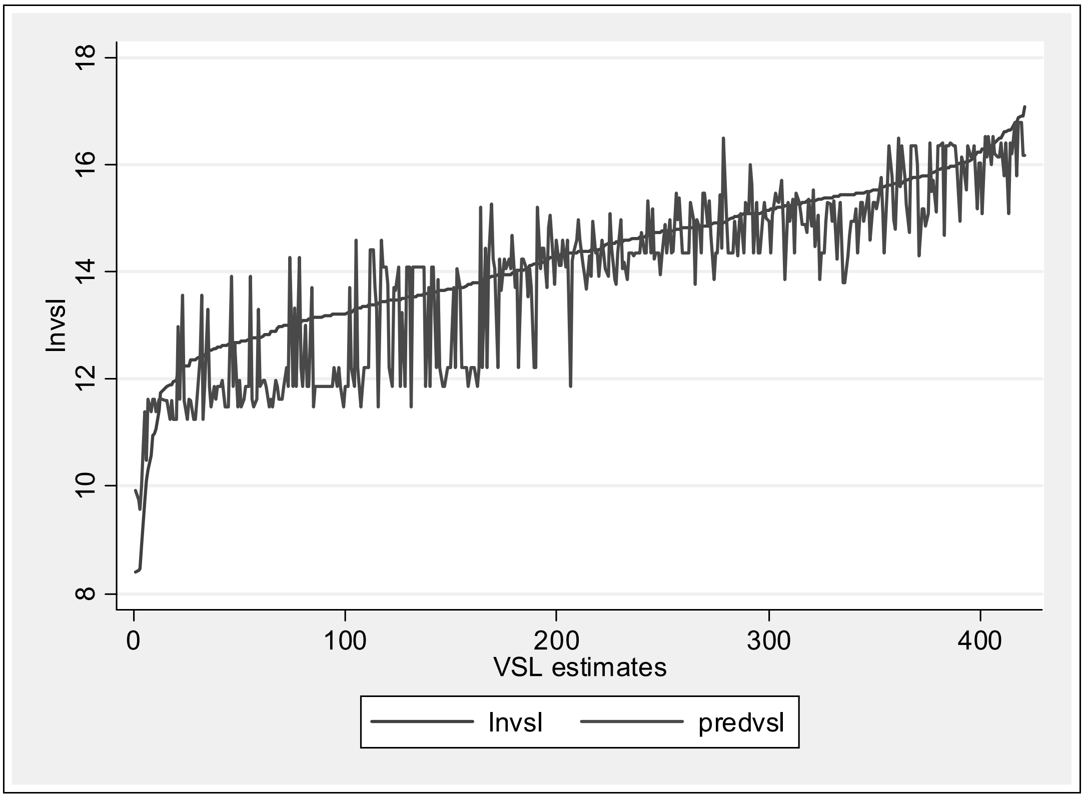
Figure 2 Plot of lnVSL and predicted or transferred lnVSL from Model I of the first level screened sample



First level screening procedure – Full Model V

87. Figure 3 shows the second BT accuracy simulation for Model V on the sample that was screened. Overall mean TE was found to be 60% and the median 49%. Accuracy increases as expected when the explanatory power increases and when more explanatory variables are included. A TE of 60% compares favourably with other such tests in the literature (Lindhjem and Navrud, 2008).

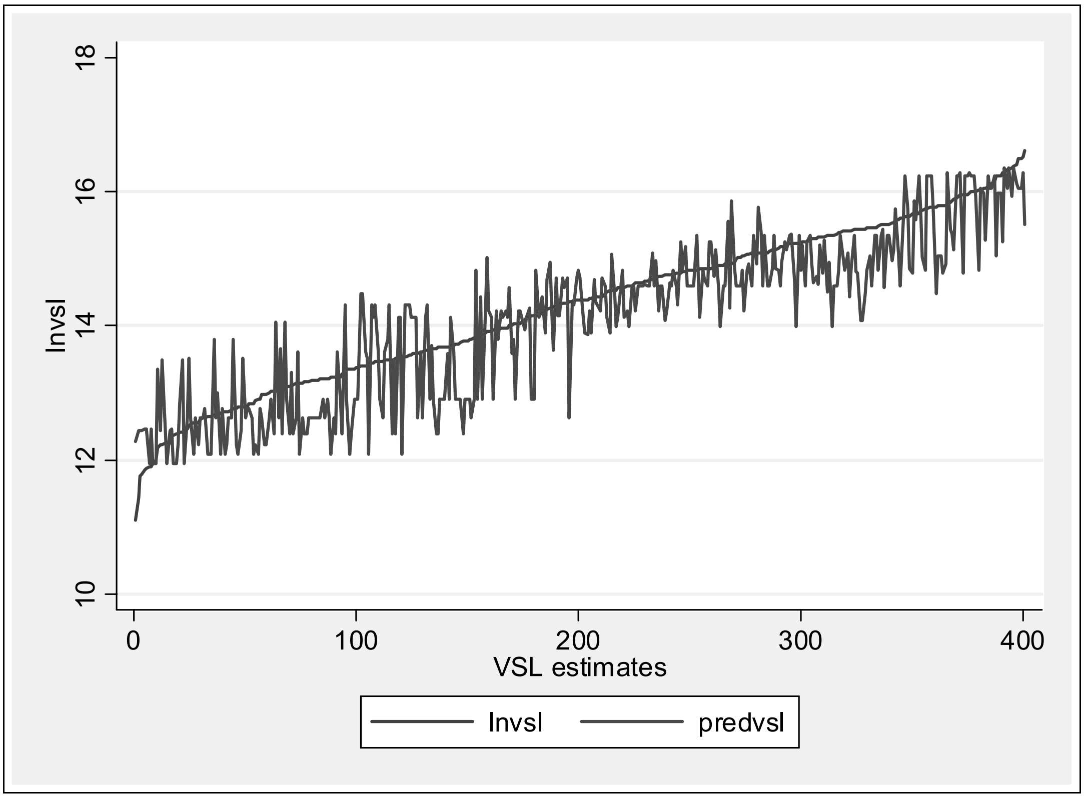
Figure 3 Plot of lnVSL and predicted or transferred lnVSL from Model V of the first level screened sample



First level screening procedure – Full Model V, trimmed

88. When trimming the same model displayed in Figure 3 (*i.e.* removing the highest and lowest 2.5% of the VSL observations), TE is reduced to 45% (see Figure 4). An unweighted version of this model was also tried: mean TE remains almost the same at 46%.

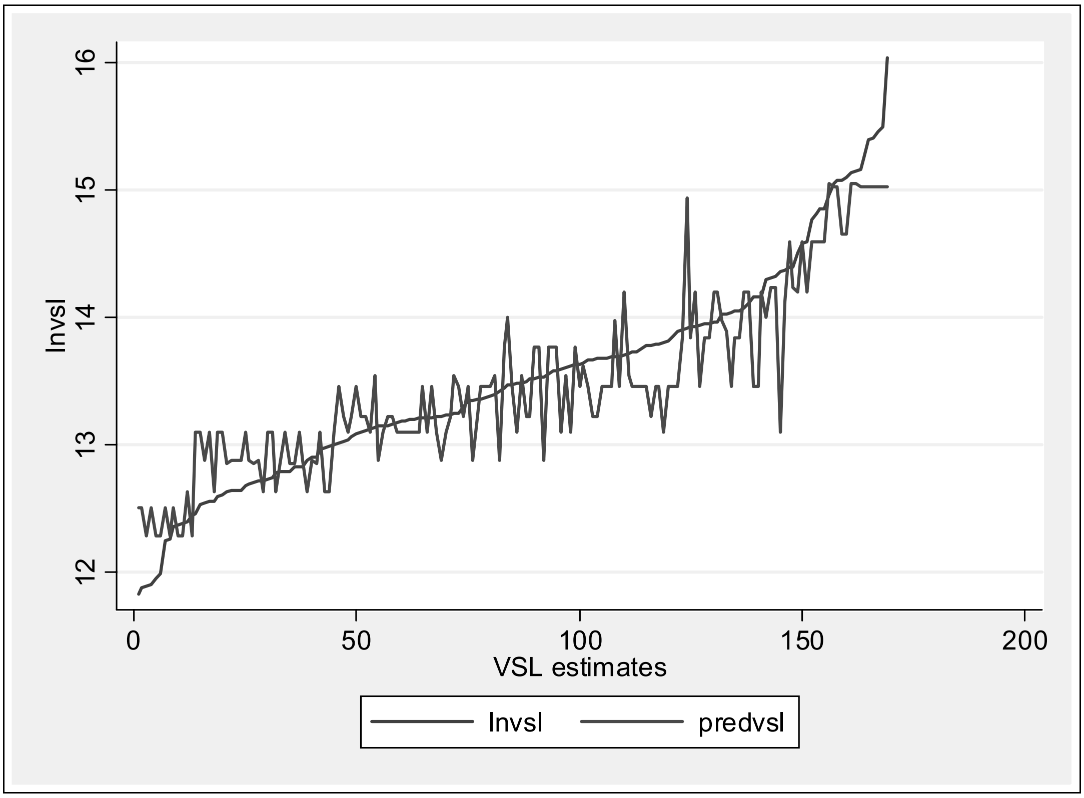
Figure 4 Plot of lnVSL and predicted or transferred lnVSL from trimmed version of Model V of the unscreened sample



Estimates from studies using same "best-practice" questionnaire

89. The same test was done for the studies that use a similar best practice questionnaire, *i.e.* Model II from Section 3.5 which has five explanatory variables. In this case much variation and heterogeneity has been eliminated by focusing on studies that are methodologically similar. One would therefore expect the model to predict out-of-sample estimates with higher accuracy than the previous models. This is also what is observed: overall mean TE is 26% and median TE 22%. The trimmed version of this experiment yields a mean TE of 25% (median 22%). That is high accuracy, approaching the low level suggested above by Kristofersson and Navrud (2007) of 20%.

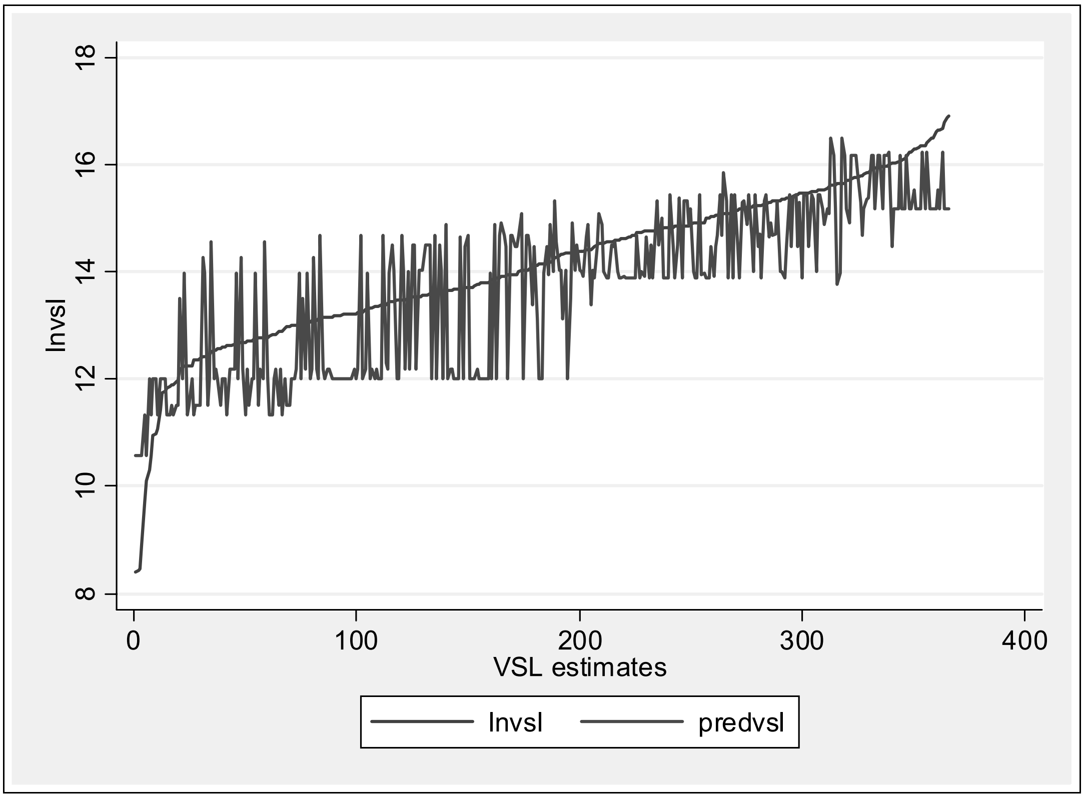
Figure 5 Plot of lnVSL and predicted or transferred lnVSL from Model II of the "best practice" questionnaire sample



Estimates recommended by authors

90. Finally, the same procedure was carried out for the full model for the sample where authors recommended values to be excluded. Note that many of the estimates authors advised to exclude were eliminated by the other screening criteria used for previous models. It is not clear what to expect in this case. Results show that mean TE is quite high at 89%, while the median is 59%. Trimming this model reduces the mean and median to 60 and 40%, respectively, approaching the high end of the accuracy interval discussed above.

Figure 6 Plot of lnVSL and predicted or transferred lnVSL from Model V of the author recommended sample



Summary points

91. An accuracy test was carried out for the four main types of screening criteria applied to the data. Removing VSL estimates one by one and estimating MA models on the remaining data to predict the out-of-sample observation, yielded the following main results:

- The unscreened dataset, with meta-regression models with the highest heterogeneity and lowest explained variation, yield the highest overall mean transfer error of around 120%.
- The mean transfer error drops to around 60% for the most comprehensive model when the first level screening criteria are applied.
- Choosing the most methodologically similar studies where values have been derived based on the same "best practice" questionnaire, yields an overall mean transfer error at a very low level of 20%.
- Following author recommendations of excluding observations do not seem to reduce transfer error. The mean TE is at almost 90%, comparable to the simple model (only GDP and risk change included) from the first level screening sample (TE of 100%).
- The more complete models (more comprehensive number of variables included) yield lower transfer errors than the simple models (only 1-2 key variables included).
- Trimming high and low values reduces transfer errors. The highest reduction is for the author recommended sample, of 30% points reduction (from 90 to 60%).
- Weighing estimates down if there are many included from one survey, does not seem to influence transfer errors much in the case where this was tried.

4.3 Comparison of BT techniques – Which one to choose?

92. To more closely resemble an actual BT situation, a single VSL estimate is drawn randomly from one study to represent a benchmark, unknown VSL value for a policy or program under assessment. This is assumed to be the "true" value for this context. The next step is to use the other studies to transfer a best VSL estimate to that policy context, based on simple and more sophisticated BT techniques. TE from the simple BT techniques is compared with the use of five MA-BT models. The choice of the latter is partly based on the accuracy assessment in the previous section.

93. This is a simple comparison based on one example of a BT situation. A comprehensive assessment for all VSL values in the dataset is not conducted, as for example done by Lindhjem and Navrud (2008) and Johnston and Thomassin (2010). Even so, this example illustrates that the choice of which BT method is not an easy one. Even if one chooses to go for a MA-BT approach, the choice of screening procedure (and other methodological choices) will influence the results.

94. Table 8 below gives an overview and explanation of different possible BT choices an analyst has when in need of a suitable VSL estimate to assess a particular mortality risk reduction policy. The first six BT techniques (N1-N6) are based on naïve transfers of mean VSL estimates that are adjusted or chosen in a certain way. The next five BT techniques (MA1-MA5) utilize the meta-regression models estimated in Chapter 3 and initially tested in Chapter 4.2, to estimate and transfer VSL estimates.

Table 8. Common BT methods tested

#	BT method for VSL	Description/Model used
N1	Naïve unit BT: mean of most similar international studies *	Pick VSL estimate from most similar study
N2	Naïve unit BT: mean of unscreened international studies	Adjusted by currency, not GDP.
N3	Naïve unit BT: mean of international studies, simple screening and GDP-adjustment	Same screening as for MA2 below. Adjusted by currency and GDP. Income elasticity set to unity.
N4	Naïve unit BT: mean of international studies with same risk change, simple screening and GDP-adjustment	Same screening as for MA2 below. Only for the studies with the same risk change. Adjusted by currency and GDP. Income elasticity set to unity.
N5	Naïve unit BT: mean of similar "best practice" studies	Same screening as for MA3 below. Adjusted by currency, not GDP.
N6	Naïve unit BT: mean of similar "best practice" studies adjusted with GDP	Same screening as for MA3 below. Adjusted by currency and GDP. Income elasticity set to unity.
MA1	Meta-analytic BT: unscreened	Model V, Section 3.3
MA2	Meta-analytic BT: simple screening	Model V, Section 3.4
MA3	Meta-analytic BT: similar "best practice" studies	Model II, Section 3.5
MA4	Meta-analytic BT: author recommendation	Model V, Section 3.6
MA5	Meta-analytic BT: simplified trimmed model	Trimmed version of Model I, Section 3.4**. (Only risk change and GDP included)

*Very few countries have enough studies domestically. Therefore the search is done for international studies.
** The same model as is displayed in Annex 2.

95. Below is presented a quick description of how each of the BT methods is used to derive a VSL estimate. At the end, the estimated values derived from each BT approach are summarized. But first a particular benchmark value is chosen which will serve as the example through this exercise.

Choice of "benchmark value"

96. It was decided to choose a study from Japan as the source for a benchmark value we want to approximate through BT techniques. The study has utilized the "best practice" questionnaire and should

represent a good quality estimate of VSL. The study reports several estimates and random VSL value of *USD 2 787 561* was chosen**.**

97. The study has valued a 1/10 000 risk change related to health (rather than environment or traffic), is immediate (not latent), is chronic and private (affects the respondent and his household only) and has been explained to respondents using a 1000 square grid. Further, the survey was conducted in 1999 using self-administration on a PC asking a dichotomous choice WTP question.

N1 – Take VSL estimate from most similar studies

98. A commonly used BT strategy is to search for a domestic study, which has valued a similar risk change and then pick one or take the mean of the most similar VSL estimates reported from that study. If a national study does not exist, an option is to choose a similar international study. It is not straight-forward to decide which "similarity criteria" should be applied as the analyst may typically not find one, unique study that match all the risk and population characteristics that define the policy context of interest.

99. On would perhaps think that the risk reduction should be the same. This reduces the number of potential VSL estimates to 91 from the full sample. Further, if we think that the type of risk should be the same ("health"), this leaves 79 potential estimates. Of these, 75 estimates are for chronic risk changes. Further, of these estimates 65 describe a private risk change which is immediate (not latent). Adding the variables from Table 3, that the risk change affects the individual (rather than the household) and is not related to cancer, leaves finally 58 candidate estimates. This search process can go on until a sufficiently similar study is found. However, it would be difficult to decide which variables should be used, in which order and when to stop the screening process.

100. The weighted mean VSL of the final 58 estimates is *USD 5 341 014*. This estimate is around double that of the benchmark value above.

N2 – Take mean of full VSL sample

101. A simpler method than picking a single study or do a detailed matching of variable characteristics with the policy context to arrive at a shortlist of similar estimates, would be instead to take a raw mean of VSL estimates of all collected studies. A weighted mean VSL (where more estimates from the same survey is weighted down) for this procedure is *USD 7 343 737*.

N3 – Take mean of screened VSL sample, adjust by GDP difference

102. Screening estimates according to the procedure discussed in Section 3.4 reduces the number of estimates from 922 to 421 at disposal for BT. Weighted mean VSL from this sample is USD 3 302 917. GDP per capita for Japan for this year was USD 30 290 while the weighted mean of the GDP per capita for the sample was USD 25 441. Assuming an income elasticity of VSL of 1 (as is roughly what is found in most of the meta-regressions), leaves a simple, income adjusted transferred VSL estimate to Japan of *USD 3 932 446*.

N4 – Take mean of screened VSL sample for same risk change, adjust by GDP difference

103. Doing the same exercise as for N3 with the exception that only studies that have the same risk reduction as the Japanese study of 1/10000 are included, reduces the number of available estimates to 41. The weighted mean of this sample is USD 3 790 811. Since the remaining estimates actually come from countries with higher mean GDP capita (USD 31 028), income adjustment yields a transferred VSL estimate for Japan of *USD 3 700 647*, when the income elasticity is set to unity.

N5 – Take mean VSL of "best practice" studies

104. Taking the mean of the estimates using "best practice" approach to VSL valuation, yields a VSL estimate of *USD 1 482 588.* This is a bit less than half of the benchmark value.

N6 – Take mean VSL of "best practice" studies, adjust by GDP difference

105. Adjusting the N5-estimate by GDP differences between the average of the sample and Japan, yields a VSL estimate of *USD 1 781 562.*

MA1 – MABT, unscreened

106. If an overall meta-regression analysis was carried out with no concern regarding screening based on objective or subjective criteria of quality, we could take as a starting point Model V in Section 3.3[13]. The estimated meta-regression function of this model is:

lnVSL = 0.790 + 1.331* lngdp + 0.159 *envir + 0.459* traffic - 0.390*public –

0.177*household + 0,908* cancerrisk – 0.489*latent +1.057*noexplan – 0.172*turnbull

107. First, this equation is used to estimate and transfer a VSL value to the policy context in Japan. Since the methodological values are unknown at the policy site (in reality), a common practice is to set the values of the methodological variables equal to some best practice value. In this case, it is good practice to use thorough explanation in explaining risk changes (hence "noexplan" is set to zero). Similarly, since the turnbull approach typically yields a lower bound on VSL, this variable is also set to zero.

108. Further, since the risk is related to health, for an individual (not a household), a private risk program, immediate and not related to cancer, all these variables are set to zero. That leaves the following simple equation:

lnVSL = 0.790 + 1.331* lngdp

109. Inserting the GDP/capita for Japan of USD 30290 and taking the antilog (inverse) of lnVSL[14] yields an estimate of VSL of USD *2 030 955*.

MA2 – MABT, first level screening

110. Instead of using the unscreened model above, the first level screening of observations from Section 3.4 is applied. Inserting values for the risk change (1/10 000) and GDP per capita yields an estimated VSL of *USD 3 199 152*.

MA3 – MABT, picking "best practice" studies

13 For simplicity and ease of comparison, for MA-BT, we use the models as they are estimated in Chapter 3 without removing the single estimate we use the models to predict and rerun the models without this observation. This has little or no impact on the tests we describe here.

14 Along with Stapler and Johnston (2009) – and to make the calculations simpler and more transparent for non-experts – no correction is made for so-called "econometric error" when converting from log, cf. Bokstael and Strand (1987). Such correction would in most cases have only a small impact on the estimated VSL values.

111. Conducting the same procedure as above, except using the most comprehensive model from Section 3.5 of the best practice studies, *i.e.* Model II, yields an estimated VSL of USD *2 143 787*. Compared to the MA-BT models above, this model also includes the variable year (of data collection). In the same way as for the previous MA-BT equations, all other variables except GDP, the risk change and study year, are set to zero to fit the policy context the estimate is transferred to.

MA4 – MABT, author recommendation

112. Finally, utilising the last screening procedure from Chapter 3, Model 5 of the author-recommended sample of Section 3.6 inserting values for the risk change and GDP for Japan, yields a VSL estimate of *USD 3 157 449*.

MA5 – MABT, simplified, trimmed model

113. A simple MA-BT option is to follow the screening procedure in Section 3.4, estimate the simplest model including only the variable risk change and GDP. Further, to eliminate impact of very high and low values, the sample can be trimmed. Using the Model I in Annex 2 where this procedure was followed, yields a VSL estimate of USD *2 386 919*.

Comparison of BT methods summarised

114. The estimated VSL values are repeated in the table below for the 11 BT methods. The first column represents the benchmark value, the true value for the Japanese policy context that we want to approximate through BT. It can be seen from the table that the simple, naïve BT methods yield higher VSL estimates and that all have higher transfer errors than for the MA-BT methods, varying from 33% to 163% (column 3). The highest TE comes from taking the raw mean from the full, unscreened sample of VSL estimates. This result is as expected. Following a searching procedure to find the most similar subset of studies (N1) also yield fairly high TE at 92%. More elaborate transfer of mean VSL in methods N3-N6 produces TE that approximate acceptable levels (Around 40%).

115. The MA-BT methods all have lower transfer error than the simple BT methods at around 13-23%. The lowest error comes from using the author-recommended screening procedure (MA4). It is also worth noting that the simple screening procedure (MA5) where a simple MA model is used where only GDP and the risk change is included, produces TE which is almost as almost as low at 14%.

Table 9. Comparison of different simple methods with meta-analytic BT for an example scenario

Method	A: "Benchmark value", policy context (USD 2005)	B:Estimated/transferred value (USD 2005)	C: Transfer error (TE, %)*	Rank in terms of TE
N1	2 787 561	5 341 014	91,60	10
N2	2 787 561	7 343 737	163,45	11
N3	2 787 561	3 932 446	41,07	8
N4	2 787 561	3 700 647	32,76	6
N5	2 787 561	1 482 588	46,81	9
N6	2 787 561	1 781 562	36,09	7
MA1	2 787 561	2 030 955	27,14	5
MA2	2 787 561	3 199 152	14,77	3
MA3	2 787 561	2 143 787	23,09	4
MA4	2 787 561	3 157 449	13,27	1
MA5	2 787 561	2 386 919	14,37	2

* C = (B-A)/A*100%, cf. definition of TE in Section 4.2.

Summary points

116. A simple example was explained where an estimate of VSL from Japan was picked to represent an unknown, true VSL value at a policy site. The task was then to use different benefit transfer techniques to derive a VSL value that could be transferred to the Japanese case. Six simple BT methods were compared with five versions of our MA models. Though no general conclusions can be drawn based on this example, the example demonstrated that:

- Transferring a raw, unadjusted mean VSL value from a full sample or a sample that has been reduced based on screening for similarity with the policy site (methods N1 and N2) produces relatively high transfer errors (92-163%).
- The transfer error for simple mean transfers can be reduced to (almost) acceptable levels (around 40%) by using our first two screening procedures and/or adjusting by GDP differences.
- The five different MA models all produce lower transfer errors (from 13-23%) than just transferring mean VSL estimates.
- The author-recommended sample used for MA-BT produces the lowest transfer error in this specific example (13%).
- The example, though just illustrative, demonstrates in our case that with two highly significant variables in the MA models of risk change and GDP, the transfer process may be simplified by including only those two variables in adjustments.

5. Summary and conclusions

5.1 Objectives and approach

117. This paper builds on and continues the work documented in Braathen *et al.* (2009). Since that study, the meta data have been updated with more recent studies (as of March 2010) and more information has been added from studies and collected from authors to arrive at a more complete and comprehensive database.

118. Based on this updated dataset, the main objectives of this paper have been to :

- Consider screening procedures for the VSL estimates based on quality of studies and other factors that will make the dataset more amenable and appropriate for statistical analysis and for use in benefit transfer (BT) applications for policy use.
- Conduct some further meta-analysis (MA) regression models on subsets of the data generated by the above screening criteria to investigate robustness and sensitivity of results.
- Consider some MA models for BT and policy use and the reliability of those models.

119. Three screening procedures were used:

- Level 1 – Risk change must be reported, small and unrepresentative samples removed
- Level 2 – In addition to Level 1 criteria, only those studies that use a similar version of a "best practice" questionnaire are included.
- Level 3 – In addition to Level 1 criteria, the authors' own recommendation related to which estimates should be excluded is applied.

120. In addition, meta-regression models for a full, unscreened dataset were run for purposes of comparison. Models were first checked for sensitivity of key explanatory variables to the screening procedures. The second part of the paper assesses the accuracy of the most promising models and compares the use of meta-analysis for benefit transfer with simple BT techniques. The purpose is to

illustrate how MA can be used for BT and that BT is many cases may be more accurate than using mean of VSL estimates from a database or a subset of a database (e.g. based on similarities with the policy site context where a VSL estimate is needed).

5.2 Main results I: Sensitivity of results to screening procedures

121. Overall, through our three different screening procedures, the following results were found:

- Gradually higher explanatory power of the models as stricter screening criteria are applied. Heterogeneity is gradually reduced in the data. The full sample also has fairly high R-squared
- Fairly robust results. Effects of income (GDP) and risk change are strong positive and negative, respectively. Income elasticity of VSL seems to be around 1 for most regressions (although lower for the "best practice" studies in Section 3.5).
- Strong indication that public risks are valued *lower* than private risks.
- Some indication that environment and traffic-related risks are valued lower than health risks.
- Some indication that latency of risk changes reduces VSL, as expected.
- No proper explanation of the risk change in the survey tends to increase WTP and corresponding VSL
- Indications that Turnbull-estimates are lower than other estimates, as expected since this non-parametric procedure gives a lower bound on WTP and VSL.

122. Further sensitivity analysis was conducted for some of the results from Sections 3.3-3.6 in the Annexes, related to the weighting strategy applied and to trimming high and low values from the samples. The results are generally found to be robust to these model changes. R-square drops somewhat for all models, but significance for the majority of variables is retained. The income elasticity seems to drop a bit, while the latency variable comes out more robustly as significantly negative than in the models in Chapter 3.

5.3 Main results II: Accuracy of MA models for benefit transfer and policy use

123. An accuracy test was carried out for the four main types of screening criteria applied to the data. Removing VSL estimates one by one and estimating MA models on the remaining data to predict the out-of-sample observation, yielded the following main results:

- The unscreened dataset, with meta-regression models with the highest heterogeneity and lowest explained variation, yield the highest overall mean transfer error of around 120%.
- The mean transfer error drops to around 60% for the most comprehensive model when the first level screening criteria are applied.
- Choosing the most methodologically similar studies where values have been derived based on the same "best practice" questionnaire, yields an overall mean transfer error at a very low level of 20%.
- Following author recommendations of excluding observations do not seem to reduce transfer error. The mean TE is at almost 90%, comparable to the simple model (only GDP and risk change included) from the first level screening sample (TE of 100%).
- The more complete models (more comprehensive number of variables included) yield lower transfer errors than the simple models (only 1-2 key variables included).
- Trimming high and low values reduces transfer errors. The highest reduction is for the author recommended sample, of 30% points reduction (from 90 to 60%).
- Weighing estimates down if there are many included from one survey, does not seem to influence transfer errors much in the case where this was tried.

5.4 Main results III: Comparison of MA-based BT and simple BT methods

A simple example was explained where an estimate of VSL from Japan was picked to represent an unknown, true VSL value at a policy site. The task was then to use different benefit transfer techniques to derive a VSL value that could be transferred to the Japanese case. Six simple BT methods were compared with five versions of our MA models. Though no general conclusions can be drawn based on this example, the example demonstrated that:

- Transferring a raw, unadjusted mean VSL value from a full sample or a sample that has been reduced based on screening for similarity with the policy site (methods N1 and N2) produces relatively high transfer errors (92-163%).
- The transfer error for simple mean transfers can be reduced to (almost) acceptable levels (around 40%) by using our first two screening procedures and/or adjusting by GDP differences.
- The five different MA models all produce lower transfer errors (from 13-23%) than just transferring mean VSL estimates.
- The author-recommended sample used for MA-BT produces the lowest transfer error in this specific example (13%).
- The example, though just illustrative, demonstrates in our case that with two highly significant variables in the MA models of risk change and GDP, the transfer process may be simplified by including only those two variables in adjustments.

5.5 Final remark

124. This paper has attempted to investigate further important issues related to the sensitivity of meta-analysis and the suitability for policy use and benefit transfer. This is an immature field where much remains to be done. However, the results show that the MA models used are not overly sensitive to the screening procedures applied or to additional sensitivity issues (trimming and weighting). The best models also do a fairly good job of predicting VSL estimates (as an indication of suitability for benefit transfer). Compared to simple benefit transfer methods, the benefit transfer example based on the best MA models perform better. They can also be simplified without much lack of precision, although this is not tested comprehensively. The two by far most important variables for benefit transfer and for explaining variation in VSL estimates are the risk change valued and the level of GDP. It is not decided here how to deal with the potentially problematic issue that VSL seems to be so sensitive to the risk change valued.

ANNEX 1: SENSITIVITY TO WEIGHTING STRATEGY

125. This annex runs through the same regressions as done in Chapter 3, except the weighting strategy is removed. In other words, each estimate counts equally and is not weighted down depending on whether there are many estimates included from one survey. This is a common procedure in the MA literature, though the weighted option is preferred in this paper, results of which are presented in the main body of the report.

A1.1 Full dataset – no screening

126. The models in Table A1.1 are comparable to the models in table 4 in Section 3.3. GDP is still highly significant through all models. The elasticity is a bit lower in the unweighted models. The cancerrisk variable is also significant as before. Interestingly, while latency was not significant in the weighted models, the variable is significant now for Models IV and V where the variable is included. However, the traffic variable is now not significant anymore. No explanation still strongly results in higher stated VSL values. Overall, the explained variation (R-squared) drops somewhat.

Table A1.1 Unweighted regression results, full sample.

	Model I	Model II	Model III	Model IV	Model V
lngdp	1.030***	1.099***	1.146***	1.039***	0.871***
	(0.218)	(0.204)	(0.207)	(0.228)	(0.229)
envir		-0.383	-0.0684	-0.0298	0.132
		(0.575)	(0.449)	(0.310)	(0.283)
traffic		-0.0159	0.0376	-0.00829	-0.127
		(0.319)	(0.318)	(0.342)	(0.297)
public			-0.222	-0.231	-0.208
			(0.275)	(0.251)	(0.224)
household			-0.329	-0.263	-0.235
			(0.297)	(0.222)	(0.194)
cancerrisk				0.859**	1.056***
				(0.399)	(0.378)
latent				-1.076***	-0.963***
				(0.347)	(0.337)
noexplan					1.183***
					(0.235)
turnbull					-0.0462
					(0.426)
Constant	4.344**	3.756*	3.359	4.456*	5.843**
	(2.173)	(2.018)	(2.025)	(2.257)	(2.322)
Observations	922	922	922	922	922
R-squared	0.302	0.311	0.322	0.420	0.487
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

A1.2 First level screening procedure

127. The models in Table A1.2 are comparable to the models in Table 5 in Section 3.4. The income elasticity also drops in this case and GDP is still highly significant. The risk change coefficient is roughly unchanged in size and significance. Traffic is still significant in one model. The public variable is less significant (drops two stars). Latency now comes out as weakly significant in one model. The "noexplan" variable is still strongly significant. Overall R-squared drops a bit.

Table A1.2 Unweighted regression results, first level screening procedure.

	Model I	Model II	Model III	Model IV	Model V
lngdp	0.625***	0.679***	0.679***	0.609***	0.525***
	(0.178)	(0.169)	(0.165)	(0.162)	(0.181)
lnchrisk	-0.385***	-0.548***	-0.572***	-0.587***	-0.515***
	(0.103)	(0.109)	(0.0884)	(0.0761)	(0.0812)
envir		-1.278***	-0.564	-0.604	-0.378
		(0.455)	(0.524)	(0.544)	(0.501)
traffic		-0.666**	-0.354	-0.450	-0.386
		(0.266)	(0.353)	(0.305)	(0.269)
public			-0.830*	-0.800*	-0.722*
			(0.424)	(0.433)	(0.407)
household			-0.372	-0.255	-0.331
			(0.542)	(0.424)	(0.408)
cancerrisk				0.260	0.349*
				(0.220)	(0.204)
latent				-0.444*	-0.311
				(0.236)	(0.236)
noexplan					0.595***
					(0.191)
turnbull					-0.642
					(0.388)
Constant	4.810***	3.199**	3.006*	3.615**	4.918**
	(1.714)	(1.550)	(1.573)	(1.675)	(1.953)
Observations	421	421	421	421	421
R-squared	0.637	0.710	0.748	0.763	0.788
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

A1.3 Estimates from studies using same "best practice" questionnaire

128. The models in Table A1.3 are comparable to the models in Table 6. The income elasticity is almost identical and significant at the same level. The risk change seems to matter a bit more as the coefficient increases (significance the same). The other variables remain (strongly) significant compared to the weighted models in Chapter 3. R-square drops a bit here as well.

Table A1.3 Unweighted regression results, "best practice" questionnaire sample

	Model I	Model II
lngdp	0.509***	0.412***
	(0.0702)	(0.0535)
lnchrisk	-0.710***	-0.684***
	(0.0485)	(0.0538)
latent		-0.177***
		(0.0296)
turnbull		-0.794**
		(0.234)
lnyear	0.546***	0.527***
	(0.117)	(0.109)
Constant	2.577**	3.853***
	(0.896)	(0.683)
Observations	169	169
R-squared	0.732	0.857
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

A1.4 Estimates recommended by authors

129. The models in Table A1.4 are comparable with the models reported in Table 7 in Section 3.6. Here, the income elasticity seems to drop more than for the other models discussed above. Environment remains significantly negative, and coefficients are larger (negative). Traffic is now also significant through all four models where the variable is included, compared to two models in the weighted version. In other words, not weighing estimates seems to make bigger differences between VSL estimates for different types of risks. Other variables also get increase in significance (for example cancer risk). The R-squared drops somewhat compared to the models in Chapter 3.

Table A1.4 Unweighted regression results, author recommended sample

	Model I	Model II	Model III	Model IV	Model V
lngdp	0.588***	0.609***	0.645***	0.559***	0.464**
	(0.177)	(0.161)	(0.171)	(0.164)	(0.179)
lnchrisk	-0.397***	-0.646***	-0.619***	-0.651***	-0.578***
	(0.111)	(0.0842)	(0.0892)	(0.0713)	(0.0776)
envir		-1.824***	-1.343***	-1.550***	-1.274***
		(0.417)	(0.396)	(0.361)	(0.344)
traffic		-0.916***	-0.664**	-0.752***	-0.685***
		(0.246)	(0.313)	(0.269)	(0.245)
public			-0.243	-0.0790	-0.0292
			(0.328)	(0.325)	(0.276)
household			-0.661	-0.623	-0.693
			(0.516)	(0.408)	(0.415)
cancerrisk				0.587*	0.668**
				(0.299)	(0.268)
latent				-0.333*	-0.194
				(0.185)	(0.187)
noexplan					0.574**
					(0.211)
turnbull					-0.684*
					(0.367)
Constant	5.042***	3.084*	2.970*	3.551**	4.961**
	(1.748)	(1.521)	(1.558)	(1.653)	(1.958)
Observations	366	366	366	366	366
R-squared	0.631	0.755	0.774	0.791	0.818
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

ANNEX 2: SENSITIVITY TO TRIMMING

130. This annex runs through the same regressions as done in Chapter 3, except we remove the 2.5% highest and lowest VSL estimates. There is no basis in theory to do this, but sometimes this strategy is applied in the MA literature to reduce the sensitivity of results to high and/or low values.

A2.1 Full dataset – no screening

131. The models in Table A2.1 can be compared to the models reported in Table 4 in Section 3.3. Note that since some further observations (46) have been screened out, the models reported here are not strictly comparable with Table 4. The first thing to note is that the income elasticity has dropped from around 1.3-1.5 to a bit above 1. Significance level is still the same. The traffic variable is less significant now. A part from these observations, significance seems to be held up for the same variables as in the untrimmed version of the models (though R-squared drops somewhat). That indicates that very high or low VSL estimates do not influence results substantially.

Table A2.1 Regression results, full sample, trimmed 2.5% high and low.

	Model I	Model II	Model III	Model IV	Model V
lngdp	1.099***	1.090***	1.169***	1.140***	1.004***
	(0.176)	(0.188)	(0.178)	(0.185)	(0.187)
envir		0.0270	0.350	0.276	0.296
		(0.375)	(0.346)	(0.261)	(0.244)
traffic		0.534**	0.581**	0.573*	0.418
		(0.256)	(0.257)	(0.291)	(0.254)
public			-0.358	-0.331	-0.305
			(0.289)	(0.299)	(0.276)
household			-0.389	-0.352	-0.269
			(0.249)	(0.254)	(0.238)
cancerrisk				0.386	0.556*
				(0.294)	(0.284)
latent				-0.594*	-0.609**
				(0.313)	(0.267)
noexplan					0.889***
					(0.228)
turnbull					0.384*
					(0.205)
Constant	3.542**	3.399*	2.744	3.027	4.156**
	(1.758)	(1.876)	(1.797)	(1.869)	(1.894)
Observations	876	876	876	876	876
R-squared	0.283	0.320	0.353	0.378	0.442
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

A2.2 First level screening procedure

132. The models in Table A2.2 are comparable to the models reported in Table 5. In the same way as above, the income elasticity drops somewhat to below one. The risk change is still highly significant at roughly the same coefficient values. Other results seem largely unaffected by the trimming strategy which in this case removes 20 estimates.

Table A2.2 Regression results, first level screening procedure. Trimmed 2.5% high and low

	Model I	Model II	Model III	Model IV	Model V
lngdp	0.645***	0.727***	0.789***	0.738***	0.609***
	(0.138)	(0.124)	(0.118)	(0.129)	(0.125)
lnchrisk	-0.371***	-0.438***	-0.447***	-0.468***	-0.436***
	(0.0857)	(0.0823)	(0.0708)	(0.0587)	(0.0530)
envir		-0.856**	-0.255	-0.366	-0.242
		(0.365)	(0.256)	(0.284)	(0.255)
traffic		-0.283	-0.0489	-0.163	-0.221
		(0.250)	(0.280)	(0.261)	(0.192)
public			-0.857***	-0.762***	-0.693***
			(0.196)	(0.198)	(0.202)
household			-0.124	-0.0645	-0.118
			(0.237)	(0.200)	(0.196)
cancerrisk				0.237	0.354
				(0.254)	(0.233)
latent				-0.622**	-0.525**
				(0.273)	(0.249)
noexplan					0.788***
					(0.159)
turnbull					-0.157
					(0.322)
Constant	4.613***	3.432**	2.757**	3.125**	4.532***
	(1.386)	(1.334)	(1.304)	(1.427)	(1.405)
Observations	401	401	401	401	401
R-squared	0.556	0.615	0.689	0.722	0.778
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

A2.3 Estimates from studies using same "best practice" questionnaire

133. The models in Table A2.3 are comparable with the models reported in Table 6 of Section 3.5. In all 8 estimates are removed by the trimming strategy. The results seem largely unaffected, including the income elasticity.

Table A2.3 Regression results, "best practice" questionnaire sample. Trimmed 2.5% high and low

	Model I	Model II
lngdp	0.562***	0.473***
	(0.110)	(0.0906)
lnchrisk	-0.511**	-0.467**
	(0.154)	(0.161)
latent		-0.237***
		(0.0491)
turnbull		-0.544**
		(0.193)
lnyear	0.432**	0.428**
	(0.144)	(0.125)
Constant	3.722*	5.066***
	(1.556)	(1.308)
Observations	161	161
R-squared	0.751	0.807
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

A2.4 Estimates recommended by authors

134. The final set of models reported here are those where author-recommended values have been excluded. Models in Table A2.4 are comparable to Table 7 in Section 3.6. In this case 18 estimates have been dropped by the trimming strategy. The income elasticity drops from around 1 to between 0.6-0.8, but retains same level of significance. The risk change variable seems to be relatively less affected. Latency turns significant in these models. R-squared drops somewhat. Apart from that, results from Chapter 3 seems fairly robust.

Table A2.4 Regression results, author recommended sample. Trimmed 2.5% high and low

	Model I	Model II	Model III	Model IV	Model V
lngdp	0.642***	0.732***	0.798***	0.726***	0.574***
	(0.147)	(0.135)	(0.125)	(0.134)	(0.129)
lnchrisk	-0.384***	-0.486***	-0.463***	-0.494***	-0.475***
	(0.0902)	(0.0862)	(0.0723)	(0.0578)	(0.0472)
envir		-1.041***	-0.383	-0.573*	-0.477*
		(0.370)	(0.270)	(0.298)	(0.255)
traffic		-0.450	-0.162	-0.259	-0.362*
		(0.286)	(0.288)	(0.270)	(0.200)
public			-0.736***	-0.603***	-0.478***
			(0.189)	(0.206)	(0.166)
household			-0.257	-0.212	-0.330
			(0.243)	(0.203)	(0.200)
cancerrisk				0.358	0.517**
				(0.275)	(0.246)
latent				-0.584**	-0.478*
				(0.272)	(0.244)
noexplan					0.833***
					(0.183)
turnbull					-0.139
					(0.299)
Constant	4.498***	3.001**	2.571*	3.031**	4.528***
	(1.468)	(1.437)	(1.367)	(1.461)	(1.475)
Observations	348	348	348	348	348
R-squared	0.564	0.644	0.711	0.746	0.805
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

REFERENCES

Alberini A, Cropper M, Krupnick A and Simon N (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: 769-792

Amiran E Y and Hagen D A (2010) The Scope Trials: Variation in Sensitivity to Scope and WTP with Directionally Bounded Utility Functions. Journal of Environmental Economics and Management 59(3): 293-301

Arrow K J, Solow R, Leamer E, Portney P, Radner R and Schuman H (1993) Report of the NOAA Panel on Contingent Valuation. Federal Register 58: 4601-4614

Biausque V (2010). The value of statistical life: a meta-analysis (available in French and English), OECD.

Bockstael N E and Strand E (1987) The effect of common sources of regression error on benefit estimates. Land Economics 63(1): 11-20

Brander L M, Florax R J G M and Verrmaat J E (2006) The Empirics of Wetland Valuation: A Comprehensive Summary and a Meta-Analysis of the Literature. Environmental & Resource economics 33: 223-250

Brander L M, van Beukering P and Cesar H (2007) The recreational value of coral reefs: a meta-analysis. Ecological Economics 63(1): 209-218

Braathen N A, Lindhjem H and Navrud S (2009). Valuing lives saved from environmental, transport and health policies: A meta-analysis of stated preference studies. ENV/EPOC/WPNEP(2008)10/FINAL. Paris, OECD.

Eeckhoudt L R and Hammitt J K (2001) Background Risks and the Value of a Statistical Life. Journal of Risk and Uncertainty 23(3): 261-279

Gilovich T, Griffin D and Kahneman D (2002) Heuristics and biases : the psychology of intuitive judgement Cambridge University Press, Cambridge,

Hammitt J K and Graham J D (1999) Willingness to pay for health protection: Inadequate sensitivity to probability? J. Risk Uncertain. 18(1): 33-62

Jacobsen J B and Hanley N (2009) Are there income effects on global willingness to pay for biodiversity conservation. Environmental and Resource Economics 43(2): 137-160

Johnston R J and Rosenberger R S (2010) Methods, Trends and Controversies in Contemprary Benefit Transfer. Journal of Economic Surveys 24(3): 479-510

Johnston R J and Thomassin P J (2010) Willingness to Pay forWater Quality Improvements in the United States and Canada: Considering Possibilities for International Meta-Analysis and Benefit Transfer. Agricultural and Resource Economics Review 39(1): 114–131

Kahneman D and Tversky A, Eds. (2000) Choices, Values and Frames. Cambridge University Press

Kristofersson D and Navrud S (2005) Validity Tests of Benefit Transfer – Are We Performing the Wrong Tests? Environmental and Resource Economics 30: 279-286

Kristofersson D and Navrud S (2007) Can Use and Non-Use Values be Transferred Across Countries? In: S. Navrud and R. Ready (ed) Environmental Value Transfer: Issues and Methods, Springer, Dordrecht, The Netherlands

Krupnick A (2007) Mortality-risk Valuation and Age: Stated Preference Evidence. Review of Environmental Economics and Policy 1(2): 261-282

Lindhjem H and Navrud S (2008) How Reliable are Meta-Analyses for International Benefit Transfer? Ecological Economics 66(2-3): 425-435

Lindhjem H and Navrud S (2009) Asking for Individual or Household Willingness to Pay for Environmental Goods: Implication for aggregate welfare measures Environmental and Resource Economics 43(1): 11-29

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

Navrud S and Ready R (2007) Review of methods for value transfer. In: S. Navrud and R. Ready (ed) Environmental value transfer: Issues and methods, Springer,

Nelson J P and Kennedy P E (2009) The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment. Environmental and Resource Economics 42(3): 345-377

Richardson L and Loomis J (2009) The total economic value of threatened, endangered and rare species: An updated meta-analysis. Ecological Economics 68(5): 1535-1548

Rosenberger R S and Johnston R J (2009) Selection effects in meta-analysis and benefit transfer: Avoiding unintended consequences. Land Economics 85(3): 410-428

SEPA (2006). An instrument for assessing the quality of environmental valuation studies, Swedish Environmental Protection Agency.

Stanley T D and Jarrel S D (2005) Meta-Regression Analysis: A Quantitative Method of Literature Surveys. Journal of Economic Surveys 19(3): 299-308

Stapler R W and Johnston R J (2009) Meta-analysis, benefit transfer, and methodological covariates: Implications for transfer error. Environmental and Resource Economics 42(2): 227-246

USEPA (2006). Report of the EPA work group on VSL meta-analysis. Report NCEE-0494. Washington, National Center for Environmental Economics, EPA.

Van Houtven G, Powers J and Pattanayak S K (2007) Valuing water quality improvements using meta-analysis: Is the glass half-full or half-empty for national policy analysis? Resource and Energy Economics 29(3): 206-228

van Houtven G, Sullivan M B and Dockins C (2008) Cancer Premiums and Latency Effects: A Risk Tradeoff Approach for Valuing Reductions in Fatal Cancer Risks. Journal of Risk and Uncertainty 36: 179-199