Guidance for the Design of Quantitative Survey-Based Evaluation

January 2nd, 2012

By Riccardo Scarpa

1   Introduction

Survey-based evaluation aims at surveying respondents from a statistically valid sample of the population of interest in order to infer from the sample statistics the range of likely values in the broader population. The population of reference can be made up of very different elements. It can be credit institutions and borrowers in policies related to credit, it can be farmers in farming policies or the entire group of SMEs in a region for policies target to this economic context. The relevant population from which to sample is defined by the nature of the policy intervention one wants to examine. In what follows I will use the term "agents" to define the elements of the target population. Survey-based quantitative evaluations are frequently used to assess how people/agents feel about the realization of policy outcomes before, during and after a given policy intervention. The term "quantitative" refers to the fact that there is normally a numerical scale behind the quantities that respondents are asked to reveal in the survey. This numerical scale can be as simple as expressing a degree of agreement or disagreement over a scale of ordinal intensities.

In these measurement efforts the common underlying assumption is that the study concerns the evaluation of measurable outcomes attributed to a generic policy intervention, which will be referred here as the "project". The scope and scale of the project may be subject to a great deal of variation. For example, it can be at the local level, such as the city quarter, the municipality or the metropolitan area. It can be at the regional level, such as the province or region or any other administrative unit overseeing the administration of a territory with many urban centres. Finally, in principle, it can be at the national and international level, when involving projects with that scope.

A second underlying assumption throughout this note is that the population of affected agents is sufficiently large for sampling to be carried out as a means to statistically infer the measurable quantities of the target population. Quantitative survey-based evaluation can also investigate quality states by using numerical descriptors. For example, some studies describe numerically the appeal of qualitatively different alternatives by asking people to rank them from least to most favourite and then use rank statistics to numerically describe the appeal of different quality combinations. A simple statistic could be the frequency with which a person of a given age, income and education is likely to prefer state A to state B if given the choice. State A and B differ qualitatively, but the strength of preference of A over B is described numerically by means of a probability. If after the policy intervention the probability of selecting A is much higher than before than we would have a quantitative indicator, although A and B differ only qualitatively.

As in every kind of policy, cohesion policies are designed to affect the life of policy recipients. The projects under study is therefore expected to have measurable consequences, some of which will be intended and some unintended. Quantitative

surveys of the affected population of agents help the policy process by measuring with some yardstick the intensity of these consequences. By measuring them before (ex-ante) and after (ex-post) the project implementation and development one can help quantify change of both intended and unintended consequences. Furthermore, they help documenting change and facilitate comparisons across time and space. For example, knowing how people reacted to air pollution policies in Rome may help predict how people in similar cities (e.g. Athens) might react to similar measures. This practice of using results of studies from one context and transferring them to other contexts is called "benefit transfer" because it was originally used to evaluate policy benefits when insufficient resources (e.g. time and/or money) were available to execute specific new studies to the case at hand.

Some surveys can be run to provide advice in terms of project planning. For example, a policy might be under discussion to reduce noise and other negative effects from traffic through a country town. This can be achieved in different ways. An expensive by-pass can be built to divert through traffic or a cheaper traffic calming scheme can be put in place to reduce these negative effects. Further, the traffic calming scheme might be engineered to achieve different results. For example, it can aim at reducing speed, noise and separation between the two parts of the community at each side of the road. A quantitative survey can be designed to evaluate how people trade off the different solutions (e.g. by-pass vs traffic calming) and even how they trade off the various goals achievable in different measures by the traffic calming (e.g. noise abatement with community separation, or either of these with increased city council tax).

Cohesion policy, especially under the European Regional Development Fund, often involves the provision of medium to large scale public goods, as well as private goods. Public goods once provided at a given level serve large number of beneficiaries, none of whom can be excluded from enjoying them. Some form of public goods can be associated with most of the categories of the EU cohesion policy projects, such as waste management, water treatment, biodiversity protection, air quality enhancement, industrial site rehabilitation and reclaiming of contaminated land, etc. Many of the effects of these projects are very valuable to the community, but they are of difficult quantification because either intangible or with no observable market. Public good's value is generally recognised as difficult to measure, since people have incentives to conceal the true value assigned to non-excludible goods. Some people say it cannot be valued. Others have worked to provide approximations of this value, based on aggregation of individual values derived from either market behaviour (revealed preference data) or from hypothetical statements (stated preference data). Both of these approaches rely on quantitative surveys for data collection. The value of a public good is an important outcome indicators and can be measured by estimates obtained from quantitative surveys on attitudes, perceptions, acceptance and a variety of other objective indicators, including economic indicators of objective well-being, such as willingness to pay for improvements and changes in overall economic wealth in the population (e.g. consumer surplus from policy development). Ideally, at the individual level the correct measure is the maximum willingness to pay an agent shows to have to enjoy the benefits of the project. These measures are used as a yardstick to describe utility changes from states in the presence and absence of the project.

When monetary estimates are not deemed adequate other measures of utility can be used. For example, recent developments in subjective well-being research in economics indicate that the Life Satisfaction Approach can be used as a non-market valuation technique. With this survey technique, which is enjoying a rapid increase in acceptance, people are not asked to express a monetary value to a change in a policy-delivered level of a public good, but more simply they are asked about their subjective level of life satisfaction. From this subjective scores indirect utility functions can be estimated by correlating the scores with other information on the individual (age, household status, education, income, level of provision of the public good etc.). Because income loss negatively correlates with subjective levels of life satisfaction it is possible to find the level of income change (e.g. decrease) that exactly balances the loss of some public good, thereby obtaining an estimate of economic equivalence to the loss of that good. Then, in the aggregate these scores are compared across communities with different levels of public good provisions to identify differences in the distribution of life satisfaction scores.

2 How to design and undertake surveys for use in Cohesion Policy monitoring and evaluation

In general, to be valid and credible, the type of survey to undertake will have to be conducted respecting the standards and the established convention defined by the state of practice in the field. While a state of the art application would be desirable, there is a considerable lag with which state of the art techniques become widely adopted by practitioners. In other words, the transition time between techniques being considered "state of art" and migrating to the "state of practice" may involve years, although the transition speed is increasing. Survey expertise can be found in every survey methodology. For example, in the field of non-market valuation one can find established protocols to conduct contingent valuation, choice experiments and travel cost surveys. Experts can be found for each category and new expertise is produced all the time by further research findings and a vivacious academic debate. In what follows I try to describe the stages that by and large every survey will have to go through. Each stage corresponds to a type of question.

2.1 What to measure in the survey?

The optimal design of a survey instrument for policy evaluation is strongly determined by the specific objectives of the policy or project in question. Agreement on what to measure and why is necessary from the onset to guide survey design. It is important to also define what needs to be obtained from the survey with respect to the evaluation needs of project management. This because different project delivery modes can be available for the same project and people may have clear preferences about the modes of realization as well as the project itself. Perhaps a frequent question in the context of the category of projects funded by the EU Cohesion Policy will refer to changes in economic welfare brought about by the improvements delivered by the policy or project (terms here used interchangeably). Economic welfare is taken here as a succinct measure of well-being expressed in monetary terms. This information will take the form of an estimated distribution of economic values in the population. Since most projects produce both winners and losers, some economic values will be negative. However, a very large variety of other indicators can be delivered by

surveys and can be as useful, if not more so, than estimates of change in welfare in guiding the public debate over the assessment of the project. For example, transport projects can be interested in the travel time savings and indicators of increased safety, such as values of reduced risk of death or severe injury (measured as subjectively perceived). Risk prevention projects can generate substantial increase in peace of mind and hence affect quality of life measures or life satisfaction measures. Ultimately, the indicators subject of survey will have to be defined case by case and will determine much of the survey methodology.

2.2 When to survey?

To ascertain and measure changes in either objective or subjective indicators linked to the implementation of policies the same measurements will need to be surveyed at two moments in time, namely before (ex-ante) and after (ex-post) the project is put in place. Ex-post surveys should be run not immediately after the completion of a project, but when this has been running for a period long enough to be considered at a stable regime. This allows all the main effects to have fully taken place.
Oftentimes the effects of project implementation cannot be completely separated from those due to other changes occurring over time and unrelated to the project under scrutiny. Hence a certain degree of confounding cannot be avoided in ex-ante versus ex-post measurements with respect to causality. This is inevitable and it is the normal condition of social sciences in which scientific experiments cannot be planned.

To establish the effect of a project, at least two surveys need to be undertaken. The first ex-ante is to establish baseline conditions and the other ex-post. However, complex or lengthy projects might benefit from survey data collected during the development of the project subject of analysis (in interim). Interim surveys can be useful to monitor those issues linked to subsequent stages of project development, especially for modular projects. Modular projects are those that can be broken down into self-contained modules. One example is the extension of the trajectory of a road to link other locations, which can be added incrementally; or the improvement of the safety standards of an existing road segment, which can be undertaken with priority given to those road segments for which safety is most valuable. The execution of modules can be prioritised according to some initial criterion, which can then be improved by information collected with interim surveys. Success of early modules of a modular project can also dictate conditions as to the prosecution of the subsequent modules. Interim surveys can be used to collect data to evaluate the indicators of such successes or failures from the relevant population of agents.

Interim surveys are also important when the prediction of negative effects linked to the execution phase of projects is subject to uncertainty. Some projects create such degree of externality (public discomfort) during their development that overall one might think that the external cost produced by the "works", when adequately factored in, may well offset even substantial benefits produced during the useful life of the project. For example, works for road broadening (e.g. addition of a lane) often require a period in which the road is actually either narrowed to allow road works or used in turns by drivers from opposite directions. When these works extend for a long time the cumulative costs can be substantial and are often inadequately quantified in the ex-post phase.

2.3 How to survey?

As mentioned in point 2.1, much of the technical details necessary to define how a survey instruments should be deployed for a given project evaluation depend on the objectives of the survey and the specific use (analysis) that the data will be supporting. Social scientists (economists, psychologists, political scientists and sociologists, amongst others) have developed sophisticated protocols for the deployment of different types of surveys for each of the methodologies. The stages into which the survey process is typically composed of include:

- The identification of a model or paradigm linking the survey statements or data to the intended objectives of the project
- Selection of survey mode (in person, by mail, telephone, web-based)
- Conduction of focus groups for the identification, clarification and communication testing of the measurable constructs
- Conduction of pilot surveys to test the survey instrument
- Sampling and conduction of field survey, possibly in sequential-stages
- Data analysis and reporting and interpreting of survey statistics
- Supply of evidence-based recommendation to policy makers

2.4 Who to survey?

It is clearly important to adequately define the population of agents affected by the project (or of interest for the estimation of the specific indicator of interest) as well as its geographical boundaries. This is dictated in part by the nature of the project to be evaluated. Some projects will have clear geographical boundaries (e.g. flood protection or waste water catchment) other will be less defined geographically (e.g. a road network development or a biodiversity conservation site). So, the jurisdiction of the project defines the target population, which in turns determines the jurisdiction of the survey and, in many instances, it dictates restrictions on the sampling procedure and the survey mode.

3 Survey-based techniques for non-market goods relevant for Cohesion Policy

In the context of policy evaluation one can think of a large variety of indicators the evaluation of which can be based on sample surveys of quantitative scores. The range is simply too large to be handled in a report of this kind. It is possibly best illustrated by a collection of cases studies. On the other hand though, it can be recognised that Cohesion Policy interventions generate substantial public good values. Because of their relevance in this context and because specific survey techniques have been developed for these goods, I illustrate here four non-market valuation methods based on sample surveys.

3.1 Contingent valuation method

This method is by far the most frequently employed stated preference (SP) method in non-market valuation of public goods. It is also the most studied and controversial. The aim of the method is to establish the distribution of economic values for a public good in the population enjoying that very public good. The decision scenario depicted to respondents in the survey is normally set-up as a hypothetical referendum vote.

Hypothetical referenda simulate real political markets, which is how public goods are allocated. In essence a scenario description is provided to respondents. The scenario presents the advantages connected to the policy proposal and it is associated with a raise in tax to be paid to the public agency in charge of providing the public good under study. Then the respondent is asked whether in a referendum of this type s/he would vote in favour or against the proposal at the proposed tax level. The variation of the tax value across respondents allows the analyst to derive the fraction of respondents voting in favour at each given tax amount. These proportions are expected not to increase as tax values go up and allow analysts to compute population estimates of mean and median WTP for the proposed project.

There are literally thousands of documented studies of this sort and there are very many widely accepted suggestions and procedural variants that have been shown to increase validity of the results. In the US the methods, under quite stringent operating conditions, has been deemed adequate to provide estimates for legal litigation over environmental liability. When the respondent believes that the survey information is going to matter for decision-making, the survey is said to be "consequential". Consequential contingent valuation surveys are reputed to be potentially "demand-revealing" mechanisms, which means that respondents have not got systematic incentives to mis-report their preferences. The results, of course, can still be biased because of the hypothetical nature of the choice, perhaps motivated by the wish to "purchase of moral satisfaction" or to comply with what is socially expected behaviour (adherence to social norms). A number of techniques have been developed and proved to reduce such bias. Calibration studies between hypothetical and real choices are also routinely conducted in various social and cultural settings to assess the correct calibration between hypothetical and real data.

The main drawback of the method is that it is quite expensive to implement because large sample sizes are necessary and it only provides estimates of economic value of a change from the status-quo to the proposed ex-post project conditions. Typically the method is not used to show how the population of interest would trade-offs between different levels of provision of the separate attributes provided by the policy. For example, it not used to inform how people would exchange one aspect (e.g. water quality improvements) with another (e.g. protection from invasive species). This because doing so would require either asking more choice question to the same respondent or to use a variety of different scenarios. Having respondents replying to more than one choice task would open the door to strategic behaviour and violate the notion of truthful demand revelation. Proposing different scenarios would increase sample size requirement even further.

3.2 Choice experiment method

When it is necessary to evaluate the values of policy interventions that can be designed to deliver varying amounts of different public goods (for example, varying degrees of noise abatement as well as levels of road safety) then it is more useful to elicit information on economic values that can be related independently to all of these different levels. In other words, rather than being interested, like in continent valuation studies, in the willingness to pay to move from the baseline to the conditions ex-post the policy intervention, the survey can be designed to generate information that can be use to evaluate all possible combinations of project outcomes

(e.g. in terms of noise abatement and road safety). One method that can deliver this is the choice experiment method. In this survey respondents are given choice tasks depicting different alternative scenarios each one with a price tag (tax cost) and asked which one they would prefer in real life. Respondents are typically asked to identify their favourite alternative (although they might also be asked to identify a full ranking or a partial one, perhaps limited to the most and least favourite alternatives in a set). With adequate use of assumptions and experimental design this method can be very cost-effective and very precise in providing policy analysts with a full preference mapping. Such mapping can then be used to identify the most cost effective policy intervention. While this method is not considered demand revealing, the general opinion is that the average respondent is too taken by the choice task details to engage in systematic strategic behaviour to push survey results to her/his own private advantage. The method has progressively been attracting more attention in the policy arena and is very frequently used in transport studies for the estimation of value of travel time savings. Its use in environmental valuation has been increasing in the recent years, including for the determination of increased risk of flooding from climate change, and the consequent insurance market associated with that.

One particular use of choice experiment is as a tool to overcome some statistical problems that typically affect revealed preference data. In many real life situations the alternatives offered by the current market structure suffer from collinearity. In other words, for some dimensions (e.g. speed limits) there is simply too little variation across options. Choice experiments are often used to break away from the limitation of collinearity in real data because in a hypothetical scenario one can introduce levels for attributes that do not exist in real life (e.g. a speed limit of 80km/h where only limits of 50 and 100 exists). Then data analysis can proceed with a merged data set from stated choice and real observed choice.

3.3 Life satisfaction method

The criticisms levelled to both contingent valuation and choice experiments are of too technical a nature to be described in depth here. Such criticisms induced some researchers (Frey et al. 2004) to promote the use of self-reported survey scores on life satisfaction as indicator of utility levels. These scores are collected regularly by many agencies. For example, in the EU (the Euro-Barometer Survey Series) the variable life satisfaction is the categorical response to the following question: "On the whole, are you: very satisfied [4], fairly satisfied [3], not very satisfied [2], or not at all satisfied [1] with the life you lead?". Apart from location of survey, additional respondent variables are collected, such as income, household size, education etc. The method consists in putting in a relationship these scores with indicators of public good supply. For example, some studies have tried to measure the effect of average temperature on life satisfaction in the attempt to predict the economic damage (measured by income equivalence) of future climate change. This approach can be used to evaluate the effect of policy intervention on life satisfaction.

3.4 Travel cost method

The travel cost method, unlike the previous two, is based on surveys of real travel behaviour to outdoor destination sites. The concept is a simple one. Values are derived from the complementarity link between cost of fuel and of transport (inclusive

of cost of travel time) and the "consumption" of visits to outdoor recreation sites (for hunting, fishing, climbing, skiing, etc.). Similar studies have been conducted to value cultural attractions and heritage sites. Travel cost studies can use surveys designed to model single site demand, multiple site demand and site destination choice. Revealed data surveys can also be supplemented with hypothetical choice questions to provide "contingent behaviour" data, in which—for example—respondents are asked how they would change their site choice behaviour if some of the quality attributes of the site changed (e.g. improved). When demand models or site choice models are estimated then all manner of simulations can be supported. Predictions of economic value from site quality change or site closure (e.g. for conservation reasons) can be obtained.

4   Examples of policy applications

In this section we report some information on selected studies organised as required by the terms of reference. Although most of them are from academic publications it is expected that all have given rise to more complete technical reports for the clients or the research funding bodies.

4.1 Environmental Infrastructure

> Scarpa R., Thiene M., Hensher D., (2010) Monitoring choice task attribute attendance in non-market valuation of multiple park management services: does it matter? *Land Economics,* 86 (4), pp.817-839.

Content: the study uses stated choice data to estimated utility functions from visitors of Cortina D'Ampezzo Natural Park and provides guidance as to what infrastructure the park management should invest in.
Study Location: In the Dolomites, Eastern Italian Alps
Method: choice experiments for 5 different categories of visitors (hickers, picnickers, mountain bikers, ferrata users and climbers) were used and design was tailored for 9 different attributes (length of trail network, new thematic itineraries, vertical signs along trails, challenge itineraries, equipped climbs, etc.) via a Bayesian adaptive design.
Indicator: WTP for two policy target levels of changes in each of the 9 attributes.
Policy: local policy on sustainable tourism management in the natural park. As a consequence of this study park managers have now available a full suite of utility function estimates for different groups of visitors and relating to Park infrastructure they directly control. These can be used to prioritize investment by using cost effective techniques.

> Ferreira, S., Gallagher L., (2010) Protest responses and community attitudes toward accepting compensation to host waste disposal infrastructure, *Land Use Policy,* 27:638-652.

Content: it investigates the effectiveness of the policy based on compensating host community to solve NIMBY (Not In My Back Yard) problems in project location. Attitudes held regarding compensation in communities directly impacted upon by final waste disposal infrastructure projects (landfill and incineration) are examined.

Study Location: Ireland

Method: two contingent valuation (CV) scenarios and a question relating to preferences for compensation delivery.

Indicator: WTP to avoid the proposed development, WTA to be compensated.

Policy: this study allowed analysts to separate residents into 'Hardcore' and 'Switcher' protesters. This illustrates a more subtle picture of the motivations behind rejection of locally undesirable facilities (NIMBY syndrome) and potentially paved the way for tailored compensation packages.

Ivehammar P., (2008) Valuing in actual travel time environmental encroachment caused by transport infrastructure, *Transportation Research Part D* 13 455–461.

Content: Most cost-benefit analyses used in transport infrastructure planning do not include the possibly significant cost of encroachment on valuable environments. Encroachment costs are difficult to transfer because of heterogeneity, uniqueness, and the importance of substitutes. The study presents an approach using actual travel time savings as payment vehicle when valuing environmental encroachment caused by transport infrastructure with two applications. Authors emphasize issues with stated preference methods in that the choice of hypothetical payment vehicle can affect the amount of stated willingness to pay, so they use two approaches to see if they converge.

Study Location: Sweden

Method: Contingent Valuation

Indicator: a) travel time savings linked to an option; b) willingness to pay for avoiding encroachment

Policy: The Swedish road administration (SRA) uses cost-benefit analysis (CBA) as a basis for decision making in road planning. When a new road and its traffic cause environmental encroachment, these effects are not currently part of these CBAs. The approach with valuing an encroachment caused by transport infrastructure in actual travel time has been tested in two studies of planned roads in cooperation with SRA and the concerned local authorities.

4.2 Waste and water supply

Willis, K, Scarpa R., Acutt M., (2005), Assessing water company customer preferences and willingness to pay for service improvements: A stated choice analysis. *Water Resources Research*, 41, p. W02019.

Content: the study estimates WTP for 14 factor services in the water supply industry and involved surveying in person both resident households and firms.

Study Location: West and North Yorkshire counties Northern England.

Method: Choice experiments for a mixture of public and private goods.

Indicator: willingness to pay for attributes related to water service factors as defined by Yorkshire Waters.

Policy: The results from this valuation study were integrated in a larger infrastructure investment study for network maintenance planning driven by customer's preferences. It motivated water tariff negotiation between the water service provider (YW) and the central government regulator (OFWAT). Following the success of the methodology OFWAT now recommends water companies to run similar studies in all areas of England.

Aadland D., Caplan A.J. (2006) Curbside Recycling: Waste Resource or Waste of Resources? Curbside Recycling: Waste Resource or Waste of Resources? *Journal of Policy Analysis and Management*, Vol. 25, No. 4, 855–874

Content: the study addresses the often contentious debate over state and local recycling policy by carefully estimating the social net benefit of curbside recycling. They claim it is a first attempt at establishing a sound economic basis for making such public policy decisions by estimating *both* the benefits and costs of curbside recycling for a wide range of communities. Benefits are estimated using household survey data from over 4,000 households across U.S.
Study Location: 40 western US cities
Method: SP (CVM) and RP
Indicator: they estimate the magnitude of the potential hypothetical bias in the WTP data by contrasting SP information (from CVM) with revealed-preference information (from actual decisions made by households in communities with voluntary CRPs).
Policy: they provide a comprehensive measure of the social net benefit of curbside recycling, in order to help answer the often contentious question: Should we be recycling? Surprisingly, several curbside recycling programs appear to be inefficient. They find that the estimated mean social net benefit of curbside recycling is almost exactly zero. On a city-by-city basis, however, the social net-benefit analysis often makes clear predictions about whether a curbside recycling program is an efficient use of resources.

4.3 Risk prevention

Rizzi L.J., De Dios Ortúzar J. (2006) Estimating the Willingness-to-Pay for Road Safety Improvements, *Transport Reviews*, Vol. 26, No. 4, 471–485.

Content: Reviews the methods of valuing risk from road-related accidents inducing death or sever injury. Reports the analysis of data from a binary stated choice study built on respondent's travel habits. Respondents had to choose between two routes for a return-trip (both are similar to the current Route 68 Santiago–Valparaíso), considering the following three factors: the toll, the travel time en route and the number of fatal crashes on each route. The latter is defined as the number of accidents per year in which at least one person travelling by car dies.
Study Location: Chile
Method: Preference-based estimation of value of risk reduction of a fatal accident
Indicator: Economic value of risk reduction of death and of severe injury from subjective probabilities of exposure
Policy: Chilean social benefit cost analysis for road safety

Shaw W. D. , Walker, Benson M. (2005) Treating and Drinking Well Water in the Presence of Health Risks from Arsenic Contamination: Results from a U.S. Hot Spot, *Risk Analysis*, Vol. 25, 6.

Content: they examined exposure to arsenic through tap water and the factors related to the household's choice to treat and drink water from private domestic wells in a rural area of the United States. The rural area is a hot spot for arsenic. A

significant proportion of households on private wells are consuming drinking water with arsenic levels that pose a health risk. The decision to treat tap water for those on private wells in this area is modeled, and the predicted probability of treatment is used to help explain drinking water consumption.

Study Location: Churchill County, Nevada US

Method: random utility models framework to investigate the decision to treat tap water or not.

Indicator: predicted probability of treatment is used to represent risk in a random utility model of drinking water behavior.

Policy: The Safe DrinkingWater Act of 1974 regulates water quality in public drinking water supply systems but does not pertain to private domestic wells, often found in rural areas throughout the country. The recent decision to tighten the drinking water standard for arsenic from 50 ppb to 10 ppb may therefore affect some households in rural communities, but may not directly reduce health risks for those on private wells. Results provide info about behaviors relating to the household's perception of risk.

Zhai GF, Ikeda S (2006) Flood risk acceptability and economic value of evacuation. *Risk Analysis* 26:683–694

Content: the study investigates the economic value of evacuation and its relationship with flood risk acceptability. That is, the extent to which people expect the occurrence of floods, in terms of scale and frequency. Shortages of information and food were found as the greatest inconveniences. Evacuation inconvenience can be regarded as an important factor causing the low rate of evacuation in Japan. The WTP for avoiding current inconvenience was approximately half of the estimated economic value of evacuation, implying that the current budget for evacuation is too small and should be increased to improve the conditions of evacuation sites.

Study Location: Japan, three flood disasters were considered.

Method: CVM

Indicator: WTP for avoiding evacuation inconvenience because of potential for certain losses as a result of evacuation.

Policy: the economic value of evacuation can be taken into consideration in the risk assessment process in terms of policy in order to evaluate the efficiency of risk reduction measures.

Luechinger S., Raschky P.A. (2009) Valuing flood disasters using the life satisfaction approach, *Journal of Public Economics* 93 620–633.

Content: The authors propose life satisfaction data to value natural disasters, by discussing the strengths compared to traditional methods (SP and RP) and by monetizing utility losses caused by floods in many European countries. Using combined cross-section and time-series data, they find a negative and significant impact of floods on life satisfaction. The estimates are comparable to price discounts found in housing markets.

Study Location: floods in 16 European countries between 1973 and 1998.

Method: life satisfaction approach, based on reported subjective well-being is used as an empirically adequate and valid approximation for individually experienced welfare or utility

Indicator: compensating surplus measures for prevention of events.
Policy: a) identify and quantify the risk towards flood at population level by means of alternative approaches than RP and SP; b) effects of different risk transfer mechanisms (i.e. mandatory insurance) are investigated. Risk-transfer mechanisms are found to have large mitigating effects that come close to fully compensating the impacts of a flood event. Consequently, the effects of flood disasters in regions without such a mechanism are larger than the effect in all regions and the effect in regions with such a mechanism is small.

4.4 Rehabilitation

> Jenkins R., Kopits E., Simpson D., (2006) Measuring the Social Benefits of EPA Land Cleanup and Reuse Programs, *Working Paper* # 06-03 September, 2006.

Content: The EPA administers many cleanup and reuse programs: i) Superfund Program addressing sites posing imminent danger and many of the most hazardous sites nationwide; ii) Brownfields Program focusing on lower risk sites. Although there is agreement about generated set of primary social benefits (reductions in health risks and ecosystem damages, and improvements in amenity values), information about the magnitude of benefits is sparse and there is a general lack of clarity about the categories of indirect benefits associated with them. They identify five different, sometimes overlapping, categories of welfare changes due to land cleanup and reuse.
Study Location: USA
Method: meta-analysis of literature and case studies.
Indicator: categories of welfare changes applied to contaminated land generally: a) primary effects on utility including health, ecosystem and amenity effects; b) indirect effects on productivity; c) depressed property transactions. Categories of welfare changes applied to contaminated land in urban settings: d) potential greenfield-saving effects of redevelopment; e) agglomeration effects.
Policy: Increasingly, public institutions and policy makers seek benefits information as policies, budgets, and legislation are developed. In response to this need, over the past couple of decades, methods have been developed to assess the social benefits and costs of most of EPA's major programs, with an important exception. Programs targeting land cleanup and reuse are still in need of accessible methodologies to assess their welfare effects.

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

Content: the authors focus on the risks of dying associated with exposure to contaminants at hazardous waste sites. Through conjoint choice approach people's preferences for delayed and permanent risk reductions are studied. In particular, preferences as those delivered by permanent remediation, as opposed to those delivered by contaminant containment mechanisms, such as capping and fencing a site, or land use restrictions are investigated. A sample selected to be representative of the residents of four cities in Italy with significant contaminated site problems was considered. Results indicate VSL to be €5.6 million for an immediate risk reduction and €1.26 million if the risk reduction takes place 20 years from now.

Area: four major cities in Italy (Naples, Venice, Milan, Bari).
Method: conjoint choice questions
Indicator: WTP for each unit of mortality risk reduction; WTP for risk reductions that continue to
occur for longer periods of time; the effect on WTP of delaying the beginning of the mortality risk reductions.
Policy: if the goal is to set cleanup standards or decide upon public programs addressing hazardous waste sites, it is useful to have information on the (monetized) value of permanent reductions in the risks to human health in order to compare it with the costs of treating contaminated soil, groundwater and surface water. To this extent it is necessary to know how much the beneficiaries of these risk reductions are willing to pay to obtain them.

4.5 Sustainable Urban Transport

Garrod G.D., Riccardo Scarpa and K.G. Willis (2002) Estimating benefits of traffic calming on through routes: a choice experiment approach. *Journal of Transport Economics and Policy* 36, 211-231.

Content: the study reports a survey of preferences from a village crossed by a trunk road and estimates WTP for traffic calming of through traffic.
Study Location: Northeast of England
Method: using stated choice from households residents in villages with trunk road the authors use random utility models to derive the value of speed reduction, noise abatement (with real noise level experience at the curbside), community severance and aesthetic effects associated with traffic calming solutions.
Indicator: marginal willingness to pay from residents to achieve targets.
Policy: Scottish office and road agencies local policy on nuisance from trunk roads

David Gaker, David Vautin, Akshay Vij and Joan L Walker, 2011, *The power and value of green in promoting sustainable transport behavior.* Environmental Research Letters, 6, doi:10.1088/1748-9326/6/3/034010

Content: the aim is to provide insight into whether, and to what extent, presenting environmental attributes of transport alternatives influences individual transport decisions. Three experiments in which subjects (University of California at Berkeley undergraduates) were presented with hypothetical scenarios of transport decisions, including auto purchase choice, mode choice, and route choice, were conducted. Decisions were modeled to determine how they value reducing their emissions relative to other attributes. Results show that subjects are willing to adjust their behavior to reduce emissions, exhibiting an average willingness to pay for emissions reduction, or value of green (VoG), of 15 cents per pound of CO_2 saved. Despite concern that people cannot meaningfully process quantities of CO_2, evidence to the contrary was found: the estimated VoG was consistent across context (the wide range of transport decisions presented) and presentation (e.g., whether the information was presented in tons or pounds, or whether a social reference point of the emissions of an average person was provided).
Area: UC Berkeley (US)
Method: lab experiments based on hypothetical scenario, choice experiments

Indicator: WTP estimate of the value of green (VoG), Value of time (VoT), value of the pound of CO2.

Policy: One aspect that has become popular in terms of externality is the damage associated with greenhouse gas emissions. In transport behavior, information regarding environmental impact has begun appearing on numerous websites. While it is increasingly popular to broadcast such information, little is known regarding the effects that this information has on human behavior. The key issue with understanding any sort of behavior is in knowing how the decision maker values and reacts to the information that is available.

Mourato M., Saynor B., Hart D., (2004) Greening London's black cabs: a study of driver's preferences for fuel cell taxis, *Energy Policy* 32, 685–695.

Content: this study investigates the preferences of London taxi drivers for driving emissions-free hydrogen fuel cell taxis, both in the short term as part of a pilot project, and in the longer term if production line fuel cell taxis become available. The results show that willingness to pay to participate in a pilot project seems to be driven mostly by drivers' expectation of personal financial gains. In contrast, however, environmental considerations are found to affect taxi drivers' longer-term vehicle purchasing decisions. The results also reveal that driving hydrogen-fuelled vehicles does not seem to raise safety concerns amongst taxi drivers.

Area: London taxi drivers.

Method: Contingent valuation (payment ladder)

Indicator: Willingness to pay for a pilot fuel cell taxi, Willingness to pay for a production fuel cell taxi (taxi already in the market)

Policy: London is considered to be a hotspot of pollution. Since road transport accounts for about a quarter of all carbon emissions in the UK, highlighting the need for low carbon alternatives to current fuels and vehicles. Running on hydrogen and virtually emissions-free, fuel cell vehicles are considered to be one of the most promising ways of reducing transport-related emissions. Understanding the user benefits of fuel cell vehicles and the determinants of demand is essential for a strategic and successful implementation of environmental policy, supporting a market penetration. Median WTP for Fuel Cell taxi GBP1500.

5 Methods, costs and utilization in policy design/assessment

5.1 Contingent valuation

In general, referendum contingent valuation studies for the quantitative estimation of economic benefits from public goods are quite expensive. This because each respondent only provides a "yes" or "no" answer to a WTP question, and as a result larger sample sizes are required to achieve a given accuracy of the estimate. The advantage of this method though is that it is thought to be the only one that is "incentive compatible". This means that respondents can be put in the condition of believing that revealing their true WTP is the best response strategy from the private utility perspective. Contingent valuation is typically preferred as a method when the public policy under investigation is well-defined in terms of its outcome and there is no interest in estimating the WTP for specific components of the policy because the policy is planned to be delivered as an "all-or-nothing" intervention. The referendum

format typically requires a sample size in excess of 600 respondents, but it can be larger when the population of beneficiaries from the project is heterogeneous. Other CVM value elicitation formats can be used though, such as a direct expression of maximum WTP (open-ended format) or the identification of the maximum WTP from within a ladder of values and combined with statements of uncertainty. These methods are less demanding in terms of sample sizes. The cost of a contingent valuation study depends on many factors (form of interview administration, quality of survey instrument, validation exercise, quality of data analysis, quality of the consultants involved, time available for administration, reliability etc.) and hence it ranges over a broad interval. In the past there have been environmental litigation studies that relied on CVM studies that cost millions of dollars. However, with the advent and growth of on-line surveying the method is bound to become less expensive. Reliable CVM surveys for small scale public goods can now be performed at a cost of ten euro per respondent plus a fixed cost of consultancy fees of about fifteen to thirty thousand Euros. The method is still the most commonly employed for public goods valuation. It is therefore deemed to be very feasible in the broad category of projects and policy interventions that produce public goods under the European Regional Development Fund.

5.2 Choice experiments

In a way, this method is closely related in terms of its utilization to the CVM method. Because repeated choice tasks are used, the statistical accuracy of the value estimate can be achieved by means of a smaller sample size. On the other hand though, various independent dimensions of a public policy are evaluated and hence this implies an increase in the sample size. The general outcome in terms of necessary sample size is a mixture between the higher and lower demand on data and it depends on the application. In practice, though, CE studies tend to be less demanding in terms of sample size, but relatively more expensive in terms of the complexity of their survey execution. The use of this method for the valuation of multiple dimensions of public projects has been growing and the prospect for its applicability under the European Regional Development Fund is good. Growing expertise had been accumulating especially with applications to health delivery, environmental benefits and quality of integrated water services (supply and waste disposal). Automated on-line surveys have the potential to further decrease the cost of this form of data collection.

5.3 Life satisfaction

Life satisfaction surveys are generally inexpensive and can be conducted in a short time. The challenge of course is to put in relation the scores with the specific policy changes put in place. Control surveys are needed to identify causal shifts in satisfaction scores, which might increase the overall cost of the exercise. Despite this the overall cost should, in the majority of the cases, still be lower than a CVM or CE survey. Altogether this method is deemed to be obviously feasible, but questions remain as to whether it can ever completely substitute direct and indirect stated preference valuation methods, such as CVM and CE.

5.4 Travel cost method

This is the method of preference for public goods that need to be visited in order to be enjoyed. Cultural events, heritage sites, areas of outstanding natural beauty, attractive landscapes etc. Can all generate a demand for visitation and such demand can be put in relation to specific policies. For example, improved transportation or sanitation or health care can all improve the relative attractiveness of a visitation site. The method is certainly feasible and the cost of collecting data to support a valuation study will depend on the visitation catchment and on the period under examination. Sites that enjoy only seasonal demand can have a very short period of visitation and hence surveys of visitors can be conducted rapidly. Other sites might be available all year round and might be subject to seasonal variation. Survey periods for these sites are longer and might be more complex. On site sampling of visitors is typically more efficient and hence less expensive. General population sampling can be expensive and it normally is made more effective by splitting surveys into separate segments. For example, a general population survey defining the rate of visitors linked to a second survey dedicated to the sub-population of visitors. This method is clearly more restricted in its scope for applications under the European Regional Development Fund. For travel cost method surveys related to specific outdoor activities, such as hunting, fishing and skiing survey data can be captured by asking respondents to keep visitation diaries for the specific outdoor activities they undertake.

6 Conclusions

This document as been developed to provide some guidance on the applicability of the various methods of quantitative surveys, especially with an eye to the economic cohesion policies funded under the European Regional Development Fund after 2013. Quantitative surveys offer great potential to inform the practice of policy development and design as well as policy outcome evaluation. They also offer the potential for a systematic form of data collection that can be used to compare results across geographical areas and across time.

7 List of abbreviations

CBA= cost-benefit analysis

CRP = curb-side recycle policy

CVM = contingent valuation method

EU = European Union

GBP = Great Britain Pound

LSA = Life Satisfaction Approach

NIMBY = Not in my back yard

OFWAT = The economic regulator of the water and sewerage industry in England and Wales

SME = small and medium enterprises

SP = stated preference

SRA = Swedish Road Administration

VoG = value of green

VoT = value of time

VSL = value of a statistical life

WTA = willingness to accept

WTP = willingness to pay

YW = Yorkshire Water