Representativity and inclusivity of Participatory Value Evaluation

A case study on relaxation of COVID-19 measures in the Netherlands

Master thesis by Selma van Delft

TU Delft

Frontpage image retrieved from (Mangelaars, 2019).

This page is intentionally left blank.

Representativity and inclusivity of Participatory Value Evaluation

A case study of relaxation of COVID-19 measures in the Netherlands

Master thesis submitted to Delft University of Technology in partial fulfilment of the requirements for the degree of

Master of Science
in Engineering and Policy Analysis
Faculty of Technology, Policy and Management

By
Selma van Delft
Student number: 4482700

To defended in public on March 16, 2021

Graduation committee

Chair	Dr. Ir. M. Kroesen, Engineering Systems and Services
First supervisor	Dr. Mr. N. Mouter, Engineering Systems and Services
Second supervisor	Dr. H.G. van der Voort, Multi-Actor Systems

Preface

This thesis is the final deliverable I submit to fulfil the graduation requirements for the master Engineering and Policy Analysis at Delft University of Technology. For the last seven months I have been working on the topic of representativity and inclusivity of Participatory Value Evaluation (PVE). I am happy I have received the chance to work on this topic, as it combined many of my interests.

I would like to use this opportunity to thank those people who helped me during this last phase of my master. Firstly, I would like to thank Niek Mouter for his enthusiasm about PVE. I really appreciated to be involved in the PVE community and all the activities such as the pub quiz and the colloquium. And of course, I want to thank Niek for all his valuable feedback. I would also like to thank Maarten Kroesen for the many meetings we had, his critical feedback really helped me to sharpen my research. Next, I would like to thank Haiko van der Voort for his refreshing perspective on the subject. Last but not least I want to thank my family and friends who have been there to support and motivate me. Special thanks to Cecile, Lisa, Roy, and Nick.

Delft, March 2021

Selma van Delft

Summary

To be able to solve the increasingly complex problems, governments need to collaborate with various parties. This way of policy development is called interactive policy making. Citizen participation is a specific form of interactive policy making and focusses on involving citizens in the policy making process. A relatively new online participation method is Participatory Value Evaluation (PVE). It aims to facilitate mass citizen participation and to be inclusive. Everyone who wants to participate in the process should get the change to do so. At the same time PVE is an evaluation method for public policy. To be able to provide an accurate advice about the preferences of citizens regarding a set of policy options, the sample under study needs to be representative. Representativity and inclusivity seem to conflict with each other as making the research available for the public very often leads to a biased sample. The aim of this study is to examine the consequences of the trade-off between representativity and inclusivity for the usability of PVE in the policy context. Therefore, the main research question of this study is:

"What are the consequences of the trade-off between representativity and inclusivity for the usability of Participatory Value Evaluation?"

A case study on relaxation of COVID-19 measures in the Netherlands was chosen as case study to answer this question. This PVE has two available datasets: a panel dataset with the responses from a representative panel and an open dataset with the responses from citizens that selected themselves to participate in the research.

Literature review showed that PVE can be used to achieve several participation goals. PVE provides citizens with information about a policy dilemma the government faces, including serval options choose from and their effects. PVE is a suitable approach to inform and educate the public, as well as fostering trust in institutions as it makes the decision-making process more transparent. An open PVE is more likely to reach these goals then a panel PVE since an open PVE has the potential to reach an unlimited amount of people, whereas the panel PVE can only be filled out by members of a panel. A panel PVE is more adequate in reaching the goals of identifying the preferences of the public and to increase the substantive quality of a decision as this sample is representative for the larger population. PVE is not a proper method to reduce conflict among stakeholders as it does not allow for direct interaction among (opposing) stakeholders. It is recommended that policy makers determine the most important participation goal per situation and based on this the choice for a panel or open PVE.

Analysing the datasets of the COVID-19 PVE experiment showed that the top three selected and the top three rejected measures were the same for both datasets. However, there were small differences in the ranking of the measures and the share of participants that supported or resisted a measure differed quite a lot. It was analysed to what extend bias in demographic variables causes these differences. It became clear that both datasets were only representative for gender. The distributions of age, education level and province in both samples statistically deviated from the population distributions. Yet, the bias for the open PVE was much larger than for the panel PVE. Looking into the effect of the demographic variables on participants' preferences showed that the sampling method influenced the relations that were found. The open PVE showed more significant relations. Yet, both the Cramer's V tests, and the binominal regression models showed that the demographic variables could only explain the participants' preferences in a very limited extend. Reweighting the panel sample did not result in different outcomes, which means that the panel provided an accurate picture of participants' preferences. Reweighting the open sample resulted in a low weighting efficiency, which indicated that the bias in the open dataset was too big to properly correct. Further research lies investigating the threshold to determine when the deviation from the population distribution becomes too large. Moreover, the weights were highly dispersed, and in combination with the low association with the dependent variables this leads to unstable estimates. It was not possible to achieve

representativity for the open dataset. For the sake of comparison, the results for correcting for representation biased in the open sample were reported and showed that the gap between the share of participants that selected or rejected a measure in the panel and in the open PVE was bridged. Yet, the correction could not account for the whole gap. This means there are differences between the respondents from the panel and the open PVE that could not be explained by the demographic variables. From a practical perspective the differences are less important, as the top three favoured measures are the same for both samples and are therefore likely to lead to the same policy decision.

Interviews with researchers showed that in the policy context a representative sample is chosen over an inclusive sample. Representativity is important to make adequate statements about the target group and is therefore seen as a basic requirement. A lack of representative results in a decrease in usability of research results. Moreover, it is seized opportunistically in the policy debate. Lastly is became clear that researchers were reluctant to use weight adjustment as method to correct for representation bias. They mention that it is a difficult and complex process that is only justifiable under specific conditions. It is recommended to examine the trade-off between inclusivity and representativity in more detail by more researchers as well as policy makers. In addition, various options should be explored to present the issue of representation bias and weight adjustment in a more accessible way.

Content

List of Tables

List of Figures

Abbreviations

CBS	Statistics Netherlands
IPF	Iterative Proportional Fitting
RIVM	National Institute for Public Health and the Environment
PVE	Participatory Value Evaluation
SCBA	Social Cost Benefit Analysis
WTP	Willingness To Pay
WTAPB	Willingness To Allocate Public Budget experiments

Chapter 1 – Introduction

1.1 Citizen participation

It is impossible to imagine Dutch society without citizen participation, a process which focusses on the involvement of citizens in public decision-making. It stems from the belief that citizens should be allowed to participate more directly in decisions that affect them (Burton, 2009).

Due to social processes such as globalization, regionalisation and the increased influence of information technology there has been a shift from 'government' to 'governance (Hajer et al., 2004). Where the government previously determined policy, policy development now involves collaboration with public, private, and social organisations. This collaboration aids in finding suitable solutions for increasingly complex problems. Governments simply do not have all the required knowledge and skills, and need the expertise of other parties (Edelenbos et al., 2006). Interactive policy making is a way of policy development whereby government bodies collaborate with other parties such as other authorities, companies, interest groups or citizens to develop and implement policies, with the purpose of making the policy process more effective (Driessen et al., 2001; Edelenbos et al., 2006). Citizen participation specifically focusses on the involvement of citizens in this process.

The interest in citizen participation has grown due to increased dissatisfaction with traditional mechanisms of political representation (Ianniello et al., 2019). There is doubt about whether the system of representative democracy as we know in the Netherlands is able to translate citizens' preferences into policy. In the Netherlands, citizens vote every 4 years for a political party based on its entire party program. To what extend this programme actually reflects the preferences of a voter in case of a specific issue is questionable. The is especially the case for topics that are not at all included in the party programme. Furthermore, the preferences of a voter can change over the years. Citizen participation provides the opportunity to translate the preferences and interest of citizens into policy in a more direct way and thereby increases the legitimacy of government decisions (Edelenbos et al., 2001).

Other benefits of citizen participation lie in the educative function for both citizens and policy makers. On the one hand, citizens can learn about the issue and the considerations behind polies that initially would not be popular in the public (Irvin & Stansbury, 2004). As a result, citizens may be more inclined to accept and support policies (Raad voor het Openbaar Bestuur, 2005). Policy makers, on the other hand, can gain insights into the preferences and interests of citizens. *"A policy that is well grounded in citizen preferences might be implemented in a smoother, less costly fashion because the public is more cooperative when the policy is implemented"* (Irvin & Stansbury, 2004, p. 56). Moreover, policy makers can learn from the knowledge and skills citizen have regarding a specific topic. The different perspectives on the issue and the diversity of knowledge, information and skills can improve the quality of the policy (Raad voor het Openbaar Bestuur, 2005).

Citizen participation comes in many shapes and forms. Citizens can be engaged in different degrees, ranging from being informed to being consulted or being able to suggest solutions or choose among policy options (Ianniello et al., 2019). Furthermore, involvement can take place at all stages of the policy process (Leyenaar, 2009). Citizen participation first took mainly place in an 'offline' environment, think of focus groups and consensus conferences. Two important drawbacks of these conventional methods are that they usually take a lot of time and allow only a select group to participate in the decision-making process. Nowadays there is an upcoming trend of online participation methods which allow involvement of more people. The main advantage of online participation is the low threshold to participate. It demands less time investment compared to conventional methods, citizens can participate wherever and whenever they want, and participation can be anonymous. As a result, participation becomes accessible to larger groups. This accessibility enables the so-called 'silent majority' to also participate in the decision-making process. And thus leads to a more diverse group

involved in the participation process since not only those who usually participate, have a lot to gain or have much spare time will have the opportunity to join (Pape & Lim, 2019; TU Delft, 2020).

1.2 Participatory Value Evaluation

A relatively new online participation method is Participatory Value Evaluation (PVE). In a PVE citizens are asked to provide advice on a government issue. The issue at stake and several policy options including their positive and negative effects are presented to the participants in an online environment. Participants have to select one or several options given a certain restriction such as a limited budget or an objective that has to be reached. Thereafter, they are asked to explain their choices. As other online participation methods, PVE allows a big group of people to participate in the decision-making process. Yet, the information provided by PVE provides in depth insights in the preferences of participants, especially compared as opinion polls and surveys which are frequently used as online participation approaches. PVE provides information about how often a policy option has been selected. Moreover, qualitative information about the motivation of participants about their decisions is given. This information gives insight in the arguments for and against each policy option. Thirdly, PVE can provide information about how participants rate different effects of the policy options and finally an optimal set of policy options can be calculated. So, PVE is both an evaluation method for public policies as well as a method to facilitate mass citizens participation (Mouter et al., 2019). Both applications are discussed in more detail.

1.2.1 PVE as evaluation method

PVE was initially developed to deal with some issues of the Social Cost-Benefit Analysis (SCBA), a policy evaluation method that is often used to support public decision-making. Within SCBA, Willingness to Pay (WTP) is often used to measure citizens' preferences. WTP is a stated preference method where respondents have to make choices about their private resources, such as income, in hypothetical scenarios. WTP assumes that the welfare effect of consumer goods and public projects can be derived from the choices people make with their private resources. The WTP valuation method is criticized for evaluating government policies as the trade-offs that individuals make between private resources and private goods may differ from the way in which these individuals believe that the government should consider trade-offs when it comes to public resources and public goods. As a result, willingness to allocate public budget (WTAPB) experiments were developed. PVE can be seen as an extension of the WTAPB experiments. In both cases participants have to make choices about the expenditure of government budget. The biggest difference is that in a PVE participants have the option to not spend the (full) budget. Furthermore, PVE is not only used in the context of a limited government budget, it can also be applied in situations where a specific objective has to be reached. (Mouter et al., 2019)

PVE is an evaluation method in the sense that it identifies the preferences of citizens regarding public policies. Based on the individual choices of the participants the optimal set of policies in terms of social value is calculated (Dekker et al., 2019). As the optimal set of policies is an aggregated output of all participants, it does not account for distribution in preferences among the participants. Yet it is likely that different groups of respondents with similar preferences can be identified. If a certain sub-group is overrepresented in the data, it can lead to a biased outcome, or in other words, the optimal set of policy options can be biased. To make sure that the optimal policy set reflects the preferences of the target group well, the response sample should be representative for the target group.

1.2.2 PVE as participation method

PVE is also used as a method to facilitate online mass participation. As PVE aims to identify the preferences of citizens regarding several policy options before the policy decision is made, this method fits in the policy preparation phase of the policy process. This is shown in Figure 1.1.

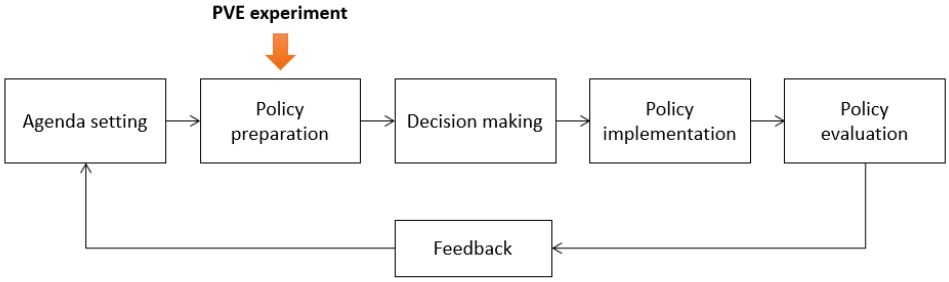

Figure 1.1 PVE in the policy process adapted from Raad voor het Openbaar Bestuur (2005, p. 10)

As PVE is a relatively new participation method it is not always clear for policy makers in which context they should choose this method and not another participation method. This is problematic as choosing a less optimal participation method leads to less effective participation. Literature provides different studies which compare several participation methods (Beierle, 1999; Leyenaar, 2009; Rowe & Frewer, 2000). However, these comparisons do not include PVE due to the novelty of this method. According to my personal awareness there is only one comparison made by Mouter et al. (2020) to improve the understanding of the strengths and weaknesses of PVE in terms of involving citizens in crisis policymaking. In their research they compare PVE to other participatory approaches which are used in crisis policymaking on the following four dimensions: practical feasibility during a pandemic and substantive, normative, and instrumental rationale for participation. A shortcoming of this comparison is that it only looks at PVE versus participation approaches that can be applied in crisis policymaking. Opportunities lie in investigating in which context, besides crisis policymaking, PVE is a suitable participation method. Moreover, there is little scientific understanding about the advantages and disadvantages of PVE compared to other participation methods.

1.3 Representativity and inclusivity

Most of the time it is not possible to sample the entire population of interest, called the target group, when doing research. Therefore, a researcher questions almost always a part of the target group. This subset is called a sample. To be able to generalize the results to the target group, the sample must be representative: the sample should be an accurate reflection of the target group. To determine if a sample is representative, researchers often look at the distribution of demographic characteristic such as gender, age, and education level in the sample. A sample is considered to be representative for a characteristic when the distribution of that characteristic in the sample matches the distribution in the population. Inclusiveness can be defined as the right to participate in the decision-making process when you are affected by the decision (Goodin, 2007).

For citizens to participate in a PVE, they should get access to the online environment in which the experiment takes place. This often happens in one of the following ways: a select group of citizens is invited to participate, or the online environment is made available to the public.
This first option often translates in use of a panel, hereinafter referred to as a panel PVE. A panel is a group of people that signed up to be questioned, often on a regular basis. Researchers make use of a panel to ensure the sample is representative. Researchers can compose the panel in such a way that it is a representative atonement of society in several aspects. However, using a panel does not guarantee that the response sample is representative for the target group. People that are invited can still decide not to participate. This is called non-response. A researcher can only check after conducting the research if the response sample is representative. A disadvantage of using a panel is that not everybody in the target population gets the chance to participate, which leads to a lower level of inclusiveness.

If the PVE experiment is freely accessible, the sample construction relies on the mechanism of self-selection of the participants, hereinafter referred to as an open PVE. The most important benefit is that everyone who wants to participate can actually participate. In this way a high level of inclusiveness can be reached. This, in turn, contributes to the credibility of the decision-making process. (Goodin, 2007; Itten, 2019) An important drawback of an open PVE is that the researcher is not in control of the selection process. People decide entirely for themselves whether they want to participate. Because the tendency to participate in the research will depend on the interest in the subject, making a PVE experiment freely accessible may lead to biased results. Citizens who choose to participate will most likely not properly represent the entire target group (Bethlehem, 2010; Lavrakas, 2008). Some groups may be over-represented, and their opinions magnified, while others may be under-represented. This can lead to inaccurate conclusions and maybe even wrong decisions (Austin, 2014).

So, representativity and inclusivity do not seem to go well together. If you want to involve everybody who wants to participate, the sample is probably not representative, and if you select a representative sample not everybody that wants to participate gets the change to do so. Yet, the underlying question is whether a panel PVE and an open PVE actually provide different results. Do the participants in a panel PVE have different preferences than the people that participate in an open PVE? This leads to the need to investigate which characteristics influence participants' preferences. In this study, the focus lies on the effect of demographic characteristics on participants' preferences because these variables are often used as a proxy to make a statement about whether a sample is representative. When you correct for the bias in these demographic characteristics, you can determine whether the panel and open PVE provide the same results. Weight adjustment can be applied to correct for bias in certain characteristics. It ensures that the sample distribution of a characteristic matches the population distribution. In this way, representativity can be restored. (Bethlehem, 2008; Engel et al., 2014; Kalton & Flores-Cervantes, 2003). For a panel PVE this means that a researcher can ensure that the sample is actually representative for the demographic characteristics in question, in case of an open PVE this means that representativity as well as inclusivity can be ensured.

The PVE on relaxation of COVID-19 measures in the Netherlands is selected as case study to investigate whether a panel version leads to different results than an open version. This PVE experiment is chosen because there are two datasets available: one panel dataset for which members of Kantar were asked to participate and one dataset which was freely accessible for the Dutch public. In the PVE experiment Dutch citizens were asked to provide an advice for the government about which COVID-19 policy measure to relax.

1.4 Research gap and research questions

PVE is a method to facilitate mass citizen participation and aims to be inclusive. Everyone who wants to participate in the process should get the change to do so. At the same time PVE is an evaluation method for public policy. To be able to provide an accurate advice about the preferences of citizens regarding a set of policy options, the sample under study needs to be representative. Representativity and inclusivity seem to conflict with each other as making the research available for the public very often leads to a biased sample. This study examines the consequences of the trade-off between representativity and inclusivity for the usability of PVE in the policy context. Therefore, the main research question of this study is:

"What are the consequences of the trade-off between representativity and inclusivity for the usability of Participatory Value Evaluation?"

In order to answer the main research, question the following sub-questions are formulated:

1. In which context is PVE a suitable participation method?

As PVE is a relative new participation method it is not always clear for policy makers in which context they should choose for this method, and whether a panel version or an open version is preferred. There is only one comparison made by Mouter et al. (2020), yet this comparison is limited. By using literature research, it is explored in which context PVE is a suitable participation method.

2. To what extend do bias in demographic characteristics influence the preferences of Dutch residents with regard to different policy measures in the PVE on relaxation of COVID-19 measures?

This sub-question investigates whether a panel PVE and an open PVE lead to difference outcomes. Firstly, the level of association between demographic characteristics and participants' preferences are determined using Cramer's' V test. Thereafter, regression models are estimated to gain an in depth understanding of the effect of each unique demographic variable on the policy choice of the respondents. Lastly, both datasets are reweighted so that the distributions of the demographic variables of each sample matches the corresponding population distributions. By comparing the outcomes statements can be made about the effect of representation bias on the participants' preferences.

3. How do experts reflect on the representativity and inclusivity of PVE, and how do they review weight adjustment as method to correct for representation bias?

This sub-question is answered by conducting serval interviews with experts. The interviews aim to gain insight in how experts view representativity and inclusivity. Moreover, it investigates what the consequences of representation bias are in the policy context and whether weight adjustment is an acceptable approach to correct for this.

1.5 Report structure

The structure of the remaining report is as follows. Chapter 3 elaborates on the PVE on relaxation of COVID-19 measures and discusses the methods used for this research. Chapter 4 discusses the outcomes of the PVE experiments, with the focus on differences between the panel dataset and the open dataset. Chapter 5 examines the effect of representation bias in demographic characteristics on the preferences of the respondents. Chapter 6 presents the outcomes of expert interviews. Lastly, chapter 7 presents the conclusion of this research and chapter 8 discusses the implications and limitations of the research.

Chapter 2 – Research approach

This chapter elaborates on the methodology used. Section 2.1 describes the PVE on relaxation of COVID-19 measures in the Netherlands. Sections 2.2 to 2.4 provide the methods used to answers the sub questions of this research.

2.1 Description of the PVE experiment on relaxation of COVID-19 measures

On the 27th of February 2020, the first official corona infection was confirmed in the Netherlands. As the situation worsened, a 'smart' lockdown of the Netherlands followed on March 23rd. The government took several measures to minimise the number of infections. These measures were based on the advice of experts from the Dutch Outbreak Management Team. After a while, the measures started to show desirable effects and there was room to relax some measures. A PVE was carried out on this topic to involve citizens in the policy decisions regarding the relaxation of lockdown measures for the period of May 20 to July 20. In a crisis situation, such as the COVID-19 pandemic, citizens participation can be very valuable. It gives people the chance to express their preferences about measures that majorly affect them. This is important because 1) citizens did not get the chance to vote on this topic during an election, so one can argue that is the 'fair' thing to do. 2) government can learn from it: not only about the preferences citizens have, but citizens may also come with new ideas. As mentioned before, decisions that are well grounded in citizens' preferences are likely to be better accepted by the public. Especially in case of the corona virus this is very important. If the measures are not accepted by the public, this will not only raise resistance but also makes the measures less effective.

In the PVE experiment on relaxation of COVID-19 measures, citizens were asked if the government should relax the current corona measures between May 20 and July 20, 2020, and if so, which measures should be preferred[1]. Participants could select a combination of measures as long as the increase in the pressure on the Dutch healthcare system did not exceed 50%. Participants could also reject each of the policy measures if they thought it should not be considered by the government. As there were various versions of the PVE with different attribute values available, the effect of a measure on the increase in pressure on the healthcare system could take a range of values. Table 2.1 provides an overview of the measures and the corresponding range of increase in pressure on the healthcare system.

Table 2.1 Overview policy measures and their effect on the healthcare system

Project	Description	Increase in pressure on healthcare system (%)
1	Nursing and care homes allow visitors	10 – 25
2	Businesses open again (except hospitality and contact-jobs)	6 – 15
3	Contact professions can open again	8 – 15
4	People younger than 18 years do not have to keep 1.5m distance	4 – 8
5	All restrictions are lifted for people who are immune	10 – 20
6	Restrictions are lifted in northern provinces Friesland, Groningen, and Drenthe	15 – 30
7	Direct family members from other households do not have to hold 1.5-meter distance	6 – 15
8	Hospitality and entertainment sector open again	15 – 25

[1] A demo-version of the PVE experiment can be found on http://pve.splicedgene.com/raadpleging-versoepeling-coronamaatregelen

Furthermore, participants received information about the effects of the measures on different attributes:

- The increase of deaths among people younger than 70 year
- The increase of deaths among people older than 70 years
- The increase of people with permanent physical injury
- The decrease of people with permanent mental injury
- The decrease in the number of households with long-term loss of income

The range of attribute values for each measure are presented in appendix A.

Available datasets

The PVE was carried out by researchers from Delft University of Technology, in collaboration with researchers from other universities and researchers from RIVM (the Dutch National Institute for Public Health and Environment). There were two datasets available: a panel dataset and an open dataset. Both were made available by the Delft University of Technology and are further described in section 4.1.

2.2 Methodology sub-question 1

The first sub-question is defined as follows: *"In which context is PVE a suitable participation method?"*

This sub question is answered by conducting literature review. The starting point for the review is the work of Mouter et al. (2020). By using a backward snowballing approach, relevant papers are selected for further reading. A first selection of relevant work is made based on the title of the document in the reference list. A second selection for further reading is based on scanning the summary, introduction, and conclusion of the papers. When the papers still seem relevant after this step, they are saved for detailed reading.

In addition, scientific databases Scopus and Google Scholar are used to retrieve papers. Search terms include, but are not limited to, (a combination of) the following terms*: "participation methods", "participation mechanisms", "participation approaches", "evaluation", "evaluation criteria", "participation criteria"*. First, the results are sorted on the number of citations in Scopus and on relevance in Google scholar. Thereafter, a same approach as described above is used to select relevant work. Finally, backward snowballing is applied to the papers that provide valuable insights.

2.3 Methodology sub-question 2

The second sub-question is: *"To what extend do bias in demographic characteristics influence the preferences of Dutch residents with regard to different policy measures in the PVE on relaxation of COVID-19 measures?"*

To answer this sub-question several steps, need to be taken, which are discussed in the following sections.

2.3.1 Descriptive results

The first step is to gain insight in the preferences of Dutch citizens with regard to the different policy measures, also called projects. The results of the panel PVE and the open PVE are reported and compared. Information about the share of participants that select a project provide insight in the support for each policy measure. Information about the share of participants that reject a project provide insight in the resistance for each project. It is important to note that not selecting a project is not the same as rejecting a project. When a respondent chooses to reject a project, he actively advises the government to not consider that policy measure, which is not the case when not selecting a project.

The following results are presented:

- Percentage of participants that select and reject a policy measure
- The number of selected and rejected projects
- The increase of pressure on the healthcare system

The McNemar's Chi-square test is used to determine whether the share of respondents that chooses a policy measure (either select or reject) is statistically different whit the share of respondents that chooses another policy measure. For example, to determine whether the share of respondents that select measure 1 is statistically different from the share of respondents that select measure 2. The McNemar's test is used since the variables of interest are binary and compared with each other within a single sample (UCLA, 2021).

2.3.2 Sampling and Representativity

The second step is to examine to what extent the samples of the panel PVE and open PVE are representative of the Dutch population. Before explaining how it can be determined whether a sample is representative, it is important to explain more about sampling and representativity in general.

Sampling

Researchers are often not able to question the entire target population. Therefore, a subset of this population is questioned instead. This subset is called a sample. Sampling is the process of selecting the group that you actually collect data from. In general, two types of sampling are distinguished: probability and non-probability sampling (de Leeuw et al., 2008). Both types consist of several sampling methods. In probability sampling, every element (person) of the target population has a known chance of being included in the sample. In non-probability sampling, the probability that each element in the population will be chosen is unknown. Besides, it cannot be guaranteed that each element even has a chance of being included in the sample. The main issue of non-probability samples is that statistical interference is not appropriate: *"With a probabilistic sample, we know the probability that we represent the population well and therefore we can estimate confidence intervals and significance tests. With a nonprobability sample, we may or may not represent the population well, but it is not appropriate to apply statistical inference to generalize to a general population. At best, we can use statistical inference to assess the precision with which we can generalize to a population consisting of whoever responded."* (de Leeuw et al., 2008, p. 9,10)

In the PVE on relaxation of COVID-19 measures, two sampling methods are used. In the panel PVE a panel is used. A panel consists of people who are willing to be questioned on a regular basis. First, the panel was composed. This can be either based on probability sampling or non-probability sampling. It is not clear which approach Kantar used to recruit her members. It is likely that the recruitment of panel members is prone to some kind of selection bias as it is unlikely that Kantar has a list of all elements in the population of interest, from which a sample may be selected, which would in this case be a population register of all Dutch citizens. In other words they do not possess a sampling frame (Lohr, 2008). After the panel is composed, the members are used as a sampling frame. People were selected based on their age and gender. This is stratified sampling: the population is divided in subgroups based on relevant characteristics. Based on the overall proportion of the population it is calculated how many elements should be sampled from each subgroup. The elements of each subgroup are then selected based on random sampling (Lohr, 2008).
For the open PVE, the sampling method used is called voluntary response sampling. Voluntary response sampling is classified as non-probability sampling. The sample consists of people who selected themselves into the survey. The open PVE can also be referred to as an unrestricted self-selected web survey. Open invitations were placed on different websites such as NOS.nl and AD.nl. The obtained sample is nonrepresentative for the population due to the lack of a sampling frame and probability sampling and due to self-selection (Manfreda & Vehovar, 2008). Self-selection bias occurs

when the tendency to self-select systematically differs between subgroups (Bethlehem, 2010; Lavrakas, 2008).

Representativity
Representativity is about the extent to which respondents of a sample are an accurate reflection of the larger population. It is important to be able to generalize and extrapolate the results to the population. As discussed in the section above, it is only appropriate to generalize the results if the sample is based on probability sampling. Nonetheless, it is still possible to check to what extend sample and population match. Demographic characteristics such as age and gender are often used as a proxy to determine whether a sample is representative. The reason for this is the availability of reference figures for these characteristics. Without proper reference figures it is not possible to check whether the distribution of a characteristic in the sample is representative for the larger population.

In this study representativity is tested based on the demographic variables that are measured in the PVE. These are the following variables: gender, age, education level, income, living situation and province. To preform representativity tests, the population distributions of the demographic variables are required. These population distributions can for example be found on the website of Statistic Netherlands (CBS). CBS provides a lot of data on demographic variables in the Netherlands.
For each variable, a table is composed with the number of observations in the sample and the number of observations that were expected based on the population distributions. Then, by using a Chi-square test, it is determined whether the number of observations in the sample significantly deviates from the expected number of observations. If the result of the Chi-square test is significant ($p<0.05$), the sample is not representative for this variable.

2.3.3 Measuring association

To determine to what extend bias in demographic variables leads to bias in preferences, the effect of the demographic variables on the preferences need to be investigated. The first step is to determine whether there is a relationship between each demographic variable and the choice of the participants to select or reject a policy measure.

There are different statistical tests to determine the level of association between two variables. Which test can be used depends on the measurement levels of the variables of interest. Four measurement levels can be distinguished:

- Nominal/categorical: the data can only be categorized (no order or direction)
- Ordinal: the data can be categorized and ranked
- Interval: the data can be categorized and ranked and the difference between two values is meaningful
- Ratio: ratio variables have all properties of interval variables, but also have a clear zero point

The measurement levels of the demographic variables and the project choice defined for this specific PVE are shown in Table 2.2.

Table 2.2 Measurement level of variables in PVE

Variable	Measurement level
Gender	Nominal
Age	Ordinal
Education level	Ordinal
Income	Ordinal
Living situation	Nominal
Province	Nominal
Selecting project X	Nominal
Rejecting project X	Nominal

To be able to perform a certain statistical test it might be needed to recategorize a variable. To determine the association between a demographic variable and the project choice (selecting and rejecting) all demographic variables are treated as nominal variables.

Cramer's V

To determine if two nominal variables are associated, one can perform a Chi-square test. Based on the p-value it can be determined if the association between two variables is statistically significant. The p-value tests the null hypothesis that no relationship exists between the categorical variables in the population. If the p-value is less than the specified significance level, the null hypothesis can be rejected. This means that there is a relationship between the variables.
To determine the strength of the association an additional test is required. In this study the Cramer's V test is used rather than the Phi test, as it is able to determine the strength of association between any two nominal variables. The Phi test can only be used if both variables only consist of two categories. The Cramer's V test is based on the Chi-square statistic of the Chi-square test and can be calculated using SPSS.

The formula for the Cramer's V is as follows (Sheskin, 2000):

$$V = \sqrt{\frac{X^2}{n(k-1)}}$$

Where:

X^2 = the Pearson Chi-square value of the Chi-square test
n = the sample size involved in the test
k = the smallest value of the number of columns or number of rows of the table

The coefficient (V) ranges from 0 to 1. The coefficient can be interpreted as follows (AcaStat Software, 2015):

- V = 0 no association
- 0 < V < 0.1 very little association
- 0.1 < V < 0.3 low association
- 0.3 < V < 0.5 moderate association
- 0.5 < V < 1 high association
- V = 1 perfect association

The p-value of Cramer's V test indicates whether the relationship is statistically significant. As Cramer's V is based on the Chi-square test, the p-values for both tests are the same. Therefore, only the results of the Cramer's V test are reported.

2.3.4 Binomial regression

Cramer's V test can determine whether there is an association between two variables and how strong this association is. However, this test does not provide insight in the direction of the association. Moreover, Cramer's V test does not control for underlying effects between the variables. Regression models provide insight in the direction of the relationship and control for underlying effects between variables. Regression models thus identify the unique effect of one or multiple independent variables on a dependent variable.

For the PVE on relaxation of COVID-19 measures, regression models can be used to gain an understanding of how demographic characteristics influence the preferences of citizens. The demographic characteristics of the participants are the independent variables. In this context there are two dependent variables of interest: whether the participant decides to select or reject a project.

Because the dependent variables only have two categories, a binomial logistic regression is applied. A binominal logistic regression *"predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable, based on one or more dependent variables that can be either continuous or categorical."* (Laerd Statistics, n.d.).

Moreover, the method of sampling, e.g., using a panel version or an open version, can be seen as moderating variable. The method of sampling may influence the effect of the independent demographic variables on the dependent variable. Therefore, separate models are estimated for the panel PVE and the open PVE. A conceptual overview of this is provided in Figure 2.1.

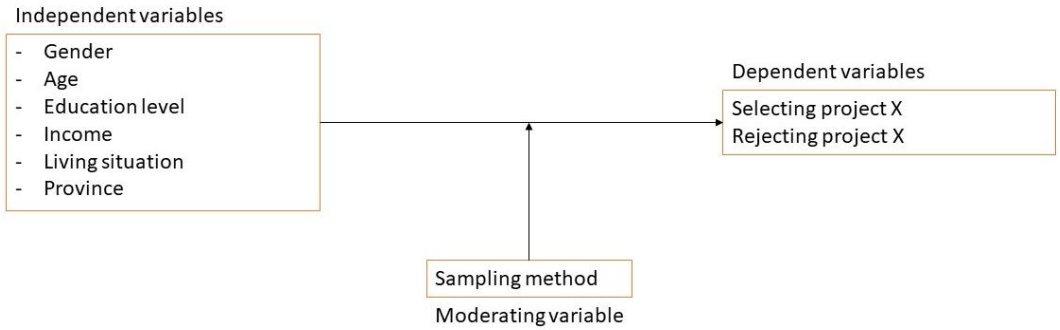

Figure 2.1 Conceptual overview binominal regression

For each of the 8 projects 2 binominal logistic regression models can be estimated: one model that predicts the probability of selecting the project, and one model that predicts the probability of rejecting the project. Furthermore, the regression analysis is applied to both datasets (panel and open). This results in a total of 32 models. These models are estimated using SPSS.

A binomial logistic regression provides several outputs. This research looks at the pseudo R^2 as well as the parameter estimates and their corresponding p-value.

Pseudo R^2

In a linear regression the R^2 represents the proportion of variance from a dependent variable that is explained by the independent variables of the model. For a binomial regression model, it is not possible to compute a R^2, so an approximation of this statistic, called the pseudo R^2, is computed instead. SPSS provides three pseudo R^2 values: the McFadden pseudo R^2, the Cox and Snell pseudo R^2 and the Nagelkerke pseudo R^2. Based on the preferences of one of the supervisors of this research the McFadden pseudo R^2 is reported.

The McFadden pseudo R^2 is calculated as follows (Bartlett, 2014):

$$R^2_{McF} = 1 - \frac{\ln(L_M)}{\ln(L_0)}$$

Where L_M refers the to the log likelihood of the logit model selected and L_0 refers to the log likelihood of the logit model if the model just had an intercept.

The McFadden pseudo R^2 can be interpreted as a normal R^2, but the values are smaller. A McFadden pseudo R^2 between 0.2 and 0.4 indicates an excellent model fit (McFadden, 1979)

Parameter estimates
The parameter estimates, also known as the coefficients of the model, provide insight in the relationship between the independent variables and the dependent variables. The coefficients can have a positive or negative sign. Given that B is the parameter estimate for the independent variable:

- A positive sign means that one unit change in the independent variable leads to an increase of B in de dependent variable
- A negative sign means that one unit change in the independent variable leads to a decrease of B in the dependent variable

The corresponding p-value of each coefficient indicates whether the relationship is statistically significant. The p-value for each independent variable tests the null hypothesis that the variables have no correlation with the dependent variable. If the p-value is less that the specified significance level, the null hypothesis can be rejected. This means that the data provides enough information for a relationship between the dependent and independent variable. Hence, a change in the independent variable leads to a change in the dependent variable.

2.3.5 Weight adjustment

When combing the insights of the regression analysis with the output of the representativity test, it is possible to indicate if a bias in certain demographic characteristic lead to over- or underestimating of selecting/rejecting each policy measure. However, it is hard to draw a conclusion on the combined effect of bias in multiple demographic variables. To demonstrate the total effect of the representativity bias, weight adjustment can be applied. The steps to perform weight adjustment are discussed below.

Step 1 – Selecting auxiliary variables
First, the auxiliary variables are selected. These are the variables on which the weight adjustment is based, in this case the demographic variables. For weight adjustment to be effective, the auxiliary variables must meet three conditions (Bethlehem, 2008):

1. The auxiliary variables must have been measured in the PVE experiment
2. The population distribution of the auxiliary variables needs to be known
3. The auxiliary variables must strongly correlate with the target variable(s)

It is likely that the demographic variables meet the first two conditions. The Cramer's' V test (section 2.3.3) reports the strength of the association. When a low association is reported, the weight adjustment is expected to not lead to major changes in the output. In other words, the preferences of the participants will not change that much. The opposite applies to a high level of association.

Step 2 – Choosing and applying the weight adjustment method(s)
Based on the selected auxiliary variables, one or multiple weight adjustment methods can be chosen. Appendix B provides a brief overview of available weight adjustment methods. In general, a researcher can choose to reweight the data on the marginal distributions of auxiliary variable or on the joint distribution of these variables. Reweighting on the joint distributions is more precise, however the data needed for this is not always available.

Step 3 – Comparing the results
The last step of the weight adjustment is to compare the policy preferences of citizens before and after weight adjustment with each other. A Chi-square test is performed to determine whether the difference is statistically significant.

2.4 Methodology sub-question 3

The third sub-question is defined as follows*: "How do experts reflect on the representativity and inclusivity of PVE, and how do they review weight adjustment as method to correct for representation bias?"*

This sub-question tries to provide insights on how experts reflect on representativity, inclusivity and weight adjustment as a method to restore representativity. These insights are obtained by conducting interviews. To reflect on the results from different perspectives, experts with different backgrounds will be invited. The background of the experts is assessed on 3 criteria:

1. Experience with PVE
2. Experience with weight adjustment
3. Experience in communicating research results with policy makers

The interview invitation can be found in Appendix E.1. Due to the circumstances of COVID-19, all interviews are conducted online, using Zoom.

When looking at how interviews are structured, three types of interviews can be distinguished: structured, semi-structured and unstructured interviews (Wildemuth, 2016). This research uses a semi-structured interview approach. A list of questions is prepared, which can be found in Appendix E.2. These questions form a guideline for the conversation, however there is still room to add or leave out questions based on the context of the answers given by the participants. Furthermore, the questions are not necessarily asked in the same order for every interview, as is the case for structured interviews. In this way semi-structured interviews allows for a more in-depth discussion of certain topics. Nonetheless, it is harder to compare the results of semi-structured interviews than for structured interviews, yet it is easier than for unstructured interviews.

Chapter 3 – PVE in the policy context

The aim of this chapter is to gain a better understanding in which context PVE is a suitable participation method and in which context a panel PVE or an open PVE is preferred. As described in the introduction, there is a trade-off between representativity and inclusivity for PVE. This trade-off affects the choice whether to use a panel PVE or an open PVE. Depending on why a policy maker wants citizens to participate or what goal he aims to achieve, either type may be more suitable.

Section 3.1 discusses the rationales for citizen participation that were used by Mouter et al. (2020). Section 3.2 discusses a set of six socials goals policy makers may aim to achieve with citizen participation. Section 3.3 discusses the potential to achieve these social goals. Section 3.4 draws a conclusion to answer sub question 1.

3.1 Rationales of citizen participation

The starting point for this chapter is the work of Mouter et al. (2020). In this research PVE is compared with other participation mechanisms based on four criteria: 1) practical feasibility during a pandemic, 2) normative rationale, 3) substantive and 4) instrumental rationale for participation. As the first criteria is context specific it is not considered in this analysis. The three rationales for participation used by Mouter et al. (2020) originate from the work of Fiorino (1990). These rationales are used in many studies such as (Leach et al., 2005; Stern et al., 1996; Stirling, 2006; Wesselink et al., 2011) and are explained below.

Normative rationale

The normative rationale suggests that that involving citizens is the righteous thing to do in a democracy. From this perspective participation does not need further justification. It argues that citizens are the best judge of their own interests and that they have the right to participate in decisions that affect them or their community. In other words, inclusivity is an important aspect for this rationale.

Substantive rationale

The substantive rationale argues that citizen participation improves the quality of the decisions made by the government. This rationale aims to gain a deeper understanding about the problem. Citizens can provide information about the issue at stake that experts or policy makers miss. From this perspective, only participants that have to additional knowledge about the issue should be included in the decision-making process.

Instrumental rationale

The instrumental rationale for participation is that it makes decisions more legitimate and improves results. It aims to restore public trust, resolve conflict, and justify decisions. This means that stakeholders that are needed for the implementation of the policy or those who can hinder the implementation should be involved in the decision-making process. For this perspective, inclusivity is less important. Only a selected set of stakeholders is included.

3.2 Social goals of citizen participation

The rationales provide reasons why policy makers want to involve citizens in the decision-making process. Another approach is to look at the goal policy makers aim to achieve with citizen participation. In the research of Beierle (1999) a framework to determine the potential of participation mechanisms to achieve a set of participation goals is presented.

In the research six social goals of participation are distinguished:

1) To inform and educate the public
2) To incorporate public values, assumptions and preferences into decision making
3) To increase the substantive quality of decision
4) To foster trust in institutions
5) To reduce conflict among stakeholders
6) To make decisions cost-effectively

These goals can be linked to the rationales for participation by asking the question why a policy maker would want to achieve this goal. Reasons for reaching the first goal, informing, and educating the public, can be normative or instrumental. Informing citizens is often seen as the neat thing to do in a democratic society. Moreover, informing the public about the issue at stake and the policy options can contribute to the acceptance of the policy decision (Li & Zhao, 2019). The second goal, incorporating public values, assumptions and preferences into the decision making, can be grounded in all three rationales. From a normative perspective, incorporating the preferences of citizens is the right thing to do. Citizens are the best judge of what they want and should therefore participate. From a substantive perspective the quality of the decision can be improved as a broader perspective of values, assumptions and preferences is taken into account. Also, reaching this goal can contribute to a better accepted decision (Irvin & Stansbury, 2004).The third goal, increasing the substantive quality of the decision, is the same as the substantive rationale. The fourth goal, fostering trust in institutions, and the fifth goal, reducing conflict among stakeholders, can both be categorized as instrumental rationales for participation. Lastly, the sixth goal, making cost-effective decisions (both in time and money) is not based on any of the rationales but is an important aspect of the participation process.

3.31 PVE and participation goals

3.3.1 Dimensions of participation mechanisms

In his research Beierle (1999) assed whether various participation mechanisms are likely to achieve the social goals described above. The participation mechanisms are compared on four dimensions: the direction of information flows, the degree of interaction among potentially opposing interests, the type of representation and the decision-making role of the public. The dimensions are discussed below.

Direction of information flows

Three types of information flows are distinguished: from government to the public, from the public to the government and two-way communication. Governments inform the public to reach the goals of education (goal 1) and can lead to greater trust (goal 4) if the decision-making process becomes more transparent. The public usually provide decision-makers with information when they want them to take public values and preferences into account (goal 2) or to contribute to the quality of the decision by providing valuable (expert) information (goal 3). Two-way communication allows for deliberation among participants. Participatory mechanisms that allow two-way communication are expected to achieve all four goals.

The degree of interaction among potentially opposing interests

The level of interaction among potentially opposing interest can range from none till high. Interaction among opposing stakeholder is crucial to the resolve disputes. Participatory approaches that bring together opposing parties are expected to provide opportunities to reduce conflict among stakeholders (goal 5).

The decision-making role of the public

The role of the public in the decision-making process can differ from none to an advisory role to a direct decisional role. Participatory approaches in which citizens have a more direct role in a decision-making process are likely to achieve trust (goal 4) than those which do not.

The type of representation
In general, three types of representation can be distinguished: citizens can represent themselves, a subgroup of 'representative' citizens may represent the larger target population, or professional interest groups/experts can represent the interests of citizens. Participatory approaches in which people directly participate are expected to better fulfil the goals of education (goal 1) and to gain trust (goal 4).

3.3.2 Dimensions of PVE

This section describes how PVE scores on the different dimensions of the framework of Beierle (1999). The direction of information flow in a PVE is two-way. On the one hand, PVE provides the opportunity for government agencies to inform and educate the public about different policy options and their effects under consideration. On the other hand, by participating citizens get the change to inform the government about their preferences regarding these policy options. Yet, the PVE experiment itself does not provide the opportunity to deliberate about the issue at stake, which is often the case for participatory approaches that have two-way communication. Participants fill out the PVE in an online environment, all by themselves. This means there is no direct interaction among participants in general and therefore neither among potentially opposing interest groups. The decision-making role of the public in a PVE may differ per situation. Yet, the idea of this participatory approach is to provide the government with an advice about the participant's preference with regard to the different policy options. So, the decision-making role can be described as advisory. The last dimension is the type of representation in a PVE. When using a panel to fill out the PVE, representation is based on a 'representative' subgroup of citizens who ought to represent the larger population. When the PVE is freely accessible (open PVE) citizens represent themselves.

3.3.3 Potential of PVE to achieve participation goals

1. Informing and educating the public
PVE is an appropriate participation method to inform and educate the public. Participants receive in-depth information about the issue at stake. An open PVE is likely to achieve this goal in a better degree than a panel PVE as the former allows (more) people to directly participate in the process. In the latter, only a group of 'representative' people can participate.

2. Incorporating public values, assumptions and preferences into decision making
A PVE experiments results in information about citizens' preferences and motivations regarding the different policy options. PVE allows citizens to transmit new ideas, arguments and values and conditions into the decision-making process. The results of a panel PVE are ought to be representative for the larger population. An open PVE is likely to lead to biased results, as the people that decide to participate are probably not an accurate reflection of the entire society. Therefore, a panel PVE is expected to be better to identify the preferences of the majority of the public. Yet, an open PVE is probably more suitable to find the rare opinions in society. Overall, the quality of preferences that people express is expected to be lower than those expressed after deliberation, such as in the case of citizen jury or consensus conference (Escobar & Elstub, 2017; Mouter et al., 2020).

3. Increasing the substantive quality of decision
Citizens are not only a source of values, assumption and preferences but they can also provide relevant knowledge or ideas that would not have been available otherwise. Based on the individual choices that participants make during the PVE experiment an optimal set of policies given a constrained public resource in terms of social value can be calculated (Dekker et al., 2019). PVE similar to participatory budgeting (PB) as it is about the allocation of a resource. Yet, in case of PB this resource is always a public budget, while a PVE can be about any public resource (Mouter et al., 2020). This evaluation aspect is a unique feature of PVE.

A panel PVE is more suitable to increase the substantive quality of the decision as representation bias in the open PVE could lead to a biased set of optimal policy options which in fact negatively effects the quality of the decision.

4. Fostering trust in institutions

PVE may also be used to foster trust in institutions. As citizens are put in the shoes of policy makers, the gain more insight in the dilemma's policy makers face. After participation they understand better which things policy makers have to consider before making a decision. Therefore, PVE makes the decision-making process more transparent and is thereby able to increase the level of trust citizen have in the government. An open PVE is expected to achieve this goal to a better extend, as more people can participate in the decision-making process.

5. Reducing conflict among stakeholders

PVE is not a suitable participatory approach to reduce conflict among stakeholders, as there is no direct interaction among these parties. Participation methods that allow for deliberation are more suitable to reach this goal. Yet, the outcomes of a PVE experiment may be used as input for other participatory approaches in which opposing stakeholders get the change to deliberate.

6. Making decisions cost-effectively

In general, online participation approaches are more cost-effectively than conventional participation methods. PVE is cost-effective in the sense that it allows a large group of citizens to participate in the decision-making process, without major costs. People can participate from their home and only need an internet connection. Filling out a PVE only takes around 30 minutes and is therefore less time-consuming than most conventional methods, but more time consuming than opinion polls.

3.4 Conclusion PVE in policy context

This chapter aimed to answer in which context PVE is a suitable participation approach. Depending on the goal a policy maker wants to achieve with citizens participation either a panel or an open PVE is preferred. This chapter showed that PVE can fulfil several participation goals. PVE provides citizens with information about a policy dilemma the government faces, including several options choose from and their effects. PVE is a suitable approach to inform and educate the public, as well as fostering trust in institutions as it makes the decision-making process more transparent. An open PVE is more likely to reach these goals then a panel PVE since an open PVE has the potential to reach an unlimited amount of people, whereas the panel PVE can only be filled out by members of a panel. A panel PVE is more adequate in reaching to goals of identifying the preferences of the public and to increase the substantive quality of a decision as this sample is representative for the larger population. PVE itself is not a proper method to reduce conflict among stakeholders as it does not allow for direct interaction among (opposing) stakeholders. Lastly, PVE is a cost-effective participation approach as a large group of people can participate without making major costs.

Chapter 4 – Data description

In this chapter, the outcomes of the PVE experiment are discussed. Section 4.1 elaborates on the available datasets. Section 4.2 presents the preferences of the participants for selecting and rejecting a project. Section 4.3 is about the number of projects participants select and reject. Lastly, section 4.4 elaborates on the effect of the selected measures on the increase in pressure on the healthcare system.

4.1 Available PVE datasets

There are two datasets available for the PVE on relaxation of COVID-19 measures, which were made available by TU Delft. Both datasets are explained below.

1) **Panel dataset**: The dataset consists of the responses of 3.358 randomly selected Dutch citizens of 18 years and older. Respondents in this sample were invited by Kantar Public based on their gender and age. Participants received a small monetary compensation. They could join from April 28 till May 3, 2020. Using this panel, the sample is expected to be representative for the Dutch population.
2) **Open dataset:** This dataset contains the responses of the PVE experiment that was made available via internet to the public. Multiple online articles to participate in the research have been published, for example by NOS, RTL, and AD. Moreover, the research got attention on social media such as LinkedIn, Facebook and Twitter (Deetman, 2020). Everyone who wanted to participate could do so. People could participate from April 28 until May 5, 2020. In total, 26.293 citizens of 18 years and older responded.

Information about the data preparation can be found in Appendix C.

4.2 Sample response selecting and rejecting policy measures

In the PVE experiment respondents were asked to select any combination of policy measures as long as the increase in pressure on the healthcare system did not exceed 50%. Furthermore, respondents could indicate which measure they thought the government should not consider, e.g., they could reject several measures. It is important to note that not selecting is not the same as rejecting a measure. For each measure, the percentage of respondents that has chosen or rejected that measure is shown in Figure 4.1.

Figure 4.1 also shows that all measures have supporters and opponents. In both datasets, policy measure 3 is selected most often. The most popular measures after that are measures 2 and 7. However, for the panel dataset measure 7 is more popular than measure 2. For the open dataset, this is the other way around. McNemar's Chi-square test shows that for both datasets the differences between the percentage of respondents that selected these measures are statistically significant.
Measure 6 is rejected the most in both datasets. The top 3 least favoured measures furthermore consist of measures 5 and 8. For the panel dataset, measure 8 is slightly more chosen than measure 5. However, McNemar's Chi-square statistics suggests that it is not a statistically significant difference (Chi-value = 1.665, p-value = 0.197). For the open dataset, measure 5 is significantly more chosen than measure 8. For all results of the McNemar's test, see appendix D.

Even though there are small differences in the ranking of the selected and rejected measures, the percentages between the open PVE and the panel PVE are quite large. For example, the most popular measure, project 3, is selected by 50.4% of the panel participants, whereas this measure is selected by 63.9% of the participants in the open version. So, the actual support or resistance for each project is quite different.

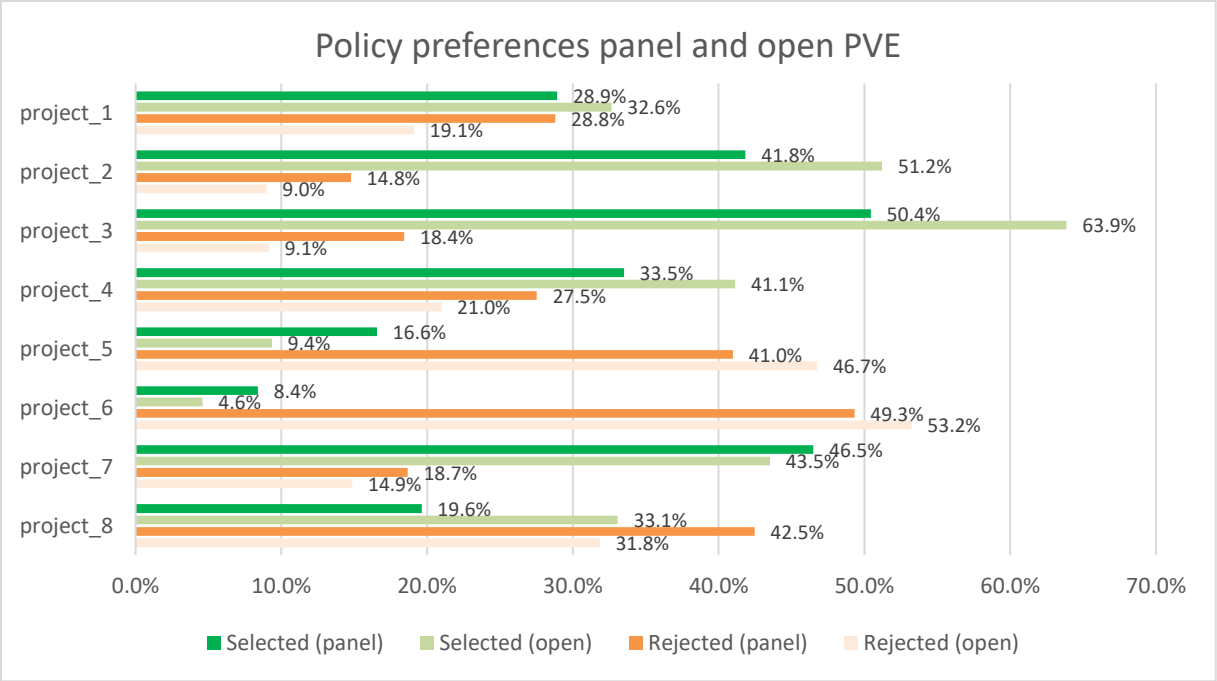

Figure 4.1 Percentage of respondents that has selected and rejected a policy measure

4.3 The number of selected and rejected measures per participant

Table 4.1 shows that in both datasets most respondents selected three policy measures and rejected two measures. Around 18% of the panel respondents and 26% from the open dataset selected more than three measures. Furthermore, 10% of the panel respondents did not select any measures at all, while this was only 5% of the respondents of the open dataset. Moreover, 66% of the respondents from the panel dataset rejected more than two measures, by the open dataset this was 61%. This leads to the conclusion that respondents from the open PVE selected on average more measures and rejected fewer measures than respondents from the panel PVE. Respondents from the open PVE were more in favour of implementing policy measures.

Table 4.1 Number of measures selected or rejected by respondents

Number of measures	**Selected**		**Rejected**	
	% respondents panel dataset	% respondents open dataset	% respondents panel dataset	% respondents open dataset
0	10.27%	4.67%	14.09%	22.35%
1	9.86%	6.53%	20.01%	16.80%
2	23.97%	22.72%	22.96%	25.56%
3	37.34%	39.97%	20.19%	18.38%
4	16.86%	23.08%	11.44%	9.11%
5	1.67%	3.00%	5.00%	4.17%
6	0.03%	0.04%	2.71%	1.80%
7	-	-	1.40%	0.91%
8	-	-	2.20%	0.91%

4.4 Increase in pressure on the healthcare system

When selecting the policy measures respondents could not exceed the maximum of 50% increase in pressure on the healthcare system. This restriction has been imposed as otherwise it would not be possible to treat all patients who have a chance of recovery. Moreover, three levels of increase in pressure could be distinguished: green, yellow and red.

Green: 0 – 25% increase. People in the healthcare sector do not have to work overtime. The chance of workers in the healthcare sector dropping out in the short and long term is about the same as in the period before the corona crisis (early 2020).

Yellow: 25 – 40% increase, which means that the health system is overloaded. On average, people in the health sector need to work 6 hours extra per week. There is a risk that workers are dropping out in the short and long term. Some treatments other than corona that are less urgent or non-emergency care should be postponed.

Red: 40 – 50% increase. In this scenario the healthcare system is heavily overloaded. Employees in this sector must work an average of 12 hours extra per week. There is a high risk of healthcare personnel dropping out in the short and long term. All treatments other than corona that are less urgent or non-emergency should be postponed. Possible shortages of protective materials arise. Nurses and doctors who normally work in different departments now have to work in the corona intensive care unit. These healthcare personnel must work in another profession for a while. This can be difficult as they have to work in a different team, which can lead to doubts about whether the right choices are being made.

Table 4.2 Increase in pressure on healthcare system for panel and open dataset

Increase in pressure on healthcare system (%)	% respondents panel dataset	% respondents open dataset
0-5	10.5	4.8
6-10	5.1	3.3
11-15	5.2	3.9
16-20	6.3	5.7
21-25	12.5	12.6
26-30	10.1	9.8
31-35	12.1	13.0
36-40	14.3	17.1
41-45	11.8	14.1
46-50	12.1	15.5

Table 4.2 shows that most participants wanted corona measures to be relaxed to a limited extent. There is little support for relaxation that will cause the healthcare system to become heavily overloaded (red). On average, participants in the open PVE recommended that the pressure on the healthcare system may increase by 32%, while participants of the panel PVE recommended that the pressure may increase by 28%. The percentage of participants who advised to relax very little (0-5% increase in pressure) is much higher for the panel PVE (11%) than for the open PVE (5%). It can be concluded that the citizens who participated in the open PVE wanted to go further in relaxing corona measures than the participants of the panel PVE. Participants in the panel PVE were more cautious in relaxing measures than participants in the open PVE. This is in line with the conclusion from section 4.3.

Chapter 5 – The effects of representation bias

This chapter examines the effect of representation bias in demographic variables on participants 'policy preferences for the various policy measures. Section 5.1 investigates to what extend the panel and open PVE are representative of the Dutch population. Section 5.2 provides the results of the Cramer's V tests, whereafter section 5.3 elaborates on the results of the binomial regression models. Sections 5.4 presents the results of weight adjustment and section 5.5 draws a conclusion for sub question 2.

5.1 Representativity

Demographic variables are used as a proxy to determine whether the samples are representative for the Dutch population. To compare the samples of the panel and open PVE with the population, the population distribution of the variables of interest should be known. The following demographic variables are used to check for representativity: gender, age, education, and province. For these demographic variables, appropriate reference figures are available at Statistic Netherlands (CBS). The reference figures for gender, age, and province could be used directly. For education level, it was necessary to recategorize the data (Appendix C). The demographic variables income and living situation are not used to check for representativity as the categories of these variables did not match at all with the data from CBS. Table 5.1 provides an overview of the CBS datasets used to determine the population distributions for Dutch citizens.

Table 5.1 CBS datasets used to determine population distributions

Demographic characteristic	Dataset CBS
Gender	Population on January 1 and average; gender, age, and region
Age	Population on January 1 and average; gender, age, and region
Education level	Population; education level; gender, age, and migration background*
Province	Population on January 1 and average; gender, age, and region

* Education level was only available for 15 years and older, and not for 18 years and older

Table 5.2 shows the distributions of the demographic characteristics in the Dutch population and the distributions in the panel and open PVE datasets. The response rate indicates the share of participants that provided information on the variables in question. Up to 20% of the participants decided not to fill out the questions about their demographic characteristics.

Based on the Chi-square test it can be determined whether a sample distribution is representative for the population. It can be concluded that both samples are only representative for gender. For the other demographic variables, the Chi-square tests show a significant difference, which means that the sample is not representative for these variables. Yet, the panel sample is more representative for the Dutch population than the open sample. This is because (for all demographic variables) the deviation from the population distributions for the panel sample is much smaller than for the open sample. This is reflected in the Chi-square value. For example, the Chi-square value for the age variable of the panel sample is equal to 40.72, whereas this value for the open sample is equal to 1605.29.

Although the panel sample is statistically seen not representative for age, education, and province, it gives a quite accurate reflection of the Dutch population for these variables. The most important difference between the panel sample and the Dutch population is that lower educated people are underrepresented in the panel sample. In the open sample, the differences are bigger. The main differences between this sample and the target population are that adults between 18 and 25 years and adults older than 66 years are underrepresented, high-educated people are largely overrepresented and that respondents from Zuid-Holland, Utrecht, and Overijssel are overrepresented.

Table 5.2 Representativity samples compared to Dutch population

Demographic variable	*Categories*	*Expected according to population (%)*	*Distribution panel sample (%)*	*Statistical test panel sample*		*Distribution open sample (%)*	*Statistical test open sample*	
Gender	Male	49.27	48.54	Response rate	83.74	49.34	Response rate	80.36
	Female	50.73	51.46	X^2	0.60	50.66	X^2	0.04
				Df	1		Df	1
				p-value	0.440		p-value	0.840
Age	18 - 25 years	12.46	14.76	Response rate	83.92	8.94	Response rate	80.58
	26 - 35 years	15.66	17.14	X^2	41.72	18.48	X^2	1605.29
	36 - 45 years	14.89	12.78	Df	5	17.07	Df	5
	46 - 55 years	18.29	15.26	p-value	0.000	22.42	p-value	0.000
	56 - 65 years	16.35	16.15			20.14		
	66+ years	22.35	23.92			12.96		
Education	Low	30.63	16.62	Response rate	82.94	4.20	Response rate	80.48
	Middle	37.36	35.44	X^2	401.76	16.03	X^2	22486.73
	High	32.01	47.94	Df	2	79.77	Df	2
				p-value	0.000		p-value	0.000
Province	Groningen	3.47	5.07	Response rate	83.92	2.66	Response rate	80.58
	Friesland	3.73	3.62	X^2	40.07	1.90	X^2	3167.25
	Drenthe	2.86	2.70	Df	11	1.69	Df	11
	Overijssel	6.59	5.57	p-value	0.000	10.89	p-value	0.000
	Flevoland	2.30	2.95			3.45		
	Gelderland	11.96	12.03			9.01		
	Utrecht	7.61	7.98			12.04		
	Noord-Holland	16.57	15.29			16.03		
	Zuid-Holland	21.13	20.12			28.41		
	Zeeland	2.24	2.31			1.08		
	Noord-Brabant	14.84	14.62			9.40		
	Limburg	6.72	7.74			3.45		

5.2 Level of association between demographic variables and policy choice

To gain insight into whether a relationship exists between the demographic variables and the project choice of respondents, and to determine how strong this relationship is, the Cramer's V test is executed. The V is interpreted as a measure of relative strength of association between two variables. If the corresponding p-value is <0.05 (95% significance level) it means that the relationship is statistically significant, and thus exists. For more information, see section 2.3.3.

5.2.1 Results Cramer's V panel PVE

Table 5.3 shows the results of the Cramer's V test for selecting a project in the panel PVE. Table 5.4 shows the test results for rejecting a project.

Table 5.3 Panel PVE - Association demographic variables and selecting projects

	Gender		*Age*		*Education*		*Province*		*Income*		*Living situation*	
	V	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**
Project_1	0.047	0.012	0.153	0.000	0.058	0.010	0.055	0.656	0.073	0.030	0.081	0.002
Project_2	0.029	0.127	0.073	0.011	0.111	0.000	0.077	0.119	0.108	0.000	0.081	0.002
Project_3	0.016	0.387	0.076	0.006	0.085	0.000	0.096	0.007	0.150	0.000	0.078	0.005
Project_4	0.010	0.603	0.049	0.239	0.054	0.018	0.092	0.014	0.094	0.001	0.113	0.000
Project_5	0.019	0.321	0.069	0.020	0.048	0.039	0.076	0.128	0.053	0.324	0.051	0195
Project_6	0.026	0.175	0.074	0.008	0.022	0.510	0.084	0.049	0.063	0.115	0.062	0.057
Project_7	0.056	0.003	0.045	0.331	0.069	0.001	0.081	0.067	0.076	0.021	0.107	0.000
Project_8	0.010	0.595	0.065	0.039	0.042	0.090	0.074	0.170	0.055	0.267	0.029	0.803

Table 5.4 Panel PVE – Association demographic variables and rejecting projects

	Gender		*Age*		*Education*		*Province*		*Income*		*Living situation*	
	V	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**
Project_1	0.073	0.000	0.115	0.000	0.023	0.490	0.083	0.054	0.137	0.000	0.066	0.031
Project_2	0.039	0.039	0.075	0.007	0.073	0.001	0.156	0.000	0.185	0.000	0.097	0.000
Project_3	0.013	0.475	0.108	0.000	0.107	0.000	0.087	0.032	0.127	0.000	0.046	0.305
Project_4	0.008	0.690	0.050	0.218	0.045	0.061	0.078	0.101	0.103	0.000	0.089	0.000
Project_5	0.034	0.074	0.102	0.000	0.028	0.328	0.101	0.002	0.085	0.004	0.077	0.005
Project_6	0.044	0.019	0.080	0.003	0.094	0.000	0.076	0.130	0.073	0.031	0.074	0.009
Project_7	0.066	0.000	0.063	0.046	0.072	0.001	0.108	0.001	0.116	0.000	0.061	0.066
Project_8	0.001	0.945	0.059	0.076	0.077	0.000	0.070	0.234	0.099	0.000	0.040	0.485

The tables above only show relationships with very little association ($0 < V < 0.1$) and low association ($0.1 < V < 0.3$). When two variables show very little association, the cell reporting the Cramer's V is marked **light green**. When two variables show a low association, the cell reporting the Cramer's V is marked **dark green**. When the relation is statistically significant, the p-value is marked **orange**.

When looking at the relationships between the demographic variables and selecting a project (Table 5.3), it becomes clear that 27 out of 48 tested relationships are statistically significant. For age and education, the relations are most often significant (for 6 out of 8 projects). This is followed by income and living situation (5 out of 8 projects), province (3 out of 8 projects), and age (2 out of 8 projects). Yet, there are only 6 relationships that show a weak association, all other associations are very low.

When looking at the relationships between the demographic variables and rejecting a project (Table 5.4), it becomes clear that 32 out of 48 tested relationships are statistically significant. For income, all 8 relations are significant. Thereafter, age shows the most significant relations (6 out of 8). Education and living situation both show significant relations for 5 out of 8 projects, for gender and province this is the case for 4 out of 8 projects. 12 out of 32 relations show a low association. This is most often between income and rejecting a project.

In conclusion, the demographic variables can be used to explain the policy choice of a respondent, yet to (very) limited extend. The demographic variables can explain better why someone rejects a measure, as there are more and stronger statistically significant relationships between the demographic characteristics and rejecting a project, than there are for selecting a project.

5.2.2 Results Cramer's V open PVE

Table 5.5 shows the results of the Cramer's V test for selecting a project in the open PVE. Table 5.6 shows the test results for rejecting a project. Statistically significant relations are marked in orange.

Table 5.5 Open PVE - Association demographic variables and selecting projects

	Gender		*Age*		*Education*		*Province*		*Income*		*Living situation*	
	V	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**
Project_1	0.131	0.000	0.153	0.000	0.006	0.658	0.045	0.000	0.060	0.000	0.080	0.000
Project_2	0.138	0.000	0.046	0.000	0.100	0.000	0.045	0.000	0.094	0.000	0.049	0.000
Project_3	0.055	0.000	0.051	0.000	0.004	0.841	0.051	0.000	0.110	0.000	0.056	0.000
Project_4	0.021	0.003	0.112	0.000	0.053	0.000	0.049	0.000	0.076	0.000	0.118	0.000
Project_5	0.045	0.000	0.027	0.010	0.016	0.059	0.030	0.056	0.040	0.000	0.028	0.005
Project_6	0.030	0.000	0.040	0.000	0.009	0.444	0.127	0.000	0.030	0.007	0.015	0.450
Project_7	0.032	0.000	0.059	0.000	0.023	0.004	0.027	0.159	0.033	0.002	0.049	0.000
Project_8	0.061	0.000	0.149	0.000	0.049	0.000	0.084	0.000	0.102	0.000	0.078	0.000

Table 5.6 Open PVE - Association demographic variables and rejecting projects

	Gender		*Age*		*Education*		*Province*		*Income*		*Living situation*	
	V	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**	**V**	**P-value**
Project_1	0.095	0.000	0.076	0.000	0.064	0.000	0.029	0.094	0.030	0.006	0.044	0.000
Project_2	0.043	0.000	0.039	0.000	0.052	0.000	0.041	0.000	0.054	0.000	0.028	0.006
Project_3	0.002	0.812	0.070	0.000	0.066	0.000	0.041	0.000	0.081	0.000	0.043	0.000
Project_4	0.008	0.238	0.081	0.000	0.073	0.000	0.049	0.000	0.064	0.000	0.090	0.000
Project_5	0.045	0.000	0.039	0.000	0.036	0.000	0.034	0.012	0.044	0.000	0.044	0.000
Project_6	0.023	0.001	0.050	0.000	0.035	0.000	0.056	0.000	0.042	0.000	0.031	0.001
Project_7	0.023	0.001	0.107	0.000	0.047	0.000	0.025	0.300	0.019	0.392	0.054	0.000
Project_8	0.017	0.012	0.159	0.000	0.053	0.000	0.067	0.000	0.076	0.000	0.081	0.000

The tables above can be interpreted in the same way as Tables 5.3 and 5.4. When looking at the relationships between the demographic variables and selecting a project (Table 5.5), it becomes clear that the majority of tested relations (41 out of 48) is statistically significant. For gender, age, and income all 8 relations are significant. For living situation this is 7 out of 8, for province 6 out of 8 and

for education only 4 out of 8. Overall, 10 of the 41 statistically significant relations show a low association.

When looking at the relationships between the demographic variables and rejecting a project (Table 5.6), it becomes clear that 43 out of 48 tested relationships are statistically significant. Yet only 2 of these relations show a low association. This is between age and project 7 as well as age and project 8.

Overall, there are more relationships between demographic variables and rejecting a project than there are for selecting a project. However, the relations between the demographic variables and selecting a project are more often stronger associated.

5.2.3 Conclusion Cramer's V

The Cramer's V tests showed that there are significant relations between the demographic variables and the project choices of respondents. However, all associations are weak or low. This means that the demographic variables have a limited effect on the participants' preferences. Lastly, it is noticeable that the Cramer's V test for the open PVE showed more statistically significant relations than for the panel PVE. This can be explained by the larger number of respondents in this dataset.

5.3 Binominal regression results

To gain insight into the direction of association between demographic variables and project choice binominal regression models are estimated. A binomial regression model estimates the unique effect of each independent variable (the demographic variables) on the dependent variables (e.g., selecting or rejecting a project). The variables age and income are included as covariates in the regression model. The variables gender, education level, living situation and province are included as factors. For each of the 8 projects 2 models can be estimated: one for selecting the project and one for rejecting the project. This is done for both datasets. For more information see section 2.3.4.

5.3.1 Regression models panel PVE

Table 5.7 provides an overview of the regression models estimated on the open dataset. First an interpretation of the pseudo R^2 is given, followed by an interpretation of the regression models for each project.

Interpretation McFadden pseudo R^2

The values for the McFadden pseudo R^2 range between 0.010 and 0.028. Models with a pseudo R^e between 0.2 and 0.4 can be interpreted as a good model fit (McFadden, 1979). When taking this into account, it can be concluded that the regression models for the panel sample have a poor model fit. Based on the demographic characteristics of a respondent it is possible to estimate whether he selected or rejected a project. However, the predictive power is limited.

Project 1 – Nursing and care homes allow visitors

The first model, estimated for selecting the project, shows significant p-values for age and gender. The older the respondent, the more likely he was to select project 1. Furthermore, females were more likely to select this project compared to males. Both effects are not surprising as women are generally more caring than men. It seems logical that older people tend to select this project more often as it is more about their peers.

The second model, the one estimated for rejecting the project, shows more significant variables than the first model. The older the respondent, the less likely he was to reject the project. Furthermore, males were more likely to reject the project compared to women. Both effects are in line with the first model, as it is the opposite. Moreover, model 2 shows that low and middle educated respondents were more likely to reject project 1, compared to high educated respondents. Yet, the higher the

income of the respondent, the more likely he was to reject the project. These effects seem contradicting because to some extend education level and income are related with each other. At first sight, it is not possible to explain these effects. Lastly, respondents living with children were more likely to reject the project than respondents that specified their living situation as 'other'.

Project 2 – Businesses open again (except hospitality and contact-jobs)
The first model, estimated for selecting the project, shows significant p-values for gender, education level, living situation, and province. Men were more likely to select this project than women. This effect was to be expected as a higher percentage of men is employed compared to women (CBS, 2020). Besides, high educated respondents were more likely to select project 2, compared to low and middle educated respondents. Respondents that live alone or with a partner were less likely to select this project than respondents that specified their living situation as 'other'. Lastly, respondents from Groningen and Drenthe were less likely to select the project compared to respondents living in Limburg.

The second model, estimated for rejecting the project, shows significant p-values for education level, living situation, and province. Respondents with a low and middle education level were more likely to reject this project than high educated respondents. Respondents that live alone or with a partner were more likely to reject project 2 compared to respondents that did specify their living situation as 'other'. Both effects are in line with the first model. Lastly, respondents from Groningen and from Friesland were more likely to reject the project than respondents from Limburg.

Project 3 – Contact professions can open again
The first model, estimated for selecting the project, shows significant p-values for age, income, education level, and province. The older the respondent, and the higher his income, the more likely he was to select this project. Low and middle educated respondents were less likely to select the project compared to high educated respondents. Lastly, respondents from Flevoland were more likely to select the projects than respondents from Limburg. These effects are surprising and cannot be explained at first glance.

The second model, estimated for rejecting the project, shows only significant p-values for education level. It shows, in line with the first model, that respondents with a low or middle education level were more likely to reject this project than high educated respondents.

Project 4 – People younger than 18 years do not have to keep 1.5 meter distance
The first model, estimated for selecting the project, shows significant p-values for living situation and province. Respondent living alone or with a partner were less likely to select this project compared to respondents that specified their living situation as 'other'. Respondents living in Flevoland and Zuid-Holland were more likely to select project 4 compared to respondents from Limburg. Again, these effects were not expected. One may expect that participants with child(ren) would have selected this project more often as their children would benefit from this measure. However, this was not the case.

The second model, estimated for rejecting the project, does not show any significant p-values, except for the intercept. This means that whether a respondent rejects this project cannot be explained by these demographic variables.

Project 5 – All restrictions are lifted for immune people
The first model, estimated for selecting the project, shows significant p-values for age and province. The older the respondent, the more likely he was to select this project. Moreover, respondents from Friesland and Zeeland were less likely to select this project compared to respondents who live in Limburg.

The second model, estimated for rejecting the project, shows significant p-values for income, gender, living situation, and province. The higher the income of the respondent, the more likely he was to reject this project. Furthermore, men were less likely to reject the project than women. Respondents that are living with partner and child(ren) are less likely to reject the project compared to those who specified their living situation as 'other'. Lastly, respondents from all provinces, except for Drenthe and Zeeland, were more likely to reject the project compared to respondents from Limburg.

Project 6 – Restrictions are lifted in northern provinces Friesland, Groningen, and Drenthe
The first model, estimated for selecting the project, shows significant p-values for age and province. The older the respondent, the more likely he was to select this project. Moreover, respondents from Groningen and Zeeland were more likely to select this project, compared to respondents from Limburg. Surprisingly, respondents from Friesland and Drenthe do not select this project more often. There is only a location effect visible for respondents from Groningen.

The second model, estimated for rejecting the project, shows significant p-values for gender, education level, living situation, and province. Women were more likely to reject this project than men. Low educated respondents were less likely to reject project 6 compared to high educated respondents. Respondents living with a partner, as well as partner and child(ren) were less likely to reject this project, compared to those who specified their living situation as 'other'. Respondents living in Flevoland, Utrecht, and Noord-Brabant were more likely to reject this measure compared to respondents from Limburg.

Project 7 – Direct family members from other households do not have to hold 1.5 meter distance
The first model, estimated for selecting the project, shows significant p-values for education level, living situation, and province. Respondents with a low education level were less likely to select this project than respondents with a high education level. Moreover, respondents living alone, with a partner, and living with roommates were less likely to select the project compared to those who specified their living situation as 'other'. Lastly, respondents from Overijssel were less likely to select this project than respondents from Limburg.

The second model, estimated for rejecting the project, shows significant p-values for gender and education. Men were more likely to reject the project than women, which is not remarkable given the nurturing nature of women. Moreover, low, and middle educated respondents were more likely to reject the project compared to high educated respondents.

Project 8 – Hospitality and entertainment sector open again
The first model, estimated for selecting the project, shows only a significant p-value for age. The younger the respondent, the more likely he was to select this project. This effect seems logical as younger people tend to be more outgoing in general.

The second model, estimated for rejecting the project, shows significant p-values for age, education, and province. The older the respondent, the more likely he was to reject this measure. This is in line with the effect of the first model. Moreover, middle educated respondents were more likely to reject this project than high educated respondents. Lastly, respondents from Friesland, Drenthe, Gelderland, and Noord-Holland were more likely to reject this project than respondents from Limburg.

Table 5.7 Binomial regression models panel PVE

	Project 1		Project 2		Project 3		Project 4		Project 5		Project 6		Project 7		Project 8	
	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected
McFadden pseudo R^2	0.026	0.023	0.017	0.028	0.017	0.015	0.017	0.010	0.017	0.015	0.025	0.017	0.016	0.016	0.011	0.011
	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
Intercept	-1.363 **	-1.134 **	0.356	-2.672 **	-0.384	-1.945 **	-0.554 *	-1.224 **	-1.718 **	-0.571 *	-3.664 **	0.202	0.510 *	-2.154 **	-1.102 **	-0.828 **
age	0.204 **	-0.155 **	-0.034	-0.036	0.077 **	-0.001	-0.012	0.051	0.078 *	-0.024	0.113 *	0.045	-0.014	0.044	-0.089 **	0.059 *
income	0.001	0.093 **	0.041	0.049	0.082 **	0.037	0.025	0.010	-0.023	0.076 *	-0.062	0.054	-0.027	0.046	0.004	-0.022
[gender=1]	-0.332 **	0.369 **	0.177 *	0.077	0.006	-0.001	0.117	-0.063	0.116	-0.190 *	0.15	-0.236 **	-0.158	0.242 *	0.104	-0.090
[gender=2]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]
[educ=1]	0.055	0.41 **	-0.507 **	0.600 **	-0.464 **	0.649 **	-0.19	0.114	0.18	-0.055	-0.027	-0.502 **	-0.320 **	0.444 **	-0.169	0.011
[educ=2]	0.047	0.222 *	-0.233 **	0.369 **	-0.259 **	0.534 **	-0.169	0.176	0.225	-0.005	0.033	-0.173	0.091	0.266 *	-0.130	0.324 **
[educ=3]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]
[living situation=1]	-0.010	0.007	-0.397 *	0.799 **	-0.153	0.122	-0.613 **	0.169	-0.053	-0.166	0.365	-0.339	-0.571 **	0.437	0.229	0.076
[living situation=2]	-0.102	0.116	-0.365 *	0.589 *	0.043	-0.016	-0.496 **	0.040	-0.116	-0.159	0.550	-0.454 *	-0.429 *	0.323	0.255	0.044
[living situation=3]	-0.097	0.053	-0.326	0.517	-0.119	-0.114	-0.172	-0.251	0.219	-0.413 *	0.642	-0.540 **	-0.330	0.393	0.263	-0.048
[living situation=4]	-0.530	0.615 *	-0.180	-0.114	0.171	0.003	-0.059	-0.129	0.24	-0.182	0.819	-0.422	0.098	0.170	0.530	0.061
[living situation=5]	-0.151	0.148	-0.161	-0.175	0.090	-0.031	-0.268	-0.256	-0.467	0.219	0.427	0.123	-0.629 **	0.532	0.149	0.048
[living situation=6]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]
[province=1]	0.05	0.355	-0.607 *	0.869 **	-0.224	0.291	0.119	0.450	-0.306	0.568 *	1.262 **	0.095	-0.161	0.166	-0.353	0.410
[province=2]	-0.056	0.029	-0.231	0.71 *	-0.052	-0.073	0.317	0.019	-1.399 **	0.889 **	0.760	0.108	-0.149	-0.195	-0.637	0.210
[province=3]	0.214	0.335	-0.581 *	0.243	0.433	0.285	0.111	0.193	-0.017	0.112	0.649	0.269	0.375	-0.223	-0.641	0.642 *
[province=4]	0.011	0.194	-0.195	0.360	0.107	0.379	0.291	0.129	-0.582	0.730 **	0.309	0.208	-0.444 *	-0.281	-0.284	0.500 *
[province=5]	-0.341	0.272	-0.039	-0.043	0.746 **	-0.076	1.046 **	-0.051	-0.361	0.535 *	0.088	0.534 *	0.149	-0.561	-0.484	0.435
[province=6]	-0.223	0.020	-0.272	-0.125	-0.049	-0.022	0.310	-0.005	-0.134	0.367 *	0.534	0.206	0.104	-0.275	-0.148	0.482 **
[province=7]	-0.090	0.199	-0.186	0.094	0.101	0.153	0.078	0.263	-0.246	0.427 *	0.433	0.430 *	-0.134	-0.117	-0.040	0.280
[province=8]	0.037	-0.032	-0.273	-0.097	0.095	0.044	0.256	-0.028	-0.293	0.360 *	0.342	0.212	0.078	-0.351	-0.064	0.454 **
[province=9]	-0.070	0.219	-0.193	-0.055	0.036	-0.105	0.363 *	0.076	-0.217	0.395 *	0.348	0.117	0.008	-0.326	-0.330	0.401
[province=10]	0.127	0.059	0.017	-0.152	-0.005	0.207	0.364	0.130	-0.966 *	0.368	0.980 *	0.016	-0.146	0.311	-0.090	0.164
[province=11]	-0.055	0.002	-0.201	-0.099	0.225	0.033	0.351	0.109	-0.328	0.666 **	0.069	0.455 **	0.112	-0.327	-0.019	0.273
[province=12]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]

Note: a) The reference category is 0, b) This parameter is set to zero because it is redundant c) *significant for p<0.05, ** significant for p<0.01

Table 5.8 Binomial regression models open PVE

	Project 1		Project 2		Project 3		Project 4		Project 5		Project 6		Project 7		Project 8	
	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected	Selected	Rejected
McFadden pseudo R^2	0.039	0.020	0.026	0.012	0.013	0.018	0.015	0.015	0.008	0.005	0.038	0.005	0.005	0.015	0.031	0.026
	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
Intercept	-1.026 **	-1.226 **	-0.623 **	-2.18 **	0.096	-1.972 **	-0.366 **	-1.184 **	-2.911 **	0.540 **	-3.948 **	0.744 **	-0.160	-2.215 **	-0.832 **	-0.726 **
age	0.244 **	-0.133 **	0.063 **	0.017	0.002	0.124 **	0.057 **	0.073 **	0.026	-0.028 **	0.095 **	-0.052 **	-0.059 **	0.185 **	-0.202 **	0.215 **
income	-0.047 **	0.006	0.073 **	-0.082 **	0.124 **	-0.129 **	0.019 *	-0.046 **	0.009	0.000	-0.065 **	0.033 **	-0.027 **	-0.005	0.122 **	-0.081 **
[gender=1]	-0.650 **	0.512 **	0.519 **	-0.280 **	0.212 **	-0.003	0.098 **	0.012	0.340 **	-0.185 **	0.316 **	-0.107 **	-0.100 **	0.065	0.304 **	-0.130 **
[gender=2]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]
[educ=1]	-0.326 **	0.585 **	-0.598 **	0.392 **	0.125	0.316 **	-0.345 **	0.284 **	0.273 *	-0.366 **	-0.141	-0.264 **	0.212 **	0.281 **	-0.047	0.034
[educ=2]	-0.149 **	0.392 **	-0.387 **	0.233 **	0.104 *	0.217 **	-0.166 **	0.274 **	0.117	-0.101 *	0.002	0.009	0.054	0.115 *	0.024	0.067
[educ=3]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]
[living situation=1]	-0.150	-0.009	-0.012	0.134	0.086	-0.135	-0.495 **	0.008	0.287	-0.110	0.104	-0.06	0.178 *	-0.070	0.273 **	-0.321 **
[living situation=2]	-0.120	0.065	-0.029	0.279	-0.057	-0.077	-0.574 **	0.056	0.142	-0.063	0.169	-0.03	0.277 **	-0.058	0.177 *	-0.250 **
[living situation=3]	-0.190 *	0.006	-0.074	0.158	0.056	-0.135	-0.104	-0.304 **	0.248	-0.219 **	0.173	-0.156	0.400 **	-0.225	0.210 *	-0.302 **
[living situation=4]	-0.228	0.036	-0.151	0.188	0.124	-0.132	-0.048	-0.4 **	0.643 **	-0.520 **	0.262	-0.267 *	0.415 **	-0.335 *	0.334 **	-0.423 **
[living situation=5]	-0.235 *	0.087	0.330 **	-0.048	-0.021	-0.282	0.217 *	-0.27 *	0.237	-0.168	0.209	-0.112	-0.126	0.200	0.437 **	-0.530 **
[living situation=6]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]
[province=1]	0.158	0.013	-0.062	0.365	-0.532 **	0.229	-0.022	0.088	-0.292	-0.097	1.745 **	-0.280	-0.222	0.098	-0.472 **	0.380 **
[province=2]	-0.009	0.109	-0.255	0.269	-0.217	-0.086	-0.066	-0.285	-0.217	-0.228	1.817 **	-0.367 **	-0.182	-0.101	-0.118	0.107
[province=3]	0.328 *	-0.118	-0.433 **	0.500 *	-0.289 *	0.049	0.035	-0.111	-0.192	0.098	1.816 **	-0.118	-0.014	-0.108	-0.481 **	0.197
[province=4]	0.255 **	-0.079	0.010	-0.035	-0.080	-0.126	-0.012	-0.018	-0.089	-0.060	0.438	-0.180 *	0.026	-0.092	-0.245 *	0.198 *
[province=5]	0.127	0.160	-0.021	0.430 *	-0.329 **	-0.061	-0.187	0.059	-0.030	0.076	0.394	-0.089	0.136	0.022	-0.203	0.134
[province=6]	0.293 **	-0.085	0.038	-0.016	-0.147	-0.015	0.184 *	-0.122	-0.023	-0.064	-0.046	0.002	0.035	-0.024	-0.025	0.006
[province=7]	0.208 *	-0.156	-0.023	0.172	-0.189 *	-0.280	0.157	-0.101	-0.113	0.070	0.383	-0.097	-0.007	-0.128	-0.052	-0.072
[province=8]	0.083	-0.001	0.047	0.203	-0.161	-0.058	0.148	-0.199 *	0.193	-0.178 *	0.444	-0.315 **	-0.049	0.015	0.120	-0.017
[province=9]	0.145	-0.014	-0.010	0.220	-0.202 *	-0.019	0.010	0.031	0.053	-0.046	0.323	-0.129	-0.001	-0.020	-0.061	0.021
[province=10]	-0.010	0.187	-0.040	-0.230	-0.048	-0.118	-0.212	0.146	0.163	-0.162	-0.015	-0.177	-0.051	-0.158	-0.116	0.087
[province=11]	0.052	-0.052	0.051	0.035	0.011	-0.289	0.06	-0.081	-0.008	-0.046	0.164	0.046	-0.034	-0.021	0.109	-0.034
[province=12]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]	0[b]

Note: a) The reference category is 0, b) This parameter is set to zero because it is redundant c) *significant for p<0.05, ** significant for p<0.01

5.3.2 Regression models open PVE

Table 5.8 provides an overview of the regression models estimated on the open dataset. First, an interpretation of the pseudo R^2 is given, followed by an interpretation of the regression models for each project.

Interpretation McFadden pseudo R^2

First, there is looked at the pseudo R^2 of the estimated models. The values range between 0.005 and 0.039. For some models, the pseudo R^2 is slightly better for the open PVE than for the panel PVE. In some cases, the pseudo R^2 is even worse. All in all, the models have low predictive power.

Project 1 – Nursing and care homes allow visitors

The first model, estimated for selecting the project, shows significant p-values for all demographic variables. The older the respondent and the lower his income, the more likely he was to select this project. Women were more likely to select project 1 than men. These effects were also found for the panel dataset and are not surprising. Respondents with a low and middle education level were less likely to select this project than respondents with a high education level. This may be explained by the fact that low and middle educated people are working more often in nursing and care homes. They could be more holding back as allowing visitors increases the risk of infection. Respondents living with partner and child(ren) as well as respondents living with roommates are less likely to select this project than respondents who specified their living situation as 'other'. Respondents from Drenthe, Overijssel, Gelderland, and Utrecht were more likely to select project 1 compared to respondents from Limburg.

The second model, estimated for rejecting the project, shows significant p-values for age, gender, and education level. The older the respondent, the less likely he was to reject the project. Besides, men were more likely to reject this project than women. Finally, low, and middle educated respondents were more likely to reject this project than high educated respondents. All effects are the opposite of, and therefore in line with, the first model.

Project 2 – Businesses open again (except hospitality and contact-jobs)

The first model, estimated for selecting the project, shows significant p-values for all demographic variables. The effects for gender and education level are the same as identified in the panel dataset: men were more likely to select the project than women and low and middle educated respondents were less likely than high educated respondents to select the measure. The older the respondent and the higher his income, the more likely he was to select this project. Moreover, respondents living with roommates were more likely to select the project compared to respondents that specified their living situation as 'other'. Finally, respondents from Drenthe were less likely to select the project compared to respondents from Limburg.

The second model, estimated for rejecting the project, shows significant p-values for income, gender, education level, and province. The higher the income of the respondent, the less likely he was to reject this project. Moreover, men were less likely than women to reject the project. Also, low, and middle educated respondents were more likely to reject the project. Lastly, respondents from Drenthe and Flevoland were more likely to reject project 2 than respondents from Limburg.

Project 3 – Contact professions can open again

The first model, estimated for selecting the project, shows significant p-values for income, gender, education level, and province. The higher the income of the respondent, the more likely he was to select the project. Besides, men were more likely to select project 3 than women. Respondents with a middle education level were more likely than respondents with a high education level to select the project. Respondents from Groningen, Drenthe, Flevoland, Utrecht, and Zuid-Holland were less likely to select this project compared to respondents from Limburg.

The second model, estimated for rejecting the project, shows significant p-values for age, income, and education level. The older the respondent, the more likely he was to reject this project. The higher his income, the less likely he was to reject the project. Lastly, low, and middle educated respondents were more likely than high educated respondents to reject the measure. This effect is remarkable as it means that low and middle educated respondents were more likely to select as well as reject the project.

Project 4 – People younger than 18 years do not have to keep 1.5 meter distance
The first model, estimated for selecting the project, shows significant p-values for all demographic variables. The older the respondent and the higher his income, the more likely he was to select the project. Also, men were more likely to select project 4 than women. Low and middle educated respondents were less likely than high educated respondents to select the project. Respondents living alone or with a partner were less likely to, and respondents living with roommates were more likely to select project 4 compared to those who specified their living situation as 'other'. Finally, respondents from Gelderland were more likely to select the measure.

The second model, estimated for rejecting the project, shows significant p-values for age, income, education level, living situation, and province. This is interesting, as the same model for the panel dataset did not show any significant relations at all.
The older the respondent, the more likely he was to reject the measure. This effect is remarkable as it means that the older the respondent, the more likely he was to select as well as reject the same project. The higher his income, the less likely he was to reject the measure. Moreover, low, and middle educated respondents were more likely to reject the project than respondents with a high education level. Respondents living with partner and child(ren), only with child(ren) or with roommates were less likely to reject the project. Lastly, respondents from Noord-Holland were less likely than respondents from Limburg to reject the measures.

Project 5 – All restrictions are lifted for immune people
The first model, estimated for selecting the project, shows significant p-values for gender, education level, and living situation. Men were more likely than women to select this project. In addition, low educated respondents were more likely to select project 5 than high educated respondents. Lastly, respondents living with child(ren) were more likely to select this project than respondents who specified their living situation as 'other'.

The second model, estimated for rejecting the project, shows significant p-values for age, gender, education level, living situation, and province. The older the respondent, the less likely he was to reject this project. Moreover, men were less likely to reject the project than women. The same effect is visible for low and middle educated respondents compared to high educated respondents. Respondents living with partner and child(ren) or only with chid(ren) were less likely to reject this project as well, compared to respondents that specified their living situation as 'other'. Lastly, respondents from Noord-Holland were less likely to reject this project compared to respondents from Limburg.

Project 6 – Restrictions are lifted in northern provinces Friesland, Groningen, and Drenthe
The first model, estimated for selecting the project, shows significant p-values for age, income, gender, and province. The older the respondent, and the lower his income, the more likely he was to select project 6. In addition, male respondents were more likely to select this project compared to female respondents. Lastly, respondents from Groningen, Friesland and Drenthe were more likely to select this project than respondents from Limburg. For this measure, a location effect is visible: the respondents from provinces who benefit from this measure select it more often.

The second model, estimated for rejecting the project, shows significant p-values for all demographic variables. The younger the respondent and the higher his income, the more likely he was to reject this

project. In addition, men were less likely to reject this project than women. This same effect is visible for low educated respondents compared to high educated respondents. Respondents living with child(ren) were also less likely to reject the project, compared to those who specified their living situation as 'other'. Finally, respondents living in Friesland, Overijssel and Noord-Holland were less likely than respondents living in Limburg to reject the project.

Project 7 – Direct family members from other households do not have to hold 1.5 meter distance
The first model, estimated for selecting the project, shows significant p-values for age, income, gender, education level, and living situation. The older the respondents and the higher his income, the less likely he was to select this project. In addition, men were less likely to select project 7 than women. Low educated respondents were more likely to select this project compared to high educated respondents. Lastly, respondents living alone, with partner, partner, and child(ren) or only with child(ren) were more likely to select this project than those who specified their living situation as other.

The second model, estimated for rejecting the project, shows significant p-values for age, education level and living situation. The older the respondent, the more likely he was to reject the project. Moreover, low, and middle educated respondents were more likely to reject the project compared to high educated respondents. This effect is remarkable as it means that low and middle educated respondents were more likely to select as well as reject the project. Lastly, respondents living with child(ren) were less likely to reject this measure compared to those who specified their living situation as 'other'.

Project 8 – Hospitality and entertainment sector open again
The first model, estimated for selecting the project, shows significant p-values for all demographic variables except for education level. The younger the respondent and the higher his income, the more likely he was to select this project. Moreover, men were more likely to select project 8 than women. For living situation, respondents from all other categories were more likely to select this project compared to those who specified their living situation as 'other'. Lastly, respondents from Groningen, Drenthe, and Overijssel were less likely to select this measure compared to respondents from Limburg.

The second model, estimated for rejecting the project, shows significant p-values for the same variables as the first model. For all effects, those who were more likely to select the project were less likely to reject the project and the other way around. The only difference is that respondents from Drenthe do not show a significant effect for rejecting the measure.

5.3.3 Conclusion regression models

The binomial regression models showed low predictive power which means that the demographic variables are only able to explain the preferences of respondents regarding the COVID-19 measures to a limited extend. Moreover, it became clear that the method of sampling is indeed a moderating variable: the sampling method influenced the effects of the demographic variables on the project choice. The binomial regression models of the open dataset showed more statistically significant effects. This can be explained by the larger number of respondents in the open dataset. The more responses, the more likely it is to find statistically significant relations. Sometimes the effects of the open dataset were opposite of the effects of the panel dataset.

5.4 Weight adjustment results

To determine the total effect of the bias in the demographic variables, the data is reweighted. When reweighting the data on a variable, called an auxiliary variable, the sample distribution of the auxiliary variable is matched with the corresponding population distribution of that variable. Data can therefore only be reweighted on variables of which the population distributions are known (Engel et al., 2014; Tarima & Pavlov, 2006). For this PVE experiment, the data can be reweighted on gender, age, education level, and province since the population distributions are known. It is not possible to reweight the data for income and living situation. As both samples were representative for age, education level, and province the samples are reweighted on these variables. As the underlying distribution may change for variables that are not considered, gender is also included in the reweighting process.

The method used to reweight the data is iterative proportional fitting (IPF), also known as ranking. IPF is the most frequently used weighting method for surveys (Mercer et al., 2018). It is based on the marginal distributions of the auxiliary variables. This means that the totals of a variable are aligned with the known population totals. In this study, IPF is implemented using the Python Quantipy package (Müller et al., 2019). For more information on the weight adjustment approach, see section 2.3.5. For more information about the different weighting methods see appendix B.

Missing values

Auxiliary variables usually do not have a category for missing values. There are different options to handle missing values in this context (Battaglia et al., 2004; Kolenikov, 2014). The easiest option is to restrict the sample to the records that do not show any missing values for all auxiliary variables. A drawback is that only a part of the sample is used. For this PVE experiment, this option is not desirable as both datasets contain a serious percentage of missing values for the demographic variables. For the panel dataset this is around 16%, and for the open dataset around 20%. A second option is to substitute the missing values before weight adjustment takes place. Yet, Battaglia et al. (2004) argue this is only suitable when the percentage of missing values is nontrivial, which is not the case for the panel sample and the open sample. This leaves the option to set the weight of the records containing missing values to 1. The weight of these records is not alert during the weight adjustment process. In this way, the preferences of these respondents are taken into account in the reweighted results.

5.4.1 Weight adjustment panel PVE

Table 5.9 shows the details of the weight adjustment process for the panel dataset. The weighting efficiency indicates how balanced the dataset is. A weighting efficiency below 80% indicates a high mismatch between the sample and the population. The weight efficiency for the panel dataset is 84%, which means that reweighting this sample was acceptable. The weight factor ratio shows the ratio between the smallest and biggest weight factor, which is in this case 9.91. This means that some records count 9.91 times as often as other records.

Table 5.9 Metrics weight adjustment panel PVE

Weighting efficiency	84.23%
Iterations required	22
Mean weight factor	1
Minimum weight factor	0.31
Maximum weight factor	3.12
Weight factor ratio	9.91

Figure 5.1 shows the preferences of the respondents from the panel PVE before and after weight adjustment.

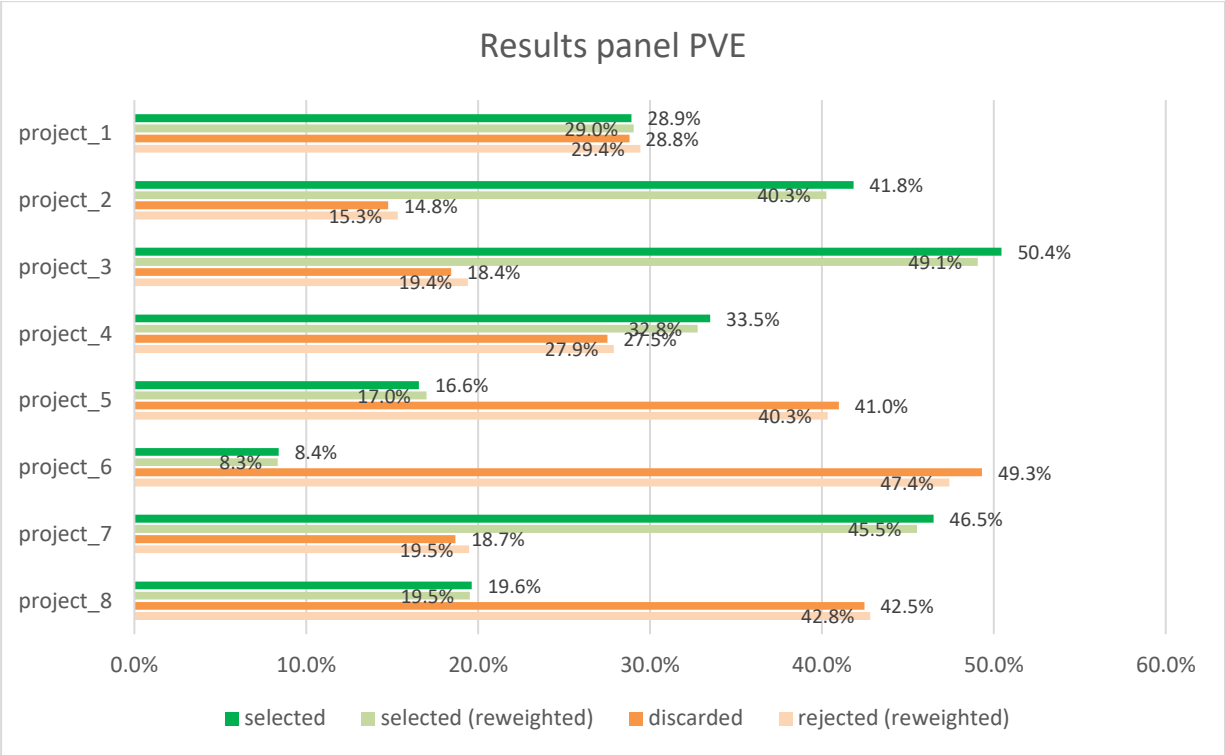

Figure 5.1 Panel PVE reweighted

Chi-square tests were performed to see whether the results before and after reweighting are statistically different. The results are shown in Table 5.10.

Table 5.10 Results Chi-square test panel dataset

Project	Selected		Rejected	
	p-value	Chi-square value	p-value	Chi-square value
1	0.868	0.03	0.417	0.66
2	0.062	3.47	0.370	0.80
3	0.107	2.60	0.154	2.03
4	0.369	0.81	0.621	0.24
5	0.491	0.47	0.444	0.59
6	0.911	0.01	0.027	4.89
7	0.267	1.23	0.244	1.36
8	0.871	0.03	0.688	0.16

From Table 5.10 can be concluded that the percentage of participants that select a policy measure before weight adjustment does not differ from the percentage after weight adjustment. So, although there were small biases for the demographic variables of the panel dataset, this does not affect the results of the PVE experiment.

From Table 5.10 can also be concluded that the percentage of participants that reject a policy measure does not differ before and after weight adjustment, except for measure 6. Here the difference is significant for $p < 0.05$, but not for $p<0.01$. So, it can be stated that the difference is significant with 95% confidence, but not with 99% confidence. Looking at Figure 5.1 it becomes clear that after weight adjustment the participants reject this measure less often (difference of 1.9%). In other words, the resistance is a bit less than in the case of the unweighted panel sample. However, this small change does not have an effect on the ranking of the options that are rejected the most: the top three rejected measures are still project 5, 6, and 8. Overall, it can be concluded that the original panel sample

provided an accurate reflection of the preferences of Dutch citizens, based on their gender, age, education level, and province.

5.4.2 Weight adjustment open PVE

Table 5.11 shows the details of the weight adjustment process for the open dataset. The weighting efficiency of 28% shows that there was a huge mismatch between sample and population. Dropping below a weight efficiency of 70% is a sign to re-examine the weight scheme specifications (Müller et al., 2019). The mismatch between the sample of the open PVE and population is mainly caused by the education level respondents (see section 5.1).

Also, there is a very high weight factor ratio of 220. This ratio is way higher than the ratio reported for the panel dataset. When weights are highly distributed and have a low association with the dependent variables, the estimators tend to be unstable. There are various approaches to reduce the variability in weights such as weight trimming, weight modelling and weight modification (Chen et al., 2017). Yet, this is out of scope for this study. For the sake of comparison, the results of reweighted sample are shown in Figure 5.2.

Table 5.11 Metrics weight adjustment open PVE

Weighting efficiency	27.93%
Iterations required	20
Mean weight factor	1
Minimum weight factor	0.10
Maximum weight factor	22.76
Weight factor ratio	220.28

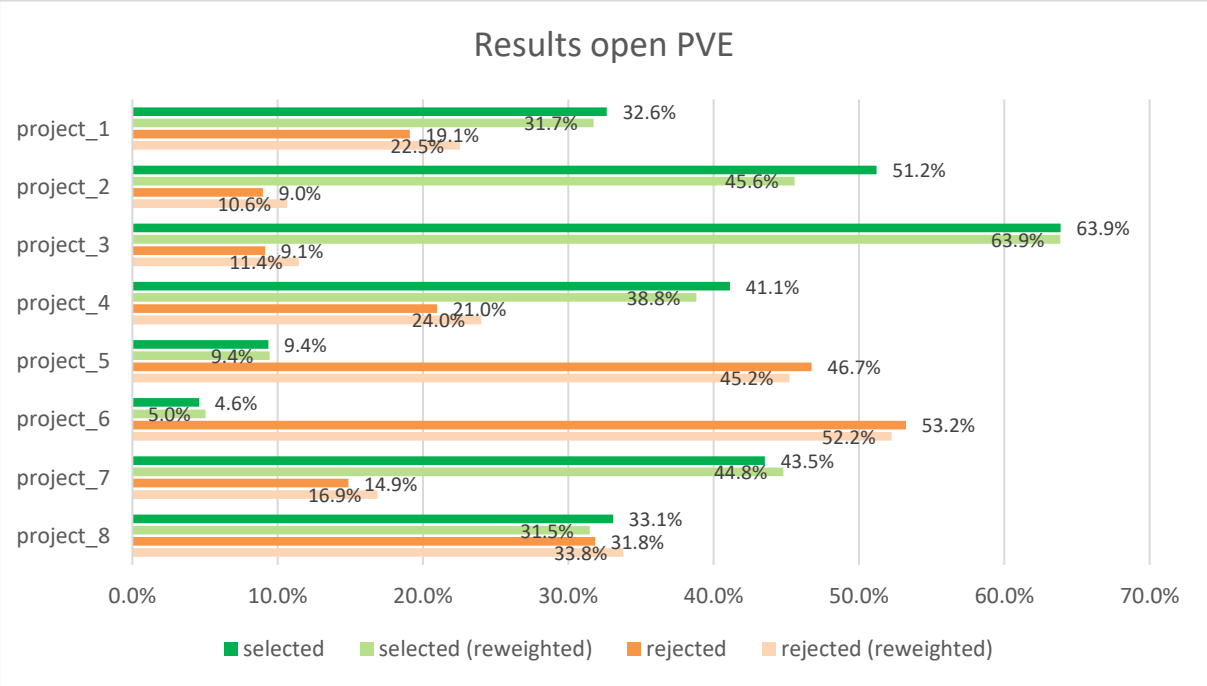

Figure 5.2 Open PVE reweighted

Again, Chi-square tests were performed to see whether the results before and after reweighting are statistically different. The results are shown in Table 5.12.

Table 5.12 Results Chi-square test panel dataset

Project	Selected		Rejected	
	p-value	Chi-square value	p-value	Chi-square value
1	0.002	10.04	0.000	178.80
2	0.000	26291.00	0.000	76.97
3	0.981	0.00	0.000	138.19
4	0.000	60.08	0.000	132.55
5	0.665	0.19	0.000	24.82
6	0.001	11.08	0.001	10.39
7	0.000	17.33	0.000	75.33
8	0.000	31.51	0.000	44.59

Table 5.12 shows that with 99% certainty there is no significant difference between selecting measure 3 and measure 5 before and after weight adjustment. For all other cases, the difference before and after weight adjustment is significant. Measures 1, 2, 4, and 8 are selected less often and rejected more after weight adjustment. Measures 3 is also rejected more often after reweighing and measure 5 less often. Measure 6 is also rejected less often but selected more often after reweighing the data. Lastly, the effects for measure 7 are remarkable: this project is selected more often as well as rejected more often after correction for the representation bias. Overall, the ranking of the policy measures does not change.

5.4.3 Comparison panel sample and open sample

Table 5.13 provides a complete overview for selecting a policy measure for both samples, before and after weight adjustment. Table 5.14 provides the same overview for rejecting the policy measures.

Table 5.13 Overview selected policy measures

Selected	Panel	Reweighted panel	Open	Reweighted open
project_1	28.9%	29.0%	32.6%	31.7%
project_2	41.8%	40.3%	51.2%	45.6%
project_3	50.4%	49.1%	63.9%	63.9%
project_4	33.5%	32.8%	41.1%	38.8%
project_5	16.6%	17.0%	9.4%	9.4%
project_6	8.4%	8.3%	4.6%	5.0%
project_7	46.5%	45.5%	43.5%	44.8%
project_8	19.6%	19.5%	33.1%	31.5%

Table 5.14 Overview rejected policy measures

Rejected	Panel	Reweighted panel	Open	Reweighted open
project_1	28.8%	29.4%	19.1%	22.5%
project_2	14.8%	15.3%	9.0%	10.6%
project_3	18.4%	19.4%	9.1%	11.4%
project_4	27.5%	27.9%	21.0%	24.0%
project_5	41.0%	40.3%	46.7%	45.2%
project_6	49.3%	47.4%	53.2%	52.2%
project_7	18.7%	19.5%	14.9%	16.9%
project_8	42.5%	42.8%	31.8%	33.8%

It can be concluded that the differences between the original panel sample and the reweighted panel sample are smaller than the differences between the unweighted datasets. However, correcting for representation bias in demographic variables still results in differences in the share of participants that select or reject a measure. This means that there are other characteristics (either observed or unobserved) account for the differences. From a practical point of view, the open and panel PVE are likely to result in the same policy decision as the top three favoured measures are the same (measure 2, 3, and 7).

5.5 Conclusion effect of representation bias on policy choice

This chapter aimed to answer the sub question to what extend bias in demographic characteristics influenced the preferences of participants in both datasets.

Analysing the datasets of the COVID-19 PVE experiment showed differences in preferences among participants in the open PVE and the panel PVE. The top three selected and the top three rejected measures were the same for both datasets. However, there were small differences in the ranking of the measures and the share of participants that supported or resisted a measure differed quite a lot.

It was analysed to what extend bias in demographic variables sample causes these differences. It became clear that both datasets were only representative for gender. The distributions of age, education level, and province in both samples statistically deviated from the population distributions. Yet, the bias for the open PVE was much larger than for the panel PVE. Looking into the effect of the demographic variables on participants' preferences showed that the sampling method influenced the relations that were found. The open PVE showed more significant relations. Yet, both the Cramer's V tests, and the binominal regression models showed that the demographic variables could only explain the participants' preferences to a very limited extend.

Reweighting the panel sample did not result in different outcomes, which means that the panel provided an accurate picture of participants' preferences. Reweighting the open sample resulted in a low weighting efficiency, which indicated that the bias in the open dataset with regard to age, education level and province was too big to properly correct. Moreover, the weights were highly dispersed, and in combination with the low association with the dependent variables this leads to unstable estimates.
For the sake of comparison, the results for correcting for representation biased in the open sample were reported and showed that the gap in the share of participants that selected or rejected an option in the panel and in the open PVE was bridged. Yet, the correction could not account for the whole gap. This means there are differences between the respondents from the panel and the open PVE that could not be explained by the demographic variables. From a practical perspective the differences in the share of participants that selected a policy measure are less important, as the top three favoured measures are the same for both samples and are likely to lead to the same decision in the policy context.

Chapter 6 – Interviews

Section 6.1 elaborates on the set-up of the interviews. Section 6.2 provides the main findings of the interviews. Lastly, section 6.3 draws a conclusion and thereby answers sub question 3.

6.1 Interview set-up

The goal of the interviews was to find out how experts reflect on representativity, inclusivity, the consequences of representation bias in a sample and how they review weight adjustment as a method to correct for this. In total five interviews were conducted in an online environment using Zoom. For each interview, a total of 1 hour was available. A semi-structured approach was used to allow for a in depth conversation with the participants.

The people that have been interviewed have a diverse background to ensure a board perspective on the topic. Their experience with PVE, weight adjustment and communicating research results with policy makers has been taken into account. Table 6.1 provides an overview of the participants and their background. For more information about the set-up, see section 2.4. The interview questions can be found in appendix E.2.

Table 6.1 Overview interview respondents

Respondents	Background
1	This respondent has a background in public participation and co-creation. He is a member of the PVE team of TU Delft, which means he has a lot of experience with PVE.
2	This respondent works for the Ministry of Health, Welfare and Sport. He was involved by only one PVE experiment and is not really experienced with weight adjustment. For his work he is often involved in communicating research results to policy makers.
3	This respondent has a background in energy governance. She is also part of the PVE team of TU Delft. She is often involved in communicating results of PVE experiments towards policy makers. In some of the PVE experiments she was involved in weight adjustment was used.
4	This respondent does a PhD about lifestyle interventions for people with a low socioeconomic status. She carried out her own PVE experiment on this subject. She is not experienced with weight adjustment nor communicating research results to policy makers.
5	This respondent has a background in labour economics, environmental economics, and measurement of preferences for public policy. She was seriously involved in two PVE experiments. In one of them weight adjustment was used. In this experiment she was also involved in communicating the results to the municipality.

6.2 Interview results

The complete interviews can be found in appendix E.3. The following sections present the main observations of the interviews. The quotes presented below have been translated from Dutch to English. The original quotes can be found in Appendix E.4.

6.2.1 General experience with PVE

Overall, the participants were positive about PVE. Interviewees pointed out that PVE is a suitable method to let participants consider different options: *"I find it especially important that it concerns concrete alternatives that you can compare and weigh against each other. Especially that you can see the impact. That you also see the disadvantages."* (Respondent 1)

Remarkable is that PVE is mainly used in the context of citizen participation and not as an alternative for a Social Cost Benefit Analysis (SCBA), for which it intentionally was developed. The participants mentioned different reasons to use PVE:

a. To measure the level of support for different policy measures
b. To gain insight in the preferences and motivations of the respondents regarding the policy options
c. To provide the respondents with information about the issue at stake
d. To involve citizens' in the decision-making process
e. As an opportunity for citizens to share their views on a topic
f. To reach a large group of participants

The respondents also mentioned various disadvantages of the method. It was mentioned that the method can demand a lot from the people that are involved in developing the experiment. A lot of information is needed about different policy options and their effects that are presented in a PVE experiment. Sometimes it is not easy for a researcher to gain this information from policy makers. Moreover, policy makers can exert a lot of influence on the content of a PVE in this process. As a result, one option may be put in a more positive light than another option.

Another point of attention is that PVE aims to be inclusive, but it is difficult to determine how inclusive the method really is. In the end there are always people who do not participate. This can be for various reasons. Firstly, it is pointed out that PVE demands quite a lot from the participants to make a well-considered choice. Not everyone can afford to invest the required time in the research, for other the task in a PVE may be too complicated to understand. Respondent 4 illustrated that her target group consists of various people with a migration background. These people do not all have a sufficient understanding of the Dutch language to understand the issue that is presented in her PVE. Respondent 5 pointed out that is hard to find the right level of complexity *"because different respondents simply want a different amount of information."* Besides some people think that others do not find their opinion important and therefore do not participate. Lastly, some people simply do not want to participate, which can be seen as a protest vote.

To involve as many people as possible, different recruitment methods were used by the interviewee respondents. This included sending letters and mails as well as distributing flyers. Moreover, the possibilities for a simplified PVE and an offline PVE are being explored within the PVE team.

6.2.2 Representativity

Definition of representativity

Different definitions of representativity are used. Researchers are interested to what extend a sample correctly reflects the entire target group, so that reliable statements about the target group can be made. Respondent 3 pointed out that the terms representativity, inclusivity and diversity are used interchangeably in public. For citizens representativity is more about legitimacy and the fact that you can choose who speaks on your behalf. So even if a sample is representative, it still can lead to resistance: *"People have the feeling that it is not right to conclude anything about a large group based on a subgroup."* She said this is because people think they are more unique than they actually are.

In contrast candidate 1 said: *"I think we are much more fragmented than we think."* He argued that a sample of 1.000 respondents (which most researcher assume is sufficient) is probably not sufficient to capture the variety of opinions. He indicates that you rather need 10.000 people. Furthermore, he mentioned that representativity has an additional dimension for him, called argumentative representativity. It is about whether all arguments of the debate are represented in the research: *"If there are only 4 options in the PVE, but there are 5 other options that everyone is discussing, then the research is not really representative for the discussion in reality."*

Approaches to determine whether a sample is representative

A sample can only be representative for certain characteristics. Often standard demographic characteristics are used to check for representativity such as age and gender. Yet, all interviewees indicated that there are other characteristics that play a role in someone's choice. In context of the PVE of relaxation of COVID-19 measures this may include whether the respondent has a disease or their perceived risk of getting infected with the virus in general. However, it is not possible to check the sample on these characteristics if you do not know how they are distributed among the target group. Some characteristics are not measurable, which means that there are no reference figures. Some characteristics are measurable, but the reference figures may not be accessible for the researcher.

To check if a sample is representative, the interviewees looked at how the characteristics are distributed in the sample. Some checked whether the population percentages visually correspond sufficiently with the sample percentages. Others checked this based on a statistical test. Respondent 2 did not check himself if the research was representative, he relied on the expertise of other researchers.

The importance of a representative sample

It is mentioned that representativity is important to make accurate statements about the target population under investigation. *"From a policy perspective representativity is a basic assumption. We assume that it is guaranteed by the researchers, scientists."* (Respondent 2).

Yet, it rarely occurs that a sample is representative for all specified characteristics. Respondent 4 stated the following: *"If your sample is not representative for the purpose of your research, I think the problem is especially that you cannot say with certainty that the results are correct for the target group you are focusing on."* A lack of representativity does not mean that the research cannot be used at all. However, the usability is limited as the results do not apply to the entire target group. Researchers must make clear remarks for which part of the population the results are valid. Furthermore, there is a danger that incorrect conclusions are drawn. Lastly, the lack of representativity is often used as a counter argument in the public debate. In literature this is referred to as opportunistic use of research outcomes. It means that the results are used in such a way to enhance one's own interests or to marginalize the interests of other stakeholders (Mouter, 2017).

6.2.3 Panel and open PVE

Interviewees mentioned that the most suitable version of PVE depends on the goal of the research. When they were confronted to make a choice between a panel PVE or an open PVE about relaxation of COVID-19 measures, they all choose the panel version as this sample is the most representative for the target group. Moreover, for policy makers it is easier to justify their choices based on the panel version. Respondent 5 said: *"Even if you don't have that many observations, you know that you are not going in a biased direction. This in contrast to when you have a lot of people from one particular group who all invited each other to click on that survey."* This described effect is called a snowball effect. An advantage of the snowball effect is that subgroups that are normally hard to reach can be identified. A drawback is that the respondents are likely to share the same characteristic, which is likely to lead to overrepresentation of certain subgroups.

Nevertheless, they all agreed that the open version also provides valuable information. Respondent 2 about the two versions: *"The representative sample of 3.500 provides the clearest picture for the whole of the Netherlands. The group of 25.000+ gives an image of those who want to contribute and be heard. It is also good to take note of that."* The open version provides the opportunity for citizens to share their point of view. The qualitative motivations in an open PVE are especially important. The more people participate, the bigger the chance you capture the 'rare' opinions and motivations. The

reactions of the open version may contain certain subgroups that are not (well) represented in the panel version. For example, respondents with diverse migration backgrounds. However, it is important to note that people who decide to participate in case of the open version often have a strong opinion, either positive or negative.

6.2.4 Weight adjustment

Weight adjustment is a method to align the sample distribution of a variable with its population distribution. By doing this, you can correct for representation bias of a sample. Respondent 3 said the following about weight adjustment: *"For us (researchers) it is a nice trick to see if it matters. If you reweight the data and you find out that it does not really matter for the general conclusions of your research. Then that is a nice bonus, it means that you don't have to worry too much about it."* She mentioned that you can restore representativity in terms of the research definition. However, you cannot make up for a lack of legitimacy or inclusivity. Reweighting does not enrich the data with new opinions nor does it allow more people to participate.

The other interviewees were more reserved about weight adjustment. They refer to the complexity of the weight adjustment process. This is illustrated by the following quotes:

- *"Intuitively, I always find weight adjustment a bit fishy, a bit dangerous. You have a small group of which you do not really know whether it is representative of the sub-population in which you are interested. And then you are going to blow it up completely."* (Respondent 5)
- *"So, it is a solution to make a skewed population, to make the results thereof, more representative. But with a lot of questions and comments in mind."* (Respondent 4)
- *"It remains a magic trick to fix something you cannot really fix."* (Respondent 2)

Conditions for weight adjustment

Interviewees mentioned different points to consider when preforming weight adjustment. Respondent 1 thought that weight adjustment is justifiable if you want to provide advice*: "When you say I just want to know what is going on, what the values, the considerations and the perspectives are. Then weight adjustment is an option to achieve a representative advice."* Yet, when the results of the research are more biding, for example in case of a referendum, it is not plausible.

Secondly, weight adjustment should not be applied to small samples or be used to try to make statements about groups that did not participate at all. And even though you have a large sample, attention must be paid to the underlying distributions of the characteristics. Some interviewees pointed out that only looking at the marginal distributions of characteristics is not sufficient. During the interviews, an example of weight adjustment on education level was shown to the respondents. Only small changes were visible in the preferences of respondents before and after reweighting on education level. Based on this example, Respondent 5 pointed out that when you want to correct for education level, you need to look at each subgroup of low, middle, and high educated respondents separately: what is the variety in the other demographic variables in this group and to what extend is this aligned with the population distributions? Respondent 1 noted that even though education level on itself does not seem to have a lot of influence on the policy choice of respondents, it does not mean that education in combination with another characteristic does not have an influence.

One of the interviewees pointed that the smaller the bias, the more weight adjustment is plausible. In addition, the bigger the group of respondents, the higher the chance that there is variation in the underlying characteristics. If there is only little variation in the underlying characteristics this results in reweighing a specific sub-group which is still not representative for the larger group. For example, when you want to reweigh on education level and you have only low educated men in your dataset, you are still not able to say something about the low educated women.

6.2.5 Communication

Lastly, some questions were asked about communicating research results to others, with the focus on how to communicate the topics representativity and weight adjustment.

Respondent 2 was involved in communicating the results of the PVE experiment on COVID-19 measures to the spokespersons of the ministry and the crisis organisation within the ministry. He mentioned that he is not worried about representativity and weight adjustment in the policy context:" *To me, they are both things that happen behind the back door, somewhere in science."* About representativity he said the following*: "In policy making, I am not so worried about representativity. This is mainly discussed in the start-up phase. When I give an assignment, I just assume that it is representative. That they arrange that at the beginning."* He also mentioned that he finds weight adjustment a difficult and complex process. However, when researchers assure him it is needed, he relies on their expertise. He mentioned: "*It is of course easier for policy and debate if something is representative than if it is not."*

Respondents 3 and 5 were involved in PVE experiments where reweighted results were presented to a municipality. Respondent 3 mentioned that it is hard to explain difficulties around representativity with the municipality. On the question how they made the weight adjustment clear to the municipality Respondent 5 answered: *"We (researchers) are very concerned about that, but I actually cannot remember whether we have discussed this with them at all. Probably yes, but it wasn't that penetrating or difficult."* Respondent 5 presented weight adjustment as a thought experiment: what would happen if you made the sample representative for the specified demographic characteristics? However, she experienced that it was difficult to understand for the municipality and that "*Ultimately, a municipality mainly wants to hear whether it is representative or not.*" Such a yes-no answer is not possible to give for a researcher.

To better explain representativity to policy makers interviewees mentioned that it is helpful to provide clear and tangible examples. Moreover, it may help to first let them explain how they define representativity and then compare that with the research definition of representativity.

When explaining weight adjustment in a policy context it is important to do this step by step. Interviewees mentioned that it is important that policy makers can understand what happened, otherwise they will not be able to explain it to others. However, as Respondent 2 mentioned it is a difficult and complex process. According to several interviewees, demonstrating the level of association of a characteristic and the preferences of participants is a good intermediate step to demonstrate whether a reweighting has an effect or not. However, the output of the statistical test itself requires quite a lot of explanation for many people. Contrary, Respondent 5 does not agree with this. She has doubt whether the level of association between variables should play a role in the reweighting process. She mentions again that researchers should rather look at the underlying characteristics of the sample. Visualizing the results before and after reweighting as presented in the interview may be a good way to explain if reweighting has an effect or not. Yet, it should be tested how different policy makers respond to this.

6.3 Conclusion interviews

To goal of this chapter was to provide an answer to sub question 3*: "How do experts reflect on representativity and inclusivity of PVE, and how do they review weight adjustment as method to correct for representation bias?"*

PVE aims to be an inclusive method, yet it is difficult to determine how inclusive the method really is. It became clear that PVE demands of a lot of participants to make a well-considered choice. Not everybody has the time and capacity to participate in a PVE experiment. So, even when the experiment is freely available, it is does not mean that the method is actually accessible for everyone.

All interviewees preferred the representative panel version of the PVE experiment over the open version as it is easier to justify in the policy context. In the end policy makers want to hear whether the research is representative. For them it is a basic requirement. For researchers it is almost never possible to say that the sample is fully representative. In addition, it is questioned whether the standard demographic characteristics are sufficient to determine whether a research is representative. There are other characteristics of respondents that play a role in their opinion. However, it can be hard to measure these characteristics and/or the required reference figures are not available to determine whether the sample is representative for these characteristics.

The consequence of a limited representative research is that the outcomes of that research are less usable than a representative research. Remarks must be made in which context the research results are applicable. Moreover, a lack of representativity is often used opportunistically in the public debate.

Lastly it became clear that weight adjustment can correct for a lack of representativeness in terms of the research definition: it can align the sample distributions of certain characteristics with the corresponding population distribution. However, weight adjustment cannot make up for the lack of legitimacy in the policy process or the lack of inclusivity, which are often part of the definition of representativity that is used by the public. Researchers were reluctant to apply weight adjustment even though it did not change the results of the PVE. It was mentioned that weight adjustment is difficult and complex. Researchers must think carefully whether it can be justified in a specific context. Moreover, more information is needed on how representativity and weight adjustment can be presented to policy makers in an understandable way.

Chapter 7 – Conclusion

The main research question of this research was:

"What are the consequences of the trade-off between representativity and inclusivity for the usability of Participatory Value Evaluation?"

In this study the Participatory Value Evaluation (PVE) on relaxation of COVID-19 measures was selected as case study. There were two available datasets: one dataset with the responses of panel members (panel PVE), the other with responses of people that voluntarily participated in the research (open PVE). The panel version was only accessible for panel members and therefore not very inclusive, however, this version was expected to be representative for the Dutch population. The open version was more inclusive as it was freely accessible, yet this version was expected to be biased for the Dutch population.

To answer the main research question, three sub questions were formulated, which are answered below.

1. In which context is PVE a suitable participation method?

   Literature review showed that PVE can fulfil several participation goals. PVE is a suitable approach to inform and educate the public. Moreover, PVE has the potential to foster trust in institutions. The open PVE is more fitting to fulfil these goals as larger amount of people can participate in the research. PVE can also be used to identify the preferences, values, and attitudes of the public and to increase the substantive quality of a decision. The panel PVE is more suitable to reach these goals as the outcomes of a panel version are representative for the larger population. PVE itself is not a proper method to reduce conflict among stakeholders as it does not allow for direct interaction among (opposing) stakeholders. Yet, the outcomes of a PVE experiment may be used as input for other participatory approaches in which opposing stakeholders get the chance to deliberate. Lastly, PVE is a cost-effectively participation method. It has the ability to include a large amount of people in the decision-making process for relatively low costs.

2. To what extend do bias in demographic characteristics influence the preferences of Dutch residents with regard to different policy measures in the PVE on relaxation of COVID-19 measures?

   Comparing the panel sample with the open sample showed that respondents from the open sample are more willing to relax COVID-19 measures. On average they selected more and rejected less measures than respondents from the panel sample. The top three selected and rejected is the same despite small differences in the ranking of the measures. However, the differences between the share of people that selected or rejected a measure in the panel version and the open version are quite large.

   Both samples showed representation bias for age, education level and province. The biases in the panel sample were relatively small compared to the biases in the open sample. Looking into the effect of demographic variables on the project choice of respondents showed that the method of sampling affects the number and direction of the relations between the demographic variables and participants' project choice. The open sample showed more, and sometimes opposite statistically significant relations compared to the panel sample.

Correcting for representation bias of age, education level and province in the panel sample showed no difference in participants' preferences. It can be concluded that the panel sample provide provided indeed an accurate reflection of the participants' preferences with regard to gender, age, education level and province. Reweighting the open sample resulted in a low weighting efficiency, which indicated that the bias in the open dataset was too big to properly correct. Moreover, the weights were highly dispersed, and in combination with the low association with the dependent variables this leads to unstable estimates. It was not possible to achieve representativity for the open dataset. For the sake of comparison, the results for correcting for representation biased in the open sample were reported and showed that the gap in the share of participants that selected or rejected an option in the panel and in the open PVE was bridged. Yet, the correction could not account for the whole gap.

It can be concluded that the demographic variables can only partly explain the differences in preferences between the samples. Other characteristics do play a role in the preferences of participants. The open PVE and the panel PVE resulted in more or less the same results, except for difference in the exact share of respondents that choose a measure. From a practical perspective the differences are less important, as the top three favoured measures are the same for both samples.

3. How do experts reflect on the representativity and inclusivity of PVE, and how do they review weight adjustment as method to correct for representation bias?

Interviews have been conducted to let experts reflect on representativity and inclusivity in the policy context. At first sight, all interviewees chose a representative sample over a larger sample based on self-selection, as the former is more reliable for the target population. Representativity is seen as a basic requirement in the policy context. A lack of representativity decreases the usability of a research as the results do not apply for the entire target group. Clear remarks must be made to whom the results are applicable. In addition, a lack of representativity is often used opportunistically in the policy debate. When showing that the bias in the open sample had only a very limited influence on the preferences of participants, still doubts arise about whether to use weight adjustment. In general, the interviewees were reluctant to apply to use it as it can only be justified under certain circumstances. All in all, panel sample is better accepted than the open sample.

To conclude, the trade-off between inclusivity and representativity manifest itself in the potential to reach certain participation goals. An inclusive sample is more suitable to educate and inform the public and to enlarge the level of trust citizens have in the government. A representative sample is more important to provide accurate results about the preferences of the Dutch population and to improve the quality of the decision. The case study analysis showed that for the PVE on COVID-19 measures the inclusive and representative sample led to more or less the same results. However, it was not possible to reweight the inclusive sample is a proper way to make it representative for certain demographic variables. This showed that weight adjustment is not always suitable to reach representativity and that the trade-off is present. For other case studies, an inclusive and representative sample might not provide the results. From a policy perspective, a representative sample is valued higher than an inclusive sample as the former is seen as a basic requirement.

Chapter 8 – Discussion and recommendations

Section 8.1 reflects on the outcomes as well as the limitations of this study and makes recommendations for further research. Section 8.2 discusses the practical implications for conducting a PVE experiment.

8.1 Discussion and recommendations for further research

8.1.1 Limitations due to scope and time frame

The scope of this research contributed to the limitations. Only one case study, the PVE on relaxation of COVID-19 measures, was used to explore the effects of demographic characteristics on participants' preferences. The outcome of sub question 2 is only applicable to this context. It is recommended to conduct more case studies to determine the influence of demographic variables on participants' preferences in other contexts. When comparing several cases to each other, a more reliable answer on the potential trade-off of inclusivity and representativity can be provided.

In addition, the limited time frame of this study posed an important limitation to the interviewing phase. Five interviews were conducted to reflect on representativity and inclusivity in the policy context. A criterium to determine if enough people are interviewed is called theoretical saturation. It indicates that no new information was obtained in answers to the interview questions. However, for this study theoretical saturation was not reached. The last interview still provided new information. Yet, due to the timeframe of this research it was not possible to conduct additional interviews. Moreover, all interview respondents were researchers. Possibilities lie in interviewing more researchers as well as policy makers about inclusivity and representativity.

8.1.2 Limitations due to available data

In this study two datasets of the PVE on relaxation of COVID-19 measures were used. Reference figures for demographic variables were obtained from Statistics Netherlands. Using these data leads to the following discussion points and limitations.

Panel sample

This research showed that the panel version of the PVE provided an accurate reflection of the preferences of the target population, as correcting for the minor representation bias in the sample did not lead to other results. However, it is important to note that it is not clear in which way Kantar recruited her panel members. It is likely that the selection of panel members is prone to some kind of selection bias, as Kantar could not be in the possession of a population register of all Dutch citizens. Moreover, even if the panel sample was based on probability sampling, this still means that only members of Kantar could participate in the PVE experiment. Lazarsfeld (1940) pointed out that members of a panel develop a critical attitude and hence over time cease to be representative of the public. A non-panel sample based on probability sampling may therefore lead to other results.

Lack of reference figures

There were no proper references figures available for the demographic variables income and living situation. As a result, it was not possible to check whether the samples were representative for these variables. Furthermore, it was also not possible to reweight the data on these variables. Moreover, not for all variables information on cell level was available, which meant that cell weighting could not be applied.

8.3 Discussion effect of demographic variables on participants' preferences

Another point of discussion is that the Cramer's V tests on demographic variables and project choice showed that the statistically significant relations have either a weak or low level of association. This is confirmed by the low predictive power of the binomial regression models. This means that the

demographic variables can only explain the preferences of respondents to a limited extend. It could be questioned whether these demographic variables are sufficiently capable of capturing the diversity in opinions in the Dutch society. And thus, whether the demographic characteristics are the correct characteristics to determine whether a sample is representative. Opportunities for further research lie investigating which non-demographic variables influence participants' preferences regarding COVID-19 measures. In the PVE experiment itself some other variables are taken into account such as the perceived risk of yourself or family/friends getting inflected with COVID-19 and whether a participant expects that his income increases or decreases in the coming months. Yet, other characteristics that are not considered in the PVE could play a role. In one of the interviews, it was mentioned that whether the participant has a disease could influence his choices. In addition, political preferences or the level of trust in governmental institutions may be important factors.

8.4 Discussion weight adjustment

This study used IPF as weight adjustment method to correct for representation bias. Reweighting the open sample showed low weighting efficiency and a high weight ratio, which meant that the overall bias in the data was too big to correct in a responsible way. So, it was not possible to achieve both inclusivity and representativity. First of all, it would be interesting to preform IPF without the variable education level as the biggest part of the overall bias in the dataset was caused by this variable. It is recommended to explore whether IPF on gender, age and province would results in a higher weighting efficiency and a lower weight ratio and therefore more reliable results. Another point of for further research is to determine when the sample deviates to much from the population distribution to preform weight adjustment. A weight efficiency below 70% indicates that sample and population distributions differ too much for all included variables. Yet, this is something you can only calculate afterwards. It would be interesting to know beforehand when the deviation becomes too large so that research can collect more responses. Literature provides guidelines for the minimum number of responses per category, for example a cell size of 50 is recommended by Kolenikov (2016). Yet, to my awareness literature does not provide guidelines for a maximum cell size or the maximum deviation between the sample and the population. Lastly, it is important to remember that representation bias and weight adjustment are complex topics. It is important to explore how these topics could be presented in a more understandable way to people without prior knowledge.

8.2 Practical implications

This study also resulted in some practical implications for conduction a PVE experiment.

When the main goal of the research is to inform and educate the public, an open PVE experiment is more suitable than a panel PVE. The reason for this is that an open PVE allows everyone who is interested in the research to participate. Moreover, PVE has the potential to foster trust in institutions as it makes the decision-making process more transparent by providing insight in the issue at stake and the effects the government must consider. The more people gain an understanding about the dilemma the government face, the better this goal can be achieved.
When the main goal of the PVE is to provide an adequate advice about the preferences of a population regarding several policy measures, a panel PVE is more fitting as it the results of this sample are representative for the larger population. In case of the COVID-19 PVE, the results of the panel and the open PVE were quite similar, so in this situation an open PVE would be suitable as well. However, it is not possible to say beforehand whether both versions would provide the same results. Thus, when a policy maker wants to carry out a new PVE with this goal in mind it is better to use the panel version.
Moreover, the panel PVE is more suitable to reach the goal of improving the substantive quality of decisions.

Another practical implication for researchers is to think in advance about the characteristics that they want to check for representativity. It is recommended that researchers look at agencies that provide reference figures, such as Statistics Netherlands, and see how they measure the variables of interest. Researcher should use the same categories where possible, to ensure the sample can be checked for representativity and so that other statistical procedures such as weight adjustment can be applied.

References

AcaStat Software. (2015). *Chi Square Measures of Association*. http://www.acastat.com/statbook/chisqassoc.htm

Austin, G. (2014, July 30). *The importance of representative samples and how to get them | Austin Research—Market Research Consultants*. https://austinresearch.co.uk/the-importance-of-representative-samples-and-how-to-get-them/

Bartlett, J. (2014, February 8). *R squared in logistic regression*. https://thestatsgeek.com/2014/02/08/r-squared-in-logistic-regression/

Battaglia, M., Izrael, D., Hoaglin, D., & Frankel, M. (2004). Tips and Tricks for Raking Survey Data (aka Sample Balancing). *Abt Associates*.

Beierle, T. C. (1999). Using Social Goals to Evaluate Public Participation in Environmental Decisions. *Review of Policy Research*, *16*(3–4), 75–103. https://doi.org/10.1111/j.1541-1338.1999.tb00879.x

Bethlehem, J. (2008). *How accurate are self-selection web surveys?* https://peilingpraktijken.nl/wp-content/uploads/2014/06/bethlehem04.pdf

Bethlehem, J. (2010). Selection Bias in Web Surveys. *International Statistical Review*, *78*(2), 161–188. https://doi.org/10.1111/j.1751-5823.2010.00112.x

Biemer, P. P., & Christ, S. L. (2008). Weighting survey data. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman, *International Handbook of Survey Methodology*. http://www.joophox.net/papers/SurveyHandbookCRC.pdf

Burton, P. (2009). Conceptual, Theoretical and Practical Issues in Measuring the Benefits of Public Participation. *Evaluation*, *15*(3), 263–284. https://doi.org/10.1177/1356389009105881

CBS. (2020). *Arbeidsparticipatie naar leeftijd en geslacht* [Webpagina]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-arbeidsmarkt/werkenden/arbeidsparticipatie-naar-leeftijd-en-geslacht

Chen, Q., Elliott, M. R., Haziza, D., Yang, Y., Ghosh, M., Little, R. J. A., Sedransk, J., & Thompson, M. (2017). Approaches to Improving Survey-Weighted Estimates. *Statistical Science*, *32*(2), 227–248.

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). The corestones of survey research. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman, *International Handbook of Survey Methodology*. http://www.joophox.net/papers/SurveyHandbookCRC.pdf

Deetman, T. (2020). *Online omgevingsanalyse—Onderzoek verspoeling COVID maatregelen*.

Dekker, T., Koster, P., & Mouter, N. (2019). The Economics of Participatory Value Evaluation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3323645

Deville, J.-C., Särndal, C.-E., & Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, *88*(423), 1013–1020. https://doi.org/10.1080/01621459.1993.10476369

Driessen, P. P. J., Glasbergen, P., & Verdaas, C. (2001). Interactive policy-making—A model of management for public works. *European Journal of Operational Research*, *2*(128), 322–337.

Edelenbos, J., Domingo, A., Klok, P. J., & Tatenhove, J. P. M. van. (2006). *Burgers als beleidsadviseurs. Een vergelijkend onderzoek naar acht projecten van interactieve beleidsvorming bij drie departementen*. Instituut voor Publiek en Politiek. https://research.wur.nl/en/publications/burgers-als-beleidsadviseurs-een-vergelijkend-onderzoek-naar-acht

Edelenbos, J., Teisman, G. R., & Reuding, M. (2001). *Interactieve beleidsvorming als sturingsopgave*. http://www.innovatieagroennatuur.nl/nl/bibliotheek/rapporten/173/Interactievebeleidsvormingalssturingsopgave.html

Engel, U., Jann, B., Lynn, P., Scherpenzeel, A., & Sturgis, P. (2014). *Improving Survey Methods: Lessons from Recent Research*. Routledge.

Escobar, O., & Elstub, S. (2017). Forms of Mini-Publics: An introduction to deliberative innovations in democratic practice. *Canberra: NewDemocracy*. https://www.newdemocracy.com.au/2017/05/08/forms-of-mini-publics/

Fiorino, D. J. (1990). Citizen Participation and Environmental Risk: A Survey of Institutional Mechanisms. *Science, Technology, & Human Values*, *15*(2), 226–243. https://doi.org/10.1177/016224399001500204

Goodin, R. E. (2007). Enfranchising All Affected Interests, and Its Alternatives. *Philosophy & Public Affairs*, *35*(1), 40–68. https://doi.org/10.1111/j.1088-4963.2007.00098.x

Hajer, M. A., van Tatenhove, J. P. M., & Laurent, C. (2004). *Nieuwe vormen van Governance—Een essay over de nieuwe vormen van bestuur met een empirische uitwerking naar de domeinen van voedselveiligheid en gebiedsgericht beleid* (p. 37).

Ianniello, M., Iacuzzi, S., Fedele, P., & Brusati, L. (2019). Obstacles and solutions on the ladder of citizen participation: A systematic review. *Public Management Review*, *21*(1), 21–46. https://doi.org/10.1080/14719037.2018.1438499

Irvin, R. A., & Stansbury, J. (2004). Citizen Participation in Decision Making: Is It Worth the Effort? *Public Administration Review*, *64*(1), 55–65. https://doi.org/10.1111/j.1540-6210.2004.00346.x

Itten, A. (2019). *Overcoming Social Division: Conflict Resolution in Times of Polarisation and Democratic Disconnection*. Routlegde. https://www.researchgate.net/publication/327238829_Overcoming_Social_Division_Conflict_Resolution_in_Times_of_Polarisation_and_Democratic_Disconnection

Kalton, G., & Flores-Cervantes, I. (2003). *Weighting methods*. *19*(2), 81–97.

Kolenikov, S. (2014). Calibrating Survey Data using Iterative Proportional Fitting (Raking). *The Stata Journal*, *14*(1), 22–59. https://doi.org/10.1177/1536867X1401400104

Kolenikov, S. (2016). Post-stratification or non-response adjustment? *Survey Practice*, *9*(3), 2809. https://doi.org/10.29115/SP-2016-0014

Laerd Statistics. (n.d.). *Binomial Logistic Regression using SPSS Statistics*. Retrieved 19 January 2021, from https://secure.livechatinc.com/

Lavrakas, P. (2008). Self-Selection Bias. In *Encyclopedia of Survey Research Methods*. Sage Publications, Inc. https://doi.org/10.4135/9781412963947.n526

Lazarsfeld, P. F. (1940). 'Panel' Studies. *The Public Opinion Quarterly*, *4*(1), 122–128.

Leach, M., Scoones, I., & Wynne, B. (2005). *Science and Citizens: Globalization and the Challenge of Engagement*. Zed Books.

Leyenaar, M. (2009). *Burgers aan zet. Vormen van burgerparticipatie: Inventarisatie en evaluatie*. https://vng.nl/artikelen/reeks-burgers-aan-zet

Li, M., & Zhao, J. (2019). Gaining Acceptance by Informing the People? Public Knowledge, Attitudes, and Acceptance of Transportation Policies. *Journal of Planning Education and Research*, *39*(2), 166–183. https://doi.org/10.1177/0739456X17732623

Lohr, S. L. (2008). Coverage and Sampling. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman, *International Handbook of Survey Methodology*. http://www.joophox.net/papers/SurveyHandbookCRC.pdf

Lovelace, R., Birkin, M., Ballas, D., & van Leeuwen, E. (2015). Evaluating the Performance of Iterative Proportional Fitting for Spatial Microsimulation: New Tests for an Established Technique. *Journal of Artificial Societies and Social Simulation*, *18*(2), 21.

Manfreda, K. L., & Vehovar, V. (2008). Internet surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman, *International Handbook of Survey Methodology*. http://www.joophox.net/papers/SurveyHandbookCRC.pdf

Mangelaars, F. (2019, February 5). Representativiteit. *Focus op Afstuderen*. https://focusopafstuderen.nl/studietips/representativiteit/

McFadden, D. (1979). Quantiative Methods for Analysing Travel Behaviour of Individuals: Some recent Development. In A. D. Hensher & P. R. Stopher (Eds.), *Behavioural Travel Modelling* (1st ed.). Routledge. https://www.routledge.com/Behavioural-Travel-Modelling/Hensher-Stopher/p/book/9780367740870

Mercer, A., Lau, A., & Kennedy, C. (2018, January 26). How different weighting methods work. *PeW Research Center*. https://www.pewresearch.org/methods/2018/01/26/how-different-weighting-methods-work/

Mouter, N. (2017). Dutch politicians' use of cost–benefit analysis. *Transportation*, *44*(5), 1127–1145. https://doi.org/10.1007/s11116-016-9697-3

Mouter, N., Hernandez, J. I., & Itten, A. V. (2020). Public participation in crisis policymaking. How 30,000 Dutch citizens advised their government on relaxing COVID-19 lockdown measures. *MedRxiv*, 2020.11.09.20228718. https://doi.org/10.1101/2020.11.09.20228718

Mouter, N., Koster, P., & Dekker, T. (2019). *An Introduction to Participatory Value Evaluation*. https://doi.org/10.2139/ssrn.3358814

Müller, K., Buchhammer, A., Eaglestone, A., & Griffiths, J. (2019). *Quantipy Documentation*. 173.

Pape, M., & Lim, C. (2019). Beyond the "Usual Suspects"? Reimagining Democracy With Participatory Budgeting in Chicago. *Sociological Forum*, *34*(4), 861–882. https://doi.org/10.1111/socf.12550

Raad voor het Openbaar Bestuur. (2005). *Burgers betrekken—Publicatie—Raad voor het Openbaar Bestuur* [Publicatie]. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. https://www.raadopenbaarbestuur.nl/documenten/publicaties/2005/05/01/burgers-betrekken

Rowe, G., & Frewer, L. J. (2000). Public Participation Methods: A Framework for Evaluation. *Science, Technology, & Human Values*, *25*(1), 3–29. https://doi.org/10.1177/016224390002500101

Sheskin, D. J. (2000). *Handbook of PARAMETRIC and NONPARAMETRIC STATISTICAL PROCEDURES* (2nd ed.). Chappman&Hall/CRC.

Stern, P. C., Fineberg, H. V., Stern, P. C., & Fineberg, H. V. (1996). *Understanding Risk: Informing Decisions in a Democratic Society*. National Academy Press. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.351&rep=rep1&type=pdf

Stirling, A. (2006). Analysis, participation and power: Justification and closure in participatory multi-criteria analysis. *Land Use Policy*, *23*(1), 95–107. https://doi.org/10.1016/j.landusepol.2004.08.010

Tarima, S., & Pavlov, D. (2006). Using auxiliary information in statistical function estimation. *ESAIM: Probability and Statistics*, *10*, 11–23. https://doi.org/10.1051/ps:2005019

TU Delft. (2020). *PVE as a participatory tool*. TU Delft. https://www.tudelft.nl/en/tpm/pve/pve-basics/pve-as-a-participatory-tool/

UCLA. (2021). *What statistical analysis should I use? Statistical analyses using SPSS*. https://stats.idre.ucla.edu/spss/whatstat/what-statistical-analysis-should-i-usestatistical-analyses-using-spss/

Wesselink, A., Paavola, J., Fritsch, O., & Renn, O. (2011). Rationales for Public Participation in Environmental Policy and Governance: Practitioners' Perspectives. *Environment and Planning A*, *43*. https://doi.org/10.1068/a44161

Wildemuth, B. M. (2016). *Applications of Social Research Methods to Questions in Information and Library Science, 2nd Edition*. ABC-CLIO.

Appendix A – Criteria policy measures

Eight policy measure are presented In the PVE on relaxation of COVID-19 measures. The policy measures are shown Table A.1.

Table A.1 Policy measures

Project	Description
1	Nursing and care homes allow visitors
2	Businesses open again (except hospitality and contact-jobs)
3	Contact professions can open again
4	People younger than 18 years do not have to keep 1.5m distance
5	All restrictions are lifted for people who are immune
6	Restrictions are lifted in northern provinces Friesland, Groningen, and Drenthe
7	Direct family members from other households do not have to hold 1.5-meter distance
8	Hospitality and entertainment sector open again

Different versions of the experiment with different attribute values for the policy measures have been used. The corresponding range of attribute values for each of the measures is shown in Table A.2.

Table A.2 Range attribute values

Project		The increase of deaths among people older than 70 year	The increase of deaths among people younger than 70 years	The increase of people with permanent physical injury	The decrease of people with permanent mental injury	The decrease in the number of households with long-term loss of income
1	min	1500	30	100	30000	50
	max	3000	300	1000	60000	200
2	min	200	150	1000	1000	10000
	max	1000	750	7500	7500	75000
3	min	200	150	1000	5000	20000
	max	1000	1000	10000	15000	750000
4	min	50	50	500	2000	50
	max	400	300	5000	1000	5000
5	min	400	300	2000	1000	5000
	max	1500	750	5000	7500	20000
6	min	600	300	5000	10000	20000
	max	2000	1000	10000	30000	75000
7	min	600	300	2000	30000	50
	max	2000	1000	10000	60000	50
8	min	200	300	1000	15000	50000
	max	1000	1000	10000	60000	100000

Appendix B – Weight adjustment methods

When a response sample is not representative for the target population weight adjustment methods can be applied to restore the representativity and thereby improve the quality of the results. The weights should account for all factors that affect the imbalance between the sample and population (Bethlehem, 2008; Engel et al., 2014; Kalton & Flores-Cervantes, 2003).

Weight adjustments are based on auxiliary information: additional information from an independent source that can be used to improve the survey estimates. Auxiliary information can be retrieved from a variety of sources and can have different forms. Examples of auxiliary information are population based survey reports, census data, results from previous experiments, expert opinions or assumptions on population parameters (Tarima & Pavlov, 2006). Auxiliary variables are the variables on which the weight adjustments are based. The weights should bring the response in line with the data from the independent source (Engel et al., 2014). Auxiliary variables must be measured in the PVE which means that the response sample distribution is known. Furthermore, their population distribution should be known. By comparing the response distribution with the population distribution of a specific variable one can determine to what extend the response is representative with respect to that specific variable.

It is important to note that the process of weight adjustment is quite subjective (Biemer & Christ, 2008). The analyst makes a lot of decisions that affect the final results: decisions have to be made about the weighting method, what data to use in the adjustment process etc. The process of weight adjustment can be very complex, and it is important that all decisions are carefully documented.

Weight adjustments can be carried out to align the responding sample with the original sample, or to align the responding sample with the known population distributions. This research focusses on the latter. Two weight adjustment approaches are discussed in the following sections.

B.1 Post stratification and calibration

Post-stratification is simply calibration on the marginal distribution of only one variable (Kolenikov, 2016). Calibration focusses on the alignment between the response sample and the target group. Calibration changes the weights of auxiliary variables based on a distance function so that the response sample marginal distributions conform to population marginal distributions (Kalton & Flores-Cervantes, 2003). However, calibration does not make sure that the cell levels align with the population totals unless the interactions are explicitly modelled as calibration targets. It is harder to model these interactions, as it requires more information: the population totals on cell level are required, e.g., the joint distributions of the variables need to be known, and these are not always available. Furthermore, the sample size of the higher order crossed cells are often small, which can lead to unstable weights.

Calibration is explained using Table B.1. Calibration makes sure that the weighted totals in the groups defined by gender and education level are equal to the known population totals (the green cells). Calibration does not guarantee that the totals at cell level (the grey cells) are equal to the population distributions. For example, after weight adjustment the percentage of females with a high education level in the sample can still differ from the percentage of females with a high education level in the population.

Table B.1 Example calibration

Education / Gender	Low	Middle	High	**Total**
Male	% male with low education level	% male with middle education level	% male with high education level	% male
Female	% female with low education level	% female with middle education level	% female with high education level	% female
Total	% with low education level	% with middle education level	% with high education level	100%

Calibration can be based on different distance functions (Deville et al., 1993). The two most known implementations are 1) iterative proportional fitting or ranking and 2) linear calibration.

B.1.1 Iterative proportional fitting or ranking

Iterative proportional fitting (IPF) is also known as raking, the RAS algorithm or proportional fitting (Lovelace et al., 2015). It is, as the name suggest, based on a repeating procedure. First the weights are adjusted so that the row totals of the sample match with the row totals of the population. As a result, the column totals of the sample will change as well. The next step is then to adjust these changed column totals according to the column totals of the population. Then the row totals are adjusted to conform and so on, until convergence is reached (Kalton & Flores-Cervantes, 2003). However, it may happen that convergence cannot be reached.

B.1.2 Linear calibration

In linear the adjustment factor is a linear combination of calibration variables (Kolenikov, 2016). A disadvantage of linear calibration is that negative weights can be produced (Kalton & Flores-Cervantes, 2003).

B.2 Cell weighting

Cell weighting is based on the joint distribution of the auxiliary variables. The weights are applied to make sure that the weighted totals on cell level are equal to the known population totals (Kalton & Flores-Cervantes, 2003). A potential disadvantage of cell weighting is that it can lead to a large variability in the distribution of the weighting adjustments thereby inflating the variances of the survey estimates. When the number of cells is relatively small and the cell sizes are reasonably large, cell-weighting is favourable over calibration since it is more precise. However, when the sample sizes in a number of cells that need to be adjusted are small, calibration may be more suitable. This is because a small sample size can lead to instable adjustments. Small cell sizes often occur when there are many cells, e.g. a large number of auxiliary variables. (Kalton & Flores-Cervantes, 2003).

Kolenikov (2016) argues that for all weight adjustment methods all cells that are being reweighted must have at least 50 respondents, because *"otherwise, you may be adjusting on something too noisy, and the weights may blow up leading to undesirably high design effects."* (Kolenikov, 2016, p. 6)

Appendix C – Data preparation

Before analysing the data, some data was recategorized (section C.1) to be able to compare the data of the PVE with data from Statistics Netherlands (CBS). Besides recategorizing, some extra variables were created (section C.2).

C.1 Recategorizing variables

Gender

To indicate their gender, participants could choose between the options 'Male', 'Female' and 'Other'. This last option is recoded as a missing value as this does not match with the population distributions provided by Statistics Netherlands.

Age

When Kantar invited people to participate in the panel PVE, they drew the sample based on the gender and age of the participants. They used the following age groups: 18 – 25 years, 26 – 35 years, 36 – 45 years, 46 – 55 years, 56 – 65 years as well as 66 years and older.

In the PVE experiment, this last age group was divided into 66 – 74 years and 75 years and older. To make an equivalent comparison, these age groups are combined into one group (66+).

Education level

The specified categories of education level in the PVE did not match the categories of Statistic Netherlands. Therefore, it is chosen to recategorize education level into the categories low, middle, and high. Table C.1 shows how the categories are recategorized.

Table C.1 Recategorization of education variable

Education level	Education level Statistic Netherlands	Education level PVE
1 (Low)	Basisonderwijs Vmbo-b/k, mbo1 Vmbo-g/t, havo-, vwo-onderbouw	Geen opleiding Basisonderwijs/lager onderwijs LBO (ambachtsschool, huishoudschool, lts, leao, etc.) VMBO, MAVO, Mulo
2 (Middle)	Mbo2 en mbo3 Mbo4 Havo, VWO	HAVO/VWO MBO
3 (High)	HBO-, wo-bachelor HBO-, wo-master, doctor	HBO WO

C.2 Adding new variables

Total number of selected measures

This variable represents the total number of projects that one participant selected.

Total number of rejected measures

This variable represents the total number of projects that one participant rejected.

Total pressure on healthcare system

This variable represents the total percentage of increase in pressure on the healthcare system per participant. It is calculated by summing up the pressures of measures the participant selected. It became clear that some respondents somehow selected a combination of measures that led to an increase of pressure of more than 50%. As 50% was set as the limit, the records of these respondents were deleted. The panel version was reduced from 3.470 to 3.358 records and the open version from 26.302 to 26.293 records.

Appendix D – McNemar's Chi-square test

Table D.1 McNemar's test panel dataset – selected policy measures

	Selected	Chi-square	p-value
project_3	50.4%		
project_7	46.5%	11.091	0.001
project_2	41.8%	15.580	0.000
project_4	33.5%	54.282	0.000
project_1	28.9%	15.340	0.000
project_8	19.6%	73.163	0.000
project_5	16.6%	9.519	0.002
project_6	8.4%	104.091	0.000

Table D.2 McNemar's test panel dataset – rejected policy measures

	Rejected	Chi-square	p-value
project_6	49.3%		
project_8	42.5%	36.266	0.000
project_5	41.0%	1.665	0.197
project_1	28.8%	110.461	0.000
project_4	27.5%	1.397	0.237
project_7	18.7%	110.766	0.000
project_3	18.4%	0.062	0.803
project_2	14.8%	22.182	0.000

Table D.3 McNemar's test open dataset – selected policy measures

	Selected	Chi-square	p-value
project_3	63.9%		
project_2	51.2%	846.581	0.000
project_7	43.5%	297.595	0.000
project_4	41.1%	35.007	0.000
project_8	33.1%	336.434	0.000
project_1	32.6%	0.911	0.340
project_5	9.4%	3691.410	0.000
project_6	4.6%	462.254	0.000

Table D.4 McNemar's test open dataset – rejected policy measures

	Rejected	Chi-square	p-value
project_6	53.2%		
project_5	46.7%	375.488	0.000
project_8	31.8%	1370.566	0.000
project_4	21.0%	942.431	0.000
project_1	19.1%	32.005	0.000
project_7	14.9%	195.366	0.000
project_3	9.1%	503.569	0.000
project_2	9.0%	0.548	0.459

Tables D.1 till D.4 show the results of the McNemar's Chi-square test. It is used to determine whether the difference between two measures is statistically significant. For example, when looking at Table D.1, the Chi-square value of 11.091 and the p-value of 0.001 are the test results of the McNemar's test between measure 3 and 7, while the Chi-square value of 15.580 and p-value of 0.000 are the results of the McNemar's test between measure 7 and 2, and so on.

For the panel dataset the McNemar's Chi-square test shows that the differences between rejecting measures 5 and 8, measures 4 and 1 as well as measures 3 and 2 are not statistically significant. For the open dataset, the McNemar's Chi-square test suggest that the difference between selecting measure 1 and measure 8 as well as the difference between rejecting measure 2 and 3 are not statistically significant.

Appendix E – Interviews

The interviews are held in Dutch. Therefore, the interview invitation, the interview protocol as well as the documentation are available in Dutch.

E.1 Invitation

Beste …,

Mijn naam is Selma en voor mijn afstudeeropdracht wil ik graag een aantal interviews houden. Via mijn begeleider, Niek Mouter, heb ik uw naam gekregen.

In mijn afstudeeropdracht focus ik mij op de representativiteit van Participatieve Waarde Evaluaties (PWE). Representativiteit is belangrijk om de resultaten van het onderzoek te kunnen generaliseren naar de volledige doelgroep. Wanneer de steekproef niet of beperkt representatief is voor bepaalde achtergrondkenmerken roept dit al gauw vragen op. Een manier om hier mee om te gaan is het herwegen van de steekproef op deze variabelen. In mijn onderzoek kijk ik dan ook specifiek naar wat de randvoorwaarden zijn om een steekproef te kunnen herwegen en wat het effect hiervan is.

Als onderdeel van mijn thesis wil ik ook graag weten hoe andere onderzoekers omgaan met (het gebrek aan) representativiteit van PWE. Hiervoor zou ik u graag willen interviewen. Het interview zal plaatsvinden tussen 6 en 15 januari via Zoom en zal maximaal 1 uur in beslag nemen. Graag hoor ik of u mee wilt werken aan mijn onderzoek en welke datum en tijdstip gelegen komt voor u.

Mocht u nog vragen hebben over mijn onderzoek dan hoor ik dat graag.

Met vriendelijke groet,
Selma van Delft

E.2 Interview protocol

Naam:
Functie:
Datum:

Persoonlijke introductie, niet opnemen
Het interview dient te beginnen met een kort informeel gesprekje om kennis te maken.

Geluidsopname
In verband met het uitwerken van het interview zou ik graag een geluidsopname maken. De opname zal achteraf verwijderd worden. Gaat u hier mee akkoord?

- Toestemming opname: ja/ nee

Deelname interview
Ik neem vrijwillig deel aan dit interview. Dit interview vindt digitaal plaats en vraagt geen andere verplichtingen van de deelnemer. Ik ga akkoord dat de uitkomsten van dit onderzoek gebruikt worden voor de afstudeeropdracht van Selma van Delft. Ik begrijp dat ik vragen kan overslaan en kan stoppen met het interview wanneer ik wil.

- Toestemming deelname interview: ja/ nee

Verder gebruik van de informatie
Na afloop van het interview stuur ik u een samenvatting van de resultaten met daarbij quotes die ik wil gebruiken voor het onderzoek. Het is mogelijk om de resultaten anoniem te verwerken. Wat is uw voorkeur?

- Ik wens dat de interview resultaten anoniem/ niet anoniem verwerkt worden.

Wellicht worden de resultaten gebruikt voor nader onderzoek, buiten de afstudeeropdracht om. Hiervoor zullen de uitgewerkte informatie van het interview alsmede de samenvatting van het interview worden opgeslagen binnen de TU Delft Geeft u hier toestemming voor?

- Toestemming opslaan uitgewerkte informatie binnen de TU Delft: ja/ nee
- Toestemming opslaan samenvatting interview binnen de TU Delft: ja/ nee

1. Inleiding Participatieve Waarde Evaluaties (PWE)
Allereerst wil ik aangeven dat ik voor dit interview geïnteresseerd bent in uw persoonlijke ervaringen en keuzes omtrent de vraagstukken.

- Wat is uw ervaring met Participatieve Waarde Evaluaties?
- Wat is de belangrijkste reden om te kiezen voor PWE als onderzoeksmethode?
- Wat is in uw ogen het belangrijkste nadeel van PWE?

2. Steekproef en representativiteit
Omdat in onderzoek zelden de gehele populatie (doelgroep van onderzoek) ondervraagd kan worden, selecteert men vaak een deel van de populatie. Dit wordt ook wel een steekproef genoemd. Hierbij speelt representativiteit een belangrijke rol.

- Hoe definieert u representativiteit?
- Wat is in uw ogen het belang van een representatieve steekproef?
- Wanneer is een steekproef volgens u representatief en wanneer niet?
- Voor welke (demografische) kenmerken van de doelgroep moet een onderzoek representatief zijn? Hoe wordt dit bepaald?
- Indien een kenmerk niet of beperkt representatief is voor de doelgroep betekent dit dan ook dat je geen representatieve uitspraken kunt doen over de doelgroep?
- Welke problemen kan een niet of beperkt representatieve steekproef met zich mee brengen in de beleidscontext?

3. Participatieve Waarde Evaluatie over het versoepelen van corona maatregelen
Tussen 28 april en 5 mei 2020 is er een PWE uitgevoerd over het verspoelen van de toen geldende corona maatregelen. Burgers konden aangeven welke corona maatregelen zij zouden verspoelen tussen 20 mei en 20 juli als ze in de schoenen van de overheid stonden. Daarnaast konden ze aangeven welke maatregelen de overheid niet zou moeten overwegen.

<u>Vraagstuk 1</u>
Er zijn twee versie uitgevoerd van deze PWE. Eén versie kon alleen ingevuld worden door panel leden van het Kantar Panel. De andere versie was vrij toegankelijk via internet. Iedereen die mee wilde doen aan het onderzoek kon ook daadwerkelijk deelnemen.

Aan de panel versie hebben bijna 3500 burgers deelgenomen en aan de open versie ruim 26.000.

- Op de resultaten van welke versie zou beleid volgens u het best kunnen worden gebaseerd?

Na het analyseren van de data blijkt dat de panel versie een betere afspiegeling is van de Nederlandse populatie op basis van de volgende demografische kenmerken: geslacht, leeftijd, opleidingsniveau en provincie.

- Op basis van deze informatie, aan welke versie geeft u de voorkeur? En waarom?

Vraagstuk 2

Voor deze casus geeft de open PWE een minder representatieve afspiegeling van de Nederlandse bevolking op basis van de kenmerken leeftijd, opleidingsniveau en provincie. Het gebrek aan representativiteit roept al gauw vragen op en kan ervoor zorgen dat de uitkomsten van het onderzoek bekritiseerd worden.

Echter brengt het opstellen van het onderzoek wel een aantal belangrijke voordelen met zich mee: burgers die willen participeren kunnen participeren (inclusiviteit), deelnemers worden geïnformeerd over de keuzes die de overheid moet maken, deelnemers zijn eerder geneigd om de maatregelen op te volgen en het uiteindelijke beleid te accepteren doordat ze betrokken zijn bij het besluitvormingsproces.

Een manier om de open versie representatief te maken voor de eerder genoemde demografische kenmerken is het herwegen van de data op deze kenmerken. Bij het herwegen van de data wordt een gewicht toegekend aan elke deelnemer. Het herwegen wordt besproken aan de hand van de variabele opleidingsniveau.

Table E.1 Education level population and open sample

Opleidingsniveau categorie	Verdeling populatie (%)	Verdeling open PWE (%)
Laag	30.63	4.20
Middel	37.36	16.03
Hoog	32.01	79.77

Zoals te zien is in tabel E.1 zijn deelnemers met een hoog opleidingsniveau oververtegenwoordigd in de open versie van het onderzoek. Deelnemers met een hoog opleidingsniveau krijgen een relatief laag gewicht toegewezen en deelnemers met een laag en midden opleidingsniveau een hoger gewicht.

- Door het herwegen van de data wordt de mening van de ene deelnemer aan het onderzoek belangrijker dan de mening van een andere deelnemer. Wat is uw mening over dit effect?

De tweede stap is het nagaan van de samenhang tussen opleidingsniveau en keuzes van de deelnemers. Hierbij gaat het om het selecteren of afwijzen van een bepaalde beleidsmaatregel. De maatregelen zijn genummerd als project 1 t/m project 8. De samenhang wordt bepaald door een statische test, de Cramér's V test. De samenhang tussen twee variabelen wordt uitgedrukt met een getal tussen de 0 en 1, waarbij 0 staat voor geen samenhang en 1 voor volledige samenhang.

Tabel E.2 geeft de samenhang tussen opleidingsniveau en het selecteren en afwijzen van een beleidsmaatregel aan.

Table E.2 Association education level and project choice

Project	Samenhang selecteren project	Samenhang afwijzen project
1	0.058	0.023
2	0.111	0.073
3	0.085	0.107
4	0.054	0.045
5	0.048	0.028
6	0.022	0.094
7	0.069	0.072
8	0.042	0.077

De lage samenhang duidt er dus op dat opleidingsniveau niet/ nauwelijks de verklaarde factor is voor welke maatregel een deelnemer kiest. Het herwegen van de data op deze variabele zal dan ook nauwelijks de uitkomsten veranderen.

- Is het aantonen van een lage samenhang tussen opleidingsniveau en de beleidskeuzes overtuigend genoeg om (representatieve) uitspraken te doen over de gehele populatie, ondanks het gebrek aan representativiteit t.a.v. opleidingsniveau.

In het Figuur E.1 is de open versie herwogen voor de variabele opleidingsniveau. Uit het figuur komt naar voren dat de rangorde van de beleidsmaatregelen niet veranderd.

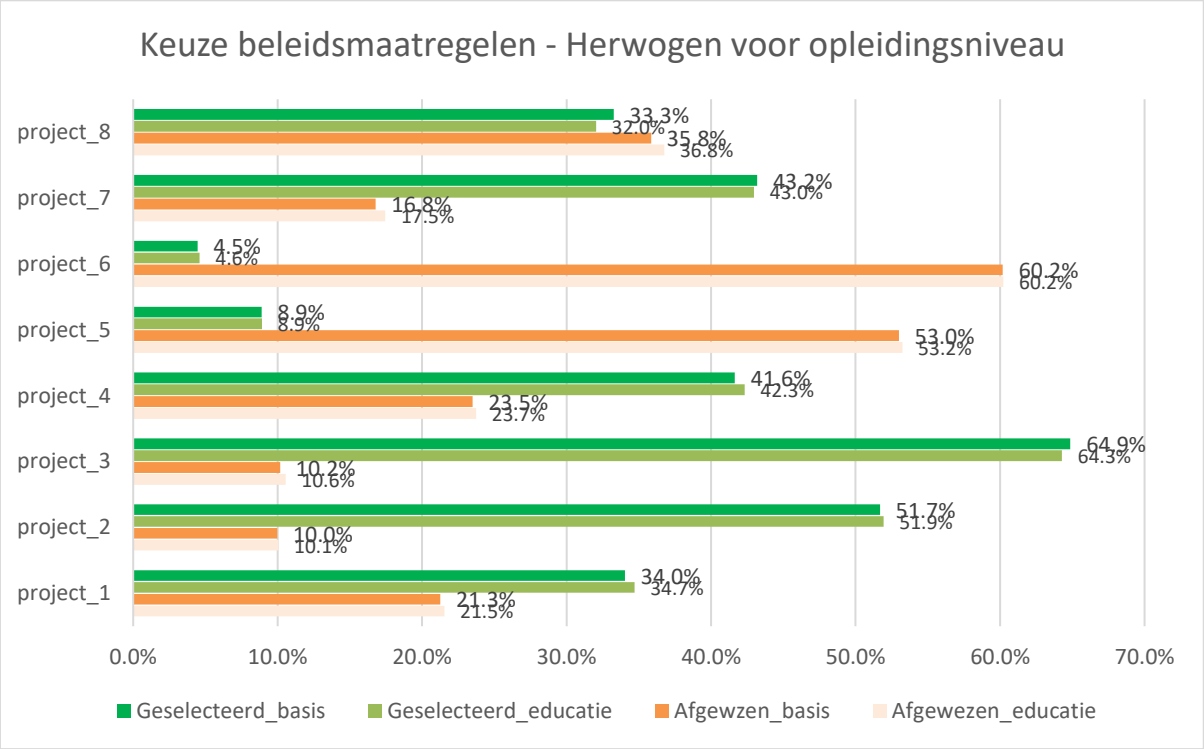

Figure E.1 Open sample reweighted for education level

- Door het herwegen is de open versie (statistisch gezien) wél representatief voor opleidingsniveau. Is herwegen een gewenste methode om de representativiteit van een steekproef te verbeteren? Waarom wel of waarom niet?
- Indien er sprake is van een lage samenhang, vind u dat er dan alsnog herwogen moet worden? Of is dit niet noodzakelijk? Waarom wel of waarom niet?

4. Verdiepende vragen herwegen

- Heeft u zelf ervaring met het herwegen van een steekproef of heeft u ervaring met situaties waarin een herwogen steekproef is voorgelegd?

Indien ja:

- Wat is uw ervaring?
- Wat zijn potentiële voordelen van herwegen?
- Wat zijn volgens u mogelijke nadelen van herwegen?
- Wat zijn de randvoorwaarden voor het herwegen van een steekproef?
- Onder welke condities in het niet mogelijk een steekproef te herwegen?

5. Communicatie

Tenslotte heb ik nog een aantal vragen over op welke manier de uitkomsten van een onderzoek gecommuniceerd worden naar andere belanghebbende, zoals bijvoorbeeld beleidsmakers of burgers.

- Bent u de persoon in kwestie die de uitkomsten van een onderzoek communiceert?

Indien ja:

- Met wie heeft u hier contact over? Aan wie communiceert u dit?
- Wat zijn voorkomende knelpunten bij het communiceren van onderzoeksresultaten?
- Is het mogelijk om de discussie rond representativiteit op een goede manier te communiceren naar mensen die geen statische kennis hebben?
- Hoe zou u de complexiteit rondom representativiteit duidelijk maken aan hen?
- Hoe zou u de uitkomsten van het herwegen communiceren aan hen?

6. Afsluiting

- Heeft u nog vragen en/of opmerkingen?

E.3 Interview results

The documentation of the interviews is available upon request by mailing Selma van Delft, selma-vandelft@hotmail.com.

E.4 Quotes

The quotes presented in chapter 6 are translated from Dutch in English. The original quotes and their translation can be found in Table E.3.

Table E.3 Interview quotes

Section	Respondent	Quote in main text	Original quote
1	1	"I find it especially important that it concerns concrete alternatives that you can compare and weigh against each other. Especially that you can see the impact. That you also see the disadvantages. "	*"Ik vind het vooral heel belangrijk dat het om concrete alternatieven gaat, die je kunt vergelijken en afwegen. Vooral dat je de impact kunt zien. Dat je ook de nadelen kunt zien."*
	5	*"Because different respondents simply want a different amount of information"*	*"Want verschillende respondenten willen nou eenmaal verschillend veel informatie."*
2	1	"If there are only 4 options in the PVE, but there are 5 other options that everyone is discussing, then the research is not really representative for the discussion in reality".	"Als je maar 4 opties in de PWE hebt, maar er zijn nog 5 andere opties waarover iedereen discussieert, dan is het eigenlijk ook niet representatief voor de discussie in de realiteit."

	1	"I think we are much more fragmented than we think"	*"Ik denk dat we veel gefragmenteerde zijn dan we denken"*
	2	"From a policy perspective representativity is a basic assumption. We assume that it is guaranteed by the researchers, scientists. "	**"***Vanuit beleid is representativiteit dus een basis aanname. We gaan ervan uit dat dat geborgd wordt door de onderzoekers, wetenschappers."*
	3	"People have the feeling that it is not right to conclude anything about a large group based on a subgroup.**"**	*"mensen hebben het gevoel dat het niet klopt om iets te concluderen over een grote groep op basis van een subgroep."*
	4	*"If your sample is not representative for the purpose of your research, I think the problem is especially that you cannot say with certainty that the results are correct for the target group you are focusing on".*	*"Dus het probleem is denk ik vooral dat als je steekproef niet representatief is voor het doel van je onderzoek, dat je dus niet met zekerheid kan zeggen dat de resultaten voor de doelgroep waar jij op focust kloppen."*
	4	"It remains a magic trick to fix something you cannot really fix. "	*Het blijft een goocheltruc om iets te repareren wat je eigenlijk niet echt kan repareren. "*
3	2	*"The representative sample of 3.500 provides the clearest picture for the whole of the Netherlands. The group of 25.000+ gives an image of those who want to contribute and be heard. It is also good to take note of that. "*	*"De representatieve uitvraag van 3.500 geeft het duidelijkst een beeld voor het geheel van Nederland. De groep van 25.000+ geeft meer een beeld van die een bijdrage willen leveren en gehoord willen worden. Het is ook goed om daarvan kennis te nemen."*
	4	"For us (researchers) it is a nice trick to see if it matters. If you reweight the data and you find out that it does not really matter for the general conclusions of your research. Then that is a nice bonus, it means that you do not have to worry too much about it. "	*Voor ons is dat een mooie truc om te kijken of het ertoe doet. Als je een herweging doet en je komt erachter dat het toch eigenlijk voor de algemene conclusies van je onderzoek niet uitmaakt. Dan is dat mooi meegenomen, dan hoef je er ook niet zo druk om te maken."*
	5	"Even if you do not have that many observations, you know that you are not going in a biased direction. This in contrast to when you have a lot of people from one particular group who all invited each other to click on that survey."	*"Ook al heb je niet zoveel waarnemingen, maar dan weet je ieder geval dat je niet hele vertekende kant opgaat. Dit in tegenstelling tot wanneer je heel veel mensen van één bepaalde groep die allemaal elkaar hebben doorgelinkt om op die survey te klikken."*
4	1	"When you say I just want to know what is going on, what the values, the considerations and the perspectives are. Then weight adjustment is an option to achieve a representative advice. "	*"Als je zegt ik wil gewoon weten wat er speelt, wat de waardes en de afwegingen en perspectieve zijn. Dan is herwegen wel een optie om een representatief advies te krijgen."*
	4	"So, it is a solution to make a skewed population, to make the results thereof, more representative. But with a lot of questions and comments in mind."	*"Dus het is een oplossing om een scheve populatie, om de resultaten daarvan, beter representatief te maken. Maar wel met een heleboel vragen en kanttekeningen in je achterhoofd. "*
	2	"In policy making, I am not so worried about representativity. This is mainly discussed in the start-up phase. When I give an assignment, I just assume that it is representative. That they arrange that at the beginning. "	*"Ik maak mij in de beleidsmaking niet zoveel zorgen over de representativiteit. Dit zit vooral in de opstartfase. Daar komt het dan op tafel. Als ik een opdracht geef, dan ga ik er gewoon vanuit dat het representatief is. Dat ze dat regelen aan het begin."*
	2	" To me, they are both things that happen behind the back door, somewhere in science.	*"Voor mij zijn het allebei dingen die voor mij achter de achterdeur gebeuren, ergens in de wetenschap."*
	4	"Intuitively, I always find weight adjustment a bit fishy, a bit dangerous. You have a small group of which you do not really know whether it is representative of the sub-population in which you are interested. And then you are going to blow it up completely. "	*"Heel intuïtief, ik vind dat herwegen altijd een beetje fishy, een beetje gevaarlijk. Je hebt een kleine groep waarvan je eigenlijk helemaal niet zo goed weet of die representatief is voor de deel populatie waarin je geïnteresseerd bent. En die ga je helemaal opblazen."*

5	2	"It is of course easier for policy and debate if something is representative than if it is not."	"Voor in het beleid en het debat is het natuurlijk makkelijker als iets representatief is dan dat het niet representatief is."
	3	*""Ultimately, a municipality mainly wants to hear whether it is representative or not."*	"Een gemeente wil uiteindelijk vooral horen of het representatief is of niet."
	5	"We (researchers) are very concerned about that, but I actually cannot remember whether we have discussed this with them at all. Probably yes, but it was not that penetrating or difficult. "	*"Maar dan maken wij ons heel druk daarom, maar ik kan me eigenlijk niet meer herinneren of we überhaupt met hen daarover hebben gehad. Waarschijnlijk wel, maar het was niet zo indringend of moeilijk."*