The Reliability of Delphi Surveys and Benefit Transfer to Predict Outcomes of Contingent Valuation Estimates

Anders Dugstad*

Postdoctoral researcher

School of Economics and Business, Norwegian University of Life Sciences (NMBU)

Chr. Magnus Falsens vei 18, 1433 Ås, Norway

anders.dugstad@nmbu.no

Ståle Navrud

Professor

School of Economics and Business, Norwegian University of Life Sciences (NMBU)

Chr. Magnus Falsens vei 18, 1433 Ås, Norway

stale.navrud@nmbu.no

*Corresponding author. Please address correspondence to: Anders Dugstad, School of Economics and Business, Norwegian University of Life Sciences (NMBU), Chr. Magnus Falsens vei 18, 1433 Ås, Norway. E-mail: anders.dugstad@nmbu.no. Voice: +4767231195.

doi:10.3368/le.99.3.051922-0037R

Abstract

Benefit transfer (BT) and valuation experts' assessments in Delphi Contingent Valuation (CV) surveys have been used as an alternative to prohibitively expensive worldwide CV surveys to obtain estimates of total non-use values of global public goods. We test the reliability of international BT and Delphi CV estimates by comparing them to a population CV survey of Norwegian households' willingness-to-pay (WTP) for Amazon Rainforest preservation plans. The Delphi CV method predicts WTP and scope effects in line with conventional BT, motivating further testing of the Delphi CV method as a time- and cost-saving way of valuing global public goods.

1. Introduction

The Amazon rainforest is the world's largest rainforest, making up as much as 40 percent of the total remaining area of tropical forest worldwide. Since the 1960s, deforestation of the Amazon rainforest has become a major global concern (Uhl, 1987). As of 2017, 17 percent of the Amazon rainforest has been disturbed (Kehl et al., 2015; Bullock et al., 2020). The main drivers of deforestation in the Amazon rainforest are cattle, crops, and timber to supply global and local markets. In addition, road networks are an essential driver (Jusys, 2016; de Oliveira et al., 2020).

The forest provides critical local, regional and global ecosystem services. Therefore, the Amazon rainforest can be defined as a global public good (Strand et al., 2017; Navrud & Strand, 2018). It provides global benefits and ecosystem services in terms of biodiversity, carbon storage, and recreational and non-use values (Andersen et al., 2002; Brouwer et al., 2022).

Deforestation causes loss of ecosystem services and biodiversity (Foley et al., 2007), which reduces human well-being for both local and distant beneficiaries of the forest. Therefore, it is reasonable to assume that distant beneficiaries are willing to pay to preserve the Amazon rainforest. Non-use values dominate among distant beneficiaries as most people globally have not visited the Amazon and do not have recreational use values of the Amazon rainforest.[i]

Non-use values represent the value of benefits people obtain by the existence of ecosystem services, the enjoyment of these services by others, and that the good is available for future generations (i.e., bequest values) (Pascual et al., 2010). Even though willingness-to-pay (WTP) per household for preserving the Amazon rainforest and its biodiversity could be small, total non-use values aggregated over the global population would be substantial. Thus, it is essential to include non-use values for distant beneficiaries in a global cost-benefit analysis of preservation plans (Navrud & Strand, 2018). Among the environmental valuation techniques, only the Stated

Preference (SP) methods, i.e., Contingent Valuation (CV) and Choice Experiments (CEs), can measure non-use values.

Only a few SP studies have estimated distant beneficiaries' WTP to preserve the Amazon rainforest. Kramer and Mercer (1997) conducted a CV study among U.S residents to determine their WTP to protect tropical rainforests. They find that U.S residents, on average, are willing to pay between $21 and $31 to preserve 5 percent of tropical rainforests in addition to what was already preserved at the time. This was a one-time voluntary payment.

Horton et al. (2003) conducted a CV study in the UK and Italy to determine households' WTP to impose preservation programs in parts of the Amazon rainforest. In the first program, 5 percent of the Brazilian Amazonia was to be preserved, with an average WTP per household of £30 as annual tax. The second program preserved 20 percent with a yearly average WTP per household of £39.

Brouwer et al. (2022) conducted a meta-analysis on the economic value of the Brazilian Amazona's ecosystem services, focusing on the Brazilian valuation literature. The study provides insights into local valuation estimates, where the estimated mean value of all ecosystem services assessed among local populations is about 410 USD/ha/year.

Navrud and Strand (2018) conducted a Delphi CV survey in 2012 to estimate WTP among households in European countries to preserve the Amazon rainforest. In the study, forty-eight valuation experts from different European countries were asked to guess the outcome of a CV survey of two preservation plans, A and B, if a CV survey was conducted in their respective countries and Europe overall. The experts were asked to state the mean and median annual WTP per household in their country and Europe overall; to be paid as an extra yearly income tax. Thus, four estimates were elicited from the experts for each of the two preservation plans; mean and

median WTP per household in their own country and the same for all European households. The Delphi survey was performed in two rounds, where the experts in Round 2 got to see the WTP distribution for their own country and Europe of what all experts stated in Round 1 and were then allowed to adjust their own four estimates. The study was later extended by Strand et al. (2017) to include USA and Canada, Australia and New Zealand, and low-income, lower-middle-income, and upper-middle-income Asian countries.

In Plan A, there would be no further loss of forest or species by 2050. Thus, 85 percent of the total area would remain in 2050. Plan B implied some forest loss, where 75 percent of the total area would remain by 2050, and 7 percent of the species would be lost. The two preservation plans were compared to a reference (business-as-usual) scenario where 60 percent of the forest would remain by 2050 and 12 percent of the species would be lost (Navrud & Strand, 2018).

Twenty-nine environmental valuation experts from Northern and Central Europe were surveyed in the European Delphi CV study, where eleven of these were from Nordic countries. Of the Nordic experts, three Norwegian experts were surveyed. The mean WTP of the predictions made by the Norwegian experts for the most ambitious preservation of Plan A in round 1 was €65 per Norwegian household (Navrud & Strand, 2018). In Round 2, where they were shown the distribution of the Round 1 responses from all experts and asked whether they would like to adjust their WTP guesses, the mean WTP of the predictions made by the Norwegian experts was €74 for Plan A. For the less ambitious Plan B, the mean WTP prediction from the Norwegian experts was €58 and €64; in Round 1 and 2, respectively (Strand et al., 2014).

A Choice Experiment (CE) survey of US and Canadian households was conducted in 2015 to estimate their marginal WTP for avoiding forest and species loss (Siikamaki et al., 2019). The CE had three scenarios with two preservation plans and a status quo option, including five choice

tasks per respondent. The status quo option is comparable to the one described in the Delphi exercise, where 30 percent of the current forested area will be lost, and 24 percent of the species will be threatened by extinction by 2050. The payment vehicle used was an annual federal income tax. Using a mixed logit model, they estimated the yearly mean WTP of North American households to be $4.97 and $3.19 for each percentage point of avoided forest area loss and species loss, respectively.

The main contribution of this paper is to test the reliability of several country groups of the Delphi CV survey by Navrud and Strand (2018) [ii] by comparing these estimates with the outcome of our CV survey of a representative sample of 300 Norwegian households valuing the same Amazon rainforest preservation plans. As an additional methodological contribution, we also compare the results from the population and Delphi CV surveys with a "conventional" international benefit transfer (BT) in terms of unit value transfers from the CE survey in North America of the similar Amazon rainforest preservation scenarios (Siikamaki et al., 2019).

These three studies are directly comparable in terms of the good being valued (i.e., the Amazon rainforest) and the scenarios/environmental changes valued. As SP techniques do not measure households´ actual WTP, this comparison of estimates from different valuation approaches does not test criterion but convergent validity (Bishop & Boyle, 2019). With only three experts in the Norwegian subsample of the Delphi CV survey, we also use expert predictions from other groups of countries which we judge to be close to Norway in terms of income level and environmental attitudes. While there have been several tests of convergent validity of transferred estimates, e.g., Ready et al. (2004) and Lindhjem and Navrud (2008), this is the first study to test the convergent validity of a Delphi CV survey where the valuation scenario in the Delphi survey and the CV survey is the same.

To assess the reliability, we estimate the transfer error (TE) between transferred mean WTP estimates and mean WTP estimates from our Norwegian CV survey, including a Monte Carlo simulation exercise as a sensitivity analysis. Results indicate that the Delphi CV estimates yield lower TEs compared to the conventional international BT exercise of the North American estimates. In most scenarios, the Delphi method provides acceptable TEs for policy use. Thus, relatively cheap and fast Delphi CV expert assessment surveys can potentially provide reliable estimates of non-use values among distant beneficiaries of global public goods. However, the comparison in reliability between the conventional international BT and the Delphi CV exercise should be interpreted with caution because of temporal, spatial, and methodological differences, which could impact and produce TEs. Thus, further comparative studies for other global public goods, contexts, and countries should be performed to see whether and under what conditions this result can be generalized and to what extent Delphi CV surveys should be used.

The paper is structured as follows. First, we provide a description of our survey design and the methods used, including the Delphi Method and BT, followed by a literature review on temporal and spatial dimensions in BT. We then present, compare, and discuss the results of the CV population survey study, the transferred Delphi CV, and the transferred North American CE estimates. Finally, the paper concludes and provides suggestions for future research in this area.

2. Data and methods

As the main contribution of this paper is to compare the outcome of an actual population CV survey with an earlier expert assessment of the outcome of such a CV survey (in terms of a Delphi CV survey), including a conventional BT exercise; we start by describing the methodology used in these three valuation approaches. However, we first would like to address some shortcomings

in the comparison. First, the three SP surveys to be compared have temporal differences. A literature review on the stability of preferences over time is thus presented in Section 3. Second, the North American survey uses the CE method, while the Delphi and the population survey use the CV method with a payment card. The difference in response format between the North American and Delphi surveys might be a source of error. Limitations, including a sensitivity analysis, are further discussed in Section 4.5. Table A.1 in the Appendix compares the three surveys efficiently.

Norwegian population CV survey

After careful piloting and revisions, an internet survey of 300 persons was conducted in April 2018. The respondents were randomly selected from the Norstat survey company internet panel to represent the Norwegian population in terms of age, gender, and education level. The survey was sent to 1451 individuals, which gives a response rate of 20 percent. This might seem low, but note that in this and many internet panel surveys, invitations are sent to members of large panels without follow-up reminders, and the survey is then closed when the number of respondents aimed for is reached.

The CV survey of the Norwegian population was constructed to be as identical as possible to the Delphi CV survey used by Navrud and Strand (2018) to make the CV survey directly comparable. In the CV survey, respondents were first asked questions regarding their preferences for public spending on a range of public services. The questions make the respondents consider their preferences regarding public spending for different public goods, avoid a focus effect on the Amazon rainforest, and train respondents for the WTP elicitation questions, as in Siikamaki et al. (2019). They were also asked if they had ever visited a tropical and Amazon rainforest.

Next, respondents were provided a definition of and information about tropical rainforests worldwide before describing the Amazon rainforest. As in Navrud and Strand (2018), maps were presented to show where the world's tropical rainforests are located and the size of Norway and other European countries relative to the Amazon rainforest.

Respondents were then asked questions to reveal their knowledge about the Amazon rainforest. Further, the two different preservation plans, A and B, were presented. They were informed that if no preservation plan for the Amazon is implemented, 24 percent of existing species and 25 percent of the current forest areas of the Amazon will be lost within 2050. This is defined as the **reference scenario**.

Just like in the Delphi CV survey; respondents were shown a slide depicting mammals in the Amazon facing potential extinction (see Figure 5 in Strand et al. (2014)), as well as maps showing the forested area in the reference scenario and both preservation plans (A and B); see Figures 1, 2, and 3 in Navrud and Strand (2018), respectively. Respondents were informed that the Brazilian government, by collaborating with NGOs, had constructed the two preservation plans, A and B. They were then told that without international funding, the costs of the preservation plans are too high to be implemented. The description was consistent with how funding was described in Siikamaki et al. (2019). **Plan A** is more extensive than Plan B and implies no further forest or species loss within 2050. **Plan B** implies 15 percent forest loss and 7 percent species loss within 2050 compared to current levels. The respondents were also reminded that 15 percent of the original Amazon rainforest has disappeared since the 1970s and will not be recovered by any preservation plans. Thus, even with the most ambitious preservation Plan A, 85 percent (and not 100 percent) of the original Amazon rainforest is preserved (this was stated in the upper right corner of the map shown to respondents; see Figure 2 in Navrud and Strand (2018)).

Respondents were then asked the most, if anything, their household is certainly willing to pay annually for preservation Plans A and B, respectively. They stated their WTP as an amount on a payment card, shown as a horizontal list of amounts from zero to 12,000 NOK (about 860 PPP-€).[iii] The payment vehicle was an extra annual national tax. Respondents were told that the tax payments would be transferred to the eight Amazon rainforest countries that have agreed to implement the preservation plan(s). The choice of payment vehicle is realistic, as recommended by Johnston et al. (2017), because Norway has already set aside money to pay Brazil to reduce deforestation. Additionally, respondents might be less skeptical about a tax earmarked for this specific purpose than a general increase in the income tax (Lindhjem & Navrud, 2009). Respondents reporting positive WTP were asked an open-ended question (i.e., no reply options provided) on *why* they are willing to pay to classify their WTP response according to their motives for paying. This information was used to check which ecosystem services people were paying to protect to avoid double counting when aggregating the value of different ecosystem services of the Amazon rainforest; see Strand et al. (2018).

Respondents stating zero WTP were also asked to select the most important reason for their zero WTP from a pre-specified list of reasons. This question was used to distinguish "true zeros" from "protest zeros": The latter group of respondents have "true" positive WTP but answer zero because they protest some part of the CV scenario. As their answer does not reflect their true WTP, they are excluded from the sample used to calculate the mean WTP (thus, we implicitly assume that the protest zeros have a WTP equal to the mean WTP of this remaining sample of respondents). If zero WTP respondents chose the reply options "Amazonian countries should pay themselves", "The Norwegian government should pay", or "Norway has already paid enough to reduce deforestation in Brazil and other countries", we classified them as protest zero responses[iv]. In

designing the CV survey, we aimed to minimize the number of protest zero respondents by making the CV scenario and the payment vehicle as realistic as possible.

Respondents were then asked: i) if they think the preservation plans will be implemented, ii) if they believe they have to pay the amounts they state, and iii) whether the results from the survey will be used as decision support for policies aiming to reduce deforestation of the Amazon rainforest. These questions were used to test the level of payment and policy consequentiality; and thus assess the truthfulness and reliability of the responses (Johnston et al., 2017). Age, gender, education, and other socioeconomic variables were also collected.

Delphi CV survey

The Delphi method gathers information on a specified subject by surveying experts about their respective opinion (Dalkey & Helmer, 1963). It was initially applied to forecast science and technology by Dalkey and Helmer (1963) and later applied in several different contexts (Sackman, 1974; Hsu & Sandford, 2007).

In valuing environmental goods and ecosystem services, the Delphi method entails asking valuation practitioners/experts how they expect households in a population of interest will value specified changes in a public good (Carson et al., 2013). Experts accumulate experience and valuation information when conducting primary valuation studies, reviewing valuation literature, and conducting BT exercises. Correspondingly, conducting a Delphi CV survey to value changes in environmental goods and ecosystem services could also be viewed as a BT technique, considering that experts utilize accumulated valuation information in a Delphi CV survey when stating their respective opinions on behalf of a population of interest (Navrud & Strand, 2018).

A Delphi survey usually consists of several rounds. In the first round, experts fill in a questionnaire and state their opinion about the specified subject without communicating with other experts. In the later rounds, the experts are shown what the other experts answered (without knowing the identity of the other experts) and are then allowed to revise their answers. Generally, predictions are believed to be more accurate later rounds (Navrud & Strand, 2018). A Delphi CV survey has the potential to provide fast and cheap WTP estimates, but the question is how accurate the method is. This is what we set out to test here by comparing it to a population CV survey. A detailed description of the Delphi CV survey for the Amazon rainforest is found in Navrud and Strand (2018).

One potential issue with the Delphi method in the context of environmental valuation might be the low number of experts with technical and country-specific knowledge and experience. This is of particular concern in low-population countries such as Norway, and in Navrud and Strand (2018), only three Norwegian valuation experts were surveyed. A potential solution to this concern would be to include expert predictions for other countries with similar characteristics. This approach could be implemented as an international unit value transfer with income adjustment based on expert assessment rather than primary valuation survey estimates.

Thus, in addition to the comparison between the mean WTP from the Norwegian expert assessment for Norway in Navrud and Strand (2018) and our population CV survey, we also compare our survey results with three other country groups defined in Navrud and Strand (2018): i) the overall European WTP prediction (including the Norwegian predictions for Europe), where the mean WTP for the preservation plans represent households in Europe as a whole, ii) the mean WTP from country-specific predictions among Nordic and Central European countries (including the Norwegian households predictions), and iii) the mean WTP from country-specific predictions

among Nordic countries (including the Norwegian households predictions). The first category consists of all forty-eight valuation experts in Europe. The Central European countries consist of twenty-nine experts from Austria, Germany, Belgium, Ireland, Netherlands, Switzerland, United Kingdom (including the Nordic countries), and the Nordic country group consists of eleven experts from Denmark, Sweden, Finland, and Norway.

Benefit transfer

The fundamental purpose of BT is to transfer valuation information from previous study sites to a new policy site (Johnston et al., 2021). There are three main BT techniques; i) unit value transfer (i.e., transferring mean WTP/household/year estimates) without or with adjustments for different incomes at the study and policy site; ii) value function transfer (i.e., transferring the WTP function from a policy site in terms of, e.g., WTP as a function of the characteristics of the environmental good valued and characteristics of the respondents), and iii) meta-analysis (i.e., transferring a WTP function estimated as a meta-regression function of data from several previous valuation studies valuing the same type of environmental good; including also the characteristics of the valuation methods in the value function to be used for BT) (Navrud & Bergland, 2004).

Our study will use unit value transfer with income adjustment when estimates are transferred across countries. In the literature, this is referred to as international BT. To evaluate the reliability of the transferred values in BT exercises, one usually calculates the associated Transfer Errors (TE) (Johnston et al., 2021). A TE is defined as the difference between transferred and estimated mean WTP as a percentage of the estimated mean WTP (KristÒFersson & Navrud, 2007). In Equation (2), WTP_{BT} is the estimate derived using BT or Delphi CV, and WTP_{E} is the estimated mean WTP derived from the population CV survey.

$$TE = \frac{|WTP_{BT} - WTP_E|}{WTP_E} \quad [1\}$$

3. Literature review on temporal and spatial dimensions in benefit transfer

There is extensive literature on TEs obtained from convergent validity tests of BT. Brouwer and Spaninks (1999) review studies where TEs are reported, concluding that unit value transfer can result in TEs as large as 56 percent. Rosenberger (2015) summarizes results from 38 studies estimating TEs using BT. Value function transfer, with a median TE of 36 percent, seems to outperform unit value transfer. The unit value function yields a median TE of 45 percent. However, this tendency is also the opposite in many studies (Johnston et al., 2021).

Time is an essential factor in BT. Valuation estimates are often transferred from studies conducted years ago (Brouwer, 2006). Meanwhile, preferences might change. Thus, temporal differences between the execution of the three surveys we compare in this study can potentially impact the size of the estimated TEs.

The literature provides evidence that preferences are stable over time. The difference in WTP estimates tends to be insignificant when similar and comparable SP surveys are conducted over different periods (Skourtos et al., 2010). However, the results are mixed. Downing and Ozuna Jr (1996) evaluate temporal BT reliability by investigating the transfer of values for eight saltwater fishing sites in Texas over three consecutive years. They find a mean error of 50 percent based on 16 observations (Boyle et al., 2010). Carson et al. (1997) evaluate the temporal reliability of estimates from two contingent valuation surveys on WTP to project Prince William Sound, Alaska, from future oil spills drawn from the same population two years apart. They find that the WTP estimates were temporally stable and not statistically different. Brouwer (2006) carried out a CV

study of the health risks with bathing water quality before and after extreme weather conditions associated with the closure of bathing locations for public health reasons. The study finds no statistical difference in mean WTP before and after the extreme event. Richardson and Loomis (2009) conduct a meta-analysis of CV studies used to value threatened, endangered, and rare species to test for changes in WTP estimates over time and use the meta-regression for BT. The meta-regression finds that average-in-sample errors range from 34 to 45 percent. A significant time coefficient in the meta-regression indicates that WTP for the conservation of different species has increased over time, but this depends on model specification and estimation procedure. Skourtos et al. (2010) review twenty CV studies addressing the temporal stability problem in WTP. Based on the review, preferences seem to remain statistically stable for two weeks to five years but not for periods of twenty years. Bliem et al. (2012) assess the temporal stability of WTP for river restoration in Austria using the same CE survey on two samples drawn from the same population one year apart. Their results suggest that preferences and WTP are robust over a short period. Lew and Wallmo (2017) test for temporal stability of preferences for species protection, using two identical CE surveys on the same population 17 months apart. They find that the WTP values from the two surveys are not statistically different.

The spatial dimension is also essential in BT. In addition to temporal differences, our comparison includes surveys conducted in different countries. Thus, parts of the comparison are international BT exercises with unit value transfer with income adjustment, where we use estimates of WTP to protect the Amazon rainforest from one country to predict actual WTP among Norwegian households. However, with international BT, some problems arise. First, valuation estimates must be converted to a common currency. Second, characteristics, cultures, and experiences with the environmental good of interest are likely to vary between the populations

(Ready et al., 2004). As pointed out by Johnston et al. (2021), the literature results are mixed but provide reasonable evidence on the validity and reliability of international transfers.

Ready et al. (2004) estimate and compare WTP for specific health improvements by conducting identical CV surveys simultaneously in five European countries. By considering a standardized respondent, they find that WTP is higher in Spain and Portugal and lower in England, Norway, and the Netherlands. Using unit value transfer with adjustment for income differences, they find a mean TE of 38 percent when transferring from a group of three countries to a fourth country. Johnston and Thomassin (2010) conduct a meta-analysis of WTP for water quality changes with impacts on marine biodiversity and recreational uses, comparing Canadian and Canadian-U.S. value surfaces. They find an absolute mean TE of 74.32 percent using unit value transfer. Bateman et al. (2011) implement controlled multi-site SP experiments in European countries to test and develop guiding principles for BT. Their findings suggest that unit value transfer with income adjustment is preferred when transferring across relatively similar sites, yielding a mean TE between 81 and 137 percent. Hynes et al. (2013) assess the reliability of different international BT techniques for coastal ecosystem services. TEs are found to range from 9.3 to 94 percent. Interestingly, according to their results, adjusting for income in unit value transfer provides similar results as adjusting for cultural differences and income combined. Czajkowski et al. (2017) investigate the reliability of international BT using data from identical and simultaneous CV studies on marine quality in nine European countries. Unit value transfer with income adjustment, yielding a mean TE of 70 percent, performs better than other approaches, such as value function transfer. Artell et al. (2019) find TEs between 46 to 108 percent by comparing a five-country CV survey of water quality in the EU. On average, unit value with income adjustment yields the lowest TEs, ranging from 46 to 60 percent.

4. Results and Discussion

Delphi CV survey results

As all the valuation experts were asked to state WTP in euros using the market exchange rate, we need to convert these amounts to NOK using the exchange rate when the experts were surveyed. We then use the Norwegian consumer price index (CPI) to convert 2012-NOK to 2018-NOK, as the Delphi CV survey was conducted in 2012. The population CV survey was conducted in 2018.[v] For the predictions made by the non-Norwegian experts on behalf of their country we use the country-specific PPP-adjusted exchange rate to NOK in 2012, as well as correcting for income differences between Norway and the countries they represent by using the PPP-adjusted GDP ratio, following best-practices e.g., Hynes et al. (2013). Table 1 reports the Delphi CV study's initial and transferred mean WTP values among the four country groups defined in Section 2.2 for Plans A and B in Round 1 and 2.

It is important to note that only the mean WTP estimates from the Norwegian experts represent the WTP of the Norwegian population. As the experts were shown the results from the other experts (without knowing their names) in Round 2 and were asked whether they would like to keep or adjust their results, we consider Round 2 replies as the best representation of expert opinion (in line with other Delphi exercises, e.g., Carson et al. (2013)). Among the three Norwegian experts, two kept his/her answer, one adjusted WTP upwards, and one adjusted WTP downwards in Round 2 for Plan A. We can see from Table 1 that the annual WTP for Plan A in round 2 ranges from NOK 505 to NOK 639, while the WTP for Plan B ranges from NOK 432 to NOK 549. The expected mean WTP estimates from the Norwegian experts are the highest.

*** INSERT TABLE 1 HERE ***

Benefit transfer results

We also test international BT of WTP estimates from the North American CE survey by Siikamaki et al. (2019) of the same preservation plans for the Amazon rainforest. The representative sample of North American households (i.e., USA and Canada) were, on average, willing to pay $4.97 and $3.19 to avoid one percentage point loss in forest area and number of species, respectively (Siikamaki et al., 2019). Multiplying these marginal WTP estimates with the avoided percentage loss of forest area and species loss for preservation Plans A and B, we obtain estimates of the mean WTP for the respective preservation plans among North American households. Unit value transfer with income adjustment is then applied to determine the mean WTP among Norwegian households for Plans A and B (Ready & Navrud, 2006; Navrud & Ready, 2007). We use a Purchase Power Parity (PPP) adjusted exchange rate to convert US dollars to NOK at the time of the survey (2015)[vi], multiply with the income difference ratio (assuming an income elasticity of WTP equal to one), and then use the Norwegian Consumer Price Index (CPI) to convert from 2015 to 2018-NOK. The mean annual WTP among Norwegian households is NOK 2280 and 1186 for Plan A and B, respectively.

CV population survey results

Table 2 reports the socio-economic characteristics of the 300 respondents in the national sample conducting the CV survey and the corresponding statistics for the Norwegian population. While the sample is representative in terms of gender, age, and distribution in different geographical regions, households with high education and high income are overrepresented.

*** INSERT TABLE 2 HERE ***

Out of the 300 respondents, 12 and 21 percent replied don't know and zero, respectively, when asked about their WTP for Plan A. 44 percent of the zero WTP respondents were identified as protest zero respondents. For Plan B, 9 percent did not know if they were willing to pay anything, and 21 percent stated zero WTP. The mean WTP estimates for Plans A and B excluded 'Don't know' answers and protest zeros; see Table 3. Overall, 220 respondents have positive WTP for Plan A, while 213 respondents have positive WTP for Plan B.

Using the midpoints (between the stated amount and the following and higher amount on the payment card), except for zero (where the "true" zeros (i.e., not protesters) were recorded as zeros), the mean WTP/household/year is NOK 945 and NOK 677 for Plans A and B, respectively. Including defined protest zeroes, these numbers change to NOK 867 and NOK 625. Using the interval unconditional censored means, the mean WTP was NOK 1136 and 796 for Plans A and B, respectively. Again, including defined protest zeroes, these numbers change to NOK 959 and NOK 750. An internal scope test was performed to test whether households´ WTP for the more extensive preservation Plan A is significantly higher than for Plan B. The bootstrapped distribution of the difference in the WTP between Plans A and B was estimated using 1000 replications. Further, we estimated the percentile-t 95 percent confidence interval of the difference (143.43, 432.05).[vii] As zero is not present, we can reject the null hypothesis of equality.[viii] This confirms that the CV survey passes the scope test, i.e., preserving larger forest areas and a higher number of species is valued higher than smaller forest areas and a lower number of species, see also Veisten et al. (2004).

*** INSERT TABLE 3 HERE ***

A sensitivity analysis was performed to evaluate the reliability of the means of WTP for Plans A and B, referred to as baseline estimates. Firstly, observations where respondents stated WTP for Plan B to be greater or equal to WTP for the more extensive preservation Plan A, were removed. The mean WTP for Plan A, estimated from the midpoints, then increased from NOK 945 to 1074. In total, 134 observations were removed. Thus, a substantial part of the sample had diminishing marginal utility or responded inconsistently with economic theory. One reason for inconsistencies could be that households find Plan B to be more realistic than A and thus state their WTP as an expected value in terms of their "true" WTP multiplied with a probability lower than one that Plan A would be implemented. This is supported by the results from a follow-up question, showing that 37 percent of the respondents found Plan B to be "very realistic", while the corresponding number for Plan A was only 15 percent. The diminishing marginal utility of increased preservation explains why several respondents state the same WTP for Plans A and B.

Respondents were asked an open-ended question about their reason for being willing to pay something for Plan A and/or Plan B. The reason for keeping this an open-ended question was to avoid influencing the respondents by providing a list of possible motives for their WTP. Siikamaki et al. (2019) provided such a list. It seemed to remind the respondents about the carbon sequestration benefits, which they might not have considered when stating their choices but maybe felt obliged to include as one of the motives for their WTP. The valuation scenarios focus on avoided biodiversity loss if the forest preservation plans are implemented. The experts in the Delphi CV survey were told explicitly to exclude carbon sequestration benefits when stating the amount they thought the households would be willing to pay. Learning from Siikamaki et al. (2019), we believe just mentioning carbon sequestration in the population CV survey, either in

terms of saying that they should exclude this benefit or list it as a possible reason for their WTP, will make it difficult to get a "clean" test of WTP for the preservation plans between the Delphi CV and the population CV surveys. As less than 10 percent mentioned carbon sequestration benefits as the primary motive for their positive WTP (see table 4), we think our approach in the population CV survey facilitates a "cleaner" comparative test.

Five motivational categories (WTP categories) were identified based on their responses: i) existence value, ii) bequest value, iii) CO2 sequestration (Carbon), iv) social responsibility, and v) don't know. As we only asked one question about why they were willing to pay something for Plan A and/or B, respondents who had positive WTP for Plan A only most likely found it difficult to answer the open question. Thus, several respondents just stated that they prefer Plan A. Therefore, we added a sixth WTP category; vi) Prefer Plan A.

*** INSERT TABLE 4 HERE ***

Existence values are the dominating motive for positive WTP, and non-use values (in terms of existence, bequest, and social responsibility values) make up nearly 2/3 of the positive WTP. Note that less than 10 percent seem to include the carbon sequestration values of the rainforest in their stated positive WTP. To assess whether these respondents have higher mean WTP for Plans A and B than the other respondents with positive WTP, a Welch's t-test of the mean difference between the two independent samples were performed. The mean WTP for Plans A and B among the "Carbon respondents" is NOK 2194 and NOK 1339 for Plans A and B, respectively. However, among the other respondents with positive WTP, the corresponding mean WTP is NOK 922 and NOK 713. Test results confirm a statistically significant difference ($p<0.01$) in mean WTP for

Plans A and B between "Carbon respondents" and other respondents with positive WTP. However, as the "Carbon respondents" make up less than 10 percent of the respondents, the mean WTP estimates should not be influenced much by these respondents. We do, however, test the effect of excluding the "Carbon respondents" to get a "cleaner" comparison of the population sample and Delphi CV survey means (see Section 4.4). We cannot completely rule out that also other respondents included carbon sequestration benefits in their WTP. Still, Table 4 clearly shows that the non-use values of the forest and its biodiversity dominate WTP.

Tables 5 and 6 report the descriptive statistics of the independent variables and results from the interval-censored WTP regression models. The dependent variable is defined as ln(wtp+1) to include zero respondents. Results are as expected, from economic theory and previous CV surveys of forest preservation, which increases the construct and convergent validity. WTP for both preservation plans increases significantly with household income, with an income elasticity of WTP of 0.35 and 0.49 for Plans A and B, respectively. Respondents that have significantly higher WTP are females (*male*), from the capital city of Oslo (*oslo*), using more than 10 minutes to complete the internet surveys (*longtime*) (and thus probably have a greater interest in the topic), stating that environmental conservation is somewhat or very important (*envlist*), and believe we must spend much more or a bit more public money on environmental conservation in South America (*moremoneySA*). Those considering the preservation plans unrealistic (*unrealplans*) have significantly lower WTP.

*** INSERT TABLE 5 HERE ***

*** INSERT TABLE 6 HERE ***

Comparison of CV, Delphi CV, and Benefit Transfer

Table 7 reports the TEs of the Delphi CV survey and the BT exercise from the North American CE study. The results show that the Delphi CV method generally provides lower TEs than the North American BT exercise. Considering Round 2 estimates from the Delphi CV survey, the TE for Plan A ranges from 37 to 56 percent. For Plan B, the TE ranges from 12 to 46 percent. The Norwegian experts seem to generate the lowest TEs for the Round 2 estimates. In most cases, the lowest TE for the Round 2 estimates would be acceptable for policy decisions (Ready & Navrud, 2006; Rosenberger & Loomis, 2017). The unit value transfer with income adjustment from the North American CE yields higher TEs, ranging from 49 to 141 percent.

*** INSERT TABLE 7 HERE ***

In the Delphi CV survey, the experts were asked to neglect the carbon sequestration benefits of preserving the Amazon rainforest. The TEs discussed from the Norwegian part of the Delphi survey from Table 7 do not exclude respondents in the CV survey that included carbon sequestration benefits in their WTP estimates. If we exclude these "carbon respondents", the interval-censored mean WTP for Plans A and B is NOK 959 (€69) and NOK 738 (€53), respectively. The corresponding midpoint means are NOK 834 (€60) and NOK 628 (€45).

Table 7 further shows that the TEs of the Norwegian part of the Delphi CV survey are substantially lower when compared to the population CV survey results excluding the "carbon respondents". Using the midpoint mean estimate in Round 2 for Plan B, the TE is as low as 13 percent. This provides further evidence of the Delphi CV survey (after Round 2) providing low, and in most cases, acceptable TEs for policy decisions.

In addition to providing lower TEs than the North American unit value transfer exercise, the Delphi CV survey also seems to better predict the scope effect in terms of the difference in WTP for the two preservation plans. The difference in mean WTP for Plans A and B is NOK 89 and NOK 1094 in the Delphi CV and unit value transfer exercise, respectively, compared to NOK 268 in the population CV survey. Consistent with our findings, León et al. (2003) found that an expert assessment helped predict households´ relative valuation of national parks in Spain.

Our results confirm the finding of previous studies (Kramer & Mercer, 1997; Horton et al., 2003; Bakhtiari et al., 2018) that distant beneficiaries are willing to pay to preserve a global public good such as the Amazon rainforest. Aggregating the mean WTP per household per year for the most comprehensive preservation Plan A over the total number of households in Norway implies that a transfer of NOK 2.5 billion annually to the Amazonian countries can be justified. Interestingly, from 2008 to 2018, the Norwegian government transferred a total of NOK 8.3 billion to Brazil to reduce deforestation of the Amazon rainforest. Brazil's goal was to reduce its annual deforestation rate by 80 percent by 2020 compared to the average annual deforestation rate from 1996 to 2005. If we interpret this as the implicit willingness-to-pay of members of the Norwegian Parliament (MPs) on behalf of the Norwegian population, it constitutes an annual payment of about NOK 350 per Norwegian household over a 10-year period[ix]. If so, the Norwegian valuation practitioners predict Norwegian households' WTP more accurately than the MPs, even though this was most likely not the goal of the MPs.

Limitations and sensitivity analysis

Following best practices for BT, our study has presented evidence that Delphi CV surveys can provide reliable estimates for non-use values among distant beneficiaries, at least equally reliable

to conventional international BT. However, our comparison has some critical limitations. Thus, our findings should be interpreted with caution. Majorly, and as acknowledged in Sections 2 and 3, our comparison includes temporal, spatial, and methodological differences.

The TEs obtained in this study were, in some cases, more significant than the average TE for international BT of 20 to 40 percent found in Ready and Navrud (2006). It is difficult to identify a specific reason why transferred estimates were not more accurate. Different timing of the surveys could generate TEs through unstable temporal preferences. The Delphi CV survey was conducted in 2012, the North American survey in 2015, and the Norwegian population CV survey was conducted in 2018. However, all surveys were conducted before the significant wildfires in the Amazon rainforest in 2019 and before the political situation in Brazil shifted towards more deforestation for economic development, eliminating the potential shock these events could have on people's preferences. The time difference is more significant for the Delphi survey than for the North American survey. As the TEs are generally lower with the Delphi survey, this favors the Delphi survey if a reliability comparison were to be made. In addition, our literature review indicates that preferences seem to be stable over time, including the time differences between the surveys (Skourtos et al., 2010).

The choice of BT technique when transferring from the North American study and the fact that it is an international transfer could also explain the higher TEs (even though parts of the Delphi exercise are also international transfers). If data had been available, a value function transfer could have allowed for corrections of population differences between the North American and the Norwegian population, potentially reducing the TEs. The SP method in the North American survey could also explain potential differences, as this is a CE. The Delphi and the population CV survey used a payment card, generating a possible difference in incentive compatibility compared to the

North American CE (Carson & Groves, 2011). Research further indicates that payment cards generate a lower WTP than the CE format (Hanley et al., 1998; Ryan & Watson, 2009; Hynes et al., 2011). Correspondingly, the response format could explain why the BT exercise from the North American survey yields, in most situations, higher TEs.

In BT exercises, numerous decisions are made by researchers to predict WTP for the policy site, following best practices. In turn, researchers can conclude the magnitude of the TEs. Even though the TEs derived in Table 7 align with the following best practices in the literature, it can be meaningful to assess under what conditions TEs are minimized. Thus, we conduct a sensitivity analysis, using a Monte Carlo simulation, with four scenarios to compare these to the results in Table 7, which we refer to as the simulated reference scenario. The sensitivity analysis is displayed in Table A.2 in the Appendix. The Monte Carlo simulation uses 5000 random draws from normal distributions with WTP means and standard deviations for the two preservation plans obtained from the three studies. For simplicity, we use the midpoint mean WTP estimates from Table 7 to simulate the reference scenario and Round 2 estimates from the Delphi study.

In our primary analysis (and in the simulated reference scenario), we assumed a scope elasticity of WTP of one for the North American BT exercise, meaning that when the quality or quantity of the environmental good increases by 1 percent, WTP will also increase proportionally (Amiran & Hagen, 2010; Whitehead, 2016; Kling & Phaneuf, 2018; Dugstad et al., 2021). The assumption was pre-determined by the linear specification of the utility functions in Siikamaki et al. (2019) to estimate WTP. However, recent research shows that when imposing more flexible utility specifications in analyzing CE data, scope elasticities can range from 0.13 to 0.58 (Dugstad et al., 2021). From our population CV results, the scope elasticity of WTP for the two preservation plans equals 0.46.[x] Correspondingly, in Scenario 1, we employ a scope elasticity of 0.46 for the

transferred estimates from the North American study. As a result, the unit value transferred WTP from Siikamaki et al. (2019) is NOK 1104 for Plan A and NOK 558 for Plan B (in 2018-NOK). Further, the simulated WTP version for Plan A and Plan B using a scope elasticity of 0.46 is NOK 1140 and NOK 577, respectively. Consequently, in this scenario, the simulated North American estimates yield the lowest TEs, with an absolute TE of 16 percent.

The results in Table 7 exclude identified protest zeroes. In Scenario 2, we include all respondents stating zero WTP. Then, the mean WTP changes from NOK 949 to NOK 867 for Plan A, and from NOK 677 to NOK 625 for Plan B. In the simulation exercise for this scenario, the respective mean WTPs become NOK 855 and NOK 621. Given this scenario, the absolute mean TE becomes lower for the Delphi estimates (28 percent) and higher for the transferred North American estimates (136 percent). The simulated Norwegian part of the Delphi exercise provides the overall lowest TE for Plan A and Plan B (18 percent).

In Scenario 3, we combine Scenario 1 and 2, i.e., use a scope elasticity of 0.46 for the transferred North American estimates and include protest zeroes for the population CV estimates. Consequently, we can see that the simulated North American estimates provide a higher TE for Plan A compared to the case in Scenario 2, but a lower TE for Plan B. The absolute TE in this scenario is 26 percent for the simulated North American estimates compared to 28 percent for the Delphi estimates.

5. Conclusion and Recommendations

The Amazon rainforest provides significant non-use values to distant beneficiaries that should be included in cost-benefit analyses to determine the optimal preservation and provision of ecosystem services from the Amazon. However, as it is prohibitively time-consuming and costly to assess

these global non-use values in SP surveys in all countries worldwide, this study tests the reliability of using international BT and expert assessment to predict distant beneficiaries´ WTP for Amazon Rainforest preservation plans. We compare BT estimates from a North American CE (Siikamaki et al., 2019) and estimates from a European Delphi CV survey (Navrud & Strand, 2018) to the outcome of a population CV survey of 300 Norwegian households.

The North American BT exercise overestimates households' WTP for the Amazon preservation plans and the scope effect in terms of the difference in mean WTP between the two preservation plans. We found that TEs of WTP for the two preservation plans ranged from 49 to 141 percent. In addition, the scope effect was about four times larger than the population CV survey (assuming a scope elasticity of one).

Valuation experts' assessments in Delphi CV surveys seem to be a fast, cheap, and relatively reliable technique for assessing the benefits of providing global public goods to distant beneficiaries. Considering the main results, the Delphi CV survey predicted the outcome of the population CV survey more accurately than the North American BT exercise, with TEs ranging from 12 to 56 percent for Round 2 estimates. However, the respective finding should be interpreted with caution. Temporal, spatial, and methodological differences could all be sources of TEs. Our comparison is not experimentally clean. Further, it is essential to note that the valuation experts consistently underestimated the WTP for both preservation plans and the difference in WTP between them. The average scope effect from the Delphi CV survey of the included country groups is lower than the Norwegian population survey. On the other hand, the average scope effect of the Delphi CV estimates is more accurate compared to the scope effect from the North American study.[xi] On average, the predicted scope effect from the country groups is NOK 110. However, in the sensitivity analysis, we saw that the transferred North American estimates are the most reliable

when imposing a scope elasticity of 0.46. The sensitivity analysis demonstrates the usefulness of using scope elasticities of WTP in BT exercises, as recognized by Kling and Phaneuf (2018). However, more research is needed to determine plausible elasticity estimates for use in specific contexts through, for example, meta-analysis (Kling & Phaneuf, 2018). An unanswered question is whether there is a non-linear scope-effect between each percentage increase of protected species and forest, as assumed in the sensitivity analysis.

The Norwegian experts provided the lowest TEs, indicating (and as expected) that the reliability of Delphi surveys in environmental valuation will increase when experts predict valuation estimates for their own country. However, the reliability of the Norwegian part of the Delphi survey is questionable due to the low number of observations. In general terms, not all countries will have a sizeable number of valuation experts to make meaningful predictions of non-use values of global public goods, which could be a general concern of the Delphi CV method in environmental valuation. Ideally, the number of experts should be high enough to make statistical analyses and tests, perhaps more than 30. However, in countries with few domestic valuation experts, our results suggest that transferring valuation estimates from predictions made by foreign experts for countries with similar characteristics can potentially provide more reliable alternative estimates (in terms of TEs) than using the conventional international unit value transfer with income adjustment. The Delphi CV method will act as a conventional international unit value transfer based on expert predictions.

We would like to end with a general conclusion and recommendations for future research. To conclude, both the Delphi and the North American estimates produce TEs. Thus, the question of which technique is the most reliable is not what we seek to answer in this study. Instead, we aim to assess whether the Delphi method can produce reliable policy site estimates. Based on our

results, the answer to this question is yes, but we recommend using the Delphi method as a complement for conventional BT, as both methods produce TEs. Further comparative studies for other global public goods, contexts, and countries should be performed to see whether and under what conditions these results can be generalized. The validity and reliability of Delphi CV studies should be further explored and tested as a method for valuing public goods when time and/or money do not allow for a new valuation study to be performed. This ideally requires new Delphi CV surveys (and conventional BT exercises) to be as similar as possible to new population CV surveys in terms of the good being valued, payment mechanism, response format, choice of SP method; and the surveys should be conducted as close in time as possible to avoid any potential shocks affecting people's preferences.

Even though the Delphi CV method will yield the most significant savings in time and money when used to value global public goods like tropical rainforests and UNESCO World Heritage sites (Carson et al., 2013), the method could also be used to value local public goods. Thus, future research could explore and compare expert predictions of local public goods with local SP surveys. However, such an approach could be more controversial among the affected population, particularly when valuing goods that often generate urban-rural conflicts. Future research should also evaluate the reliability of the Delphi CV method compared to other BT techniques, such as value function transfer. More applications and testing will increase our experience with the Delphi CV method in environmental valuation and could potentially increase the method's reliability and identify the best areas of application of this time- and cost-saving BT method.

Acknowledgments

We would like to thank Jon Strand (The World Bank, Washington D.C.) for providing input to the design of the survey and an anonymous reviewer for very helpful comments and suggestions that improved the paper significantly. Figure 3 in Strand et al. (2014) was provided by Jeff Vincent (Duke University). Britaldo Soares-Filho (Federal University of Minas Gerais, Belo Horizonte, Brazil) provided the originals for the maps in Figures 1, 2 and 3 in Navrud and Strand (2018). This study follows up on a Delphi Contingent Valuation study initiated by Jon Strand and the World Bank to document the global economic benefits of preserving the Amazon rainforest; see Strand et al. (2017) and Navrud and Strand (2018). All results and opinions expressed in the paper and any remaining errors are the authors' sole responsibility.

References

Amiran, E. Y., & 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.

Andersen, L. E., Granger, C. W., Reis, E. J., Weinhold, D., & Wunder, S. (2002). *The dynamics of deforestation and economic growth in the Brazilian Amazon*. Cambridge University Press.

Artell, J., Ahtiainen, H., & Pouta, E. (2019). Distance decay and regional statistics in international benefit transfer. *Ecological Economics*, *164*, 106383.

Bakhtiari, F., Jacobsen, J. B., Thorsen, B. J., Lundhede, T. H., Strange, N., & Boman, M. (2018). Disentangling distance and country effects on the value of conservation across national borders. *Ecological Economics*, *147*, 11-20.

Bateman, I. J., Brouwer, R., Ferrini, S., Schaafsma, M., Barton, D. N., Dubgaard, A., Hasler, B., Hime, S., Liekens, I., & Navrud, S. (2011). Making benefit transfers work: deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe. *Environmental and Resource Economics*, *50*(3), 365-387.

Bishop, R. C., & Boyle, K. J. (2019). Reliability and validity in nonmarket valuation. *Environmental and Resource Economics*, *72*(2), 559-582.

Bliem, M., Getzner, M., & Rodiga-Laßnig, P. (2012). Temporal stability of individual preferences for river restoration in Austria using a choice experiment. *Journal of environmental management*, *103*, 65-73.

Boyle, K. J., Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). The benefit-transfer challenges. *Annu. Rev. Resour. Econ.*, *2*(1), 161-182.

Brouwer, R. (2006). Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event. *Ecological Economics*, *60*(2), 399-406.

Brouwer, R., Pinto, R., Dugstad, A., & Navrud, S. (2022). The economic value of the Brazilian Amazon rainforest ecosystem services: A meta-analysis of the Brazilian literature. *PloS one*, *17*(5), e0268425.

Brouwer, R., & Spaninks, F. A. (1999). The validity of environmental benefits transfer: further empirical testing. *Environmental and Resource Economics*, *14*(1), 95-117.

Bullock, E. L., Woodcock, C. E., Souza Jr, C., & Olofsson, P. (2020). Satellite-based estimates reveal widespread forest degradation in the Amazon. *Global Change Biology*, *26*(5), 2956-2969.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Carson, R. T., Conaway, M. B., & Navrud, S. (2013). Preliminary valuation of a cultural heritage site of global significance: a Delphi contingent valuation study. In *Handbook on the economics of cultural heritage*. Edward Elgar Publishing.

Carson, R. T., & Groves, T. (2011). Incentive and information properties of preference questions: commentary and extensions. In *The international handbook on non-market environmental valuation*. Edward Elgar Publishing.

Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Rudd, P. A., Smith, V. K., Conaway, M., & Martin, K. (1997). Temporal reliability of estimates from contingent valuation. *Land economics*, 151-163.

Czajkowski, M., Ahtiainen, H., Artell, J., & Meyerhoff, J. (2017). Choosing a functional form for an international benefit transfer: evidence from a nine-country valuation experiment. *Ecological Economics*, *134*, 104-113.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science*, *9*(3), 458-467.

de Oliveira, G., Chen, J. M., Mataveli, G. A., Chaves, M. E., Seixas, H. T., Cardozo, F. d. S., Shimabukuro, Y. E., He, L., Stark, S. C., & dos Santos, C. A. (2020). Rapid recent deforestation incursion in a vulnerable indigenous land in the Brazilian Amazon and fire-driven emissions of fine particulate aerosol pollutants. *Forests*, *11*(8), 829.

Downing, M., & Ozuna Jr, T. (1996). Testing the reliability of the benefit function transfer approach. *Journal of Environmental Economics and Management*, *30*(3), 316-322.

Dugstad, A., Grimsrud, K. M., Kipperberg, G., Lindhjem, H., & Navrud, S. (2021). Scope Elasticity of Willingness to pay in Discrete Choice Experiments. *Environmental and Resource Economics*, *80*(1), 21-57.

Foley, J. A., Asner, G. P., Costa, M. H., Coe, M. T., DeFries, R., Gibbs, H. K., Howard, E. A., Olson, S., Patz, J., & Ramankutty, N. (2007). Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. *Frontiers in Ecology and the Environment*, *5*(1), 25-32.

Hanley, N., MacMillan, D., Wright, R. E., Bullock, C., Simpson, I., Parsisson, D., & Crabtree, B. (1998). Contingent valuation versus choice experiments: estimating the benefits of environmentally sensitive areas in Scotland. *Journal of agricultural economics*, *49*(1), 1-15.

Horton, B., Colarullo, G., Bateman, I. J., & Peres, C. A. (2003). Evaluating non-user willingness to pay for a large-scale conservation programme in Amazonia: a UK/Italian contingent valuation study. *Environmental conservation*, *30*(2), 139-146.

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical assessment, research, and evaluation*, *12*(1), 10.

Hynes, S., Campbell, D., & Howley, P. (2011). *A choice experiment versus a contingent valuation approach to agri-environmental policy valuation* (Working paper no. 0173).

Hynes, S., Norton, D., & Hanley, N. (2013). Adjusting for cultural differences in international benefit transfer. *Environmental and Resource Economics*, *56*(4), 499-519.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., & Scarpa, R. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319-405.

Johnston, R. J., Boyle, K. J., Loureiro, M. L., Navrud, S., & Rolfe, J. (2021). Guidance to enhance the validity and credibility of environmental benefit transfers. *Environmental and Resource Economics*, *79*(3), 575-624.

Johnston, R. J., & Thomassin, P. J. (2010). Willingness to pay for water 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.

Jusys, T. (2016). Fundamental causes and spatial heterogeneity of deforestation in Legal Amazon. *Applied Geography*, *75*, 188-199.

Kehl, T. N., Todt, V., Veronez, M. R., & Cazella, S. C. (2015). *Real time deforestation detection using ANN and Satellite images: The Amazon Rainforest study case*. Springer.

Kling, C. L., & Phaneuf, D. J. (2018). How are scope and adding up relevant for benefits transfer? *Environmental and Resource Economics*, *69*(3), 483-502.

Kommuneprofilen. (2018a). *Befolkning fordelt på kvinner og menn. Antall og andel. 2000 –2018. Landet i alt. (In Norwegian: Population distributed by females and males. Quantity and shares. 2000-2018. National Statistics, for Norway overall).* http://www.kommuneprofilen.no/profil/

Kommuneprofilen. (2018b). *Gjennomsnittlig inntekt, formue, gjeld og skatt per innbygger*. Landet i alt. (In Norwegian: Average income, wealth, debt and tax per inhabitant, National Statistics, for Norway overall).* http://www.kommuneprofilen.no/Profil/Inntekt/DinRegion/innt_hovedposter_region.aspx

Kommuneprofilen. (2018c). *Gjennomsnittlig alder. 2000-2018. Landet i alt. (In Norwegian: Average age 2008-2018. National Statistics, for Norway overall)* http://www.kommuneprofilen.no/Profil/Befolkning/DinRegion/bef_alder_region.aspx

Kramer, R. A., & Mercer, D. E. (1997). Valuing a global environmental good: US residents' willingness to pay to protect tropical rain forests. *Land economics*, 196-210.

KristÒFersson, D., & Navrud, S. (2007). Can use and non-use values be transferred across countries? In *Environmental value transfer: issues and methods* (pp. 207-225). Springer.

León, C. J., Vázquez-Polo, F. J., & González, R. L. (2003). Elicitation of expert opinion in benefit transfer of environmental goods. *Environmental and Resource Economics*, *26*(2), 199-210.

Lew, D. K., & Wallmo, K. (2017). Temporal stability of stated preferences for endangered species protection from choice experiments. *Ecological Economics*, *131*, 87-97.

Lindhjem, H., & Navrud, S. (2008). How reliable are meta-analyses for international benefit transfers? *Ecological Economics*, *66*(2-3), 425-435.

Lindhjem, H., & Navrud, S. (2009). Asking for individual or household willingness to pay for environmental goods? *Environmental and Resource Economics*, *43*(1), 11-29.

Navrud, S., & Bergland, O. (2004). Value transfer and environmental policy. *The international yearbook of environmental and resource economics*, *2005*, 189-217.

Navrud, S., & Ready, R. (2007). Review of methods for value transfer. In *Environmental value transfer: Issues and methods* (pp. 1-10). Springer.

Navrud, S., & Strand, J. (2018). Valuing global ecosystem services: What do European experts say? Applying the Delphi method to contingent valuation of the Amazon rainforest. *Environmental and Resource Economics*, *70*(1), 249-269.

Pascual, U., Muradian, R., Brander, L., Gómez-Baggethun, E., Martín-López, B., Verma, M., Armsworth, P., Christie, M., Cornelissen, H., & Eppink, F. (2010). The economics of valuing ecosystem services and biodiversity. *The economics of ecosystems and biodiversity: Ecological and economic foundations*, 183-256.

Ready, R., & Navrud, S. (2006). International benefit transfer: Methods and validity tests. *Ecological Economics*, *60*(2), 429-434.

Ready, R., Navrud, S., Day, B., Dubourg, R., Machado, F., Mourato, S., Spanninks, F., & Rodriquez, M. X. V. (2004). Benefit transfer in Europe: how reliable are transfers between countries? *Environmental and Resource Economics*, *29*(1), 67-82.

Richardson, L., & 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. (2015). Benefit transfer validity and reliability. In *Benefit transfer of environmental and resource values* (pp. 307-326). Springer.

Rosenberger, R. S., & Loomis, J. B. (2017). Benefit transfer. In *A primer on nonmarket valuation* (pp. 307-326). Springer.

Ryan, M., & Watson, V. (2009). Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. *Health economics*, *18*(4), 389-401.

Sackman, H. (1974). *Delphi assessment: Expert opinion, forecasting, and group process*. The Rand Corporation.

Siikamaki, J. V., Krupnick, A. J., Strand, J., & Vincent, J. (2019). *International willingness to pay for the protection of the Amazon rainforest* (World Bank Policy Research Working Paper, 8775).

Skourtos, M., Kontogianni, A., & Harrison, P. (2010). Reviewing the dynamics of economic values and preferences for ecosystem goods and services. *Biodiversity and conservation*, *19*(10), 2855-2872.

SSB. (2017a). *Befolkningens utdanningsnivå. (In Norwegian: Norwegians´ Education level).* https://www.ssb.no/utdanning/statistikker/utniv

SSB. (2017b). *Familier og husholdninger.(In Norwegian: Families and households)* https://www.ssb.no/familie/

SSB. (2017c). *Folkemengd, etter land, kjønn og alder. Nordiske land. 1. januar 2017. Tal og prosent (SÅ 63). (In Norwegian: Population by country, gender and age. The Nordic countries. January 1st 2017).* https://www.ssb.no/314742/folkemengd-etter-land-kjonn-og-alder.nordiske-land.1.januar-2017.tal-og-prosent-sa-63(1)

SSB. (2017d). *Inntekts- og formuesstatistikk for husholdninger. (In Norwegian: Income and Wealth statistics)* https://www.ssb.no/inntekt-og-forbruk/statistikker/ifhus

Strand, J., Carson, R. T., Navrud, S., Ortiz-Bobea, A., & Vincent, J. (2014). *A" Delphi exercise" as a tool in Amazon rainforest valuation* (World Bank Policy Research Working Paper, 7143).

Strand, J., Carson, R. T., Navrud, S., Ortiz-Bobea, A., & Vincent, J. R. (2017). Using the Delphi method to value protection of the Amazon rainforest. *Ecological Economics*, *131*, 475-484.

Strand, J., Soares-Filho, B., Costa, M. H., Oliveira, U., Ribeiro, S. C., Pires, G. F., Oliveira, A., Rajao, R., May, P., & van der Hoff, R. (2018). Spatially explicit valuation of the Brazilian Amazon forest's ecosystem services. *Nature Sustainability*, *1*(11), 657-664.

Uhl, C. (1987). Factors controlling succession following slash-and-burn agriculture in Amazonia. *The Journal of Ecology*, 377-407.

Veisten, K., Hoen, H. F., Navrud, S., & Strand, J. (2004). Scope insensitivity in contingent valuation of complex environmental amenities. *Journal of environmental management*, *73*(4), 317-331.

Whitehead, J. C. (2016). Plausible responsiveness to scope in contingent valuation. *Ecological Economics*, *128*, 17-22.

Table 1: Delphi CV survey results for Norway. Mean WTP per household (hh) / year (y) for Preservation Plans A and B in Rounds 1 and 2, in euro (€) and Norwegian kroner (NOK).

Country Group **Plan/Round**	**Mean WTP/hh/y from Delphi CV Survey (in 2012€)**	**Mean WTP/hh/y (in 2012-NOK)**	**Mean WTP/hh/y (in 2018-NOK)**
Norway (n=3)			
Plan A/Round 1	€65	NOK 486	NOK 561
Plan A/Round 2	€74	NOK 553	NOK 639
Plan B/Round 1	€58	NOK 436	NOK 503
Plan B/Round 2	€64	NOK 475	NOK 549
Nordic countries (n=11)			
Plan A/Round 1	€41	NOK 440	NOK 508
Plan A/Round 2	€42	NOK 438	NOK 505
Plan B/Round 1	€36	NOK 390	NOK 450
Plan B/Round 2	€36	NOK 375	NOK 432
Central and Northern European countries (n=29)			
Plan A/Round 1	€43	NOK 577	NOK 667
Plan A/Round 2	€39	NOK 517	NOK 596
Plan B/Round 1	€34	NOK 452	NOK 522
Plan B/Round 2	€31	NOK 401	NOK 462
Europe overall (n=48)			
Plan A/Round 1	€32	NOK 602	NOK 695
Plan A/Round 2	€28	NOK 519	NOK 599
Plan B/Round 1	€26	NOK 490	NOK 566
Plan B/Round 2	€21	NOK 396	NOK 457

Table 2: CV survey sample vs Population Characteristics

		Sample	Norwegian Population
Gender			
	Male	50.33%	50.39%
	Female	49.67%	49.61%
Income			
	Mean household gross income	NOK 773 171	NOK 518 313
Education			
	Below upper secondary education (< 11 years)	5%	26.5%
	Upper secondary education (11-13 years	29.33%	37.8%
	Tertiary vocational education	12%	2.8%
	Higher education, short (Bachelor degree)	34%	23.4%
	Higher education, long (Master or PhD degree)	19.66%	9.5%
Age categories	Classification A:		
	15-24	11%	12.7%
	25-49	39.33%	34.4%
	50-64	19%	18.4%
	65-79	30%	12.4%
	≥80	0.67%	4.2%
	Classification B:		
	15-49	50.33%	47.7%
	50 or above	49.67%	52.9%
Geographical regions			
	Mid-Norway	12.33%	8.6%
	Northern Norway	9%	9.3%
	Southern Norway	8.67%	5.7%
	Western Norway	19.33%	26%
	Eastern Norway	50.66%	50.4%

Sources: SSB (2017a, 2017b, 2017c, 2017d), Kommuneprofilen (2018a, 2018c, 2018b)

Table 3: Mean and Median Willingness-to-Pay (WTP) from the Norwegian CV survey calculated from the amount they ticked on the payment card (Payment card value), midway to the next, higher amount on the payment card (Midpoint value) and the estimated interval censored value.

	Mean WTP Plan A	Mean WTP Plan B	95% CI Plan A		95% CI Plan B	
Payment card value	730	525	572	889	413	637
Midpoint value	945	677	746	1145	531	823
Interval censored value	1136	796	994	1279	697	825
	Median WTP Plan A	Median WTP Plan B	95% CI Plan A		95% CI Plan B	
Payment card value	300	200	134	466	89	311
Midpoint Value	550	250	345	755	95	405

Note: the confidence interval for the unconditional interval censored means are obtained by the Delta-method.

Table 4: Percentage of respondents with positive willingness-to-pay (WTP) for Plan A and/or B distributed on their main motive for being willing to pay something for Amazon rainforest preservation

	Main motive for positive WTP						
	Existence	Bequest	Carbon	Social Resp.	Don't know	Prefer Plan A	Total
Percentage	45.45	8.64	9.55	12.73	17.73	5.90	100

Table 5: Description of Independent Variables

Variables	Description	N	Mean	Std.dev	Min	Max
lnhhinc	Natural log of annual gross household income (in NOK), defined as the midpoint of the income range	300	13.386	0.579	11.513	15.202
higheduc	dummy, 1 if bachelor degree or higher	300	0.523	0.500	0	1
male	dummy, 1 if male	300	0.503	0.501	0	1
lnage	Natural log of age	300	3.844	0.413	2.890	4.407
oslo	dummy, 1 if respondent lives in the city of Oslo	300	0.123	0.329	0	1
longtime	dummy, 1 if survey completion time is more than 10 minutes	300	0.300	0.459	0	1
envlist	dummy, 1 if believe environmental conservation is fairly or very important	300	0.703	0.458	0	1
moremoneySA	dummy, 1 if believe we must spend much more or a little more public money on environmental conservation in South America (SA)	300	0.277	0.448	0	1
unrealplans	dummy, 1 if believe none of the preservation plans are realistic	300	0.140	0.348	0	1
envmember	dummy, 1 if member of an environmental organization	300	0.087	0.079	0	1
visitamazon	dummy, 1 if have visited the Amazon rainforest	300	0.070	0.256	0	1
planvisitrain	dummy, 1 if quite or very sure to visit a tropical rainforest	300	0.140	0.348	0	1

Table 6: Willingness-to-Pay (WTP) regression models.

	Interval Regression		OLS Regression	
Variables	Plan A	Plan B	Plan A	Plan B
lnhhinc	0.350*	0.491**	0.348*	0.491**
	(0.195)	(0.224)	(0.200)	(0.230)
higheduc	0.169	0.301	0.170	0.298
	(0.251)	(0.273)	(0.257)	(0.279)
lnage	0.268	0.426	0.266	0.421
	(0.314)	(0.323)	(0.322)	(0.331)
Male	-0.503**	-0.438*	-0.507**	-0.437*
	(0.228)	(0.247)	(0.234)	(0.253)
Oslo	1.075***	0.887***	1.073***	0.884***
	(0.288)	(0.321)	(0.296)	(0.330)
longtime	0.593***	0.767***	0.593**	0.768***
	(0.242)	(0.244)	(0.248)	(0.250)
envlist	0.763***	0.869***	0.756**	0.862***
	(0.308)	(0.311)	(0.316)	(0.319)
moremoneySA	1.097***	0.815***	1.094***	0.813***
	(0.221)	(0.248)	(0.227)	(0.254)
unrealplans	-1.519***	-2.111***	-1.520***	-2.117***
	(0.514)	(0.493)	(0.528)	(0.506)
envmember	0.640	0.487	0.648	0.495
	(0.429)	(0.447)	(0.441)	(0.458)
visitamazon	0.288	0.00885	0.288	0.0134
	(0.650)	(0.630)	(0.666)	(0.646)
planvisitrain	0.412	0.375	0.408	0.372
	(0.281)	(0.347)	(0.288)	(0.355)
Constant	-1.092	-4.026	-1.028	-3.995
	(2.787)	(3.413)	(2.857)	(3.503)
Log Likelihood	-756.977	-749.010	-	-
Observations (N)	242	243	242	243
Adjusted R^2	-	-	0.259	0.266
McFadden's Adj. R^2	0.036	0.038		
AIC	1541.954	1526.02	-	-
BIC	1590.799	1574.922	-	-

Notes: Robust standard errors in parentheses, *** p≤0.01, ** p≤0.05, * p≤0.1.

Table 7. Transfer Error (TE) for the Delphi CV survey and the North American CE survey

Country Group **Plan/Round**	**Midpoint Mean WTP TE**	**Unconditional Censored Mean WTP TE**
Norway (N=3) with "carbon respondents"		
Plan A Round 1	41%	51%
Plan B Round 1	26%	37%
Plan A Round 2	32%	44%
Plan B Round 2	19%	31%
Norway (N=3) without "carbon respondents"		
TE Plan A Round 1	33%	42%
TE Plan B Round 1	20%	32%
TE Plan A Round 1	23%	33%
TE Plan B Round 2	13%	26%
Nordic countries (N=11)		
Plan A Round 1	46%	55%
Plan B Round 1	34%	44%
Plan A Round 2	47%	56%
Plan B Round 2	36%	46%
Central and Northern European countries (N=29)		
Plan A Round 1	30%	41%
Plan B Round 1	23%	34%
Plan A Round 2	37%	48%
Plan B Round 2	32%	42%
Expert prediction for Europe (N=49)		
Plan A Round 1	26%	38%
Plan B Round 1	17%	29%
Plan A Round 2	37%	47%
Plan B Round 2	12%	43%
North American unit value transfer with income adjustment		
Plan A	141.3%	100.7%
Plan B	75.2%	49.0%

Note: "Carbon-Respondents" are those that stated that their WTP for the Amazon rainforest preservation plans was motivated by carbon sequestration benefits; see Table 4.

Endnotes

[i] and carbon storage values can be measured using measures of the social cost of carbon.

[ii] Delphi CV surveys can be classified as a benefit transfer technique, as all benefit transfer techniques depend on experts' opinions and assessment of how previous studies can be used to estimate benefits or costs in new policy contexts (León et al., 2003; Navrud & Strand, 2018; Strand et al., 2017).

[iii] The Purchase Power Parity (PPP) corrected exchange rate between euro and NOK in 2018; 1 € = 13.89.

[iv] We are aware that the reply option "Norway has already paid enough to reduce deforestation in Brazil and other countries" could indicate that respondents have a true marginal zero WTP as they feel they have already paid. However, as we aim to elicit the welfare gain for the two specific preservation plans A and B (provided that they have not already being paid for), these zero WTP answers should be treated as protest zeros and they obviously do not represent the welfare gains of these people.

[v] The Norwegian experts in the Delphi CV were surveyed in April (Round 1) and June (Round 2) 2012, and they were asked to state the amount in euro using the market exchange rate. The average exchange rate for 2012 was 1 euro= 7,47 NOK (https://www.norges-bank.no/Statistikk/Valutakurser/valuta/EUR). The Norwegian Consumer Price Index (CPI) from 2012 to 2018, which increased by 15.4 percent (Statistics Norway), was used to convert these amounts to 2018-NOK, and make them comparable to the population CV survey results.

[vi] The North American survey was conducted in 2015, where 1 USD = NOK 9.933 using the Purchase Power Parity PPP) -adjusted exchange rate at the time of the survey (https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm).

[vii] In comparison with the percentile method, the percentile-t method has asymptotic refinement (Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.)

[viii] In addition, a paired t-test and a non-parametric sign test of two dependent samples were estimated. The null hypothesis of equality was rejected in each scope test.

[ix] For simplicity, this is the annuity at 0 percent discount rate.

[x] The scope elasticity is calculated as follows: $\left(\frac{\text{NOK } 1136 - \text{NOK } 796}{49-22}\right) \cdot \left(\frac{\frac{49}{2}+\frac{22}{2}}{\frac{1136}{2}+\frac{6962}{2}}\right) = 0.46$, where 49 and 22 are the sum of percentage of saved forest and species from Plan A and Plan B, respectively.

[xi] This also holds if we change the scope elasticity to 0.46 for the North American estimates. In addition, the predicted scope effects of the Norwegian part of the Delphi survey are also more accurate than the transferred scope effect form the North American CE.