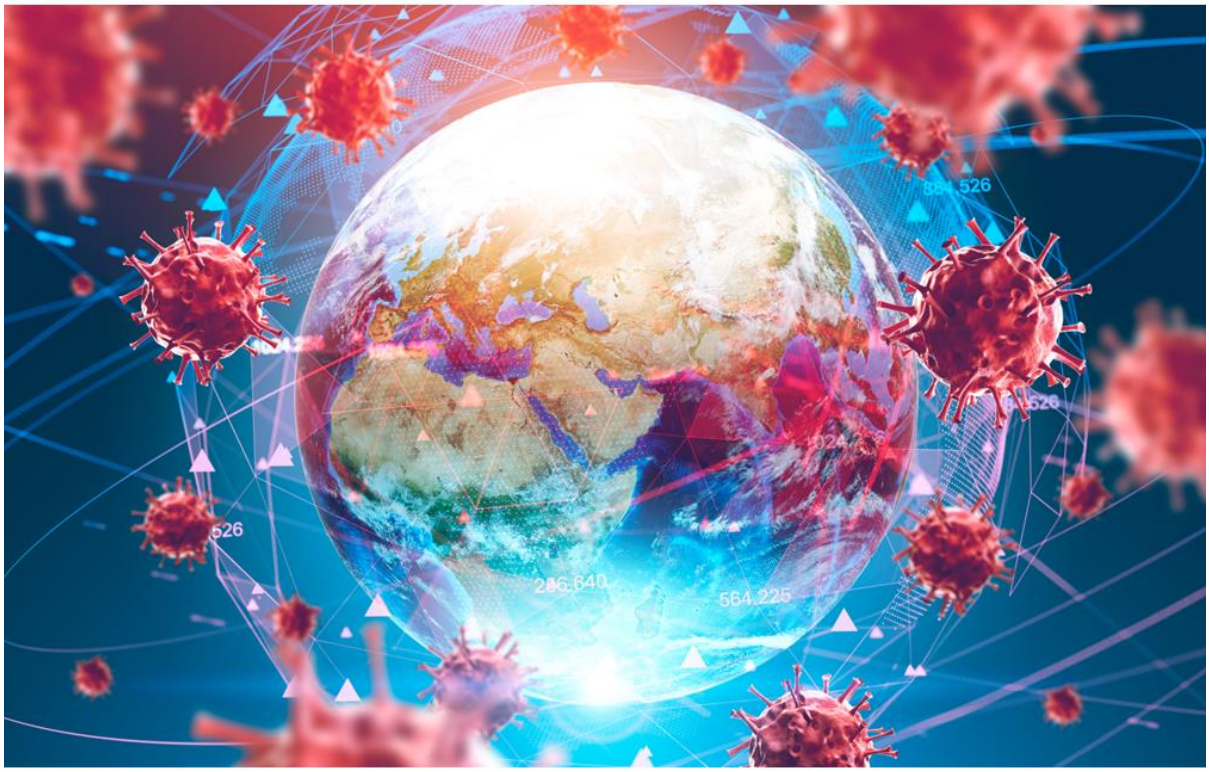
Added value of Choice Models and Discrete Choice Experiments for future pandemic policy



POPULYTICS
What would you do?

TU Delft

Added value of Choice Models and Discrete Choice Experiments for future pandemic policy

(Dis)-advantages of Mixed Logit and Latent Class models for analyzing (un)-labeled Discrete Choice Experiments that weigh societal impacts of COVID-19 policy during different pandemic phases

---

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in {**Complex Systems Engineering and Management**}

Faculty of Technology, Policy and Management

by

{Daniël Korthals}

Student number: 5187729

To be defended in public on August 24, 2023

Graduation committee

Chairperson	: { Dr. mr. N. Mouter}, { Section Transport and Logistics}
Second Supervisor	: {Dr. I. Grossmann}, {Safety and Security Science}

Contents

Preface

This report presents the results of my master thesis. It is the last hurdle in obtaining my master's degree in Complex Systems Engineering and Management at Delft University of Technology. This thesis could not have been written without the help of a few people.

Firstly, I would like to thank Niek and Irene for their supervision and advice. Furthermore, I would like to thank the Populytics team for the pleasant and educational internship. Also, I would like to thank all interviewees for their time and effort.

Excecutive summary

The first registered infection with the COVID-19 virus occurred in China at the end of 2019. The virus spread all over the world, turning a single infection into a pandemic. Most countries implemented policies to limit the spread of the virus. In general, these policies can be divided into pharmaceutical and non-pharmaceutical interventions. Vaccines are pharmaceutical interventions. Examples of non-pharmaceutical interventions are temporary closure of restaurants and bars or a limit to the number of people that are allowed to gather for social events. Although, these non-pharmaceutical interventions were effective in slowing down the spread of the COVID-19 virus, they also severely affected people's social and economic life. For example, freelancers who worked in restaurants and bars lost their income and students experienced a decrease in mental health due to the lack of social engagement. These societal impacts of COVID-19 policy gradually reduced the public support and adherence to these COVID-19 measures over the course of different pandemic waves. Measuring how people weigh the societal impacts of COVID-19 policy during these waves, provides interesting insights that can improve the effectiveness of pandemic policies for a future pandemic. A way to measure how people weigh the societal impacts of COVID-19 policy during different waves of the pandemic is by conducting Discrete Choice Experiments (DCE). These DCEs are often analyzed with Multinomial Logit (MNL) models. The MNL model is able to quantify the relative importance of different societal impacts for the population as a whole. However, there also exist other models, such as the Mixed Logit (ML) model and Latent Class (LC) model, and different types of DCEs, such as the labeled and unlabeled DCE, that have their own advantages and disadvantages in eliciting the relative importance of societal impacts during different phases of the pandemic. For instance, ML models are able to elicit how preference heterogeneity is distributed among individuals and the LC model is able to divide people into different classes with their own preferences. Also, the labeled DCE is able to measure how people weigh societal impacts when the COVID-19 measures that cause these impacts are explicitly mentioned. Unlabeled DCEs do not explicitly mention these measures, but only weigh the societal impacts.

Over the course of the end of 2022 and the beginning of 2023, the pandemic transitioned into an endemic. Therefore, this study will not only evaluate the advantages and disadvantages of using these models and DCEs during the pandemic, but also during the endemic, by asking the following question: *What are the (dis)-advantages of using ML and LC models over MNL models to analyze (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the pandemic and endemic?* The eventual goal of this study is to find out where the added value lies of ML and LC models compared to MNL models and labeled compared to unlabeled DCEs in informing future pandemic policy makers during different pandemic waves and the endemic. To answer the main research question, this study formulated three subresearch questions. The first subresearch question is defined as follows: *What are the differences in using ML, LC and MNL models to analyze (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the pandemic according to the literature?* To answer this question, the study conducted a literature review on the results obtained from labeled and unlabeled DCEs with MNL, ML and LC models in different waves of the pandemic. In the second part of the study, the following subresearch question is addressed: *What are the differences between the results produced by ML, LC and MNL models obtained from (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the endemic?* This question is answered by conducting a labeled and unlabeled Discrete Choice Experiment during the endemic. The third and last sub research question asks: *What are the (dis)-advantages of using ML, LC and MNL models to analyze (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the pandemic and endemic according to experts?* To answer this question, three expert interviews were conducted.

The results of this study show that the main advantage of the ML model is its ability to test for the existence of preference heterogeneity among individuals in a sample. This can help to check the reliability of the estimates produced by the MNL model. The reason for this is that the existence of preference heterogeneity means that the estimates of the MNL model do not adequately represent the majority of the sample, if the heterogeneity is high. Additionally, the existence of preference heterogeneity is a reason for further research into the origins of this heterogeneity. In this case, the LC model is able to estimate different classes with different preferences and characteristics that represent different subgroups in society. Providing such a classification is the main advantage of the LC model. The main disadvantage of the ML model is that it is time consuming to estimate the model. For the LC model the main disadvantage is its sensitivity to changes in covariates and initial values. With regards to the DCEs, the study shows that the main advantage of the labeled DCE is its ability to measure the effect of COVID-19 measures on how respondents weigh societal impacts. For the unlabeled DCE, the main advantage is its ability to measure how people view different societal impacts without explicitly mentioning the COVID-19 measures that caused these impacts.

With regards to the added value of DCEs for informing future pandemic policy in different waves of a pandemic and in an endemic, the study shows that unlabeled DCEs are most suitable to give a baseline estimation of the preference for societal impacts at the beginning of a pandemic. So that, these insights can be taken into consideration when the first pandemic policy package is created. Further, the study shows that labeled DCEs should be applied during and in between pandemic waves to evaluate and adjust implemented pandemic policy. Finally, a labeled DCE can be implemented during the endemic, that follows the pandemic, to evaluate the impact of COVID-19 measures. The insights obtained from this can help to inform a future pandemic. With regards to the added value of the ML and LC model in informing future pandemic policy in different waves of the pandemic and the endemic, the study shows that the ML model can be used to test the reliability of the mean coefficients of important societal impact attributes. This helps to verify if a specific societal impact attribute is a good target for mitigation with generic pandemic policy during pandemic waves. In between pandemic waves, the LC model can be used to elicit the origin of preference heterogeneity among people in the sample for different societal impact attributes. These insights can be used to create customized pandemic policy with increased public support and adherence for when a next pandemic wave hits.

The most important limitation of this research is the lack of unlabeled DCEs from both the pandemic and endemic and the lack of labeled DCEs from the endemic that are included in the literature review. A drawback of the included DCEs is that these experiments do not include exactly the same societal impact attributes and that these experiments are conducted during different waves of the pandemic in different countries. These factors make it difficult to compare the results of the studies. Also, this thesis recommends that future studies conduct further research on the societal impacts of COVID-19 measures with models that are extensions of the models that are used in this study, such as the logit or probit model and the LC model with distributed preference coefficients. Furthermore, it would be valuable to analyze the societal impacts of COVID-19 measures with models outside the domain of choice modeling, such as data driven models. Furthermore, this study emphasizes that a lot of research can be done into how and by whom the insights of this research should be implemented in the pandemic policy decision making process. For instance, questions for future research could be, should the insights of this study function as advice for the creation of pandemic policy or as directive? And does the government or the National Institute of Public Health decide upon this?

Introduction

In December 2019, the first outbreak of the COVID-19 virus took place in Wuhan, China (Risk assessment: Outbreak of acute respiratory syndrome associated with a novel coronavirus, Wuhan, China; first update, 2020). Three months later, the first Dutch citizen was infected with the virus (Man diagnosed with coronavirus (COVID-19) in the Netherlands, 2020). This was the start of several waves of infections across the country that resulted in increased mortality rates and overcrowded hospitals. The Dutch government tapered these waves down by implementing several measures. These measures consisted of administering vaccines and implementing non pharmaceutical interventions. Although, these measures were effective in slowing down the rate of spread of the COVID-19 virus, they also severely affected people's social and economic life. It is relevant for society to know how to measure these dentrimental effects to mitigate them in a future pandemic. Measuring people's preferences with regards to COVID-19 measures that impacted their social and economic life might even increase the effectiveness of pandemic policy in the future.

The remainder of this introduction will focus on the relationship between societal impacts and COVID-19 measures and it will explain how this relationship can be measured. After that, the knowledge gaps this study strives to fulfill will be discussed and consequently the main research question that encapsulates these knowledge gaps. Further, the subresearch questions and research approach will be discussed to answer this main research question. The last part of this introduction will discuss the relevance of this study to society, science and the COSEM master's programme.

Societal impacts of COVID measures

In this study, we focus on non pharmaceutical measures that aim to slow down the transmission of the COVID-19 virus by implementing infection protection and social distancing measures. Examples of these interventions are mask use in public, temporary closures of restaurants and stay-at-home orders. The effectiveness of these measures rely on the compliance of the public to them. Public support for the measures and actual adherence of the public to the measures is of the upmost importance for slowing down the spread of the virus. Both public support and public adherence are affected by the effects that the COVID-19 measures have on the public's social and economic life, in other words the societal impacts of the measures. There exist many examples of societal impacts due to the COVID-19 measures. For instance, freelancers that worked in restaurants and bars experienced a loss in income because their client's restaurant had to close due to a lockdown. Another example is the decline in mental health of students that were not able to get in touch with eachother physically, due to the lockdown closing universities. Also, physically disabled people that were in need of medical treatment were impacted, as their treatments were postponed due to overcrowded hospitals. On the other hand, people with a weak immune system supported many forms of COVID-19 measures, regardless of the impacts on their social life, in fear of COVID-19 infection and the detrimental effect of this infection on their health. These examples show that different societal aspects of people's life are impacted by COVID-19 measures. Some people might dislike certain impacts on their life more than others. They might therefore prefer some degrees of COVID-19 measures more than others.

Optimal COVID measures

To get the most effective COVID-19 measures, it is important to find out where the optimum lies between the degree of strictness of COVID-19 measures on the one hand, and the degree of impact these measures have on key societal impacts on the other hand. To illustrate the importance of finding this optimum, think about a pandemic wave in which no COVID-19 measures are implemented. At the beginning, some citizens might show support for this strategy but when the spread of COVID-19 starts to accelerate, more and more people will ask for the implementation of COVID-19 measures, due to

the increased morbidity and mortality rate and the overcrowded intensive care units (ICU). On the other hand, if strict COVID-19 measures are implemented, the rate of spread of the virus is low. However, as time goes by, more and more people will stop adhering to the measures because of the detrimental effect of the measures on societal aspects. This will in turn result in an increase in the rate of spread of the virus. Where the optimum lies also depends on the virus variant's infection rate and death rate during a wave in the pandemic. If the variant's death rate is very high, people are more likely to accept strict COVID measures even if the impact on their social life is significant.

Discrete Choice Experiments

A way to find where the optimum, as experienced in the population, lies during a specific wave in a pandemic is by using Discrete Choice Experiments. DCEs are surveys in which people are asked to make trade-offs between strictness of measures and degree of societal impacts. In this way, you retrieve the importance of certain degrees of measures relative to their impact on several aspects of people's social life. Furthermore, you also retrieve the relative importance of these societal impacts with one another. In a DCE, a respondent is asked to subsequently answer several choice tasks. In this case, these choice tasks present several policy scenario alternatives to these respondents. Each scenario consists of the same set of attributes that represent different societal impacts and in most cases an attribute that represents the strictness of COVID-19 measures. If the strictness of COVID measures is explicitly mentioned in the choice experiment it is called a labeled Discrete Choice Experiment, otherwise it is called an unlabeled Discrete Choice Experiment. The policy scenario alternatives differ in the level (degree) coupled to each attribute. For example, one policy scenario alternative might include a complete lockdown as COVID-19 measure attribute and a delay of less urgent surgeries as one of the societal impact attributes. While another scenario might only include wearing masks in public areas as COVID-19 measure and a delay of urgent surgeries as a societal impact.

Choice models

The choices that respondents make between the policy scenario alternatives in the choice tasks are used to estimate weights for all attributes of the policy scenario alternatives. These weights express the relative importance of the attributes. Determining the size of the different attribute weights is often done by estimating a so-called Multinomial Logit Model (MNL). This model estimates what the most likely weights are for each of the attributes for the total group of respondents. This is based on all the subsequent trade-offs these respondents had to make between the different attribute levels of the policy scenario alternatives they had to choose from in subsequent choice tasks. This method is especially useful to produce quick and reliable homogeneous results. During the pandemic quick and reliable homogeneous results were important for swift and effective implementation and adaptation of COVID-19 measures based on public support and adherence of the largest most homogeneous group possible.

Knowledge gap

Other choice models, such as mixed logit (ML) models and Latent Class (LC) models are able to identify heterogeneous results, which provide more information on the distribution of the relative importance of the different attributes among individuals and can help identify groups of individuals that give differing weights to the attributes. One of the main reasons to use these models over the MNL model is because a population naturally consists of respondents with different sociodemographic characteristics and views that result in different preferences for the trade-off between COVID-19 measures and societal impacts as discussed in the examples before. Therefore, identifying these heterogeneities in the population may provide valuable information for policy makers. Therefore, it would be interesting to study the specific advantages and disadvantages of using these models over

the MNL model to analyze DCEs, to see where exactly the added values lies in informing pandemic policy makers in the future. During the COVID-19 pandemic, many subgroups in society felt neglected and valuable information about their preferences was not utilized by policy makers. This could be a missed chance for developing customized measures that might have been more effective in slowing down transmission of the COVID-19 virus, while reducing damaging effects on citizen well being and economic prosperity.

At the same time it would be interesting to study the advantages and disadvantages of labeled and unlabeled DCEs. To see where its value lies in informing future pandemic policy. The reason for this is that labeled and unlabeled DCEs both have their own distinct advantages. For example, an advantage of unlabeled DCEs is that they purely measure the relative importance of the societal impacts. These measurements are not affected by explicitly mentioning the COVID-19 measures that cause these impacts. Some respondents might be biased towards certain COVID-19 measures, due to bad experiences. This can affect how they weigh the societal impacts. On the other hand, an advantage of labelled DCEs is that it allows citizens to show their preferences towards a specific COVID-19 measure in relation to the measure's societal impacts (Mouter et al., 2021). More importantly, a practical advantage of these different forms of DCEs is that each is able to inform policy makers in different waves of a pandemic. Policy makers can deploy an unlabeled DCE to already gather information on citizen's preferences for societal impacts without knowing what package of COVID-19 measures will be implemented (Chorus et al., 2020). This is useful at the start of the pandemic. Policy makers can deploy a labeled DCE during a pandemic wave to analyze what people think of the COVID-19 measures, but also what they thought of the COVID-19 measures in hindside, during the endemic.

The reason it is important to study DCEs that are conducted in different phases of the pandemic is because the phase affects respondent's choice behaviour (Chorus et al., 2020; Loría-Rebolledo et al., 2022). For example, at the beginning of the pandemic the morbidity and mortality rate of the virus were not clear yet. The fear of infection was high and people embraced COVID-19 measures that slowed down the rate of spread, regardless of the impact on their social life. On the other hand, at the end of one of the lockdowns, people were fed up with the seemingly endless prolongation of COVID-19 measures and the effect it had on their social life. Both timepoints in the pandemic would probably have yielded significantly different results. The literature substantiates these claims. At the beginning of the COVID-19 pandemic, three DCEs researched how respondents weigh societal impacts of COVID-19 measures (Chorus et al., 2020; Krauth et al., 2021; Reed et al., 2020). In the study of Reed et al. (2020) most individuals were not willing to accept higher COVID-19 infection risk even if it meant less financial security. Only 13 percent of the individuals of this study had a strong opinion in favour of re-opening non-essential businesses in the short run, to increase financial security. Krauth et al. (2021) state that for society the health of the public was more important than the economic health of the country. Individuals were willing to accept 20% unemployment rate for the upcoming two years to avoid insufficient intensive care unit capacity. These studies show that at the beginning of the pandemic, avoiding excess deaths and infections were the most important attributes. The DCEs that were conducted in a later stage of the pandemic showed increased importance for other societal impacts and increased negative preference for lockdowns. For example, Mühlbacher et al. (2022) showed that financial effects of COVID-19 policy such as deterioration in personal income had a significant negative impact on preferences of German people for lockdowns. A study conducted in the United Kingdom in this phase of the pandemic (Loría-Rebolledo et al., 2022), showed that 80% of the respondents were willing to accept an increase in surplus deaths for relaxations in lockdown restrictions. For instance, the average UK citizen was willing to accept around 14,000 surplus deaths to avert a very firm lockdown.

The review of literature in the previous alinea leads to the introduction of two knowledge gaps that this study strives to fill. Both knowledge gaps are situated in the context of DCEs that weigh societal impacts of COVID-19 policy. The first knowlede gap is: The advantages and disadvantages of ML and LC models over MNL models in this context. The second knowledge gap is: The advantages and disadvantages of labeled and unlabeled DCEs in this context. Both knowledge gaps are captured by the following main research question:

What are the (dis)-advantages of using ML and LC models over MNL models to analyze (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the pandemic and endemic?

The eventual goal of this study is to explain where the added value lies of ML and LC models over MNL models and labeled and unlabeled DCEs in informing future pandemic policy in different pandemic waves.

Research approach

This study will answer three subresearch questions. The subresearch questions and consequently the main research question are solved with a mixed methods research approach, consisting of qualitative and quantitative research. The mixed methods research approach of Creswell & Plano Clark (2017) is followed. The subquestions are formulated as follows:

- *What are the differences in using ML, LC and MNL models to analyze (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the pandemic according to the literature?*
- *What are the differences between the results produced by ML, LC and MNL models obtained from (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the endemic?*
- *What are the (dis)-advantages of using ML, LC and MNL models to analyze (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the pandemic and endemic according to experts?*

The first step in this mixed methods research approach focuses on answering the first subresearch question through qualitative research, in the form of a systematic literature review. The review will focus on the different results produced by ML, LC and MNL models from Discrete Choice Experiments that weigh societal impacts of COVID-19 policy. In addition, the review focuses on the difference in results due to the pandemic wave in which the DCE was conducted and the type of DCE that was conducted, namely unlabeled or labeled. The results from the literature review substantiate the results that are gathered during the quantitative research that is used to answer sub research question two.

The second subresearch question is answered using a quantitative research approach. A labeled and unlabeled DCE are conducted that analyze how respondents weigh different societal impacts of COVID-19 pandemic policy in the Netherlands during the endemic phase. Three different choice models will be used to analyze this data, namely the MNL, LC and ML model. An advantage of this method is that it will provide valuable practical insights on how to construct and analyze Discrete Choice Experiments. These insights create a better understanding of the different advantages and disadvantages of the models.

The last subresearch question is anwered with a qualitative research approach. The goal is to elicit what the advantages and disadvantages are of the different models and DCEs in weighing the societal impacts of COVID-19 policy, according to experts. The results of the literature review and the DCEs are presented to different experts and discussed in interviews.

Scientific relevance

Although, several DCEs were conducted during the first, second and third wave of the pandemic. No DCEs have been conducted in the endemic that we find ourselves in today. Therefore, this study adds to the scientific literature by conducting a labeled and unlabeled DCE in the Netherlands during the endemic. The results of these DCEs give insights on the current view among Dutch citizens on the impacts of COVID-19 measures on societal aspects. This can be used to compare to similar studies that were conducted in the Netherlands during the pandemic, such as Chorus et al. (2020), to see how Dutch citizen's choice behaviour changed compared to times in which COVID-19 measures were implemented. Also, the results can be used by researchers in other countries to compare their results to, when they deploy a DCE during the endemic. Furthermore, the only unlabeled DCE on the societal impacts of COVID-19 measures was conducted by Chorus et al. (2020). Therefore, the unlabeled DCE that was conducted by this study adds to the study of Chorus et al. (2020). In addition, the fact that a labeled and unlabeled DCE were conducted under the same conditions, provides additional insights to the scientific community on the differences between these types of DCEs in this context. This study also systematically reviews the literature on the results of labeled and unlabeled DCEs on the societal impact of COVID-19 measures during three waves of the pandemic in different countries. The results give information on the most important societal impacts and COVID-19 measures that were included in the studies. These societal impacts and COVID-19 measures can be used as initial attributes in DCEs that are conducted in future pandemics, because of the high probability that these attributes are relevant. The review also gives a substantial overview of the changes in choice behaviour that occur during subsequent pandemic waves. This emphasizes the importance of the time of measurement for the DCEs to researchers. Lastly, this study will include valuable insights from several experts on the difficulties of conveying the results from DCEs to policy makers and the public and how to solve these difficulties.

Societal relevance

The literature review will explain the development of preferences for the most important societal impacts of COVID-19 policy across several DCEs conducted in different countries at different timepoints. This information helps policy makers in future pandemics to understand what societal aspects are valued the most by citizens and will help them to anticipate on how the preferences for these aspects will develop during a future pandemic. This knowledge can help policy makers to adjust measures during the pandemic to alleviate the pressure on these aspects or to take precautionary actions to limit the impacts on these aspects. The labeled and unlabeled DCE conducted by this study during the endemic in the Netherlands elicit how different subgroups and individuals in society evaluate the impact of COVID-19 measures on social and economic impacts of their life. The value that these experiments elicit might stimulate policy makers to conduct similar experiments in a future pandemic to develop customized pandemic policy that might be more effective in slowing down transmission of a virus, due to increased public support and adherence, while reducing damaging effects on citizen's well being and economic prosperity. Lastly, the experts give advice on better ways of conveying the results produced by choice models to policy makers and citizens. This advice can be used to better inform citizens. Improved communication can enhance people's understanding for the need of pandemic policy. This may also result in increased public support and adherence.

Relevance to master's programme

At the heart of this study you will find the Discrete Choice Experiments that were designed to be presented to Dutch citizens. According to the Senior Advisor of the Corona behaviour unit of the RIVM, these DCEs are helpful in simulating a simplified version of the pandemic policy decision making process. A process in which policy makers have to deal with the complex nature of Dutch society that

include multiple stakeholder groups with widely diverging preferences. This fulfills the aim of the COSEM master thesis project to design a solution for a sociotechnical problem. This study also assessed the impact that technical solutions such as the Latent Class and Mixed Logit model can make on the pandemic policy decision making process by producing valuable insights from these DCEs. This fulfils the aim that COSEM students should assess the impact of technical solutions. Furthermore, the study discussed strategies with experts to effectively manage the complexity of the results that are produced by these models. In specific, the experts gave advice on how to convey these results to policy makers.

In the next chapter of this study, the different research methods that are used to answer the three subresearch questions will be discussed. This will be followed by the results to the three subresearch questions and finally the conclusion and discussion.

Research methods

This section discusses the research methods this study uses to answer the subresearch questions. It explains the selection process for the papers of the literature review and it explains what the aim of the review is. After that, it discusses the setup of the Discrete Choice Experiments, the data collection process, and the characteristics of the data. Followed, by an explanation of the data analysis process. This process describes the working mechanisms of the models this study uses and how the model performance is measured. Finally, this section discusses how the interviews are conducted and with whom the interviews are conducted. An overview of used research methods is given in the research flow diagram in Figure 1.

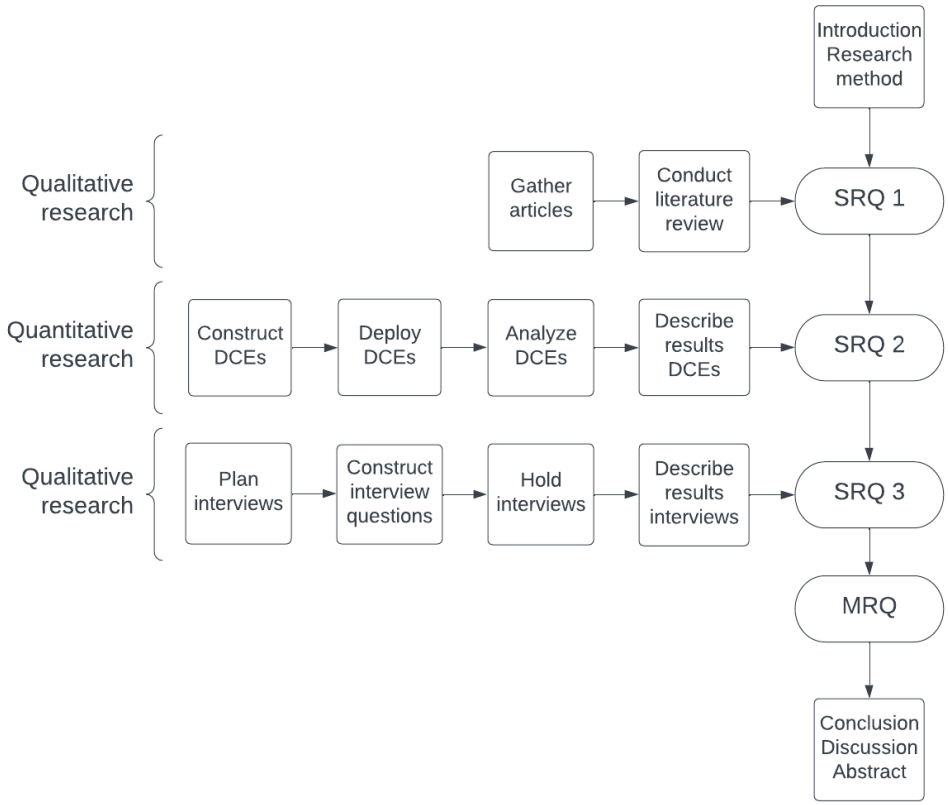


Figure 1: Research flow diagram.

Subresearch Question 1

Thist study answers the first subresearch question by systematically reviewing relevant literature. The literature discussed in the description of the knowledge gaps entails seven articles to which another seven articles are added by snowballing through the first seven articles or by using the following search query in the Google Scholar database:

"Discrete Choice Experiment" AND COVID AND (logit OR logistic OR "latent class")

The articles for the review were published between the beginning of the pandemic, at the end of 2019 until now. The articles discuss the use of MNL, ML and LC models to analyze Discrete Choice Experiments that weigh societal impacts of COVID-19 policy. This study manually checks if the articles focus on the societal impacts of COVID-19 policy. The reason that the query does not include 'the societal impacts of COVID-19 policy' is because this sentence can be formulated in many different ways. If this study excludes a possible formulation of this sentence it might also exclude relevant articles. Also, there exist many synonyms for the MNL and ML model. The query only contains the words logit and logistic to prevent that it excludes articles that use other terms. This means this study also manually checks if the articles produced by the query use the correct models.

After this study checks the relevancy of the papers, the papers are divided based on the choice model used to analyze the DCEs. This division is visualized by using a matrix in which the papers are ordered on the y-axis and the different choice models are mentioned on the x-axis. The DCE results of each paper are discussed in different paragraphs of the review. Each paragraph refers to a different choice model. Within each paragraph the results are chronologically ordered based on the wave in which a study was conducted. Additionally, for studies that used both ML and LC models to analyze their DCE, the review discusses the differences in results produced by these models. Because, this study only includes two unlabeled DCEs in total, the review also looks at literature that compares labeled and unlabeled DCEs outside the context of the societal impacts of COVID-19 policy. The results of the review are used to answer the first subresearch question. Furthermore, the results are compared to the DCEs that this study conducted to answer subresearch question two.

To be clear, the articles included in the review discuss how people view the impacts of COVID-19 measures on societal aspects of their life. The articles do not dicuss the actual impacts of COVID-19 measures on these societal aspects. To clarify, the actual impact of the measures on for example, the income of citizens, is the measured decline in their income in euros in reality. This study focuses on how much people dislike this decline in income. A hypothetical decline in income is part of the research, as it is presented to respondents through the attribute levels in the choice task. However, the models analyze how people in society perceive these hypothetical declines in income.

Subresearch Question 2

This study answers sub research question two by analyzing an unlabeled and labeled DCE that weigh the societal impacts of COVID-19 measures during the endemic by using several different MNL, LC and ML models. This section explains the data development process and the data analysis process. For instance, the part on the data development process, discusses the selection process of the attributes of the DCEs and its levels, the experimental design of the DCEs and the data collection process. The part on the data analysis process explains how MNL, LC and ML models work. For example, it discusses utility theory, estimation of attribute weights, accounting for heterogeneity, checking model performance and choosing model settings.

Data development

Attribute selection

Composing a Discrete Choice Experiment starts with choosing the attributes. The study of Chorus et al. (2020) was used as a starting point to choose these attributes. The DCE of Chorus et al. (2020) was conducted in the Netherlands during the first pandemic wave and also focused on the societal impacts of COVID-19 measures. The research included the following attributes; an increase in mortalities, an increase in psychological problems, an increase in physical health problems, an increase in financial problems, an increase in children with educational disadvantages, degree of strain on the medical system and a one-off COVID-19 tariff. After careful consideration with the research team of Populytics and the Societal Impact Team (SIT) of the Dutch government, it was decided that the attributes related to educational disadvantage and the one-off COVID-19 tariff were excluded, because of their irrelevance at the timepoint that this study's DCEs were conducted. Furthermore, the definition of some of the included attributes was changed. For instance, instead of using 'working pressure experience by health care workers' to describe the strain on the medical system, it was decided to describe this attribute as; the extent to which surgeries are delayed. Lastly, the study includes an extra attribute that is defined as; the stringency of COVID-19 measures. This attribute is included to measure how respondents make trade-offs between different levels of COVID-19 measures and societal impacts. Following the attribute selection process, these study's DCEs contain the following attributes; increase in COVID-19 related deaths, physical complaints and mental issues, increase in citizens with financial problems, delay of (non)-urgent surgeries and stringency of COVID-19 measures. The specific definitions of all attributes are listed in Table 1.

Attribute level selection

The levels for the six different attributes of this study's DCEs were chosen based on desk research and advice of experts. The decision was made to choose five levels for each attribute, to allow for considerable variation in choices. The middle level for the attribute on additional deaths due to COVID-19 infection is 7,000 citizens. This number is based on the COVID-19 mortality in 2022, as registered by Statistics Netherlands (Excess mortality for the third consecutive year in 2022, 2023). The other four levels are extrapolated from the middle level and range between 4,000 and 10,000 citizens.

Secondly, the attribute levels for the attribute that relates to an increase in physical health problems ranges from 150,000 to 550,000. The reason this range was chosen is because the RIVM estimated that one out of eight people infected with the the virus would have longlasting physical complaints. In 2021, circa five million people contracted the virus. Therefore, it was calculated that 625,000 people had longstanding physical complaints due to the infection. However, because the coronavirus variant in the winter of 2021/2022 resulted in less severe physical complaints, the attribute levels were set to be lower than 625,000.

Thirdly, the attribute that relates to an increase in mental issues has a middle level of 450,000. This level was chosen from a study of (Ervaren impact corona op mentale gezondheid en leefstijl, 2021) that shows that the number of Dutch citizens with psychological problems increased with 12% from 2014 to 2020. In 2021 alone, this number increased to 15%. This increase of three percent relates to 450,000 Dutch citizens in the group of people that are twelve years and older. The other levels range from 150,000 to 750,000.

Fourthly, the middle level of the attribute that is defined by an increase in citizens with financial problems is 300,000. The reason for this number is the following: The past ten years before the pandemic, the number of Dutch people with financial problems was one million per year on average (Kansrijk armoedebeleid, 2020). In some years, the number of people with financial problems dropped to 900,000 and in other years it rose to 1,200,000. Hence, the decision was made to choose attribute levels of which the overall variance was as large as the variance of people with financial problems in the past decade. This resulted in attribute levels ranging from 0 to 600,000.

Fifthly, the levels of the attribute that relates to surgery postponement were chosen based on advice from RIVM experts that stated that at the end of October 2022, non-urgent operations were delayed for approximately one to five months. The experts also added that at the height of the COVID-19 pandemic, urgent surgeries were delayed for approximately one month. For this reason the lowest attribute level is equal to a one month postponement of non-urgent surgeries and the highest attribute level is equal to a one month postponement of urgent surgeries and a five month postponement of non-urgent surgeries. Lastly, the attribute levels for the stringency of COVID-19 measures were based on the first wave of the pandemic and range from wearing a mask and testing for COVID-19 infection to closure of restaurants and nightclubs. An overview of all attribute levels can be found in Table 1.

The chosen attribute levels were tested in a pilot survey. The analysis of the pilot survey showed that the attribute for psychological issues was statistically insignificant. There was no measurable change in preference from respondents when confronted with different attribute levels of this attribute. To solve this, the variation in attribute levels for this attribute was increased. Furthermore, the variation in the attribute levels for delaying medical treatments was reduced as some respondents non-traded on this attribute.

	Level 1	Level 2	Level 3	Level 4	Level 5
Additional deaths in 2023 due to the COVID-19 pandemic	4,000	5,500	7,000	8,500	10,000
Additional number of citizens with physical complaints longer than 3 months in 2023 due to the COVID-19 pandemic	150,000	250,000	350,000	450,000	550,000
Additional number of citizens with mental health issues longer than 3 months in 2023 due to the COVID-19 pandemic	150,000	300,000	450,000	600,000	750,000
Additional number of citizens who have difficulty making ends meet in 2023 due the COVID-19 pandemic	0	150,000	300,000	450,000	600,000
Will surgeries have to be postponed in 2023 because there are many COVID-19 patients in hospital?	There is no need to postpone surgeries	Hospitals have to postpone some surgeries for about 1 month. This happens only for surgeries that are not so urgent, such as knee surgeries and cataract surgeries.	Hospitals have to postpone some surgeries for about 3 month. This happens only for surgeries that are not so urgent, such as knee surgeries and cataract surgeries.	Hospitals have to postpone some surgeries for about 5 month. This happens only for surgeries that are not so urgent, such as knee surgeries and cataract surgeries.	Hospitals postpone surgeries that are not as urgent (such as knee surgery and cataract surgery) by about 5 months. Some surgeries that are urgent but not life-threatening, such as some heart surgeries, are also postponed by about 1 month.
Are there any COVID-19 measures taken that will affect the daily lives of citizens in 2023? (only in unlabeled DCE)	There are no measures that affect our daily lives	The measures have minor effects on our daily lives. For	The measures affect our daily lives. For example	The measures have big implications for our daily lives. For	The measures have very big impacts on our daily lives.

		example, compulsory wearing a mouth mask in the supermarket and Public Transport	compulsory wearing a mouth mask. And taking a COVID-19 test to go to concerts and sports events	example, fewer people are allowed in a restaurant or café	Nightclubs, restaurants and cafes have to close, for example

Table 1: Overview of different attributes and attribute levels as included in both DCEs.

Experimental design

The selected attributes were used to create two DCEs, one labeled and one unlabeled DCE. The labeled DCE includes the attribute that relates to the strictness of COVID-19 measures. A so-called D-efficient experimental design was used to create 20 choice tasks for each Discrete Choice Experiment. Each choice task presents two alternatives to respondents that represent different policy scenarios. These policy scenarios have the same attributes but different levels. An example of a choice task is shown in Figure 1.

The D-efficient experimental design follows established methodologies for conducting healthcare-related Discrete Choice Experiments (Johnson et al., 2013). The D-efficient design ensures that the variability of choice model estimates is minimized. This is done by carefully selecting attribute levels for the policy scenarios of the 20 choice tasks. The design's primary objective is to reduce the D-error, which determines the expected variance-covariance matrix of a choice model given a fixed number of choice tasks and analyst-defined prior parameters. By utilizing D-efficient designs, statistical efficiency is maximized, leading to smaller required sample sizes during data collection.

The creation of the DCE's D-efficient design consisted of two stages. In the first stage, 20 choice tasks were constructed for each DCE in the pilot survey. Attribute priors were assigned small values, and their signs were determined based on previous studies on COVID-19 preferences. The design aimed to minimize D-error for a MNL model with linear utility functions. Additionally, the experimental design was restricted to exclude dominant and dominated alternatives within choice situations. This restriction was necessary to ensure that the alternatives provided relevant information regarding respondents' trade-offs for attributes, thereby preserving the statistical efficiency of the final model. If a DCE contains dominant alternatives this mean that a significant number of respondents choose this alternative without considering the attributes and their levels, this results in biased attribute coefficients.

In the second stage, the responses from the pilot survey were used to estimate an MNL model, and the resulting estimates were used as priors to construct the final set of 20 choice situations for each DCE. The final choice tasks utilized the attribute levels presented in Table 1 while maintaining the same restrictions to prevent dominant or dominated alternatives. All experimental designs were generated using Ngene, a software specifically designed for constructing experimental designs for Discrete Choice Experiments.

Figure 2: Example of a choice task from the labeled DCE.

Data collection

The respondents for both DCEs were sampled from an internet panel of a market research company called Dynata between November 24, 2022 and December 12, 2022. The respondents are representative for the Dutch adult population with regards to age, gender and education. The labeled DCE was answered by 1106 respondents. Every respondent answered five choice tasks in the labeled DCE. Therefore, the total number of observations for this DCE sums up to 5530. The unlabeled DCE was answered by 1070 respondents. Every respondent had to answer six choice tasks. In total, there are 6420 observations for this DCE.

Besides answering the choice tasks of the DCEs, the respondents also provided data on several socio-demographic characteristics such as their gender, age and educational level. In addition, the respondents answered several statements. Both the sociodemographic characteristics and statements are used as covariates in the estimation of classes for the LC model. All different covariates are mentioned in Table 2.

Covariates
Gender
Educational level
Age
Income Issues
Chronic Disease
Roommate with Chronic Disease
Vaccination
Can't live desired life due to COVID-19
Social life deteriorated due to COVID-19
Feeling worse due to COVID-19
COVID-19 would make me very ill
I would be hospitalised due to COVID-19
I would die of a COVID-19 infection

Table 2: Covariates used to estimate the LC models in this study.

Data analysis

Both DCEs are analyzed with the following models; a regular MNL model, a MNL model with dummy coded 'surgery delay'-categorical variables and 'stringency of COVID-19 measures'-categorical variables, a LC model with covariates and a ML model. The dummy coded model is produced with the help of the Biogene package in Python. The other models are produced with the help of the Apollo package in R studio.

First, a regular MNL model is estimated. The estimates are used to compare preferences between the different samples that answered the labeled and unlabeled DCEs. The estimates will also be used to calculate the marginal rate of substitution (MRS) for an increase in deaths over an increase in one of the other attributes. The MRS provides insights on the willingness to accept increases in negative societal impacts to avoid an increase in deaths.

Secondly, a dummy coded MNL model will be estimated to elicit what the attribute weights are for the different attribute levels of the attributes; surgery delay and COVID-19 measures. These estimates can again be used for comparing the preferences of people that answered the unlabeled and labeled DCEs and to calculate the MRSs for the different levels of the categorical variables.

Secondly, a LC model is estimated with the covariates from table two. To determine the optimal number of classes for this LC model, several models with one to six classes are fitted without covariates. The chosen number of classes depends on the goodness-of-fit statistics. The goodness-of-fit statistics used in this study are the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC) and the adjusted rho squared. Besides this, the covariate estimates are used to define the different classes and to estimate class probabilities. The estimates are also used to determine class membership probabilities. These probabilities explain the chance that a respondent with certain characteristics belongs to a certain class. The goal of fitting a latent class model is to identify (discrete) heterogeneity among groups in the sample.

Thirdly, a ML model is estimated that includes random coefficients. The goal of fitting the ML model is to identify (continuous) heterogeneity among individuals in the sample. The mean and standard deviation estimates provide information on the dispersion of preference among individuals for the different DCE attributes. The model estimates also provide insights on the differences in the samples that answered the labeled and unlabeled DCE.

The estimation results and goodness-of-fit statistics of all models are used to compare results and model fit. The eventual goal of using different choice models to analyze the same DCEs is to find out what the differences are between the results obtained.

MNL model

The first model that is used to make estimations is the MNL model. This model is build on the premises that respondents that are asked to make a decision between alternatives from a discrete choice task, try to maximize their utility. Every alternative's utility is presented by a linear additive function. This function consists off a sum of multiplications between the different attribute level values for the alternative in the choice task times the attribute's weight. This linear additive function is shown in Equation one.

$$U_{nj} = V_{nj} + \varepsilon_{nj} = \beta' X_{nj} + \varepsilon_{nj},$$

Equation 1: The total utility of alternative (j) for respondent (n)

In Equation one, V_{nj} is the systematic utility for an alternative from a specific choice task, which is equal to $\beta' X_{nj}$. Where X_{nj} is the vector of attributes of alternative j from a choice task and β is a vector of weights for these attributes. ε_{nj} Is an error term with Extreme Value (Gumbel) distribution that includes all unobserved factors and errors of measurement. The error term entails individual specific preferences and omitted variables. For example, someone's aversion against delaying surgeries or excluding unemployment rate from the model. The difference in utility per person is due to this individual and task specific error term, as the systematic utility is the same for all individuals. Train. (2009) shows that the probability of choosing alternative i by respondent n is given by the following equation:

$$p_{ni,MNL} = P(U_{ni} \geq U_{nk}, \forall i \neq k) = \frac{\exp(V_{ni})}{\sum_j \exp(V_{nj})}$$

Equation 2: The probability of choosing a certain alternative (i) from a choice task by respondent (n) , where V refers to the total utlity utility minus the error term.

Equation two shows that the probability of the chosen alternative for a specific choice task presented to an individual given certain attribute weights is obtained by taking the exponential function of the log-odds of the systematic utility of the chosen alternative over the systematic utility of all alternatives. To clarify, the probability is calculated based on the observed factors that are part of the systematic utility. A particular individual in a particular choice situation can still decide to choose an alternative that does not have the highest systematic utility, but does have the highest total utility (McFadden., 2000). Furthermore, it is important to note that equation two does not give the probability that an individual chooses a certain alternative. The data on what alternative is chosen by respondents is already collected. In this study, the probability is not used to predict, It is used to obtain the likelihood that the given answer is obtained for specific values of the β vector.

$$L(\beta) = \prod_{t=1}^{Tn} P_{j^*,n,t}$$

Equation 3: Likelihood function given beta, as the product of the probabilities of chosen alternatives (j) over all choice tasks of one individual (Tn).*

The product of the probabilities of all the chosen alternatives from the different choice tasks that an individual answered are shown in equation three, the likelihood equation for that individual. The data contains multiple successive choices per individual and therefore can be seen as panel data, this is accounted for by multiplying the probabilities across individual choice observations for the same individual. The multinomial logit model applies equation three to all choice tasks of all individuals. After that, the model estimates the set of values for vector beta that maximize the log-likelihood. In other words, the model finds the set of betas that make the data the most likely (McFadden., 2000).

Model settings

The initial values of all betas for the multinomial logit model in this study are set to zero to obtain initial estimates for the choice model, without constraining the estimation process to what is to be the expected direction of the attribute weight values based on the attribute levels in the Discrete Choice Experiment. Furthermore, the distribution of the error term is closed form which ensures that the choice probabilities can be computed without simulation. Lastly, The Broyden Fletcher Goldfarb Shanno estimation method is used to obtain the parameter values for which the likelihood is maximized. This estimation method was chosen as it is quicker than Newton's methods. Broyden Fletcher Goldfarb Shanno estimation is a quasi-newton method that uses hessian approximation instead of actual calculation like the Newton's methods do. This saves a lot of time.

LC model

Another model that is used to analyze the unlabeled and labeled DCE is the LC model. The LC model estimates several MNL models with differing attribute weights for each model. Each model, with its own weights, belongs to a class of respondents. The number of classes is based on the model fit. The probability that a respondent n chooses an alternative i is given by the following equation:

$$p_{ni,LC}(\beta) = \sum_{c} \pi_{nc} \cdot p_{ni,MNL}(\beta_c)$$

Equation 4: Sum of class probabilites (π) times the probabilities for choosing an alternative ($p_{ni,MNL}$) given the attribute weights for that class (β_c).

In Equation four, π_{nc} is the probability that respondent n is part off class c. β_c is a vector of estimated attribute weights that belong to class c. The sum of class probabilities over all classes is given by $\sum_c \pi_{nc}$ and adds up to 1. Equation four shows that the choice probabilities of a LC model are a weighted sum of MNL choice probabilities per class. Each class represents a subgroup of the sample that is defined by their own set of class attribute weights β_c.

The class probabilities are calculated with the following equation:

$$\pi_{ni} = \frac{\exp(\delta_i)}{\sum_c \exp(\delta_c)}$$

Equation 5: Exponential function of log-odds of one latent class constant (δ) over all other latent class constants to calculate the probability for a respondent (n) to belong to class (i).

Equation five shows that the probability for the different classes is calculated in the same way as the choice probabilities. In this equation δ stands for a constant that represents a certain class. If covariates are added to the LC model, each class is not only presented by a constant. In addition to the constant, each class is presented by a linear additive function of covariates and their weights.

Model settings

The initial values of the LC model are determined by an algorithm named the Searchstartvalue Algorithm from the apollo package. The distribution of the error term is closed form and the Broyden Fletcher Goldfarb Shanno estimation method is used to obtain the parameter values for which the likelihood is maximized, just like for the MNL models.

Mixed Multinomial Logit model

The study also includes a ML model. By turning the Multinomial Logit Model into a Mixed Multinomial Logit model with random coefficients it is possible to include the heterogeneity that is captured as an unobserved factor in the error term of the Multinomial Logit Model and make it observable by fitting it to attribute coefficients. This can give a better fit of the model to the data.

In a ML model, the utility of a respondent n for alternative j is a function of individual-specific attribute weights:

$$U_{nj}^{MXL} = V_{nj} + \varepsilon_{nj} = \beta_n' X_{nj} + \varepsilon_{nj}$$

Equation 6: Utility function for the ML model with idiosyncratic attribute weight β_n.

where β_n is a vector of idiosyncratic attribute weights. However, estimating idiosyncratic attribute weights is difficult, thus the researcher assumes that β_n is a vector of random coefficients with a known distribution, such that:

$$U_{nj}^{MXL} = V_{nj} + \varepsilon_{nj} = \beta^{r\prime X_{nj}} + \varepsilon_{nj}$$

Equation 7: Utility function for the ML model where β^r is a vector of random coefficients.

where β^r is a vector of random coefficients. The researcher fits a distribution to the coefficients and estimates the parameters that characterize such distribution, such as the mean and standard deviation. In this way, the researcher can account for heterogeneity across respondents.

It is important to note that not necessarily all attribute weights are turned into random coefficients. The ML model is built by turning crisp attribute weights from the MNL model into random coefficients, step-by-step. If the random coefficient for a certain attribute is significant, it is kept random. Otherwise, it is turned back to a crisp parameter without distribution.

The chosen distribution for the different random coefficients depends on the size of the sample and the expected sign of the attribute weight. For example, in this study the sample size is large. Therefore, choosing a normal distribution is appropriate. Furthermore, all attributes are expected to have a negative sign because respondents are expected to dislike increases in the attributes of the DCEs. Knowing this, fitting a negative lognormal distribution to the attribute weights might result in a better fit of the model. However, in first instance, a ML model with normal distribution is fitted. So that, the results of this model can be compared to the results from the model with negative lognormal distributions.

The Mixed Logit probability of individual n's chosen alternative for every choice task given beta is given by equation eight. For this equation β is the set of true but unobserved coefficients for individual n which is distributed according to $g\,(\beta \mid \Omega)$, where Ω contains the mean and variance of the distribution. Because there are multiple observations per individual in each dataset, the assumption is made that sensitivities vary across people, but stay constant across individuals. Note that, since the same sensitivities apply to all choices by a given individual, the integration over the density of β applies to all the consumer's choices combined (McFadden & Train, 2000).

$$P_n\,(\Omega) \;=\; \int_{\beta} \prod_{t=1}^{Tn} P_{n,t}\,(j*_{n,t} \mid \beta)\;\; g\,(\beta \mid \Omega)\; d\beta$$

Equation 8: The Mixed Logit probability for the chosen alternative (j) of individual (n) over all choice tasks (t) given β and distribution parameters Ω.*

The log-likelihood function of the mixed logit model is given in equation nine. The sum of the log of the likelihoods is taken for all individuals over all their choice tasks. The integrals do not take a closed form, but are approximated via simulation (McFadden & Train, 2000).

$$LL\,(\Omega) \;=\; \sum_{n=1}^{N} ln\left(\int_{\beta}\left[\prod_{t=1}^{Tn}(P_{n,t}\,(j*_{n,t} \mid \beta))\right] g\,(\beta \mid \Omega)\; d\beta\right)$$

Equation 9: The log-likelihood function of the mixed logit model as the sum over all individuals of the log of the likelihoods of an individual's sequence of choices.

The simulation of the loglikelihood is given in equation ten. In this equation R is the number of draws. In this equation the product over choice situations is calculated for each draw. The product is averaged over the draws and after that the log of the average is taken. The number of logit probabilities that are computed is RT_n (McFadden & Train, 2000).

$$SLL\ (\Omega)\ =\ \sum_{n=1}^{N} ln\left(\frac{1}{R}\sum_{r=1}^{R}\left[\prod_{t=1}^{Tn}(P_{n,t}\ (j\ *_{n,t}\ \mid \beta_{r,n}))\right]\right)$$

Equation 10: The simulated log likelihood is the sum over individuals of the log of the average (across draws) of products over choice situations. Where R is the total number of draws and r is a specific draw.

Model settings

The lognormally distributed random coefficients have an initial mean of -3 and sigma of 0.01. These are the starting values of the mean and sigma of the logarithm of a beta. Starting the mean of log beta at -3 is the same as startig the median of beta close to zero, which would be the starting value for a normally distributed beta. All mixed logit models use 1000 halton draws to simulate the log-likelihood. Increasing the number of draws even further had limited effect on the accuracy of the model but did greatly affect the runtime of the model.

Model performance

Several goodness-of-fit statistics are used in this study. This section explains how the BIC, adjusted rho squared and AIC work.

$$BIC = -2 \cdot LL + \ln(N) \cdot k$$

Equation 11: The Bayesian Information Criterion that corrects the log-likelihood (LL) for the size of the sample (N) and the number of parameters used in the model (k)

Equation 11 shows the formula for calculating the BIC, where LL is the maximized log-likelihood of the model and k is the number of estimated parameters of the model. The BIC accounts for both the number of estimated parameters and the sample size. The model with the lowest BIC value is deemed to be the most optimal model. Nonetheless, a model with somewhat higher BIC but a smaller number of classes can be chosen if the researcher can get a more rich interpretation of results from such model. In general, the BIC value gets priority over other goodness-of-fit statistics as it applies the most stringent correction measures.

The adjusted ρ^2 was also included as a measurement-of-fit statistic for the different models. The adjusted ρ^2 is calculated with the following equation:

$$\bar{\rho}^2\ =\ 1\ -\ \frac{LL(\hat{\theta})\ -\ k}{LL(0)}$$

Equation 12: shows the adjusted rho squared with the log-likelihood of the estimated model $LL(\hat{\theta})$ and the log-likelihood of the nul model $LL(0)$, corrected for the number of estimated parameters.

The adjusted rho squared gives a value from zero to one, with zero being a fit as bad as the nul model and one for a model that is equal to the true data generating process. The statistic explains how much of the variance of the results is explained by the model (Miles., 2005).

The Akaike information criterion, given in equation 13, is a statistic that corrects the log-likelihood of the estimated model for the number of estimated parameters. The lower the value of the statistic the better. The AIC scores of the models are only useful in comparison of the models (Ben-Akiva & Swait, 1986).

$$AIC = -2LL(\hat{\theta}) + 2K$$

Equation 13: shows the Akaike Information Criterion.

All measurement-of-fit statistics correct for the number of estimated parameters. This is done, because increasing the number of parameters in a model almost always improves the fit of the model, which would result in overfitting.

Sub Research Question 3

Three experts are asked to evaluate the advantages and disadvantages of the results produced by the DCEs and the literature review by conducting semi-structured interviews. This form of interviews is chosen as it is a combination of systematically asking several key questions based on the results of the previous research questions and asking several follow-up questions that are dependent on the interviewee's responses. The three experts are a representative of Populytics, the National Institute of Public Health and the Environment and the TU Delft. All three experts are familiar with the results of the DCEs on societal impacts of COVID-19 policy but can provide different angles. For example, the TU Delft researcher has worked with the different models used in this study on many occassions. Because of this, the researcher is able to explain the differences in results of the models by understanding the nature of the models. The representative of the National Institute of Public Health and the Environment can provide feedback on the interpretation of the results from the perspective of a regulatory body. The eventual goal of the interviews is to elicit the added value of the models in producing useable insights for creating effective COVID-19 measures.

Results

This section discusses the results obtained with the different research methods. These results help to answer the sub research questions. Subsequently, the results from the literature review, Discrete Choice Experiments and interviews are discussed. The results from the review and the DCEs provide information on how the MNL, LC and ML model are used during a pandemic and endemic. The expert interviews help to understand how these results can be used to inform pandemic policy.

Sub Research Question 1

This segment of the study answers the first sub research question by reviewing the literature on weighing societal impacts of COVID-19 measures with DCEs during different waves of the pandemic in different countries. The review categorizes the DCE results from the different papers based on the models that were used to analyze the DCEs. The review aims to elicit the differences in the results produced by different models. This provides insights on the different ways these models can be used. Furthermore, it creates understanding about the development of preferences over different phases of the pandemic.

Pandemic phase	Country	Article	MNL	ML	LC	Unlabeled
First wave	United States	Reed et al. (2020)			X	
	Singapore	Ozedemir et al. (2021)	X			
	Germany	Krauth et al. (2021)	X		X	
	Australia	Degeling et al. (2020)	X	X		
	Italy	Belle & cantarelli (2022)	X			
	The Netherlands	Chorus et al. (2020)	X		X	Yes
Second wave	Germany	Mühlbacher et al. (2022)		X		
	United Kingdom	Loría-Rebolledo et al. (2022)	X			
	France	Sicsic et al. (2022)		X		
	Australia	Manipis et al. (2021)		X	X	
	The Netherlands	Mouter et al. (2021)		X		
	United States	Li et al. (2021)		X		
Third wave	France, India, Italy, UK, US	Fink et al. (2022)	X			
	Portugal	Filipe et al. (2022)	X		X	

Table 3: Articles includeed in the literature review

Table three shows a matrix with 14 articles on the y-axis and different model subgroups on the x-axis. The crosses in the columns of the matrix show if an article used a certain choice modeling technique to obtain their results. The last column of the table shows if the DCE used in each article is unlabeled or labeled. The first two columns show the pandemic wave and country in which the study was conducted. This table is used as a guideline to structure the review. The review starts by analyzing the articles that use the MNL model, followed by, the ML model and the LC model. The paragraph on articles that use the MNL model is structured based on the wave in which the article was conducted and discusses the estimates of the most important attributes. The paragraph on articles that use the ML model is divided into a discussion of the mean estimates and standard deviation estimates. The last paragraph discusses the class preferences, class probabilities, and class membership profiles that are discussed in the papers that use the LC model. In addition, all model paragraphs discuss the results for the marginal rates of substitution for the attribute coefficients. The second part of the review looks at the additional insights of the only three articles that used both a MNL and LC model. Lastly,

literature on labeled and unlabeled DCEs outside the context of this study is analyzed, due to the lack of unlabeled DCEs that weigh the societal impacts of COVID-19 measures.

MNL

Mean estimates

First wave

In the first wave of the pandemic, the studies of Chorus et al. (2020), Ozedemir et al. (2021), Belle & Cantarelli (2022) and Degeling et al. (2020) all identify COVID-19 related deaths as crucial factors in determining the acceptance of COVID-19 measures. While Chorus et al. (2020) conducted an unlabeled DCE that only considered various societal impacts such as income loss, physical health problems, mental issues, educational disadvantage and tax increase besides COVID-19 related deaths. The other studies conducted labeled DCEs that explicitly mentioned different types of COVID-19 measures, besides societal impacts. For instance, Belle & Cantarelli (2022) focused on income loss and COVID-19 related deaths, due to an implemented lockdown. Ozedemir et al. (2021) focused on several COVID-19 policies such as lockdowns and school closures in combination with the number of COVID-19 related deaths and infections. Degeling et al. (2020) and Ozedemir et al. (2021) emphasize that the number of COVID-19 related deaths was more important than the number of infections in their study. Belle & Cantarelli (2022) state that an increased loss in income is more disliked than a similar percentage increase in deaths. This is in contrast with the study from chorus et al. (2020). Indicating that the Italians value income loss more than the Dutch in the first wave of the pandemic. Another interesting finding from the first wave of the pandemic comes from Krauth et al. (2021). This study did not consider attributes such as COVID-19 related deaths and infections, but focused on the effects of COVID-19 measures on available ICU capacity and unemployment rate. An interesting finding of this study is that, only a large increase in unemployment rate was significantly disliked in the first wave of the pandemic. A small increase in unemployment was not, suggesting that the elasticity of this attribute was large in Germany at the beginning of the pandemic.

Second wave

In the second wave of the pandemic Loría-Rebolledo et al. (2022) conducted a DCE in which an increase in deaths is the most important attribute. Citizens from all parts of the United Kingdom had a significant negative association with this attribute. Although, the citizens of England cared the least about this attribute. These respondents were most willing to accept a higher increase in deaths to get less strict lockdown restrictions. Furthermore, this study showed that respondents throughout the UK disliked an increase in unemployment.

Third wave

In the third wave of the pandemic Filipe et al. (2022) launched a DCE in Portugal. The respondents disliked an increasing number of deaths the most. Respondents also disliked an increase in educational disadvantage and loss of income for households. Chorus et al. (2020) also showed a negative association of respondents with an increase in educational disadvantage in the first wave. Filipe et al. (2022) showed that a small increase of students with educational disadvantage is more disliked than a small increase in income losses. However, a larger increase of income losses is more disliked than a larger increase of students with educational disadvantage. This indicates that respondents in this study have a non-linear relationship with an increase in income loss. Chorus et al. (2020) showed in their study that an increase in income loss is less disliked than an increase in educational disadvantage in the first wave of the pandemic in the Netherlands.

In conclusion, Ozedemir et al. (2021), Chorus et al. (2020), Degeling et al. (2020), Belle & Cantarelli (2022), Loría-Rebolledo et al. (2022) and Filipe et al. (2022) highlight the importance of COVID-19 related deaths as a significant factor in the acceptance of COVID-19 measures. However, Belle & Cantarelli (2022) show that for their respondents an increase in deaths is not the most important attribute. For this article the loss of income was considered to be more disliked. This is an interesting finding as the study's DCE was conducted in the first wave of the pandemic in Italy. Other studies in the first wave show that the loss of income is an important attribute but not more important than an increase in deaths. Furthermore, Krauth et al. (2021) and Loría-Rebolledo et al. (2022) showed that an increased unemployment rate was disliked by their respondents. Although, Krauth et al. (2021) showed that this was only significant for high unemployment rates in the first wave in Germany. Loría-Rebolledo et al. (2022) showed that in the third wave in the UK all levels of increase in unemployment rate were disliked. Filipe et al. (2022) and Chorus et al. (2020) considered an increase in educational disadvantage. Both showed that a small increase in educational disadvantage was more disliked than a small increase in income loss. Although, Filipe et al. (2022) showed that large decreases in income were more disliked than large increases in educational disadvantage. This could be because Filipe et al. (2022) was conducted in the third wave of the pandemic.

Marginal rate of substitution

First wave

During the first wave of the pandemic, several articles looked at the marginal rate of substitution for various attributes. While Belle and Cantarelli (2022) and Chorus et al. (2020) focused on the willingness to accept trade-offs related to deaths and societal impacts, Krauth et al. investigate the trade-offs between specific lockdown measures and the unemployment rate. Belle and Cantarelli's findings reveal that respondents have a preference for avoiding income losses over a reduction in the number of COVID-19 deaths. The researchers also demonstrate that the relative preference for saving income over saving lives widens as the percentage of income and human losses at stake increases. The researchers were able to demonstrate this by estimating the MRS for the different categorical variables that constitute their income loss attribute. Chorus et al. (2020) shows that citizens are most willing to sacrifice the income of households and least willing to sacrifice people's mental health. These citizens were also more likely to accept a decrease in income for households than an educational disadvantage to children. Krauth et al. (2021) shows that people are most willing to accept a steep increase in unemployment rate for the prevention of a mandatory tracing app and after that, prevention of ICU overload. Citizens are least interested in incurring higher unemployment rate for re-opening schools.

Second wave

In the second wave Loría-Rebolledo et al. (2022) looked at the willingness to accept several degrees of lockdowns to prevent an increase in deaths. The maximum number of acceptable deaths is smaller when lockdown measures become longer and more strict and postpone routine medical treatments. This indicates that respondents are willing to accept stricter and longer lockdowns. But, only if these lockdowns keep the number of deaths small.

Third wave

In the third wave of the pandemic two studies took place. Fink et al. (2022) investigated the willingness of individuals across France, Italy, India, the US and the UK to sacrifice a portion of their income to avoid specific COVID-19 measures. The researchers found that respondents are willing to give up a significant percentage of their annual salary to prevent school closures and closures of restaurants,

bars, and clubs. However, their study also reveals lower willingness to pay to prevent measures such as travel restrictions and wearing masks in public. Filipe et al. (2022) studied the willingness to sacrifice different attributes to avoid an increase in deaths. Respondents were least willing to risk an increase of poverty among the population to avoid deaths. Secondly, respondents were least willing to sacrifice a decrease in income to avoid deaths. On the other hand, people were more willing to accept an increase in students with educational disadvantage and a strict lockdown to avoid excess deaths. Due to the use of categorical variables in the calculation of the MRS, Filipe et al. (2022) is able to show that a 20% increase of lost income is only accepted if it saves roughly 4,5 times as much lives as a 10% increase does. To compare, for people in this study to accept high life restrictions instead off medium life restrictions, only 24 lives per day have to be saved instead off 19. This shows how strongly people in Portugal value income. It is also interesting to see that the willingness of people in Portugal to incur income loss due to COVID-19 measures in the third wave of the pandemic is much lower than for people in the Netherlands in the first wave of the pandemic (Filipe et al., 2022; Chorus et al., 2020).

To summarize, the literature shows that the MRS in the context of societal impacts of COVID-19 measures often calculates the willingness to accept changes in different attributes, such as lockdown measures to prevent an increase in deaths (Belle & Cantarelli., 2022; Chorus et al., 2020; Loría-Rebolledo et al., 2022; Filipe et al., 2022). Other ratios that are used are the willingness to trade-off different attributes, such as lockdown measures versus a loss in income (Fink et al., 2022) or an increase in unemployment (Krauth et al., 2021). Chorus et al. (2020) shows that citizens were most willing to sacrifice income to avoid deaths. Furthermore, Fink et al. (2022) and Filipe et al. (2022) conducted their DCEs in the third wave of the pandemic. Fink et al. (2022) and Filipe et al. (2022) show that respondents were less willing to pay for avoiding certain lockdown measures and deaths than studies in earlier waves of the pandemic.

ML

Mean estimates

First wave

The only study conducted during the first wave of the pandemic and before the pandemic hit was the study of Degeling et al. (2020). The study by Degeling et al. (2020) focused on participants' preferences for a surveillance system that was able to track the spread of infectional diseases. The study revealed that respondents placed the greatest importance on the ability of the system to prevent excess infections and deaths. Respondents cared less about keeping personal autonomy about the data that the system gathered. After the onset of COVID-19 had taken place, the importance of personal autonomy became insignificant. Due to the onset of COVID-19, the spread of an infectious disease was not hypothetical anymore but reality. This changed people’s choice behaviour towards their personal autonomy.

Second wave

In the second wave the study by Muhlbacher et al. (2022) found that individual income decrease was the most important attribute. The attribute with the largest positive impact and the largest negative impact on respondents’ choice behaviour both belong to this attribute. The study was able to distinguish impacts between different levels of the same attribute by estimating separate weights for each category of the variable. Other factors, such as excess mortality and individual risk of infection, were also important. The study conducted by Sicsic et al. (2022) explored the trade-offs between the overload of hospitals and stricter control measures. Postponing non-urgent surgeries to avoid overcrowded ICUs was disliked by respondents. The study reveals that the closure of public spaces,

targeted lockdowns, homeschooling, and medically prescribed self-isolation were seen as better alternatives to prevent overcrowded ICUs. The research by Manipis et al. (2021) investigated public preferences for policies related to infections, deaths, and economic consequences. The article states that respondents strongly preferred to avoid infections and deaths. They also preferred shorter durations for restrictions. Additionally, respondents exhibited a preference for policies that reduced unemployment and government expenditure. Also, the study by Mouter et al. (2021) focuses on the uptake of a tracing app and its attributes. Mouter et al. (2021) states that the prevention of deaths and financial problems of households had a very strong influence on the uptake of a tracing app. It identifies two distinct clusters of respondents with different preferences: Cluster 1 prioritizes reducing societal impacts and favors an app that effectively reduces the negative consequences of COVID-19, while Cluster 2 emphasizes privacy and freedom and expresses concerns about potential privacy infringements. This is interesting as the study from Degeling et al. (2020), that was conducted just before and at the beginning of the pandemic, showed that their respondents in Australia gave low importance to privacy and personal autonomy when asked about the introduction of a tracking system.

To summarize, Degeling et al. (2020) showed that respondents found prevention of excess deaths and infections the most important attributes in the first wave of the pandemic. Mühlbacher et al. (2022) showed that in the second wave, income loss was more important than excess deaths and infections for the respondents in their study. Furthermore, Degeling et al. (2020) showed that at the start of the first wave of the pandemic, personal autonomy was not important. However, Mouter et al. (2021) showed that during the first wave of the pandemic, respondents did find personal autonomy important. This difference could be caused by the fact that Degeling et al. (2020) was conducted during the onset of the COVID-19 pandemic, when little was known about the mortality rate of the COVID-19 virus. During this time people were not willing to take risks. At the moment the study of Mouter et al. (2021) was conducted, people were already familiar with the mortality rates of subsequent virus variants.

Standard deviation estimates

First wave

Degeling et al. (2020) shows that several attribute levels, of attributes in their study, show prove of preference dispersion among the sample, due to the significant standard deviation estimates for these levels.

Second wave

Furthermore, this review already stated that in the study of Muhlbacher et al. (2022), the respondents showed the strongest negative and positive preference for two attibute levels of the attribute about income loss. Besides this, these attribute levels also have large standard deviation estimates, which indicates a large dispersion of preference among individuals. This shows that the preference given by respondents to this attribute is strong but not unanimous. The same can be said for Manipis et al. (2021), where a strong preference was given by the respondents against the most strict COVID-19 measures, but with a large dispersion of preference among the entire sample. Sicsic et al. (2022) showed significant standard deviations for all normally distributed random coefficients for the attributes. In addition, for all attributes, a fraction of the individual part-worth utilities had opposite sign from the mean coefficients. This means that a significant proportion of the sample attributed positive utility to an attribute with a negative mean coefficient and vice versa. The same can be said for Li et al. (2021). Li et al. (2021) showed relatively large dispersion for people's

preference for an increase in unemployment benefit claims. Although, the mean estimate of this attribute was positive, some respondents viewed an increase in these claims as negative.

To summarize, Muhlbacher et al. (2022) and Manipis et al. (2021) show that a strong preference among respondents for an attribute does not mean that there exists no dispersion among the sample for such attribute. Although the preferene in Muhlbacher et al. (2022) for income loss is strong, it is not unanimous. The same can be said for the attribute that relates to the most stringent COVID-19 measures in the study of Manipis et al. (2021). Furthermore, the standard deviation estimates of Sicsic et al. (2022) and Li et al. (2021) show that a part of the respondents can have a negative preference for an attribute while on average the sample's association with this attribute is positive and vice versa.

Marginal rate of substitution

Second wave

Finally, the MRS in this section concentrates mostly on willingness to accept more risk of infection to avoid increase in deaths and decrease in income (Sicsic et al., 2022; Muhlbacher et al., 2022; Li et al., 2021). Sicsic et al. (2022) states that people are willing to accept a larger infection risk to prevent more deaths and a lowered individual income. People were willing to accept more infection risk to avoid deterioration in these two attributes than to avoid a decrease in GDP or the installation of curfews (Li et al., 2021). Respondents would accept a risk of infection of about 37% to prevent a 75% decrease in income. Respondents would accept a lower additional risk of infection to avoid GDP decrease, and curfews (Li et al., 2021).

LC

Class preferences and class probabilities

First wave

In the first wave, Reed et al. (2020) conducted a DCE in the United States to explore preferences regarding the reopening of nonessential businesses. The experiment consisted of attributes for the reopening timeline, percentage of infected citizens, time for economic recovery, and the percentage of households below the poverty threshold. Using latent-class modeling, the analysis revealed a four-class model. The largest class (36%) prioritized minimizing the risk of COVID-19. Approximately 26% belonged to the "waiters" class, who preferred delaying the reopening of nonessential businesses, emphasizing the poverty factor. Another 25% represented the "recovery-supporters" class, prioritizing faster economic recovery over the timing of reopening. The remaining class, called "openers," strongly advocated for immediate reopening, accepting higher COVID-19 risks. In a study conducted by Krauth et al. (2021) in Germany, two latent classes were identified, with class shares of 46% and 54%, respectively. These classes exhibited clear differences in their preferences and characteristics. Class 1 showed strong preferences for a rapid reopening of schools and gastronomy, while rejecting the implementation of a mandatory tracing app and isolation of the elder. On the other hand, class 2 members preferred keeping schools and gastronomy closed for a longer period, with similar preferences against the tracing app and isolation of elderly individuals, albeit to a lesser extent compared to class 1. Class 1 prioritized avoiding long-term unemployment but did not mind overloading ICU capacities. In contrast, class 2 demonstrated reversed preferences, emphasizing the importance of providing sufficient ICU capacities while not considering the unemployment rate. The last study that used LC analysis in the first wave of the pandemic was Chorus et al. (2020). Chorus et al. (2020) opted for a latent class (LC) model with three classes. The fit of this model surpassed that of the logit model, primarily due to its capacity to incorporate heterogeneity and

account for the panel structure of the data. Analyzing the coefficients given to each class, it can be interpreted as follows: Class 1 shows limited sensitivity to changes in the number of deaths, but shows significant negative association with the other policy impacts, particularly tax increases. Class 1 constitutes circa 20% of the sample. Class 2 shows high sensitivity to all policy effects, except for tax increases, to which these people respond similarly to the mean respondent. This class comprises around 29% of the sample. Class 3, representing about 51% of the sample, demonstrates average sensitivity to fatality numbers and working pressure in the healthcare sector, but displays lesser sensitivity to other policy effects.

Second wave

In the second wave, only one study was conducted. Manipis et al. (2021) fit a latent class model with two classes. Class one's class probability is 57% and the class probability for class 2 is 43%. For class 1, the only significant coefficients were those associated with medium restrictions over low restrictions, with tracking bracelets preferred to no tracking, with 500,000 deaths being less acceptable than 10,000 deaths and with either 5% and 3% tax levies compared to a levy of 1%. Four class 2, the coefficients for three months duration of restrictions and tracking using a bracelet or mobile phone were insignificant, all others were significant. For these significant coefficients, the study states that, tighter restrictions were less preferred than lose restrictions. Also, restrictions lasting six months were less preferred than those lasting one month. Class 2 had strong preferences against policies that resulted in a high burden of disease, with the largest decrements being for the highest infection-related death level with additional government spending.

Third wave

Filipe et al. (2022) was the only study conducted in the third wave and identified three classes for their study. People with Class 1 preferences, who represent 28% of the sample, gave the lowest importance to COVID-19 deaths. These people gave higher relative value to education, life restrictions, and risk of poverty than the other two classes. People with Class 2 preferences, representing 41% of the sample, were characterized by a higher valuation of COVID-19 deaths, when compared to people with Class 1 preferences. They also had a lower valuation for household income losses, when compared with people from Class 1 and Class 3. They gave a relatively lower importance to education, life restrictions and poverty, than people with Class 1 preferences. Individuals with Class 3 preferences, representing 31% of the sample, had the strongest valuation towards all levels of the deaths' attribute.

In conclusion, all articles in this section show that the LC model is particularily useful to divide the sample of respondents into groups of respondents with their own distinct characteristics, based on their preferences for the attributes of the DCE. Furthermore, the proportions of these groups within the sample can be estimated. The LC analysis of the studies differ in the amount of classes included and the sizes of these classes.

Class membership

First wave

In the study of Reed et al. (2020) class membership correlated with respondent characteristics, with political affiliation being the strongest association. Democratic and Republican affiliations were more likely to be "risk-minimizers". Independents were positively associated with the "recovery-supporters" and "openers" classes, and negatively associated with "waiters." Higher income was linked to "recovery-supporters," while lower income was associated with "waiters." Salaried individuals were less likely to be risk minimizers or openers but more likely to be waiters. Non-white respondents were strongly associated with the waiters class. Krauth et al. (2021) states that regarding socio-demographic

characteristics, both classes did not significantly differ in terms of gender, age groups, region, and parenthood percentages. However, differences were observed in migration background, chronic diseases, attitudes, risk perception, and behavior. Class 1 members exhibited lower trust in public institutions, media, and science compared to class 2. They were more likely to reject restrictions aimed at containing the pandemic and had a lower risk perception. Additionally, they had more social contacts with individuals outside their own households. Chorus et al. (2020) states that in class 1, older people are over-represented and in class 2, higher educated people are over-represented.

Second wave

Manipis et al. (2021) shows young people and the ones who state that their self-rated health is excellent were more likely to have class 1 preferences, whereas older people and those who state that their self-rated health is poor were more likely to have class 2 preferences.

Third wave

Filipe et al. (2022) States that people with Class 1 preferences were more likely to be male, work remotely, and to be part of a household with children. People with Class 2 preferences were less likely to be of old age.

To summarize, all studies show that the LC model is able to link a variety of socio-demographic characteristics to its classes. Chorus et al. (2020), Manipis et al. (2021) and Filipe et al. (2022) all showed that the covariate age often is a determining factor for class membership. Krauth et al. (2021) states that, although significant, age did not necessarily differ that much between classes in their study. Covariates that showed significant differences in this study were migration background, chronic diseases, attitudes, risk perception and behavior. Chorus et al. (2020) also shows risk perception is a significant covariate that can explain class membership.

Marginal rate of substitution

First wave

The study of Reed et al. (2020) investigated the maximum acceptable risks for each class. The "risk-minimizers" prioritized avoiding COVID-19 risks. The "waiters" found no level of COVID-19 infection risk acceptable for earlier reopening but were willing to tolerate a slight increase to prevent greater poverty. The "recovery-supporters" and "openers" were willing to accept higher COVID-19 infection risks for faster economic recovery. The former would accept a significant increase in COVID-19 infection risk to shorten economic recovery time, while the latter would accept a lesser increase for the same gain.

MNL vs LC results

The literature review includes three articles that estimated a MNL model and a LC model. Krauth et al. (2021) was conducted during the first wave in Germany, Chorus et al. (2020) during the first wave in the Netherlands and Filipe et al. (2022) during the third wave in Portugal. All three studies applied both models on the same dataset. For this reason, the results of these studies possess valuable information on the additional insights that the LC model provides compared to the MNL model.

The LC analysis of Krauth et al. (2021) provides additional insights on how the respondents in the sample can be grouped. The study identifies two classes of roughly the same size. In addition, the LC analysis provides the differences in preferences between these groups. For instance, the study shows that one class has a strong preference for rapid reopening of schools and restaurants while the other class prefers to keep both closed for a longer period. Also, the study showed that one of the classes prioritized avoiding long-term unemployment while the other class did not consider the unemployment rate at all. The MNL model that was conducted in this study did not provide any information on the differences in preferences among respondents in the sample or the way respondents could be grouped. Furthermore, the MNL model showed that a high increase in unemployment rate was disliked by all respondents throughout the sample. On the contrary, the LC analysis showed that this strong dislike towards increasing unemployment rate came from one class of respondents. In short, the LC analysis for this study provided additional insights on differences in preferences between respondent groups, size of these groups and the effect of strong class preferences for an attribute on the sample mean of the attribute.

The MNL model results from the study of Chorus et al. (2020) showed that loss of income was the most important attribute for the entire sample. The LC analysis from this study showed that the sample could be divided into three classes and that only for the respondents in class two, the loss of income attribute was the most important. In class three, the attribute related to deaths due to COVID-19 infection was most important. In class two, the attribute related to physical injuries was most important. In addition, the analysis shows that the largest class in the sample is class three, with a 51 percent chance that respondents belong to that class. Income loss is the most important attribute for the entire sample because the combined preference of class one and two for a loss of income is stronger than the preference of class three for the attribute related to deaths. The article of Chorus et al. (2020) clearly shows that LC analysis is able to provide additional insights on the subgroups that exist in a sample, the sizes of these subgroups and the difference in preferences between the respondents in these subgroups. In addition, the relative coefficient sizes for the preferences in these classes and the different class sizes provide information on why certain attributes are more important than other attributes for an entire sample, even if these attributes are not important for a specific class.

Filipe et al. (2022) also showed that LC analysis is able to provide information on classes with distinct preferences in a sample and their class sizes. In addition, all three studies also provided information on the sociodemographic characteristics that can be linked to the respondents in the different classes. These characteristics provide additional insights on the reason for certain preferences in specific classes. For example, Filipe et al. (2022) showed that respondents in class one were more likeliy to be male, work remotely and to be part of a household with children. Class one in this study included respondents that give a lot off value to education. This is in line with the characteristic that these respondents are likely to have kids. On the contrary, the LC analysis of Chorus et al. (2020) shows that respondents in class one are more likely to be old. Which is not necessarily in line with the fact that the respondents in this class value income loss more than risk of death, as these old respondents are more susceptible to infection. These insights on socio-demographic characteristics are not brought

forward by the MNL model results of these studies. The added value that LC models offer by eliciting subgroups from samples that can also be identified in the population cannot be provided by MNL models.

Labeled and unlabeled DCE

Unfortunately, the studies from Chorus et al. (2020) is the only study that conducted an unlabeled DCE in the literature review. Also, this study only conducted an unlabeled DCE and did not deploy a labeled DCE to compare the results. It is not wise to compare this single unlabeled study to the labeled studies in the literature review because for instance the experimental design, chosen attributes, time of deployment and constitution of the sample are different. Hence, this paragraph will discuss literature on labeled and unlabeled DCEs outside the scope of the societal impacts of COVID-19 measures. This is done to substantiate the results of the labeled and unlabeled DCEs for subresearch question two.

De Bekker-Grob et al. (2020) states that in their study on cancer screening choices the use of labels influenced the decisions of respondents and lowered the attention individuals gave to the attributes. The researchers also state that adding labels may yield more valid results, as these labels exist in the real life decision making situation the DCE thrives to simulate. Unlabeled DCEs are more effective in measuring the relative importance between attributes, as respondents are not distracted by the different labels when making their decision. In addition, the study states unlabeled DCEs are also convenient when individuals have little knowledge of what the labels actually mean. Janssen et al. (2017) states that unlabeled DCEs are more suitable for analyzing attribute nonattendance. Futhermore, the study states that in a labeled DCE it is possible that respondents choose an alternative that is primarily based on the label of that alternative. Respondents that make decisions this way often have non-compensatory preferences. In other words, these individuals will choose an alternative based on the label, regardless of the levels of the attributes. Jin et al. (2017) research the influence of labeling or not labeling alternatives in a mode choice study. The labeled DCE produces slightly higher WTP estimates than the unlabeled DCE in their mode choice study. This suggests that respondents care less about the attributes of an alternative if an alternative is labeled. This is in line with the statement that De Bekker-Grob et al. (2020) made on the use of labels lowering attention of individuals to attributes. Moreover, Doherty et al. (2013) state that based on their study on recreational site choice, a significant portion of respondents made a decision for an alternative that was solely based on the label. In other words, their decisions were not influenced by the attributes. Jansen et al. (2017) made a similar argument. Another interesting finding from Doherty et al. (2013), that studied different walking trails in rural Ireland, was that the group of people that made their decision solely on the base of the label was predominantly from urban areas. Suggesting that the people that have the least understanding about the attributes of the alternative rely more heavily on the labels of the alternative to make their decision. Van Rijnsoever et al. (2015) describe in their study on the public acceptance of energy technologies that vastly dfiferent answers are yielded from a labeled DCE in comparison to an unlabeled DCE. In short, the researchers show that when labels are hidden, respondents choose nuclear energy or biomass as preferred alternatives. When labels are shown of the different possible energy alternatives, the respondents choose renewable and natural gas technologies. Suggesting that other facors besides the attributes play a role in decision making. This can for example be, the public perception of the energy source.

According to the aforementioned literature, it can be expected that the coefficient estimates for the attributes of a labeled DCE are lower than the coefficient estimates of an unlabeled DCE, due to the absence of alternative labels that draw attention of the respondents. Furthermore, unlabeled DCEs are more suitable for determining the most important attributes in an experiment, as the results are not disturbed by the labels. However, the most realistic results will be produced by the labeled DCE, as these DCEs are closest to real life situations.

Sub Research Question 2

This section will discuss the results that were obtained by estimating the coefficients for the attributes of the labeled and unlabeled DCEs with different MNL, ML and LC models. In this way, this section strives to give insights to answer sub research question two. Firstly, this section will discuss the characteristics of the labeled and unlabeled DCE sample. After which, the results from the MNL model, MNL model with dummy coded categorical variables, the ML model and the two three class-LC models with covariates will be presented and discussed.

Data collection

	Unlabeled DCE	**Labeled DCE**	**Percentage of the Dutch adult population (CBS, 2020)**	**Chi-square test (2-sided)**
Sample size	1070	1106		
Gender				
Male	49,1% (523)	48,8% (536)	49,3%	0,77
Female	50,9% (543)	51,2% (562)	50,3%	0,65
Age				
<34 years	29,0% (310)	31,9% (352)	26,7%	0,53
35 – 64 years	48,5% (518)	46,9% (517)	49,5%	0,22
<65 years	22,5% (240)	21,1% (233)	23,8%	
Education Level				
Low	20,1% (214)	18,7% (205)	28,5%	0,0
Medium	40,6% (431)	43,5% (477)	36,8%	0,0
High	39,3% (418)	37,8% (414)	34,6%	
Vaccination Status				
Vaccinated	84,7% (906)	83,6% (925)	82,3%	0,04 0,24

Table 4: Socio-demographic characteristics of the respondents of the unlabeled DCE, labeled DCE and the Dutch adult population.

In total, 2187 people completed the survey. The total sample of 2187 individuals was representative for the Dutch population when it comes to level of education, gender and age. The socio-demographic characteristics of the sample can be found in table three. 11 respondents of the total group of 2187 people were excluded from the study because these respondents completed the survey too fast and gave the same anwer to every subset of questions. The threshold for completing the survey too fast was ‘quicker than one third of the median time for survey completion by the respondents of the entire sample’. The survey included two slightly different DCEs. One DCE was an unlabeled experiment and the other a labeled experiment. Each individual had to answer the unlabeled DCE or the labeled DCE. In total, 1070 respondents filled in the unlabeled DCE and 1106 respondents filled in the labeled DCE.

Multinomial logit models

	Unlabeled DCE (5 attributes)			Labeled DCE (6 attributes)		
Estimates	**Estimate**	**Std. Err.**	**T-statistic**	**Estimate**	**Std. Err.**	**T-statistic**
Death (per 1,000)	-0.143	0.010	-14.484	-0.073	0.020	-3.586
Physical injuries (per 100,000)	-0.130	0.015	-8.897	-0.201	0.024	-8.384
Mental Injuries (per 100,000)	-0.093	0.007	-14.144	-0.117	0.009	-12.334
Income issues (per 100,000)	-0.182	0.009	-20.755	-0.145	0.009	-15.752
Delay surgeries	-0.265	0.018	-14.627	-0.089	0.015	-5.781
Stringency measures	-	-	-	-0.063	0.024	-2.565
Marginal rates of substitution:						
Death/Physical injuries	1.103			0.363		
Death/Mental injuries	1.537			0.625		
Death/Income issues	0.788			0.501		
Death/Delay surgeries	0.542			0.817		
Death/Stringency measures				1.160		
Model outputs:						
Number of observations	6,420			5,530		
Log-likelihood (null)	-4,450.00			-3,833.10		
Log-likelihood (final)	-4,074.20			-3,584.34		
AIC	8,158.41			7,180.68		
BIC	8,192.24			7,220.39		
Rho-squared	0.08			0.06		

Table 5: Estimation results of the MNL model.

Estimates

Table five shows the estimates of the MNL model of both the unlabeled and labeled DCE. For both DCEs, the estimates, standard errors and significance are presented in the table. The attributes for deaths, injuries and income issues were scaled before the MNL model was used to estimate the attribute coefficients. This was done to omit numerical overflow and to ensure correct interpretation. All coefficients are significant at the 1% level, except the estimate for 'stringency measures'. This attribute is significant at the 5% level. All attribute estimates have the expected sign, namely a negative sign.

In general, the results of both DCEs show that respondents do not prefer an increase in deaths, additional physical or mental injuries and an increase in households with income issues. Furthermore, the results from the labeled DCE show that respondents do not prefer increasingly stringent COVID-19 measures. To be more specific, the unlabeled DCE shows that respondents care most about not delaying surgeries, the estimate for this attribute is most negative. The respondents care the least about an increase in mental health problems among society. After the postponement of surgeries, respondents also care a lot about an increase of households with income issues. The estimates from the labeled DCE show that respondents have the largest negative association with an increase in physical injuries, followed by an increase in household income issues. These respondents care the least about the stringency of measures and an increase in the number of deaths caused by COVID-19.

Marginal rate of substitution

The attribute coefficients from the previous table that were estimated with the help of the MNL model can be used as input to calculate the marginal rates of substitution between two coefficients. In this study, the marginal rate of substitution (MRS) is calculated by dividing the attribute for an increase in deaths by one of the attributes from the study. The MRS gives an insight into the willingness to prevent an increase in deaths while accepting an increase in one of the other attributes.

For the unlabeled DCE, the results of this study show that Dutch people are willing to prevent an increase in one COVID-19 death by; accepting an increase of 110 cases of physical injuries, accepting an increase of 153 cases of mental health injuries and accepting an increase of 79 households with income issues. The labeled DCE shows that Dutch citizens are willing to prevent an increase of one COVID-19 death by; accepting an increase of 36 cases of physical injuries; accepting an increase of 62 cases of mental health injuries and accepting an increase of 50 households with income issues. The relative importance of avoiding delays in surgeries versus preventing an increase in COVID-19 deaths has also been calculated. An increase of one step in the delay of surgeries relates to an increase of 1846 COVID-19 deaths in the unlabeled DCE. However, for the labeled DCE, a similar stepwise increase in surgery delays relates to 1219 extra COVID-19 deaths. Furthermore, an increased stringency level of the COVID-19 measures relates to 863 extra COVID-19 deaths in the labeled DCE. These results show that when respondents are confronted with an explicitly mentioned stringency level of COVID-19 measures, respondents attribute a lower value to avoiding deaths and surgery delay and more value to physical and mental health problems.

	Unlabeled DCE (5 attributes)			Labeled DCE (6 attributes)		
Estimates	**Estimate**	**Std. Err.**	**T-statistic**	**Estimate**	**Std. Err.**	**T-statistic**
Death (per 1,000)	-0.145	0.010	-14.191	-0.088	0.024	-3.638
Physical injuries (per 100,000)	-0.129	0.015	-8.633	-0.181	0.031	-5.780
Mental Injuries (per 100,000)	-0.095	0.007	-14.248	-0.106	0.014	-7.618
Income issues (per 100,000)	-0.182	0.009	-19.903	-0.145	0.011	-13.495
Delay surgeries 1	-0.318	0.063	-5.057	-0.138	0.093	-1.476
Delay surgeries 2	-0.470	0.066	-7.156	-0.292	0.095	-3.086
Delay surgeries 3	-0.844	0.069	-12.194	-0.165	0.103	-1.606
Delay surgeries 4	-1.070	0.081	-13.277	-0.428	0.095	-4.487
Stringency measures 1	-	-	-	0.028	0.104	0.266
Stringency measures 2	-	-	-	0.011	0.064	0.167
Stringency measures 3	-	-	-	-0.107	0.098	-1.095
Stringency measures 4	-	-	-	-0.442	0.121	-3.651
Marginal rates of substitution						
Death/Physical injuries	1.124			0.484		
Death/Mental injuries	1.534			0.826		
Death/Income issues	0.797			0.604		
Death/Delay surgeries 1	0.456			0.635		
Death/Delay surgeries 2	0.309			0.300		
Death/Delay surgeries 3	0.172			0.531		
Death/Delay surgeries 4	0.136			0.205		
Death/Stringency measures 1	-			-3.174		
Death/Stringency measures 2	-			-8.264		
Death/Stringency measures 3	-			0.819		
Death/Stringency measures 4	-			0.198		
Model outputs:						
Number of observations	6,420			5,530		
Log-likelihood (null)	-4,450.01			-3,833.10		
Log-likelihood (final)	-4,071.67			-3,576.14		
AIC	8,159.35			7,176.27		
BIC	8,213.49			7,255.69		
Rho-squared	0.09			0.07		

Table 6: Estimation results of the MNL model with dummy coded categorical variables.

Estimates

Table five and table six show the goodness-of-fit statistics of the two MNL models included in this study. The final log-likelihood of the MNL model without categorical variables is smaller for both DCEs than for the MNL model with categorical variables. The same can be said for the adjusted rho squared value of the MNL model without categorical variables. Further, the AIC values suggest that the model with categorical variables is a better fit for the labeled DCE and the model without categorical variables for the unlabeled DCE. However, the BIC values of the model with categorical variables are lower for both DCEs. The BIC values get priority over the other measurement-of-fit statistics. Therefore, the BIC concludes that the MNL model with categorical variables has the best fit. Apparently, a separate coefficient for each attribute level of the two categorical variables fits the true DGP better than one attribute estimate for all levels.

Table five and Table six show almost no differences between the estimates for the non categorical variables. Furthermore, the t values are very similar. The results from the unlabeled DCE show an increasingly negative association of respondents with increased delay of urgent and non-urgent surgeries. For the unlabeled DCE all levels of the 'delayed surgeries'-attribute are significant.

For the labeled DCE, only the categorical variable 'delayed surgeries 2' and 'delayed surgeries 4' are significant. Level two and four correspond to postponing non-urgent surgeries for three months, and postponing non-urgent surgeries for five months and urgent surgeries for one month. It seems that the respondents do care about medium and long-term postponement of non-urgent surgeries and short-term postponement of urgent surgeries but not about short-term postponement of non-urgent surgeries. Furthermore, the labeled DCE shows that respondents only have a strong negative preference against the most stringent COVID-19 measures, but not to lighter versions. The most stringent set of COVID-19 measures includes measures that have major impacts on our daily lives such as a lockdown in which nightclubs, restaurants and cafes have to close.

Mixed Multinomial Logit model

	Unlabeled DCE (5 attributes)			Labeled DCE (6 attributes)		
	Estimate	**Std. Err.**	**T-statistic**	**Estimate**	**Std. Err.**	**T-statistic**
Estimates:						
Death (per 1,000)	-2.187	0.157	-13.962	-3.169	0.474	-6.685
SD Death (per 1,000)	1.418	0.170	8.358	2.067	0.340	6.083
Physical injuries (per 100,000)	-0.152	0.018	-8.356	-0.265	0.031	-8.670
Mental Injuries (per 100,000)	-2.346	0.122	-19.225	-2.638	0.235	-11.206
SD Mental Injuries (per 100,000)	0.831	0.139	5.983	1.673	0.269	6.210
Income issues (per 100,000)	-1.694	0.086	-19.766	-2.008	0.131	-15.292
SD Income issues (per 100,000)	0.938	0.093	10.047	1.219	0.144	8.447
Delay surgeries	-1.746	0.170	-10.262	-2.311	0.276	-8.376
SD Delay surgeries	1.215	0.149	8.145	1.079	0.224	4.812
Stringency measures				-5.229	1.127	-4.639
SD Stringency measures				3.675	0.685	5.368
Marginal rates of substitution:						
Death/Physical injuries	14.436			11.964		
Death/Mental injuries	0.932			1.202		
Death/Income issues	1.291			1.579		
Death/Delay surgeries	1.253			1.371		
Death/Stringency measures				0.606		
Model outputs:						
Number of observations	6,420			5,530		
Log-likelihood (null)	-4450.00			-3833.10		
Log-likelihood (final)	-3942.27			-3411.09		
AIC	7902.53			6844.18		
Rho-squared	0.11			0.11		

Table 7: The estimation results of the ML model.

Estimates

Table seven shows the results of the ML model. All mean and standard deviation estimates are significant for both DCEs. Also, all measurement-of-fit statistics show that the ML model has a better fit with the data than the two MNL models. Furthermore, the 'physical injuries'-attribute estimate entered the model as a fixed parameter instead off as a random parameter. This estimate did not have a significant SD estimate. Therefore, it can be stated that this estimate does not hold any significant heterogeneity among the sample. In other words, the vast majority of the sample holds a small and negative preference towards physical injuries. It is important to note that the estimations values of the ML model are the logvalues, as the distribution used was a negative lognormal distribution.

Latent Class model

	Class 1			Class 2			Class 3		
Class size	**37.4%**			**30.3%**			**32.4%**		
Estimates	**Est.**	**S.E.**	**T-stat.**	**Est.**	**S.E.**	**T-stat.**	**Est.**	**S.E.**	**T-stat.**
Death (per 1,000)	-0.481	0.063	-7.656	-0.123	0.038	-3.273	0.037	0.027	1.372
Physical injuries (per 100,000)	-0.261	0.053	-4.898	-0.339	0.072	-4.715	-0.001	0.032	-0.036
Mental Injuries (per 100,000)	-0.272	0.033	-8.159	-0.140	0.027	-5.182	-0.001	0.017	-0.033
Income issues (per 100,000)	-0.279	0.035	-7.916	-0.646	0.086	-7.525	0.072	0.031	2.350
Delay surgeries	-0.526	0.087	-6.028	-0.398	0.057	-6.945	-0.053	0.043	-1.241
Class membership parameters									
Intercept	-0.787	0.676	-1.163	0 (fixed)	-	-	0.954	0.674	1.414
Gender	-0.097	0.218	-0.447	0 (fixed)	-	-	-0.105	0.224	-0.468
Education level	0.138	0.155	0.886	0 (fixed)	-	-	-0.065	0.163	-0.400
Age	0.403	0.163	2.479	0 (fixed)	-	-	-0.415	0.173	-2.394
Income Issues	0.000	0.123	0.004	0 (fixed)	-	-	-0.064	0.129	-0.499
Chronic Disease	0.219	0.231	0.951	0 (fixed)	-	-	0.453	0.256	1.768
Roommate with Chronic Disease	-0.266	0.292	-0.910	0 (fixed)	-	-	-0.585	0.303	-1.931
Vaccination	0.098	0.291	0.338	0 (fixed)	-	-	0.298	0.291	1.023
Can't live desired life due to COVID-19	0.041	0.103	0.400	0 (fixed)	-	-	-0.047	0.105	-0.444
Social life deteriorated due to COVID-19	0.113	0.115	0.986	0 (fixed)	-	-	0.307	0.114	2.681
Feeling worse due to COVID-19	-0.021	0.106	-0.196	0 (fixed)	-	-	-0.172	0.109	-1.573
COVID-19 would make me very ill	0.124	0.146	0.849	0 (fixed)	-	-	0.523	0.154	3.405
I would be hospitalised due to COVID-19	-0.207	0.201	-1.033	0 (fixed)	-	-	-0.257	0.209	-1.229
I would die of a COVID-19 infection	0.191	0.185	1.034	0 (fixed)	-	-	-0.313	0.183	-1.711
Model outputs									
Number of observations	6,330								
Log-likelihood (null)	-4,387.62								
Log-likelihood (final)	-3,803.11								
AIC	7,991.36								
BIC	7,694.22								
Rho-squared	0.13								
Model profile (only significant variables)									
Age									
Younger than 35 year	18.8%			29.0%			40.8%		
35 - 64 year	50.9%			47.0%			46.0%		
65 years and older	30.1%			23.9%			12.9%		
Social life deteriorated due to COVID-19									
Totally agree	22.9%			28.0%			13.0%		
Agree	15.7%			16.1%			21.5%		
Neutral	18.4%			18.4%			26.6%		
Disagree	37.3%			31.4%			27.4%		
Totally disagree	5.7%			6.1%			11.4%		
COVID-19 would make me very ill									
Extremely high risk	2.2%			3.6%			2.6%		
High risk	8.4%			9.5%			8.6%		
Average risk	41.2%			42.8%			37.4%		
Low risk	41.2%			37.7%			38.4%		
No risk	7.0%			6.4%			12.9%		

Table 8: Results from the LC model of the unlabeled DCE.

Estimates

Table eight shows the estimates for the LC model that was estimated for the unlabeled DCE. The measurement-of-fit statistics show that this model has the best fit of all models for the unlabeled DCE mentioned in this section of the study. Besides the measurement-of-fit statistics the table also shows the different class sizes for the three classes of the model. In percentage, the class sizes for the first, second and third class are 37.4%, 30.3% and 32.4%. Each class in the table includes the different attribute coefficients. These estimated coefficients are all significant at the 5% level and have the expected negative sign for class 1 and 2. Therefore, it can be stated that on average, the respondents that belong to these classes do not prefer increases in any of the attributes included in this model. For the third class, only the coefficient that relates to an increase in households with income issues is significant at the 5% level. Moreover, the respondents in this class prefer an increase in the number of households with income issues. The size of this coefficient is 0.072, which is relatively much smaller than the same coefficient for class one and two, with a value of -0.279 and -0.646. Therefore, it can be stated that although the positive coefficient for class three respondents is significant it is of less importance to these respondents than to the respondents of the other two classes.

Class preferences and probabilities

The size of the attribute coefficients holds information that helps to define the different classes. The respondents in class one (37.4%) care most about not postponing surgeries and after that avoiding an increasing number of deaths. These respondents attach approximately the same value to physical injuries, mental injuries and income issues. Class two (30.3%) is defined by respondents who attribute the highest value to the prevention of an increase in households with income issues. These respondents also care a lot about preventing the postponement of surgeries and avoiding an increase in physical injuries. The least importance in this class is given to an increase in deaths and mental health problems. The respondents of class three (32.4%) do not have a strong preference for an increase in any of the mentioned health problems. The respondents in this class show a positive association with an increase in households with income issues. The estimates from the different classes clearly show that the respondents in class one care most about avoiding the delay of surgeries. The respondents of class two care more about preventing an increase in physical injuries and income issues.

Class membership

The covariate estimates for age in class one and three are statistically significant. Also, the covariate estimates of class three of the statement ‘social life deteriorated due to COVID-19’ and the statement ‘COVID-19 would make me very ill’ were significant. These findings can be used to create profiles of class characteristics. Based on these findings, it can be stated that class one is characterized by middle-aged and older respondents. On the other hand, class three is characterized by middle-aged and young respondents. The respondents in class three are neutral towards the statement that their social life degraded due to COVID-19. The same respondents believe that there exists average-to-low risk that COVID-19 would make them very ill.

Class size	46.1%			38.6%			15.4%		
Estimates	**Est.**	**S.E.**	**T-stat.**	**Est.**	**S.E.**	**T-stat.**	**Est.**	**S.E.**	**T-stat.**
Death (per 1,000)	-0.082	0.038	-2.162	-0.212	0.097	-2.198	0.374	0.165	2.270
Physical injuries (per 100,000)	-0.095	0.046	-2.063	-0.288	0.092	-3.131	-1.713	0.543	-3.157
Mental Injuries (per 100,000)	-0.016	0.021	-0.729	-0.308	0.044	-7.051	-0.457	0.214	-2.137
Income issues (per 100,000)	-0.072	0.020	-3.571	-0.489	0.069	-7.082	0.229	0.237	0.965
Delay surgeries	-0.168	0.030	-5.542	-0.244	0.053	-4.566	2.048	0.571	3.585
Stringency measures	-0.019	0.055	-0.347	-0.530	0.134	-3.960	2.675	0.794	3.370
Class membership parameters									
Intercept	2.498	0.651	3.835	0 (fixed)	-	-	0.40201	0.7267	0.553
Gender	-0.083	0.200	-0.416	0 (fixed)	-	-	-0.253	0.227	-1.116
Education level	-0.101	0.151	-0.667	0 (fixed)	-	-	0.016	0.164	0.096
Age	-0.625	0.164	-3.802	0 (fixed)	-	-	-0.106	0.173	-0.612
Income Issues	-0.136	0.124	-1.092	0 (fixed)	-	-	-0.143	0.128	-1.118
Chronic Disease	-0.058	0.231	-0.253	0 (fixed)	-	-	0.594	0.266	2.233
Roommate with Chronic Disease	-0.481	0.276	-1.741	0 (fixed)	-	-	-0.278	0.341	-0.818
Vaccination	0.019	0.238	0.079	0 (fixed)	-	-	-0.214	0.292	-0.730
Can't live desired life due to COVID-19	0.121	0.106	1.144	0 (fixed)	-	-	0.165	0.116	1.421
Social life deteriorated due to COVID-19	0.046	0.113	0.412	0 (fixed)	-	-	-0.154	0.132	-1.163
Feeling worse due to COVID-19	-0.100	0.098	-1.012	0 (fixed)	-	-	-0.005	0.120	-0.039
COVID-19 would make me very ill	0.068	0.137	0.492	0 (fixed)	-	-	-0.027	0.164	-0.167
I would be hospitalised due to COVID-19	-0.216	0.185	-1.168	0 (fixed)	-	-	-0.283	0.213	-1.325
I would die of a COVID-19 infection	-0.189	0.172	-1.098	0 (fixed)	-	-	-0.051	0.189	-0.271
Model outputs									
Number of observations	5,530								
Log-likelihood (null)	-3,767.25								
Log-likelihood (final)	-3,303.39								
AIC	7,011.01								
BIC	6,700.78								
Rho-squared	0.12								
Model profile (only significant vars.)									
Age									
Younger than 35 year	40.9%			23.3%			26.1%		
35 - 64 year	42.9%			48.6%			51.2%		
65 years and older	16.0%			27.7%			22.1%		
Chronic Disease									
Yes	34.4%			29.0%			22.8%		
No	63.6%			67.5%			71.0%		

Table 9: Results from the LC model for the labeled DCE.

Estimates

The estimated coefficients for the LC model of the labeled DCE can be found in table nine. The measurements-of-fit of this model are better than for the other models. The class sizes for the three classes are 46.1% for class one, 38.6% for class two and 15.4% for class three. For class one, only the attribute coefficients for an increase in mental health problems and the stringency of COVID-19 measures are insignificant. All other parameters are significant and have the expected negative sign. For the second class, all estimated attribute coefficients are significant and have the expected negative sign. The only attribute coefficient that is insignificant for class three, is the coefficient for an increase in households with income issues. However, in class three, the estimates for the attributes have

positive or negative signs. The attribute estimates that relate to physical and mental health problems have the expected negative signs, but the estimates related to an increase in deaths, delayed surgeries and stringency of COVID-19 measures are positive.

Class preferences and probabilities

Each class can be defined by the attribute estimates. Class one is defined by respondents that dislike postponing surgeries the most and care about an increase in households with income issues the least. These respondents care equally about avoiding an increase in deaths and physical injuries. Class two is defined by respondents that care the most about not to stringent COVID-19 measures and care the least about an increase in the number of deaths. These respondents care somewhat equally about an increase in physical and mental injuries and postponing surgeries. People in class three regard more stringent measures as positive. For these people stringent measures are the most important. After that, these people are most in favour of postponing surgeries. Furthermore, these people have a negative association with increasing number of physical injuries. These respondents care the least about increases in deaths. When comparing classes, the estimates of class one show low sensitivity to increases in attributes compared to the other two classes. The respondents in class three care more about avoiding increases in mental and physical injuries than the respondents in class two. The people in class one and two have opposing preferences to the people in class three for increases in deaths, increases in households with income issues, postponement of surgeries and stringency of COVID measures.

Class membership

With regards to covariate estimates, the table shows that the covariate estimate related to age in class one is significant. Also the covariate estimate related to chronic disease in class three is significant. These findings show that class one is related to respondents who are of middle-age or younger. Class three is related to respondents without a chronic disease.

Sub research question 3

This section contains the results of the expert interviews that are used to answer the last sub research question. This question focuses on the opinion of experts on the (dis)-advantages of using choice models to analyze DCEs that measure the societal impacts of COVID-19 measures in different phases of the pandemic. Three interviews have been conducted with the following experts; a head researcher of Populytics, a choice modeling researcher connected to the TU Delft and a senior advisor from the Corona behaviour unit of the RIVM. The interview protocol for all interviews can be found in Appendix A.1. Each interview is summarized based on the topics that were discussed. The results in this section discuss the views of all experts on the different topics. The main topics that are discussed are the advantages and disadvantages of DCEs and the advantages and disadvantages of the different modeling techniques. The different advantages and disadvantages are subdivided into methodological advantages, practical advantages and advantages related to implementation. Methodological advantages are defined as advantages that stem from the model or DCE characteristics. The practical advantages are defined as advantages that come from different uses of the results produced by the models. The implementation advantages are defined as the advantages of the results in informing policy decision making and/or informing society.

Advantages of DCEs

The researcher of Populytics and the advisor of the RIVM both state that a methodological advantage of deploying a DCE is that it is able to elicit the respondent's sensitivity to changes in different attributes at the same time, as every choice task includes multiple attributes. The advisor to the RIVM states that this is a substantial advantage of DCEs over other survey forms, such as a questionnaire, in which usually only one attribute can be tested per question. Furthermore, both praise the DCE data's suitability to draw generalized conclusions on the differences and similarities in attribute preferences between individuals and among one individual's choices. Drawing these statistically sound conclusions is possible because a multitude of different respondents answer a similar subsequent set of choice tasks. Lastly, the advisor of the RIVM states that another methodological advantage of deploying a DCE is that it is a relatively easy way to ask how respondents judge complex policy decision making situations. More specific, the ability of DCEs to extract the relative importance of key aspects of such a complex situation is of high value. This relative importance is brought forward through the decisions that respondents make when confronted with a trade-off between different attributes of a DCE in a choice task. This information cannot be elicited through other survey forms, the advisor states.

According to the researcher from Populytics and the choice modeler from the TU Delft, the results produced by DCEs have several practical advantages. The DCE results can be used to produce valuable metrics. For instance, coefficients can be estimated by using choice models that show the importance of attributes and the type of association (negative or positive) that respondents have with an attribute. In addition, the ratio between the estimated coefficients of the attributes can be used to calculate the so called marginal rate of substitution, which shows the relative importance between the coefficients of attributes. On the subject of labeled versus unlabeled DCEs, the choice modeler from the TU DELFT states that a practical advantage of a labeled DCE is that it gathers information on the trade-offs respondents make between certain measures and their societal impacts. The choice modeler says: "A labeled DCE is most useful if policy makers are certain about the measures that need to be implemented and the policy makers want to analyze the relative impact of each measure on the different societal attributes. On the other hand, if it is unknown what COVID-19 measures will be implemented, it is not possible to include the measures in the experiment. In this case, it is still possible to deploy an unlabeled DCE in which the relative importance of different societal impacts can be measured, regardless of the implemented measures."

The researcher of Populytics and the advisor of the RIVM both emphasize that a major advantage for the implementation of DCE results is that the estimated attribute coefficients help policy makers to prioritize the attributes that are part off the different policy alternatives in each choice task. This helps policy makers to understand what aspects of the policy can be adjusted without losing public support for the policy. Lastly, the advisor from the RIVM states that one of the most important advantages of conducting DCEs is that it creates understanding among respondents for the complexity that policy makers face in making decisions on what policy to implement. The reason it is so important to ask respondents to make trade-offs is because the choices that policy makers have to make are never absolute choices. The choices policy makers have to make always ask for a compromise between different aspects of a policy. In other words, every policy option has advantages and disadvantages. For instance, the advisor says; "A DCE is able to show through its alternatives that when COVID-19 measures become stricter, this will result in less deaths and infections due to COVID-19. But, it will also mean less freedom."

Disadvantages of DCEs

According to the choice modeler from the TU DELFT, the practical disadvantage of unlabeled DCEs is that they do not measure the importance of the different societal impacts relative to the implemented COVID-19 measures. Because of this, it is not known how the preferences of respondents for the different societal impacts change if certain measures are implemented.

Both the Populytics researcher and the RIVM advisor state that there exist some disadvantages with regards to the implementation of DCE results, because clients and citizens find it hard to understand the results. It is difficult for people to understand the meaning of the different metrics such as the coefficient estimates and marginal rates of substitution that are often expressed in tables that mention their mean, standard deviation and other statistical quantitites. The researcher of Populytics says; "The company tries to convey these estimates by using bar charts, as this gives a visual presentation of the coefficient's sizes without the relevant but confusing statistical quantitites." The advisor of the RIVM explains that the best way to transfer the information from the estimated coefficients to policy makers is by putting the coefficients into context. The advisor states; "Policy makers love it if you contextualize the coefficients by saying, to avoid 100 deaths respondents are prepared to accept that 300.000 children will not be able to go to school for a given time period." This is the best way of formulating what the relative importance of the different coefficients for the attributes are.

The head researcher of Populytics states that another concept that is difficult to understand for clients is that different categorical coefficients are measured relative to a reference level. A way to improve the client's understanding about the relative nature of categorical coefficients is by expressing these coefficients in percentages. This increases understanding as respondents are more familiar with percentages as a way to express the relative size of numbers.

Advantages of ML models

The choice modeler from the TU DELFT explains that the ML model is an extension of the MNL model that includes random coefficients. A major methodological advantage of these random coefficients is that these coefficients show the dispersion of preferences among individuals of a sample, by fitting a probability distribution to the preferences of the individuals in a sample. This distribution represents the likelihood for individuals in the sample to have a certain preference based on the choices these respondents made in the choice tasks of the DCE. When comparing the ML model with the LC model, a methodological advantage of the ML model is that the model almost always reaches a global optimum for the likelihood function. If the likelihood function for the LC model is maximized, it can get stuck in local optima that are smaller than the global maximum.

A practical advantage of fitting a distribution to the coefficients of a ML model, is that this distribution can be used to calculate the percentage of people from a sample who have a negative preference and the percentage of people that have a positive preference for an attribute. The choice modeler states: "This is the main advantage of ML models over MNL models, as these percentages cannot be derived with an MNL model. This is because, in an MNL model the entire population is described in terms off one parameter per attribute, this parameter is the mean estimate. The ML model includes two parameters per attribute, which are the mean and standard deviation estimates of the assumed distribution."

Disdvantages of ML models

A methodological disadvantage of ML models is that the estimation process is more time consuming than for MNL models. The reason for this is that a ML model draws a customized set of attribute coefficients for every choice task through simulation to maximize the likelihood of the chosen alternative. For the MNL model one set of attribute coefficients is estimated to maximize the likelihood of all chosen alternatives over all choice tasks at the same time. The choice modeler says: "Nowadays, the computing time has been reduced substantially compared to two decades ago. However, if it takes seconds to estimate one MNL model and around ten minutes to estimate one ML model, estimating a multitude of models per day takes much more time." Secondly, the choice modeler says: "Because simulated distributions are used, there exists no guarantee that the results are the same all the time. In theory, if an infinite number of simulations are used, the results will always converge to the same value. However, due to time restrictions and limited computer processor capacity, oftentimes only 500 to 1000 draws are used." Another methodological disadvantage of using ML models is that the fitted distribution for ML models is an assumed distribution. This means that a distribution is chosen based on the size of the sample and the underlying relationships between the attribute coefficients in the data. For example, a normal distribution can only be used when the sample size is sufficient. If the assumed distribution is incorrect this can produce biased results. For example, if a lognormal distribution is chosen to represent a sample's collective preference for an attribute, the mean of that attribute coefficient is either forced to be positive or negative. The reason for this is that a lognormal distribution can only contain positive or negative values and not both. If the sample is in fact normally distributed, this will produce incorrect results. The same can be said for fitting a normal distribution to a random coefficient that all respondents have a negative association with. Eventually, the goal of the modeler is to find the model that fits the data best and is most behaviourally sound. Furthermore, when comparing LC and ML models, the methodological disadvantage of ML models over LC models is that ML models do not provide the benefits of describing the population in terms of subgroups. The classes that are estimated with LC models have their own characteristics and serve as valuable input to construct profiles of distinct subgroups in a population. This can help policy makers to customize their policies accordingly.

A practical disadvantage in the use of distributed coefficients to calculate metrics such as the MRS, is the risk of dividing through zero when a normally distributed coefficient is used as denominator. Dividing by zero results in an infinite MRS. A possible solution for this problem is to keep the coefficient in the denominator fixed. This means the coefficient in the denominator only has a mean estimate and no standard deviation estimate. In other words, the coefficient is not distributed but is expressed as a number. This strategy is commonly used to calculate the WTP, which is an example of the MRS that uses a cost parameter as denominator. This solution is generally accepted for a cost parameter as it is assumed that the cost parameter usually has a small variance. The reasoning behind this is that almost everybody has a negative preference for higher cost. In general, it is therefore important to choose a coefficient with a small variance to present the denominator of the MRS ratio, so that this coefficient can be fixed.

Advantages of LC models

All interviewees agree that the most important methodological advantage of the LC model is its ability to estimate different classes with distinct preferences. The Populytics researcher explains that these classes can be used to represent subgroups in the population. The researcher states that an additional advantage is that specific socio-demographic characteristics of respondents in the sample can be linked to these classes. Aforementioned, makes this model more suitable than the other models for the production of tangible results.

The Populytics researcher also mentions a specific practical advantage of the LC model. The researcher states that the LC model can be used check the significance of covariates such as gender, age and educational level. The researcher from Populytics says; "clients often ask for the significance of these covariates in explaining the differences in results obtained by the LC model. Contrary to believe, these covariates are often not significant and do not explain the distinction in classes." The RIVM advisor also provides a specific context in which dividing the population into subgroups is valuable. Currently, in the endemic phase of the COVID pandemic there exists more time to reflect for citizens on what happened during the pandemic. This provides policy makers with the opportunity to assess what should be done if a new pandemic occurs. Different groups in society have their own view on this. A practical advantage of LC models is that the RIVM can use these models to assess these different views among groups in society.

Disdvantages of LC models

One of the biggest methodological disadvantages of the LC model, according to the choice modeler from the TU DELFT, is that the estimated likelihood function is not well-behaved, as it often has multiple local optima. The log-likelihood function of the MNL model is well behaved, as the function always has one maximum. The results that are produced by the multinomial logit model are always the only possible results. The results produced by the LC model could be based on a local optimum that is not the global maximum, which delivers suboptimal results. The choice modeler says: "The way to resolve this problem for the LC model is by trying different starting values until the best model fit is reached. This is a cumbersome and iterative process, and it is plausible you may never find the best results possible." The advisor of the RIVM and the researcher of Populytics mention that a methodological disadvantage of the latent class model is that the class sizes and coefficients that are estimated with the latent class models change when new covariates are added to the model. Also, the constitution of the respondents that belong to a class change. In short, the results produced by the LC model are sensitive to changes. The Populytics researcher also adds that another methodological disadvantage is that the latent variable(s) that usually explain the difference in classes are difficult to identify. The latent variable(s), also called covariates, are unknown and can only be identified by testing all possible combinations of these variables.

The RIVM advisor states that it is important for advising policy makers that the results from a model are reliable and robust. The reliability of results stands for the consistency of the results generated by a model during different runs of the model under similar conditions. Robustness of results is defined by how consistent a model performs under varying conditions. The LC model does not perform well when changes are made to the covariates included in the model. For this reason, it can be said that the model is not robust. The advisor to the RIVM states that a solution for surpassing the issue of the LC model's sensitivity to changes in covariates is to fix the covariates before running the models. The researcher of Populytics gives a similar argument in a different context and also provides a solution. The researcher states that as a company it is important to send an unambiguous message to the client,

which is difficult when the model is not robust. The best way to approach this problem is to construct a clear scope and run multiple models before presenting the results to the client.

Furthermore, the researcher of Populytics states that a practical disadvantage of the LC model is that the class membership profiles, that are based on the significant covariates per class, sometimes do not differ a lot in terms of their constitution. For example, the researcher says; "age could be a significant covariate for all three classes of a LC model. If the class membership profiles of these classes show that the probability of young people belonging to a certain class is equally probable in all classes, this does not provide valuable insights."

Key insights

To summarize, the experts state there exist several general advantages of DCEs that also apply to weighing societal impacts of COVID-19 policy. In short, the experts praise the DCE's ability to analyze multiple attributes at the same time and the DCE data's suitability to draw statistically sound conclusions from. It is also emphasized that the ability of DCEs to deduce complex situations into a set of key aspects is of high value, as are the practical insights obtained from for example, the MRS. Lastly, deploying DCEs creates understanding among respondents for the complexity of policy decision making on the one hand. On the other hand, DCEs aid policy makers in implementing policies with high public support. Nonetheless, the experts also indicate that it is difficult for policy makers and citizens to interpret DCE results. It is best to visualize these results with charts or to put the results into context.

In addition, the use of labeled and unlabeled DCEs have specific value in weighing the societal impacts of COVID-19 measures. One of the experts states that a labeled DCE is most valuable if there exists certainty on the COVID-19 measures to be implemented. It provides insights on the relative importance between COVID-19 measures and societal impacts. An unlabeled DCE can still provide valuable information on the relative importance of different societal impacts. This is useful at the start of a pandemic when there is uncertainty on what COVID-19 measures to implement. However, a disadvantage of unlabeled DCEs is that it is not known how the relative importance of the impacts change if COVID-19 measures are implemented.

Concerning the different choice modeling techniques, the experts praise the robust results from the ML and MNL models, because these results help to give confident and unambiguous advice to policy makers, which is needed to create clear policies for the public. Also, the experts state that both the ML and LC model are useful for visualizing the different preferences among individuals. The distribution that is fitted to the coefficients of the ML model visualize what percentage of a sample like or dislike an attribute. This helps to predict how effective a policy might be when implemented. If the mean coefficient estimate for implementing masks is slightly negative, but the fitted distribution shows that a large part of the sample has a positive association with this measure this might indicate some support for the measure that was not clear from the MNL results. The LC model is able to produce the most tangible results for policy makers by estimating classes with distinct preferences that represent subgroups in the population. In addition, these classes can be linked to specific socio-demographic characteristics. One of the experts emphasizes that, at this moment in the endemic phase, the LC model is useful for evaluating the different views of subgroups in society on the impact of COVID-19 measures during the pandemic. The insights on this can be used to produce customized policies that fit multiple subgroups in a future pandemic.

However, the models also have several disadvantages that need to be addressed. For instance, the use of ML models is time consuming due to the simulation of distributions, which is a disadvantage at the peak of a pandemic when urgent decisions have to be made. Also, the fitted distributions for the model are based on assumptions that are made about the distributions by the choice modeler. Incorrect assumptions cause biased results. In addition, estimation results of the model might deviate slightly because of the number of draws that are used to simulate the distributions. Equal results for every run of the model can only be achieved by making an infinite number of draws. A practical disadvantage of the ML model is that metrics such as the MRS are difficult to calculate when using distributed coefficients. A solution for this is to fix the attribute in the denominator of the MRS. However, the variance of this attribute must be small for it to be assumed fixed. The LC model is sensitive to changes in intial values and included covariates. Therefore, the model is not reliable and robust. A solution for overcoming the model's sensitivity to changes in covariates is by either fixing the covariates or checking the LC results with the results of other models. The model's sensitivity to changing initial values can be overcome by trying multiple sets of initial values and keeping the values for which the model fit statistics are the best. Lastly, one of the experts mentioned that there no particular model is better than the other models in analyzing labeled or unlabeled DCEs.

Conclusion

This study strives to find out where the added value lies of Latent Class and Mixed logit models over MNL models, and labeled and unlabeled Discrete Choice Experiments, in informing future pandemic policy in different phases of a pandemic. The study does this by answering the following main research question: *What are the (dis)-advantages of using ML and LC models over MNL models to analyze (un)-labeled DCEs that weigh societal impacts of COVID-19 policy during the pandemic and endemic?* The study answers the main research question by analyzing the two knowledge gaps that are included in the main research question. The two knowledge gaps both revolve around DCEs that weigh the societal impacts of COVID-19 policy in the pandemic and endemic and are formulated as follows: The first knowledge gap is: The advantages and disadvantages of ML and LC models. The second knowledge gap is: The advantages and disadvantages of labeled and unlabeled DCEs. The analysis consists of discussing the most important advantages and disadvantages from the results section and explaining the implication of these results for policy makers.

Mixed Logit and Latent Class model

Advantages

The results from this study show that the main advantage of the ML model is, the ability to elicit preference heterogeneity among individuals for the societal impact attributes, by fitting a distribution to the mean coefficient of each attribute. Fitting a distribution to the mean coefficient of an attribute provides several advantages that can create valuable insights for policy makers. The first advantage is the possibility to test if there exists preference heterogeneity for a societal impact attribute. This is possible because the ML model estimates a mean and standard deviation for the distribution of the coefficient of a societal attribute. If the standard deviation estimate is insignificant, this means there exists no dispersion of preferences. For example, the standard deviation estimates produced by the ML model for this study's labeled and unlabeled DCE showed that the standard deviation estimate for the attribute on physical injuries was insignificant. This means that all individuals in the sample equally dislike an increase in people with physical complaints due to COVID-19 infection. Firstly, this is a valuable insight, as it implies that policy makers can implement a pandemic policy that is the same for the entire population. Secondly, policy makers do not have to conduct further research into the origins of preference heterogeneity for this attribute. On the other hand, if the standard deviation estimate for the distributed preference of an attribute is significant, this means that there exists preference heterogeneity among individuals in the sample. For example, The ML model estimates a significant standard deviation for the 'mental health'-attribute of this study's labeled DCE. This is a valuable insight that tells policy makers that it might be worthwile to conduct further research into the preference heterogeneity among people in this sample and the reasoning behind their preference. A second advantage of estimating a mean and standard deviation for the preference of a societal attribute, is the possibility to check how reliable the mean coefficient of a societal attribute is in representing the majority of the sample. If the standard deviation is small, the preference of the majority of the sample is centered around the mean. If the standard deviation is large, the preference of the majority of the sample is much more dispersed. In this last case, the mean coefficient is much less reliable in representing the majority of the sample. For example, Mühlbacher et al. (2022) showed that a high degree in income loss has a large standard deviation for its preference distribution. This insight tells policy makers that implementing pandemic policy based on the mean coefficient of this societal attribute might not be effective as the mean coefficient is not a good representation of the majority of the population. It would result in a minor increase in public support and adherence for the policy. The ML model also has two distinct advantages that do not relate to their application but to

the model´s working mechanism. Firstly, the studies that used a ML model and a MNL model, such as Degeling et al. (2020) and the labeled and unlabeled DCEs of this study both show that the model fit of the ML model is always better than the fit of the MNL model when the assumed distribution that is chosen is correct, the choice modeler from the TU Delft also confirms this. This means that the ML model describes the true data generating process better than the MNL model. This is valuable for policy makers, as insights on the preferences of people in society produced by the ML model are more in line with reality than the insights produced by the MNL model. Another advantage of the ML model over the LC model, according to the choice modeler from the TU Delft, is that the ML model almost always reaches a maximum that is the global optimum for the likelihood function. This ensures policy makers that the model produces valid and reliable results for every run of the model. This is important for policy makers, says the senior advisor of the RIVM, as the policy makers want to convey clear and unambiguous pandemic policy to citizens of a country, to avoid confusion and increase effectiveness of the policies.

The main advantage of the LC model is that it is able to divide the sample of respondents into different classes with their own preferences and sociodemographic characteristics. The information on different classes, their preferences and underlying respondent characteristics can be used by policy makers to develop customized pandemic policy that is targeted towards different subgroups in the population. An additional advantage of the LC Model is that the class preferences produced by the LC model help to substantiate the results that are obtained in the MNL model. For example, Krauth et al. (2021), Chorus et al. (2020) and Filipe et al. (2022) are three studies from the review that show that the sensitivity to a societal attribute from one group of respondents can substantially influence the importance of that attribute for the entire sample. For example, in the study of Filipe et al. (2021), the results produced by the MNL model showed that a loss in income was the most important attribute overall. While, the LC analysis showed that only one out of three classes thought that this was the most important attribute. However, the sensitivity of the respondents in this class to the income attribute was so high that it greatly affected the importance of this attribute for the entire sample. First of all, this is a valuable insight because this shows that the mean coefficient of the entire sample is not a good representation of the majority of the sample. Secondly, these insights can help to make the pandemic policy more effective. For example, with regards to the example of Filipe et al. (2021), the government can subsidize the loss of income caused by the implemented COVID-19 measures to the identified subgroup for which a loss in income is substantially more important than for the other groups. In this way, the government can ensure that public support and adherence of this group for the COVID-19 measures does not decline. An additional advantage is that the subsidy does not have to be distributed to all citizens of a country. Another advantage of the LC model is that the model can be used to construct class membership profiles. A class membership profile calculates how likely it is that respondents in a specific class have certain socio-demographic characteristics. For example, Chorus et al. (2020) shows that respondents in one of their classes value avoiding tax increases more than an increase in the number of deaths. The membership profile of this class shows that people in this class are more likely to be old. In a second class of this study, people care less about tax increases. The people in this class are likely to be highly educated. This is helpful information for policy makers to create customized pandemic policy that alleviates tax increases for older people and increases tax for highly educated people. By customizing pandemic policy in this way, instead of implementing a flat tax rate for the entire population, public support and adherence to the measures can be optimized. Which in turn makes the implemented pandemic policy more effective in mitigating morbidity and mortality rate and makes the policy more cost effective.

Disadvantages

The ML model also has several disadvantages. Firstly, the Choice modeler from the TU Delft states that the estimation process for the ML model is more time consuming than for the MNL model. The Senior advisor from the RIVM states that this is a major disadvantage over the MNL model. The advisor states that during the peaks of the waves in a pandemic, quick results have to be produced to ensure effective pandemic policy that can be implemented and adapted swiftly. A solution for making the estimation process less time consuming according to the choice modeler from the TU Delft is by limiting the number of draws that are made from the simulated distribution for each run of the model. A consecutive disadvantage of using a limited number of draws is that multiple runs of the ML model on the same dataset can produce slightly different results. The results can only be exactly the same if an infinite number of draws is used for every run, according to the choice modeler from the TU Delft. This is a disadvantage to policy makers as policy makers do not only want quick results but also reliable results. Another disadvantage, according to the choice modeler from the TU Delft is that, if the assumed distribution of preferences for the sample is wrong, this will cause biased results. Luckily, this problem is easy to solve. As an incorrectly assumed distribution will result in a ML model fit that is worse than the model fit of the MNL model. Another disadvantage of the ML model over the LC model that is brought forward by this study's DCEs and the literature review, is that the ML model is not able to elicit what groups or individuals are linked to different parts of the distributed preference. This is a disadvantage to policy makers as it is difficult to implement pandemic policy that addresses differences in preferences without knowing who the target group is for the pandemic policy.

A disadvantage of the LC model according to the head researcher of Populytics and the Senior advisor of the RIVM is that the model lacks robustness and resiliency. A model lacks robustness if the results produced by the model severely change when minor changes to the input of the model are made. For example, both experts state that the lack of robustness is shown by the results obtained in this study's DCEs. The DCEs show that a minor change in the covariates that are included in the LC model yields substantially different attribute coefficient estimates for the different classes. In addition, the choice modeler from the TU Delft stated that it is sometimes necessary to change the initial values for the LC model because the likelihood function of the model easily gets stuck in local optima. This shows the lack of resiliency of the model, as the model is not able to get out of a local optimum by itself. Resiliency, in this case, refers to the ability of a model to recognize that it is stuck in a local optimum and to take the necessary steps to find the global optimum. The likelihood function for the MNL and ML model do not have this disadvantage as these models only have one optimum, which is the global maximum, for this reason these models cannot get stuck. The reason robustness and resiliency of a model are important is because policy makers want to implement unambiguous pandemic policy to avoid confusion among citizens of a country, according to the senior advisor of the RIVM. For this reason, the input for the creation of pandemic policy from different models should also be unambiguous. A way to cope with the LC model's sensitivity to changes in covariates, according to the Senior advisor of the RIVM, is by predetermining what covariates will be added to the LC model. This yields more reliable results than changing the constitution of covariates for the LC model every time the model is estimated again. The head researcher of Populytics agrees it is best to have a clear scope on what covariates you want to study with the latent class model before running the model. In addition, he mentions it is good to compare the results from the LC model with the results from other models. The choice modeler from the TU Delft explained that the model's sensitivity to changes in initial values can be overcome by trying multiple sets of initial values and keeping the values for which the likelihood is highest. This study's DCEs used an algorithm that searched for the correct starting values and that was able to get the model out of local optima. Using such a 'searchstartvalue'-algorithm improved the resiliency of the model.

Labeled and unlabeled DCEs

Advantages and disadvantages

First of all, the literature review, this study's unlabeled and labeled DCE analysis and the expert interviews have not provided any evidence that would suggest that one of the models in this study is better at analyzing labeled or unlabeled DCEs than the other models. In fact, the choice modeler from the TU Delft confirms that this is true. With regards to the advantages and disadvantages of labeled and unlabeled DCEs, it is important to know if there exist any measurable differences between the results obtained by these DCEs in the context of this study. The results from the MNL and ML model that were used to analyze this study's DCEs showed that when respondents are confronted with a labeled DCE that includes an attribute that corresponds to the stringency level of COVID-19 measures, respondents attribute a lower value to avoiding deaths and surgery delay and more value to physical and mental health problems than in an unlabeled DCE. This example shows that a labeled DCE indeed does obtain different results than an unlabeled DCE.

Knowing this, the question arises when it is advantageous to use a labeled DCE and when an unlabeled DCE. The main advantage of a labeled DCE in the context of measuring the societal impacts of COVID measures, is that it is able to measure the effect the implementation of the measures has on the importance of the different societal impacts. The TU Delft choice modeler and Mouter et al. (2021) substantiate these findings by stating that a labeled DCE is most useful if policy makers are certain about the measures that will be implemented and these policy makers want to analyze the impact of these measures on societal attributes. The TU Delft choice modeler and Chorus et al. (2020) state that the main advantage of an unlabeled DCE is that it can still measure the relative importance of different societal impacts when it is unknown what COVID-19 measures will be implemented. This study's unlabeled DCE confirms that it is possible to measure the relative importance of different societal impacts without explicitly mentionining implemented COVID-19 measures.

Added value of models for the future pandemic policy decision making process

DCEs in different waves and phases of the pandemic

In the aforementioned parts of the conclusion, the advantages and disadvantages of the labeled and unlabeled DCEs in weighing societal impacts of COVID-19 policy and the implications of these insights for policy makers were thoroughly discussed. The analysis of these insights and their implications provides information on the timing of the use of these DCEs and models and their results in a future pandemic. For instance, it was discussed that unlabeled DCEs are able to measure the relative importance of different societal impacts when it is unknown what COVID-19 measures will be implemented. At the start of the COVID-19 pandemic it was unknown what COVID-19 measures would be effective in reducing the spread of the COVID-19 virus. Based on these insights, an unlabeled DCE would be a suitable tool to get a baseline estimation of the value of different societal aspects at the beginning of a future pandemic. The insights provided by the unlabeled DCE can be taken into consideration when the first bundle of pandemic policies is created by policy makers. Also, in the endemic phase, an unlabeled DCE can be implemented to evaluate how societal aspects have been affected by the pandemic. The results produced by both unlabeled DCEs, at the start of the pandemic in the endemic phase, can be compared to get an indication of the degree to which several societal aspects of everyday life have deteriorated over the course of a pandemic.

It was also discussed that labeled DCEs are able to elicit the effect that the implementation of COVID-19 measures has on the importance of the different societal impacts. Therefore, during the first wave of the pandemic, policy makers can deploy labeled DCEs that weigh the societal impacts of the implemented pandemic policy. The insights produced by the DCE can help adjust the pandemic policy in such a way that it increases public support and adherence to the policy and alleviates the burden of the policy on society. During and between the waves that follow, policy makers can successively implement multiple labeled DCEs. The insights produced by these DCEs can help to evaluate and adjust COVID-19 measures that have been implemented, during the pandemic wave. Or these insights can help to evaluate and adjust COVID-19 measures in between waves to prepare the policy package for the next wave. In addition, in the endemic phase, a labeled DCE can be implemented by policy makers to extensively evaluate the impact of COVID-19 measures that have been implemented during the pandemic waves. The reason this can deliver interesting insights according to the senior advisor of the RIVM, is because in the endemic phase there exists more time to reflect for citizens on what happened during the pandemic. This provides policy makers with the opportunity to assess what should be done if another new pandemic occurs.

Models in different waves and phases of the pandemic

The MNL model is best used during waves of the pandemic. The reason for this is the model´s ability to produce quick, reliable and valid results. Specifically, during different waves of the pandemic these characteristics are important because pandemic policy has to be devised, implemented and adjusted quickly. The advantages and disadvantages of ML and LC models in analyzing DCEs and their implications to policy makers that were previously discussed, provide valuable insights on the timing of the use of these models. For instance, it was discussed that the ML model has the ability to test for the existence of dispersion of preference among individuals for a societal attribute. This can help to ensure that the mean coefficients produced by the MNL model are reliable targets for policy makers to focus on when devising generalized pandemic policy. For example, if the results from the MNL model show that a loss in income is the most important attribute and policy makers decide to ensure that pandemic policy mitigates the impact on this attribute, it is important to know if the importance of this attribute is shared by a majority of the sample. Therefore, if the standard deviation estimate of this attribute is insignificant according to the ML model, this shows that the attribute is a good target for generalized pandemic policy as the attribute is equally important to all respondents in the sample. During different waves of the pandemic, when results need to be produced urgently, it is best to estimate ML models with a limited number of draws. It was previously discussed that a disadvantage of using a limited number of draws to estimate a ML model is that estimation results for every run of the model will deviate slightly. However, this does not matter in this case as the goal of the estimation process is to identify if dispersion of preferences exists for an attribute based on the significance of the standard deviation estimate, not the exact size of this estimate. If the standard deviation estimate produced by the ML model is significant, policy makers can decide to focus on another societal attribute that carries no or less dispersion among the sample.

In between different waves of the pandemic, when mild COVID-19 measures are implemented and the pressure for quick results is a bit lower than during the waves, the important attributes that showed to have significant dispersion among the sample can be researched more carefully by estimating multiple successive ML models with an increasing number of draws. It was discussed that only an infinite number of draws per run of the model will yield exactly the same results. This would take an inmeasurable number of time. Therefore, it is common practice to start with 500 draws and double the number of draws for every run of the model until the change in results is so minor that it

does not weigh up against the increase of time to run a model with double the number of draws. If the final estimates from the ML model for the attributes are established, two choices can be made based on the size of the standard deviation estimate. If this estimate is small, this means that the majority of respondents in the sample have a preference surrounding the mean preference. This means implementation of generic pandemic policy is effective. On the other hand, if the standard deviation estimate is large, this means the preference of the majority is dispersed. In this case, the LC model can be used to study why the preference for this attribute is so dispersed. As previously discussed, the model is able to divide the respondents in the sample into different classes with their own preferences and characteristics. In this way, policy makers are able to link different preferences that are visible in the distribution produced by the ML model to specific subgroups in society. The characteristics of these subgroups, based on the covariates that are added to the LC model, can help to get a better understanding of what people in society have certain preferences. This information can help policy makers to develop customized pandemic policy for the next pandemic wave that is focused on these different subgroups. As discussed earlier, this can provide more public support and adherence and for this reason more effective pandemic policy that also is more cost effective. Every DCE conducted in a subsequent wave of the pandemic can help to evaluate if the adjustment to pandemic policy based on the analysis of the data from the previous wave has been effective in improving the view that people in society have on the impacts of these policies on societal aspects of everyday life.

Lastly, as stated by the senior advisor of the RIVM, in the endemic phase there exists more time to reflect on the pandemic as a whole. This provides policy makers with the opportunity to evaluate the policy implementation and adjustment steps that were taken based on the insights of the different models throughout the pandemic. All models have value in evaluating the changes in preferences for different societal impacts between different waves of the pandemic. In addition, the LC model can help to analyze the changes in the constitution of classes and characteristics between different waves. This can provide insights on how to improve pandemic policy for the next pandemic.

Discussion

This chapter discusses the limitations of this study and possible angles for further research. This sections starts by discussing the limitations that are related to the investigation of the two main knowledge gaps, namely the advantages and disadvantages of labeled and unlabeled DCEs and ML and LC models. The section on angles for further research discusses possible extensions of the models used in this study, other models and implementation of the insights developed by this study in the pandemic policy decision-making process.

Labeled and unlabeled DCEs

This study strived to elicit the advantages and disadvantages of labeled and unlabeled DCEs in weighing societal impacts of pandemic policy in the pandemic and endemic phase. This study investigated these advantages and disadvantages by comparing the MNL and ML model results from three unlabeled DCEs and 13 labeled DCEs. Ideally, these studies would have included the same societal attributes and COVID-19 measures. In this way, it would have been possible to compare the mean and standard deviation estimates for the preference of these societal attributes from different studies. Although, many studies included attributes that correspond to an increase in deaths or infections. Oftentimes, the other attributes were different. The reason why it is better if all societal attributes included in the analyzed DCEs are exactly the same is because the estimated coefficients for the preference of these attributes depend on one another. In other words, changing one attribute affects the estimated importance of another attribute.

To ensure for the labeled DCEs that the changes in mean and standard deviation estimates were solely due to the label, all other conditions should have been equal. Unfortunately, the DCEs included in this study were conducted in a variety of countries and during different waves and phases of the pandemic. Different countries provide different samples that have varying preferences and therefore produce different coefficient estimates. Also, the different sentiments that exist towards certain societal attributes and COVID-19 measures during different waves of the pandemic affects people's preferences. Moreover, there are also other factors that influence the estimation results of a DCE that are not considered in the literature review of this study. For instance, the experimental design of the DCE, the estimation method chosen to maximize the log-likelihood function for the different models, the number of draws used to simulate distributions for the ML model, allowing for correlation between coefficients etc… By extension of this argument, it would have been ideal for the comparison of labeled and unlabeled DCEs if every study in the review had included a labeled and an unlabeled DCE with the same societal attributes and COVID-19 measures, using the same experimental design etc... Unfortunately, the only study that did this was this study.

Furthermore, only one unlabeled and labeled DCE were conducted in the endemic phase. Namely, this study's DCEs. Therefore, the advantages and disadvantages of using labeled and unlabeled DCEs in specifically the endemic phase, had to be based on only one study. Also, the advantages and disadvantages of unlabeled DCEs in the pandemic phase had to be based on two articles from the literature review. The results discussed in the conclusion would have been better substantiated if this study would have included more unlabeled DCEs conducted in both the pandemic and endemic phase and more labeled DCEs conducted in the endemic phase. It is possible that there exist more unlabeled DCEs in the context of this study than included in this study. This study only used the Google scholar database to search for suitable articles.

This study tried to compensate for the lack of unlabeled DCEs and labeled DCE that weigh the societal impacts of COVID-19 measures by reviewing literature on these DCEs outside the scope of this study. This revealed that labeled DCEs in other domains are constructed differently. This study explicitly included COVID-19 measures as an attribute in the labeled DCE. Other labeled DCEs keep the same constitution of attributes for the DCE and unlabeled DCE. The difference in these studies is that the label refers to the title of the alternative in the choice task. The idea behind this is that you can measure what happens to the value respondents contribute to the different attributes when the alternatives in a choice task are given a label. The advantage of labeling the alternative instead of adding an attribute is that the estimation of coefficients for the attributes is not affected by a change in the constitution of attributes. On the other hand, an advantage of adding the label as an attribute is that it is possible to measure the preference for this attribute.

ML and LC model

With regards to the use of the models in this study, there are several remarks that need to be made. The conclusion mentions that a disadvantage of the ML model over the LC model is that the ML model is not able to elicit what groups or individuals are linked to different parts of the distributed preference for each societal attribute. Most articles in the review and this study's DCE use the LC model to classify groups of respondents, their preferences and their sociodemographic characteristics because the model is so convenient for this purpose. However, some studies such as Mouter et al. (2021) show that it is possible to do this with a ML model. The study shows that by splitting the dataset into two and estimating two separate ML models it is possible to obtain two subgroups of respondents with their own preferences. It is also possible within any of these two datasets to elicit the difference in preferences based on covariates that correspond to sociodemographic characteristics. By dummy or effects coding this covariate into the utility function of the ML model, it is possible to estimate the preference of a group of respondents with a certain sociodemographic characteristic, such as gender. Nonetheless, testing multiple covariates this way is a cumbersome process that asks for the estimation of multiple random coefficients which will result into colinearity issues that do not occur when estimating the LC model.

The conclusion also mentions that a disadvantage of the LC model is its sensitivity to changes in the included covariates. A solution that is mentioned by both the senior advisor of the RIVM and a head researcher of Populytics is that it is best to fix the covariates before estimating the model. The problem with this approach is that it forces the model to classify the respondents in a way that is predetermined by the chosen covariates. The goal of using the LC model is to find the classification of respondents that best fits the sociodemographic characteristics of the sample. The researcher should try to estimate multiple LC models with differing covariate constitutions. The model with the best model fit, holds the covariates that represent the sample of respondents the best. To find the LC model with the best fit it is important to try all possible combinations of covariates. This study's DCE did not use this approach as this would have taken too much time. To elaborate, this study's dataset included 13 covariates. Testing all possible combinations of these covariates to get the LC model with the best possible fit, would mean testing all 13 single covariates separately, all sets of two covariates, all sets of three covariates all the way up to a single set of 13 covariates. This takes a lot of computing time. Furthermore, this excludes the time needed to find the correct set of initial values per combination. However, if the researcher devises an algorithm to search for the best starting values and to run through all possible combinations it is possible to prove what covariates are most suitable to include. The covariates that the algorithm suggests to give the best model fit, are the latent variables that best explain the difference in class preferences.

Recommendations for future research

Supplementary modeling approaches

This study investigated the added value of LC and ML choice models for informing pandemic policy makers. As mentioned in the conclusion, the LC model is able to visualize the dispersion of preferences in a sample by dividing the sample into discrete classes with descrete class preferences. The ML model is able to present dispersion of preferences in the sample as a continuous distribution. Afcourse, this is not the only way that the preferences of respondents for societal impacts can be elicited. There exist models within the choice modeling domain and outside the choice modeling domain that can provide different insights on the preferences of respondents, that might be as valuable or more valuable to policy makers than the insights provided by this study. For example, an interesting modeling approach would be to combine the characteristics of the ML model and the LC model, by estimating a LC model with continuous class preferences. It is possible to fit a distribution to the preferences for the societal attributes of each class. By doing this, the researcher can check if important societal attributes that have a sizeable mean coefficient also enjoy the support of the majority of the sample. As illustrated before, the sensitivity of respondents to a societal attribute and the unanimity of this sensitivity among respondents are both important factors that determine if adjustments to pandemic policy that aim to mitigate the impact of the policy on a societal aspect are succesful. Recall that the LC model is able to elicit what the probability is that someone that belongs to a specific class with specific class preferences has a certain age or gender, by estimating class membership profiles. What if a policy maker uses reverse logic and wants to reason the other way around. For example, the policy maker would like to know how likely it is that a male has a high preference for wearing mouth masks. In this case, a researcher can use an ordered probit or logit model. There also exist models outside the domain of choice modeling that are able to elicit preferences of respondents for societal impacts of pandemic policy. For instance, the choice modeling researcher from the TU Delft mentioned that it would be interesting to analyze the same data with data driven models. The reason this is interesting is because choice models are based upon the utility maximization paradigm. Utility maximization assumes linearity in parameters. The choice modeler from the TU Delft states that this assumption is often falls. For example, several articles in the literature review showed that an increase in income loss showed non-linear increase of disutlity. The studies from the review were able to extract this by estimating a mean coefficient for every attribute level of the 'income'-loss attribute. The increase between these estimates was non-linear.

Implementation of modeling results in pandemic policy decision making process

This study has been able to design a simplified simulation of a piece of the complex pandemic policy decision making process. These designed simulations are the Discrete Choice Experiments that weigh the societal impacts of pandemic policy that were presented to Dutch citizens. These Discrete Choice Experiments were used to gather data on the choices Dutch citizens would make when confronted with different trade-offs. Dutch citizens had to make trade-offs between the implementation of different COVID-19 measures to save lifes, and the impact these measures would have on Dutch social and economic life. Different models were used to extract valuable insights from these trade-offs. The conclusion of this study discusses these insights and their possible implications for policy makers. The conclusion also discusses when the best moment is to extract these insights with these models during different waves and phases of a pandemic. However, what this study did not investigate is where, when and how these insights should intervene in the pandemic policy decision making process. Conducting further research into this and applying this research during the next pandemic would transform Dutch citizens from advising stakeholders into participating shareholders of the pandemic policy decision making process.

References

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. The lancet, 395 (10228), 931-934.

Belle, N., & Cantarelli, P. (2022). Your Money, Your Life, or Your Freedom? A Discrete-Choice Experiment on Trade-Offs During a Public Health Crisis. Public Administration Review, 82(1), 59-68.

Blayac, T., Dubois, D., Duchêne, S., Nguyen-Van, P., Ventelou, B., & Willinger, M. (2021). Population preferences for inclusive COVID-19 policy responses. The Lancet Public Health, 6(1), e9.

Chorus, C., Sandorf, E. D., & Mouter, N. (2020). Diabolical dilemmas of COVID-19: An empirical study into Dutch society's trade-offs between health impacts and other effects of the lockdown. PLoS One, 15(9), e0238683.

Creswell, J.W. & Plano Clark, V.L. (2017). Designing and Conducting Mixed Methods Research. 3rd edition. London: SAGE Publication Ltd.

Degeling, C., Chen, G., Gilbert, G. L., Brookes, V., Thai, T., Wilson, A., & Johnson, J. (2020). Changes in public preferences for technologically enhanced surveillance following the COVID-19 pandemic: a Discrete Choice Experiment. BMJ open, 10(11), e041592.

Ervaren impact corona op mentale gezondheid en leefstijl. (2021). Research of Statistics Netherlands. https://www.cbs.nl/nl-nl/longread/statistische-trends/2021/ervaren-impact-corona-op-mentale-gezondheid-en-leefstijl?onepage=true.

Excess mortality for the third consecutive year in 2022. (2023, January 25). Statistics Netherlands. https://www.cbs.nl/en-gb/news/2023/04/excess-mortality-for-the-third-consecutive-year-in-2022

Filipe, L., de Almeida, S. V., Costa, E., da Costa, J. G., Lopes, F. V., & Santos, J. V. (2022). Trade-offs during the COVID-19 pandemic: A Discrete Choice Experiment about policy preferences in Portugal. Plos one, 17(12), e0278526.

Fink, G., Tediosi, F., & Felder, S. (2022). Burden of COVID-19 restrictions: National, regional and global estimates. EClinicalMedicine, 45, 101305.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., ... & Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value in health, 16(1), 3-13.

Kansrijk armoedebeleid. (2020). Bureau for Economic Policy Analysis and Netherlands Institute for Social Research. https://www.scp.nl/publicaties/publicaties/2020/06/18/kansrijk-armoedebeleid.

Krauth, C., Oedingen, C., Bartling, T., Dreier, M., Spura, A., de Bock, F., ... & Robra, B. P. (2021). Public preferences for exit strategies from COVID-19 lockdown in Germany—a discrete choice experiment. International Journal of Public Health, 66, 591027.

Li, L., Long, D., Rouhi Rad, M., & Sloggy, M. R. (2021). Stay-at-home orders and the willingness to stay home during the COVID-19 pandemic: a stated-preference Discrete Choice Experiment. PLoS One, 16(7), e0253910.

Loría-Rebolledo, L. E., Ryan, M., Watson, V., Genie, M. G., Sakowsky, R. A., Powell, D., & Paranjothy, S. (2022). Public acceptability of non-pharmaceutical interventions to control a pandemic in the UK: a discrete choice experiment. BMJ open, 12(3), e054155.

Manipis, K., Street, D., Cronin, P., Viney, R., & Goodall, S. (2021). Exploring the trade-off between economic and health outcomes during a pandemic: a Discrete Choice Experiment of lockdown policies in Australia. The Patient-Patient-Centered Outcomes Research, 14, 359-371.

McFadden, D. (2000). Economic Choices. Nobel Prize Lecture. URL: https://www.nobelprize. org/uploads/2018/06/mcfadden-lecture.pdf.

McFadden, D., Train, K. (2000). Mixed MNL Models for discrete response. Journal of Applied Econometrics 15, 447–470.

Miles, J. (2005). R-Squared, Adjusted R-Squared. Encyclopedia of statistics in behavioral science.

Man diagnosed with coronavirus (COVID-19) in the Netherlands. (2020, February 28). Ministerie van Algemene Zaken. https://www.government.nl/latest/news/2020/02/27/man-diagnosed-with-coronavirus-COVID-19-in-the-netherlands.

Mouter, N., Jara, K. T., Hernandez, J. I., Kroesen, M., de Vries, M., Geijsen, T., Kroese, F., Uiters, E. & de Bruin, M. (2022). Stepping into the shoes of the policy maker: Results of a Participatory Value Evaluation for the Dutch long term COVID-19 strategy. Social Science and Medicine. 314, 115430.

Mühlbacher, A. C., Sadler, A., & Jordan, Y. (2022). Population preferences for non-pharmaceutical interventions to control the SARS-CoV-2 pandemic: trade-offs among public health, individual rights, and economics. The European Journal of Health Economics, 23(9), 1483-1496.

Ozdemir, S., Tan, S. N. G., Chaudhry, I., Malhotra, C., & Finkelstein, E. A. (2021). Public preferences for government response policies on outbreak control. The Patient-Patient-Centered Outcomes Research, 14, 347-358.

Reed, S., Gonzalez, J. M., & Johnson, F. R. (2020). Willingness to accept trade-offs among COVID-19 cases, social-distancing restrictions, and economic impact: a nationwide US study. Value in health, 23(11), 1438-1443.

Risk assessment: Outbreak of acute respiratory syndrome associated with a novel coronavirus, Wuhan, China; first update. (2020b, January 22). European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/publications-data/risk-assessment-outbreak-acute-respiratory-syndrome-associated-novel-coronavirus

Sicsic, J., Blondel, S., Chyderiotis, S., Langot, F., & Mueller, J. E. (2023). Preferences for COVID-19 epidemic control measures among French adults: a discrete choice experiment. The European Journal of Health Economics, 24(1), 81-98.

Train, K. (2009). Discrete Choice Methods with Simulation. second edition ed., Cambridge University Press, Cambridge, MA.

Wep, G. I. P. (2019). Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza. www.who.int. https://www.who.int/publications/i/item/non-pharmaceutical-public-health-measuresfor-mitigating-the-risk-and-impact-of-epidemic-and-pandemic-influenza.

Appendix

Appendix A.1 Interview protocols

This appendix includes the expert interview protocols. The setup of the different protocols is the same with regards to the line of questioning for sub interview questions. The sub interview questions help to answer the different parts of sub research question three. In addition to these questions, each protocol contains expert specific questions.

Interview protocol 1

19-05-2023

Head researcher

Populytics

Worked on analyzing societal impacts of COVID policy for RIVM on behalf of Populytics.

Introduction

My name is Daniel Korthals, master student Complex Systems Engineering and Management at the TU Delft and intern at Populytics. Currently, I am writing my thesis on the added value of choice modeling techniques for the analysis of DCEs that weigh the societal impact of COVID-19 measures. Your interview will be anonymized and summarized. The summary will be part of the appendix of the thesis.

Goal of the interview

To get an answer to sub research question three from the perspective of the head researcher of a company.

Introductory question

Could you introduce yourself and your work?

Sub interview questions

1. What are the (dis)-advantages of using ML, LC and MNL models in your work in general?
2. What are the (dis)-advantages of using labeled vs. unlabeled DCEs in your experience?
3. Is there any reason to prefer the use of ML, LC or MNL models to analyze labeled or unlabeled DCEs? If so, what reasons?
4. Which of the three models would you prefer to use to analyze DCEs during a pandemic and why?

Expert specific questions

5. What does populytics usually use the results from the DCE for?
6. What do clients of Populytics usually think of the results that are obtained with DCEs conducted by Populytics?
7. What is your opinion on the use of DCEs to elicit preferences of respondents?
8. What is the added value of DCEs to inform (COVID-19) policy decision-making processes?

Closing question:

Thank you very much for the interview. Do you have any questions for me?

Interview protocol 2

19-05-2023

Choice modeling researcher

TU Delft

Worked on analyzing societal impacts of COVID policy for RIVM on behalf of TU Delft.

Introduction

My name is Daniel Korthals, master student Complex Systems Engineering and Management at the TU Delft and intern at Populytics. Currently, I am writing my thesis on the added value of choice modeling techniques for the analysis of DCEs that weigh the societal impact of COVID-19 measures. Your interview will be anonymized and summarized. The summary will be part of the appendix of the thesis.

Goal of the interview

To get an answer to sub research question three from the perspective of a choice modeler.

Introductory question

Could you introduce yourself and your work?

Sub interview questions

1. What are the (dis)-advantages of using ML, LC and MNL models in your work in general?
2. What are the (dis)-advantages of using labeled vs. unlabeled DCEs in your experience?
3. Is there any reason to prefer the use of ML, LC or MNL models to analyze labeled or unlabeled DCEs? If so, what reasons?
4. Which of the three models would you prefer to use to analyze DCEs during a pandemic and why?

Expert specific questions

5. What did you think of the results obtained in the paper on the societal impacts of COVID policy?
6. Are there other models that are more suitable to analyze DCEs during a pandemic?
7. What is the added value of analyzing DCEs with LC, ML models over MNL models to inform the COVID-19 policy decision making process?

Closing question:

Thank you very much for the interview. Do you have any questions for me?

Interview protocol 3

05-06-2023
Senior advisor behaviour and communication (Senior advisor to the Corona behaviour unit)
RIVM (National Institute for Public Health and the Environment)

Introduction:

My name is Daniel Korthals, master student Complex Systems Engineering and Management at the TU Delft and intern at Populytics. Currently writing my thesis on the added value of choice modeling techniques for the analysis of DCEs on the societal impact of COVID-19 measures.

Introductory question:

Could you introduce yourself and your work?

Goal of interview:

To get an answer to the main reserch question from the perspective of a representative from the RIVM.

Abbreviations:

Multinomial Logit = MNL
Latent Class = LC
Mixed Multinomial Logit = ML
Discrete Choice Experiment = DCE

Sub questions:

1. What are the (dis)-advantages of using ML, LC and MNL models in your work in general?
    - ML vs MNL
    - LC vs MNL
    - ML vs LC
2. What are the (dis)-advantages of using labeled vs. unlabeled DCEs in your experience?
3. Is there any reason to prefer the use of ML, LC or MNL models to analyze labeled or unlabeled DCEs? If so, what reasons?
4. Which of the 3 models would you prefer to use to analyze DCEs during a pandemic and why?

Expert specific questions:

5. What is your opinion on the use of DCEs to elicit preferences of respondents?
    - Do you have a preference for labeled or unlabeled DCEs, and why so?
6. What is the added value of DCEs to inform the (COVID-19) policy decision-making processes?
7. What did you think of the results obtained in the paper on the societal impacts of COVID policy?
    - The decision was made to exclude the results from the latent class model and mixed logit model, why was this decision made?
8. What criteria do the results of stated preference studies (such as DCEs) have to satisfy to be useful for the RIVM?
9. When looking at the literature that was produced with stated preference studies during the different waves of the pandemic and the endemic phase we transitioned to currently, what are the differences between the results obtained in the different phases in general according to your experience?

Close interview

Thank you very much for the interview. Do you have any questions for me?