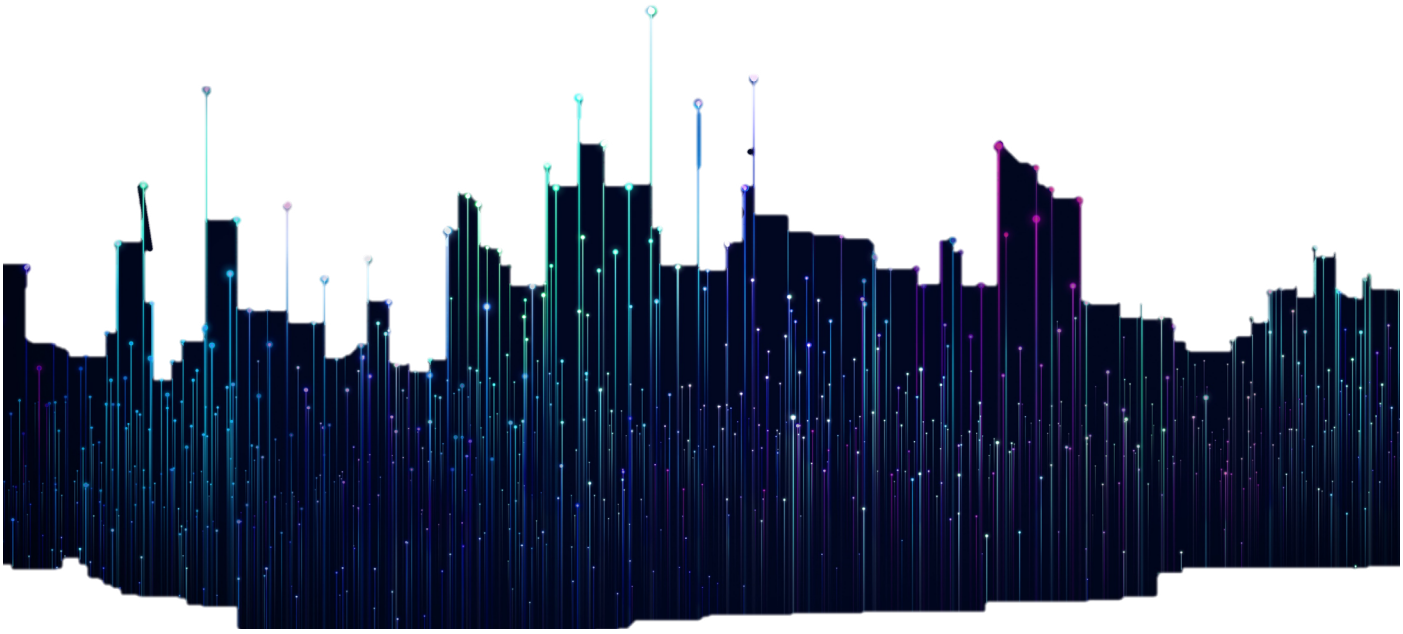
BIDA
UTS: Business Intelligence & Data Analytics



WORKING PAPER
BIDA-WP-2002

The role of constants in discrete choice models: It's not a constant sum game

By:
John M. Rose[1]*, Antonio Borriello[1], Andrea Pellegrini[1], Daniel Masters[1]

August 2020

UTS

TITLE:

The role of constants in discrete choice models: It's not a constant sum game

AUTHORS:

John M. Rose[1]*, Antonio Borriello[1], Andrea Pellegrini[1], and Daniel Masters[1]

ABSTRACT:

Discrete choice experiments undertaken within the health economics domain involve respondents having to choose from amongst two or more alternatives. An examination of the literature suggests that models applied to unlabelled choice data often do not involve the estimation of a constant term, or when a no choice or status quo alternative is present, a single constant linked to the no-choice alternative. In this paper, we argue that choice models estimated using unlabelled choice data should always involve the estimation of constant terms for all but one alternative, without which significant biases can occur with respect to parameter estimates and willingness to pay outputs. We argue that constants should be included for all but one alternative, even if they are not statistically significantly different from each other, or from zero. We posit that despite not offering a behavioural interpretation for such data, alternative specific constants can account for different data issues that may go undetected. We also recommend that papers reporting models estimated using discrete choice data provide more detail about the descriptive statistics of the data than currently appears to be the case. Finally, we propose that future efforts exploring the external validity of models estimated using discrete choice experiments focus on real markets in which observed changes are likely to occur, rather than attempt to simply predict the market shares of an existing market at one point in time.

KEY WORDS:

Constants; Alternative specific constants; Discrete Choice Experiments; Prediction, forecasting

CONTACT:

*John M. Rose
The Centre for Business Intelligence and Data Analytics (BIDA)
The University of Technology, Sydney
PO Box 123, Sydney, NSW 2007, Australia
Telephone: +61 2 9514 5994
Email: bida@uts.edu.au
Website: uts.edu.au/BIDA Research

[1]The Centre for Business Intelligence and Data Analytics (BIDA)
UTS Business School, University of Technology, Sydney


UTS
WORKING PAPER
BIDA-WP-2002

1.0 Introduction

Hensher et al. (2005) suggested that when dealing with unlabelled discrete choice experiment (DCEs), constant terms offer no substantive behavioural meaning and hence should not be estimated as part of the model. In a later edition, Hensher et al. (2015) contradict this original advice and suggest that alternative specific constants (ASCs) should be incorporated in discrete choice models dealing with unlabelled DCEs. Despite this retraction, it is evident that many researchers fail to estimate constant terms for all but one alternative. Indeed, it appears to be common practice, particularly when dealing with unlabelled binary choice experiments used to capture dimensions of quality of life such as the EQ-5D (e.g.,), not to include any constant terms, or for trinary unlabelled DCEs involving a status quo or no choice alternative (the two terms appear to be used interchangeably), to estimate a single constant term linked to the no choice option.

We argue that unlabelled choice experiments should always involve the estimation of ASCs, irrespective of the number of alternatives and/or presence of a no choice alternative. We present five possible causes that may result in an imbalance of preferences across alternatives within unlabelled DCEs, that can be at least partially dealt with only with the inclusion of ASC terms. The presence of one or more these issues, which are unlikely to be unknown to the analyst, may result in biased estimates if not dealt with appropriately. Whilst more advanced econometric models may be able to deal with one or more of these issues directly, such models are not in wide-spread use, and even when used, unless they account for all of five issues, may not fully overcome the problem with preference imbalance across alternatives. Further, at issue is the fact that many papers present results without estimating ASCs for all but one utility function, making it impossible to determine if they are reporting models based on data sets with preference imbalance across the alternatives, and hence models with potentially biased estimates. We argue that aside from additional degrees of freedom, the inclusion of alternative specific constants does not cost the analyst anything, yet can prevent specific effects from biasing the estimates of discrete choice models.

To demonstrate the impact of not including alternative specific constants for all but one alternative, we present both a theoretical argument as well as demonstrate the outcomes from models estimated on a real-world empirical data set. In addition to understanding the role that constants play in model estimation, we further discuss the role of constants in model prediction and forecasting.

The remainder of the paper is organised as follows. In the next section, we discuss the theoretical role that constants play in the estimation of discrete choice models, and how the standard practice of estimating a single constant for a status quo alternative when more than three alternatives are present can result in biased outputs. Section 3 of the paper presents an empirical data set, after which results are presented in Section 4 demonstrating the issues raised earlier in Section 2. Section 3 further discusses issues around the use of constants in forecasting. Section 4 also presents further analysis showing that more advanced choice models also fail to deal with the issue of incorrect specification of the constant terms within discrete choice models. Finally, Section 5 presents a brief discussion and outlines a series of recommendations related to the use of constants in studies modelling discrete choice outcomes.

2.0 The role of the constants in discrete choice models

Independent of the data type, constants terms play multiple roles in discrete choice models. The most common role attributed to constants is that they represent the mean of the unobserved effects of the alternative(s) to which they are assigned, after accounting for the contribution of non-constant variables contained within the modelled system of utility functions. The second role assumed, related to the first, is that the modelled constants reflect preferences for the related alternatives, once more after accounting for the impact of the other variables contained within the model. This second role is usually confined to interpreting the outputs of models estimated on data containing labelled alternatives or where one alternative is a no-choice or status quo alternative. A third much less stated role relates to the fact that the estimation procedures used to obtain the parameter estimates of discrete choice models are designed to ensure that the predicted aggregate choice probabilities obtained from the model match as closely as possible the market shares contained within the data. That is to say that $P_j \approx \sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{j=1}^{J}\hat{P}_{nsj}\Big/NS$ where P_j is the sample market share for alternative *j*, $\hat{P}_{nsj}$ is the choice probability obtained from the model for respondent *n* in choice situation *s* associated with *j* alternative, and *NS* reflects the total number of choices made within the sample data.

Whilst the first two roles primarily relate to the interpretation of the constant terms of the model, the latter relates to their function in model estimation, and more importantly, determines the values that they will take. To demonstrate, let U_{nsj} denote the utility of alternative *j* perceived by respondent *n* in choice situation *s*, which consists of an observed component V_{nsj} and an unobserved component ε_{nsj},

$$U_{nsj} = V_{nsj} + \varepsilon_{nsj}. \qquad (1)$$

As is common practice, the observed component is assumed to be described by a linear relationship of *k* observed attribute levels of each alternative, *x*, and their corresponding weights (parameters), β, such that utility can be expressed as

$$U_{nsj} = \sum_{k=1}^{K}\beta_k x_{nsjk} + \varepsilon_{nsj}. \qquad (2)$$

In case a certain parameter β_k appears in the utility function of multiple alternatives *j*, it is said to be generic over these alternatives. Otherwise, the parameter is called alternative-specific. In our notation, if a certain attribute *k* does not appear in the utility function of a certain alternative *j*, then we assume that $x_{nsjk} = 0$. To aid in interpretation, we will denote constants as α_j to differentiate them from non-constant terms.

Now assume the existence of a discrete choice experiment involving three alternatives representing options defined by attributes and attribute levels. Following standard practice, the system of utility equations for this experiment would be written as

$$
\begin{aligned}
U_{ns1} &= \alpha_1 + \sum_{k=1}^{K} \beta_k x_{ns1k} + \varepsilon_{ns1}, \\
U_{ns2} &= \alpha_2 + \sum_{k=1}^{K} \beta_k x_{ns2k} + \varepsilon_{ns2}, \\
U_{ns3} &= \sum_{k=1}^{K} \beta_k x_{ns3k} + \varepsilon_{ns3},
\end{aligned}
\tag{3}
$$

where we include alternative specific constants for *J*-1 of the alternatives.

Assuming that the error terms are IID Extreme Value (EV) type 1 distributed, the mean of the error term will be $E\left(\varepsilon_{nsj}\right) = 0.57721/\sigma_n$ and variance $\text{var}\left(\varepsilon_{nsj}\right) = \frac{\pi^2}{6\sigma_n^2}$, where a positive scale factor that is typically normalised to one in most applications. Ignoring the systematic component of utility, adding a constant term to a distribution shifts all elements of the distribution by the same amount, shifting the mean of the distribution, but not affecting the variance, such that the mean of the error term for a given alternative will be equal to $\alpha_j + 0.57721$. Note that, if two or more constants are not estimated, then the means of the error terms for these alternatives are assumed to be equal. As we discuss later, it is common in many discrete choice experiments to estimate models with less than *J*-1 alternatives. This is particularly the case, but not limited to, studies involving a no-choice or status quo alternative, where a constant is estimated only for the no or status quo option, as shown in the system of utility equations given in Equation (4).

$$
\begin{aligned}
U_{ns1} &= \sum_{k=1}^{K} \beta_k x_{ns1k} + \varepsilon_{ns1}, \\
U_{ns2} &= \sum_{k=1}^{K} \beta_k x_{ns2k} + \varepsilon_{ns2}, \\
U_{nsSQ} &= \alpha_{SQ} + \varepsilon_{nsSQ}.
\end{aligned}
\tag{4}
$$

As per the discussion above, Equation (4) assumes that the means of error terms associated with alternatives 1 and 2 are simultaneously equal to 0.57721, assuming scale is normalised to 1.0.

In addition to representing the means of the error terms, it is important to understand that constant terms are parameters that enter into utility, and hence also impact on the choice probabilities of estimated models. To understand this, assume a standard multinomial logit (MNL) model is to be estimated. Other more advanced models can be similarly assumed, by appropriate adaption of the functional form of choice probabilities shown. As such, the discussion that follows is not specific to the MNL model and can be generalised to any discrete choice model. Assuming the system of utility functions given by equation (4), the choice probabilities for the model would be

$$\hat{P}_{ns1} = \exp\left(\sum_{k=1}^{K}\beta_k x_{ns1k}\right)\Big/\left(\exp\left(\sum_{k=1}^{K}\beta_k x_{ns1k}\right)+\exp\left(\sum_{k=1}^{K}\beta_k x_{ns2k}\right)+\exp\left(\alpha_{SQ}\right)\right) \quad (5)$$

$$\hat{P}_{ns2} = \exp\left(\sum_{k=1}^{K}\beta_k x_{ns2k}\right)\Big/\left(\exp\left(\sum_{k=1}^{K}\beta_k x_{ns1k}\right)+\exp\left(\sum_{k=1}^{K}\beta_k x_{ns2k}\right)+\exp\left(\alpha_{SQ}\right)\right)$$

$$\hat{P}_{ns3} = \exp\left(\alpha_{SQ}\right)\Big/\left(\exp\left(\sum_{k=1}^{K}\beta_k x_{ns1k}\right)+\exp\left(\sum_{k=1}^{K}\beta_k x_{ns2k}\right)+\exp\left(\alpha_{SQ}\right)\right).$$

This is important for the following reason. Firstly, the parameter estimates of discrete choice models are obtained using maximum likelihood techniques. For the MNL model, the log-likelihood function can be represented as

$$LL = \sum_{n=1}^{N}\sum_{s=1=rest}^{S}\sum_{j=1}^{J} y_{nsj}\ln\left(P_{nsj}\right), \quad (6)$$

where y_{nsj} represents the observed choice for respondent *n* in choice situation *s*, equal to 1 if alternative *j* is chosen, or 0 otherwise, and P_{nsj} represents the choice probabilities given in Equation (5).

Within Equations (5) and (6), x_{nsjk} and y_{nsj} are data, and hence given. Thus, in maximising equation (6), only α_{SQ} and β_k are free to be estimated. Now assume that data is collected using an unlabelled choice experiment, such that the non-constant parameter terms should be estimated as generic estimates. Issues will then arise if for whatever reason the means of the error terms for alternatives 1 and 2 are not equal, all else being equal. In the aggregate, this will mean that the actual observed market shares within the data are such that $P_1 \neq P_2$, with the only mechanism available to the model to ensure that $\hat{P}_1 \approx P_1 \neq \hat{P}_2 \approx P_2$ is via the preference parameters, β_k, as the utility functions for these alternatives do not contain a constant term. As such, the preference parameters in such an instance reflect not just the influence of the attributes on choice, but also are forced to take on the role of alternative specific constants in order for the model to better reflect the known choice shares within the data. Only if $P_1 \approx P_2$, will it be possible to estimate α_3 so that over the sample, $\hat{P}_1 \approx \hat{P}_2 \approx P_1 \approx P_2$, and $\hat{P}_3 \approx P_3$, after accounting for the influence of the attributes, x_{nsjk}, on choice.

3.0 Possible causes of preference imbalance across alternatives

There two broad types of survey error that trouble researchers, sampling error and non-sampling error. In the simplest terms, sampling error occurs as the analysis is often unable to collect data from the whole population and must settle for a subset drawn from the population. Non- sampling error describes all other sources of error including response, and non-response errors. The most significant cause of response error, referred to as order effects, relates to the ordering of questions or responses categories to questions (Sanjeev and Balyan, 2014). Non-response errors, as the term suggests, result from participants not responding to questions or declining to provide information.

Both sources of error, and the resulting biases, are relevant to DCEs.

In this section, we discuss five possible issues that may result in an alternative being selected more or less than other alternatives within a DCE. Note that these issues can arise in any DCE, irrespective of the number of alternatives, the presence or absence of a no choice or status quo alternative, or the type of design generation process employed. We further note that the issues discussed are not mutually exclusive meaning that a data set may have more than one issue leading to preference imbalance across alternatives.

3.1 Response Bias

Researches have long been aware of the potential influence of survey design and structure to affect responses of participants and result in the type of response errors noted above (for example see Sayer, 1939; Cantril, 1944; Bradburn and Mason, 1974). The most common of these influences, referred to as order effects, result from how questions or responses categories to questions are sequenced within the survey instrument. That is, the response to Question A may differ if it is presented after Question B than if it were presented after Question C, or alternatively, if the order of the questions is reversed. As this type of error is result of question sequence, Strack suggested 'question-order effects' or simply 'question effects' was a more suitable term (Strack, 1992). Chrzan (1994) presents three additional order effects for the choice analyst to consider; (1) choice set order, within the set of choice sets; (2) profile order within choice sets; and (3) attribute order within profiles.

In addition to 'question-order effects' and those noted by Chrzan (1994), the potential for 'left-right' order effects should also be considered by the choice analysts. The assumption is that in most Western countries, respondents read left to right, and all else being equal, will tend to select options that they see first, these being those on the left-hand side in a DCE. This 'left-right' bias may be more likely to occur if the respondent finds the choice task difficult and employs 'left-right' decision heuristic. Alternatively, Krosnick (1999) suggests that the first item perceived and examined by the respondent will be subject to deeper cognitive effort and may for the basis of comparison. For similar reasons, top-bottom response biases may also exist. The most cited reason for estimating alternative specific constants is to test and account for possible order effect biases, and particularly left-to-right biases, that may exist in choice experiments.

3.2 Block imbalance in data collection

It is common for researchers to rely on designs where that the total number of tasks within the design is greater than the number of tasks any individual respondent can practically answer when completing a survey. As a result, it is necessary to determine how the various tasks can be assigned to respondents. In some instances, analysts randomly assign questions to respondents to overcome the type of biases noted by Chrzan (1994) and discussed above. However, a far more common approach is to generate an additional column when constructing the design, called a blocking column. The levels of the blocking column are then used to assign tasks to respondents. Again, these blocking variables may be assigned to respondents randomly. An example is shown in Table 1 for a simple design problem involving two alternatives, each described by two attributes with two levels. Overall, the design has 8 choice tasks. An additional blocking column has been generated with four levels.

Table 1: Example design with blocking column

Set	A1	B1	A2	B2	Block
1	0	0	0	0	0
2	1	1	1	1	0
3	0	1	1	0	1
4	1	0	0	1	1
5	0	1	0	1	3
6	1	0	1	0	3
7	0	0	1	1	4
8	1	1	0	0	4

Assuming more respondents are to be sampled than levels used in constructing the blocking column, multiple respondents will need to be assigned to complete questions associated with each block. Unless specific controls are put in place, it is not uncommon for different numbers of respondents to complete each block, meaning that in the data, each question in the design will not have equal representation. For studies using orthogonal designs, this means that the data will not be orthogonal. For studies using efficient designs, unless a heterogeneous design strategy was used (see Sandor and Wedel 2005), the efficiency of the design will not carry through to the data (it is not necessary that the data will be less efficient than assumed however, as the efficiency may be improved under certain circumstances).

Depending on how the design is generated and blocking column constructed, it is possible that the attribute levels will also be unbalanced within the data. If an orthogonal design is used with an orthogonal blocking column, this is less likely to be the case. To demonstrate, consider the design shown in Table 1, which is an orthogonal design with an orthogonal blocking column. If three respondents were to complete a survey constructed using the design shown answering questions from blocks 1, 2 and 4 respectively, despite the data no longer being orthogonal, each level in the data would still appear three times for each attribute. This is because the attributes and levels in the design are orthogonal to the blocking column. For efficient designs which are not orthogonal, it is likely to be impossible to generate an orthogonal blocking column for the design, and hence an imbalance in the number of blocks within a data set will likely result in an imbalance in the number of times each level appears for each attribute within the data. If this occurs at different rates across alternatives, then where certain levels are more or less desirable within a design, certain alternatives may be observed to be more or less likely to be chosen than others.

3.3 Missing responses, or missing data

Depending on the data collection process employed, some studies allow respondents not to respond to different choice tasks (e.g., paper and pencil surveys). This means that each question within the design will not be equally replicated in the data, even if the blocks are equally replicated within the data. This will not only impact orthogonality or efficiency of the data, but will likely result in an imbalance in the attribute levels across alternatives also, which may, within the data, make one or more alternatives slightly more, or less, attractive than others.

Other issues, rarely considered, relate to missing data of supplementary questions (those not directly related to the DCE) and the impact of 'data cleaning'.

In the case of missing data from supplementary questions, the analyst may be interested in the effect of attitudes, income or some other socio-demographic characteristic on choice behaviour.

Where a respondent declines to provide a response or is provided the option of 'I would rather not say', that response will be treated as 'missing' for that respondent. Depending on how the researcher's chosen software handles such data, it is possible the entire data for that respondent will be (list-wise or case-wise) removed from the analysis.

Thus, even if all blocks are equally represented within a data set, this does not mean that they will be equally represented within the analysis. During the data cleaning process, the researcher may remove respondents considered of 'poor' quality. There are several legitimate reasons for a researcher to remove respondent data such as the respondent 'speeding' though the survey or those that 'flat-line' supplementary questions. If an online survey is used, the researcher may also elect to remove responses identified as 'bots', as 'duplicate respondents' or surveys completed by IP addresses outside of the study area. With the growth of online surveys, the need for the researcher to remove these types of responses is also growing. As these issues may only be detected by the researcher after data collection, removing these respondents will result in an unequal representation of choice tasks and the potential loss of orthogonality in the data. In the author's own experience, it is not unusual for 5% of the data collected to exhibit quality issues of the type described above.

3.4 Dominance in designs

A fourth possible issue relates to the possible presence of dominated alternatives in the experimental designs that are used to assign the attribute levels to the tasks underly discrete choice experiments. Dominated alternatives occur when the levels of all attributes of an alternative are worse than those of one or more of the other alternatives within the design. In such cases, the dominant alternative will likely be chosen more often than the dominated alternative (we say more likely than opposed to will be given that we assume choices are consistent with random utility theory, meaning that there is an error attached to each choice. That is choices are not deterministic). To determine if an alternative is dominated, it is necessary to assume a priori knowledge about the preference order of all attribute levels within the design. Consider for example that for both attributes A and B in Table 1, that 0 is preferred to 1. In such a case, then alternative 1 will dominate alternative 2 in choice task 7, whereas the opposite is true in choice task 8. In this example, each alternative has an equal number of dominated alternatives within the design, however this need not hold in all cases. If within a design, one alternative is dominated more or less than all other alternatives, then even in data sets where blocks are equally replicated and there are no missing responses or data, it is likely that that alternative will be chosen more or less than others, all else being equal.

Note that the presence of dominant alternatives is not limited to orthogonal designs, although orthogonal designs will more likely result in such alternatives occurring within the design. Whilst efficient designs that assume non-zero priors explicitly account for the preference ordering of the levels when generating the design and attempt to avoid tasks that have dominated alternatives, the absence of dominated alternatives cannot be guaranteed, particularly when constraints such as attribute level balance are imposed during the design generation process. This is because the feasible set of tasks that do not display non-dominance yet retain attribute level balance may be less than the total number of tasks required by the analyst when generating the design. Further, efficient designs that assume zero priors do not account for preference ordering of the attribute levels during design construction, and hence do not directly or indirectly account for this issue when generating the design.

3.5 Attribute non-attendance

It is finally worth noting that an assumption in generating designs is that all of the attributes used in the design will be considered when respondents answer the survey. Unfortunately, there exists a vast literature on attribute non-attendance that shows that this is not the case (see e.g., Ryan et al. 2009 or Hole 2011). In such cases, it is possible that far more choice tasks than assumed will display alternative dominance than is assumed, and that such dominance will only occur for a subset of respondents. For example, assume that 10 percent of respondents answering questions based on the design shown in Table 1 do not use attribute B. In that case, assuming 0 is still preferred to 1 for attribute A, then for these respondents, alternative 1 will be dominated in choice task 4 and 8, whilst alternative 2 in choice tasks 3 and 7 will be dominated. If different respondents ignore or process different attributes in the sample, then dominance may vary significantly within a data set across different alternatives.

4.0 Is this really an issue?

In order to understand how researchers are treating constants for DCEs within the Health Economics literature, Table 2 presents the results of a literature review for the years 2015 to 2020 drawing from the journals Health Economics, PharmacoEconomics and Value in Health. Whilst other journals report DCE studies, these journals were selected as being representative of the literature. The search criteria was limited to papers that use the term discrete choice experiment, and only papers that report studies involving a three or more alternatives are included. It is important to note that the issues discussed are not limited to studies involving three or more alternatives, and can equally impact the results of binary choice experiments (we found over 60 papers in these journals which used a binary choice experiment). Studies involving labelled alternatives were also excluded.

In total, twenty-five papers were identified based on the above outlined strategy. No papers were found with more than three alternatives. Five papers report studies that utilised a force choice experiment (that is, did not include a status quo or no choice alternative). Of the 26 papers identified, five report results from models that did not include any constants (19.23 percent), whereas 16 (61.54 percent) report models with a single constant associated with a status quo or no choice alternative. One paper did not provide any information about the modelling undertaken making it impossible to determine whether constants were estimated or not. As such, only four of the 26 papers (15.83 percent) examined estimated constants for all but one of the alternatives present within the study.

Also shown in the table are the sample sizes, number of observations, the size of the design used, the number of tasks each respondent was asked to complete, and how many blocks were used for the design. Whilst all papers report the final sample size used in the analysis, a large number of papers do not report information either about the number of observations used in the study (making it impossible to determine if missing responses are present) or about the design itself (making it impossible to determine if blocks or choice tasks are equally replication within the data). Were information is provided, dividing the total number of observations reported by the design size indicates that it is impossible for seven of the 26 papers (26.92 percent) to have used data where every choice task is replicated equally, whilst two others report having used random blocking which will also likely result in a similar outcome. Lancsar et al. (2017) actually report that "The first and second blocks of 16 choice-sets were answered by 37 and 39 participants, respectively". Seventeen papers do not provide sufficient evidence to determine if data issues may be present, whereas we can conclude for one paper only, van de Wetering et al. (2015), that no such issues exist given that these authors gave the exact same nine choice tasks to all respondents.

Sicsic et al (2018) represents the only other paper where it is possible where all choice sets and blocks are equally replicated, although the authors do not explicitly state that the two blocks of their design were equally replicated within their data set. Based on the above findings, we can conclude that the majority of papers examined either exhibit one or more of the issues identified in Section 3, or do not provide sufficient information to determine whether such issues exist. Papers that fall into this latter category are particularly problematic given, as there is no way to verify whether a problem exists or not.

Table 2: Papers with three or more alternatives and the use of constants

Authors	Year	Number of alts	Sample size	Observations	Design size	Tasks	Blocks
No constant							
van Dijk et al.	2016	3 SP	429	NR	NR	8	200 random
Doiren and Yoo	2017	3 SP	241	NR	NR	8	NR
Mohammadi et al.	2017	2 SP + None	194	NR	NR	10+2*	12
Mulhern et al.	2017	2 SP + None	366	3646	120	10	12
Ostermann et al.	2020	3 SP	403	6432	96	16	6
Opt-out constant only							
Chen et al.	2015	2 SP + None	838	NR	15	5 + 2*	Random
van de Wetering et al.	2015	2 SP + None	1,205	NR	72	8+2*	9
Dong et al.	2016	2 SP + None	189	NR	32	8	4
Mühlbacher et al.	2016	2 SP + None	1,301	NR	36	6	6
Veldwijk et al.	2016	2 SP + None	1,045	9,405	9	9	1
Lancsar et al.	2017	2 SP + None	76	2,432	32	16	2
Wright et al.	2017	2 SP + None	702	NR	40	10	4
Heidenreich et al.	2018	2 SP + None	443	4,625	32	11 or 10	3
Quaife et al.	2018	2 SP + SQ	244	2,440	NR	10	NR
Ryan et al.	2018	2 SP + SQ	58	2,807	NR	12	NR
Vass et al.	2018	2 SP + None	1,018	11,198	44	11	4
Wong et al.	2018	2 SP + None	482	NR	8	8+1*	1
de Bekker et al.	2019	2 SP + None	418	6,688	160	16	10
Allanson et al.	2020	2 SP + None	48	575	48	12+2*	4
Norman et al.	2020	2 SP + None	503	NR	NR	NR	NR
Krucien et al.	2019a	2 SP + SQ	200	4,200	NR	10 or 12*	NR
Alternative specific constants							
Flynn et al.	2016	3 SP	525	NR	32	4	8
Holte et al.	2016	2 SP + SQ	934	4,670	20	5	4
Sicsic et al.	2018	2 SP + SQ	812	6,496	16	8	2
Krucien et al.	2019b	3 SP	311	3,732	NR	14	NR
Indeterminant constants							
Marshall et al.	2017	2 SP + None	193	NR	NR	6	NR

*NR = not reported, SQ = Status quo. * Additional choice tasks were employed to test for consistency in choices*

If is finally worth noting that of the four papers that report estimating constants for two of the three alternatives within their data (all four estimated a SQ plus one other constant term), two (Flynn et al. 2016 and Holte et al. 2016) report statistically significant constants in at least one model for the constant associated with the non-status quo alternative.

This suggests that the mean of the error terms may be different for the non-status quo alternatives in these models, indicating that the above identified problems may exist, but have been corrected for via the use of alternative specific constants.

5.0 Empirical data set example

For the current study, we utilise data obtained from an online survey designed to capture the preferences of Australian citizens for a vaccine specifically targeted at immunising against SARS-CoV-2. Respondents completing the survey undertook a DCE consisting of four alternatives, of which three hypothetical vaccines defined by seven attributes, and a no-choice alternative. Table 3 lists the attributes and their respective levels used to describe the various vaccines over the course of the experiment.

Table 3: attributes and attributes level

Attribute	Attribute description	Levels
Mild side effects	Number of incidences per 10,000 citizens	10, 20, 100, 200
Major side effects	Number of incidences per 10,000 citizens	1, 2, 10, 20
Vaccination effectiveness	The percentage of individuals given the vaccine who become immune to the virus	84%, 89%, 94%, 99%
Mode of administration	How the vaccine is administered	Oral, Injection
Location	Where the vaccine is administered	Doctor's office, Hospital, Pharmacy
When available	How long (in months) until the vaccine becomes available	0, 2,4,6,8,10,12, 14
Cost	The out of pocket expense to the respondent	$0, $20, $40, $60, $80, $100, $120, $140

A Bayesian D-efficient design constructed using Ngene [2] was used to assign the attribute levels of the choice experiment to each of the 40 choice tasks generated. Priors for the design were obtained from a small pilot study consisting of 10 respondents, with uninformative priors in the form of uniform distributions used to generate the final design. The design was programmed to avoid dominated alternatives, and was generated so that all respondents saw a common set of four choice tasks, plus four additional choice tasks drawn from one of nine blocks. Two thousand Sobol draws were employed in constructing the design. An example choice task is shown to respondents is reproduced in Figure 1.

	Vaccine A	Vaccine B	Vaccine C	None
Mild side effects	10 per 10,000 individuals	200 per 10,000 individuals	100 per 10,000 individuals	
Major side effects	2 per 10,000 individuals	2 per 10,000 individuals	20 per 10,000 individuals	
Vaccination effectiveness	84.00%	84.00%	89.00%	
Administration	Injection	Injection	Injection	
Location	Dr's office	Pharmacy	Pharmacy	
When will it be available (months from now)	10 month(s)	14 month(s)	12 month(s)	
Price	$100.00	$60.00	$120.00	
I would choose	○	○	○	○

Figure 1: example of choice task

The survey was administered to 2,151 Australian citizens drawn from all states and territories between the 27th and 31st March 2020 with survey eligibility restricted to persons aged 18 years or older. Respondents were recruited using the online survey panel Online Research Unit (http://www.theoru.com/index.htm). Data from 15 respondents were removed due to inconsistent responses to questions or a completion time less than two minutes resulting in a data set consisting of 17,088 choice observations obtained from 2,136 respondents.

The top segment of Table 4 shows the number of times each alternative was chosen both in absolute and terms and as choice shares. The lower segment of the table shows the average attribute levels for the same data. As shown in the table, the first and second alternative were chosen 38 and 39 percent of the time respectively, whilst the third alternative was selected only 16 percent of the time. The no choice alternative was chosen as the most preferred alternative in only six percent of choice tasks. Comparing the average attribute levels within the data, significant differences become readily apparent, despite the experimental design being balanced in the attributes. This is likely the result of an imbalance in the blocks collected over the sample.

On average, the second alternative is much lower in price than the other two non-no choice alternatives, has less mild reaction indications, and has choice tasks with a lower average number of months until the vaccine becomes available. The third alternative appears to have more vaccines administered using more needles and a higher number of severe reactions, to be more likely to be administered at a doctor's surgery or pharmacy, and to take longer to become available relative to the other vaccine alternatives. Of note is that out of all attributes presented, the efficacy attribute is most similar on average across the three vaccine alternatives. This observation will become important later. Independent of the differences between alternatives however, it is the combination of attribute levels within choice tasks and not the averages over the data that impact on choices.

Table 4: Summary statistics of data

	Alt A	Alt B	Alt C	No choice
Chosen	6,524	6,718	2,748	1,098
Share	38.00%	39.00%	16.00%	6.00%
Average attribute level by alternative				
Price	79.71	63.93	70.82	0.00
Mild reactions (*n* in 10,000)	95.32	77.50	95.69	0.00
Severe reactions (*n* in 10,000)	6.13	8.25	9.25	0.00
Efficacy	91.99	90.73	90.11	0.00
Mode of Administration (Needle)*	-0.50	-0.39	0.22	0.00
Performed at doctor's surgery**	0.11	0.06	0.39	0.00
Performed at pharmacy**	-0.03	-0.01	0.15	0.00
Months till Available	7.47	5.92	7.83	0.00

* Effects coded (base is pill)
** Effects coded (base is performed at hospital)

6.0 Empirical Example

6.1 MNL model outcomes

In this section, we present the results for four sets of models, two multinomial logit (MNL), two mixed multinomial logit (MMNL), and two error components (EC) models. We estimate the MNL and MMNL models as these models represent the two dominant models estimated within the Health Economics literature. Soekhai et al. (2019) report an expanded use of MMNL models to estimate discrete choice experiments (DCE) within the health sphere. Specifically, it was reported that only one out of 34 (2.94 percent) published papers reported using a MMNL model between 1990 and 2000, six out of 114 (5.26 percent) papers between 2001 and 2008, 45 out of 179 (25.14 percent) papers between 2009 and 2012, and finally to 130[1] out of 301 (43.19 percent) papers between 2013 and 2017. The MMNL was the most widely used model in the health economics literature during the 2013 to 2017 period. In this same period, the next most widely used model was the multinomial logit model, being used in 116 of the 301 (representing 38.54 percent). Although not a commonly reported model in the Health Economics literature, the EC model is heavily applied within the Environmental Economics literature to account for likely substitution patterns between non-status quo alternatives in DCEs (see Scarpa et al. 2005). We prefer the EC model to the Nested Logit model used by Campbell and Erdem (2019) insofar as both models capture the same effects, however the EC model allows for the pseudo panel nature of DCEs, whereas the NL model does not. We report the EC model to demonstrate that the effects of preference imbalance may not necessarily be dealt with by models that are primarily developed to deal with the presence of a status quo or no choice alternative.

Table 5 presents the results from two MNL models, one with a constant associated only with the no-choice alternative, and the second where alternative-specific constants are estimated for the three vaccine alternatives. With respect to the first model, the expectation is that constant be negatively signed given that this alternative has the lowest share of choices (i.e., six percent) out of the four alternatives shown in the experiment, and the fact that respondents should prefer to have a vaccine available, all else being equal. As shown in Table 3 however, the sign of the constant term is positive. Given the result obtained, we interpret the constant as meaning that all else being equal, respondents would prefer to not have a vaccine than to have one. To explain this somewhat counter-intuitive result, we note that the average estimated utilities for the first three alternatives over all observations within the data are 5.859, 5.917 and 5.368 respectively, which are computed purely based on the attributes of the design with no constant terms. Given that utility is relative, and the fact that estimation of the parameters within the model will be such that the predicted choice shares will replicate as best as possible the known market shares of the data, the single constant term associated with the final alternative in this case is forced to be positive, against our *a priori* expectations. This later requirement that the predicted shares from the model equal the actual shares in the data is also the reason that the constants in the second model are negative. For this second model, the vaccine alternatives return positive utilities on average, producing average utilities of 1.139, 1.326 and 0.589. Independent of the reason why, the strict interpretation of the constants for this second model is that after accounting for the design attributes, the sample population would once again prefer not to have a vaccine.

1 Soekhai et al. (2019) separate this into 118 papers reporting using MMNL and 12 GMNL models. Following Hess and Rose (2012) and Hess and Train (2017), we do not view the two as being different models and consider the GMNL model to be a specific functional form of the MMNL model.

Comparing the parameter estimates of the two models, whilst the signs of the coefficients from two models are the same, the magnitudes are very different. To demonstrate how big these differences are in practical terms, we present the marginal willingness to pay (WTP) estimates for the two models at the base of the table. As shown in the table, the WTP estimates for the second model are almost half those obtained from the first model, being anywhere between $10 and $20 different for all but the mild reactions attribute. We note however that these WTP estimates are statistically significantly different only for the efficacy attribute. As noted previously, the efficacy attribute is the attribute that is most similar on average across the three vaccine alternatives within the data (see Table 2). In any case, to highlight concerns related to attempts to meaningfully interpret the constants from DCEs, the first model suggests that the sample is willing to pay $1,297.75 to avoid having a vaccine, all else being equal. Results from the second model suggest that people are willing to pay in the vicinity of $650 to avoid having a vaccine, all else being equal.

Table 5: MNL model results

	Par.	(rob. t-rat.)	Par.	(rob. t-rat.)
Model constants				
ASC no choice	4.592	(15.64)	-	-
ASC Alt A	-	-	-3.298	(-10.50)
ASC Alt B	-	-	-3.459	(-11.36)
ASC Alt C	-	-	-3.672	(-11.95)
Attributes				
Price	-0.004	(-7.21)	-0.005	(-10.66)
Mild reactions (*n* in 10,000)	-0.002	(-7.86)	-0.001	(-6.26)
Severe reactions (*n* in 10,000)	-0.063	(-24.35)	-0.058	(-22.01)
Efficacy	0.084	(27.70)	0.071	(22.59)
Mode of Administration (Needle)*	-0.164	(-11.77)	-0.150	(-10.48)
Performed at doctor's surgery**	-0.101	(-6.25)	-0.081	(-5.02)
Performed at pharmacy**	0.168	(9.19)	0.152	(8.26)
Months till Available	-0.144	(-24.11)	-0.136	(-21.76)
Model fit				
LL(0)	-23688.998		-23688.998	
LL(β)	-13358.696		-13284.862	
ρ^2	0.436		0.439	
Adj ρ^2	0.564		0.561	
AIC	26735.392		26591.724	
BIC	26805.107		26676.931	
N	2136		2136	
K	9		11	
Willingness to Pay estimates				
	WTP	**95% con. int.**	**WTP**	**95% con. int.**
ASC no choice	-$1,297.75	(-$1717.01 - -$878.49)	-	-
ASC Alt A	-	-	$617.19	($439.16 - $796.13)
ASC Alt B	-	-	$647.33	($467.25 - $827.93)
ASC Alt B	-	-	$687.28	($501.4 - $872.39)
Mild reactions (*n* in 10,000)	$0.44	($0.27 - $0.61)	$0.24	($0.15 - $0.33)
Severe reactions (*n* in 10,000)	$17.93	(-$12.92- $22.93)	$10.84	($8.62 - $13.06)
Efficacy	-$23.62	(-$30.5 - -$16.74)	-$13.32	(-$16.13 - -$10.5)
Mode of Administration (Needle)*	$46.30	($30.39 - $62.21)	$28.05	($20.14 - $35.96)
Performed at doctor's surgery**	$28.61	($15.97 - $41.24)	$15.09	($8.20 - $21.98)
Performed at pharmacy**	-$47.39	(-$63.45 - -$31.32)	-$28.53	(-$36.46 - -$20.59)
Months till Available	$40.65	($27.60 - $53.70)	$25.43	($19.21 - $31.65)

* Effects coded (base is pill)
** Effects coded (base is performed at hospital)

6.2 MMNL and EC model outcomes

Table 6 presents the results of four models, two Mixed Multinomial Logit (MMNL) models and two Error Components (EC) estimated on the same data as previously reported. Both MMNL models assume normal distributions for the non-cost attributes and constants, and log-normals for the cost attributes. All four models were estimated in Pythonbiogeme [1] using 2000 MLHS draws. The first MMNL and EC models assume only a single constant for the no-choice alternative whilst the second model allows for three alternative specific constants. With respect to the MMNL models, examining the means of the constant terms only, it is noticeable that the signs of the constants are now the reverse of those reported earlier. Indeed, the signs of constants from both models now conform to our *a priori* expectations. We note however that for the first model, the standard deviation parameter for the constant is excessively large, being twice the magnitude of the mean estimate. As such, we conclude that the model under this specification with non-uniform choice shares may trade-off the mean estimate of the constant for increased heterogeneity. Whilst the WTP estimates (confidence intervals are computed for the mean estimates only using the delta method) obtained between the two MMNL model specifications are much more similar, this has come at the cost of increased heterogeneity for the constant. In this way, the model is able to reproduce the choice shares however the analyst may conclude that there exists significant preference heterogeneity where very little really exists, as per the second model results.

Table 6: MMNL model results

		MMNL 1		MMNL 2		EC 1		EC 2	
	Moment	**Par.**	**(rob. t-rat.)**	**Par.**	**(rob. t-rat.)**	**Par.**	**(rob. t-rat.)**	**Par.**	**(rob. t-rat.)**
Model constants									
ASC no choice	Mean Std Dev.	-285 4.34	(-3.03) (8.04)	- -	- -	-0.033 -	(0.07) -	- -	- -
ASC Alt A	Mean Std Dev.	- -	- -	2.61 0.559	(3.41) (6.65)	- -	- -	1.250 -	(2.66) -
ASC Alt B	Mean Std Dev.	- -	- -	2.44 0.284	(3.22) (1.29)	- -	- -	1.080 -	(2.33) -
ASC Alt C	Mean Std Dev.	- -	- -	2.34 0.686	(3.07) (8.18)	- -	- -	0.921 -	(1.98) -
Attributes									
Price	Mean Std Dev.	-5.700 1.840	(-30.39) (29.05)	-5.650 1.850	(-29.17 (23.54	-0.005 -	(-8.93) -	-0.006 -	(-11.50) -
Mild reactions (*n* in 10,000)	Mean Std Dev.	-0.003 0.005	(-10.31) (10.31)	-0.003 0.006	(-9.40) (10.44)	-0.002 -	(-10.54) -	-0.002 -	(-9.28) -
Severe reactions (*n* in 10,000)	Mean Std Dev.	-0.105 0.095	(-21.71) (17.98)	-0.107 0.102	(-20.29) (18.12)	-0.067 -	(-23.73) -	-0.062 -	(-21.20) -
Efficacy	Mean Std Dev.	0.126 0.088	(23.93) (12.78)	0.121 0.089	(20.09 (15.70)	0.086 -	(26.55) -	0.074 -	(21.76) -
Mode of Administration (needle)	Mean Std Dev.	-0.165 0.198	(-7.19) (2.18)	-0.181 0.236	(-7.70) (3.82)	-0.119 -	(-7.71) -	-0.109 -	(-7.00) -
Performed at doctor's surgery**	Mean Std Dev.	-0.068 0.419	(-2.74) (9.25)	-0.066 0.391	(-2.57) (7.35	-0.107 -	(-6.14) -	-0.085 -	(-4.94) -
Performed at pharmacy**	Mean Std Dev.	0.208 0.319	(7.84) (4.89)	0.219 0.347	(7.61) (5.15)	0.142 -	(7.46) -	0.126 -	(6.47) -
Months till available	Mean Std Dev.	-0.255 0.236	(-21.13) (23.15	-0.278 0.249	(-21.15) (21.72)	-0.144 -	(-22.66) -	-0.137 -	(-20.54) -
Error component									
Alt A, B and C		-	-	-	-	4.470	(16.16)	4.700	(16.4)
Model fit									
LL (0)		-23688.998		-23688.998		-23688.998		-23688.998	
LL (β)		-10,512.352		-10,449.106		-11,562.246		-11,506.667	
ρ2		0.556		0.559		0.512		0.514	
Adj ρ2		0.439		0.435		0.486		0.483	
AIC		21060.704		20942.212		23144.492		23037.334	
BIC		21162.704		21066.879		23201.159		23105.334	
N		2136		2136		2136		2136	
K		18		22		10		12	

	MMNL 1		MMNL 2		EC 1		EC 2	
	Par.	(rob. t-rat.)	Par.	(rob. t-rat.)	Par.	(rob. t-rat.)	Par.	(rob. t-rat.)
Willingness to Pay								
ASC no choice	$156.73 ($31.25 - $517.25)		- -		$6.81 (-$174.25 - $187.88)		- -	
ASC Alt A	- -		-134.03 (-$490.25 - -$23.99)		- -		-$195.38 (-$339.17 - -$51.60)	
ASC Alt B	- -		-125.3 (-$463.65 - -$20.46)		- -		-$168.12 (-$309.19 - -$27.05)	
ASC Alt C	- -		-120.17 (-$447.66 - $17.84)		- -		-$143.86 (-$285.65 - -$2.07)	
Mild reactions (*n* in 10,000)	$0.18 (-$2.45 - $3.20)		$0.17 (-$2.45 - $3.22)		$0.17 (-$2.45 - $3.22)		$0.33 ($0.24 - $0.42)	
Severe reactions (*n* in 10,000)	$5.77 (-$1.68 - $23.98)		$5.49 (-$1.78 - $24.43)		$13.67 ($10.64 - $16.71)		$9.68 ($8.11 - $11.25)	
Efficacy	($6.93) (-$27.28 - $0.80)		($6.21) (-$26.66 - $1.48)		-$17.64 (-$21.95 - -$13.34)		-$11.61 (-$14.03 - -$9.19)	
Mode of administration (needle)	$9.07 (-$8.15 - $43.77)		$9.30 (-$6.71 - $45.44)		$24.22 (15.016 - $33.42)		$17.07 ($13.87 - $20.30)	
Performed at doctor's surgery**	$3.72 (-$17.73 - $32.37)		$3.37 (-$17.71 - $31.82)		$21.80 ($12.97 - $30.64)		$13.28 ($7.26 - $19.30)	
Performed at pharmacy**	($11.44) (-$50.68 - $6.16)		($11.25) (-$52.24 - $5.92)		-$29.10 (-$39.07 - -$19.13)		-$19.70 (-$22.88 - -$16.51)	
Months till available	$14.02 ($1.41 - $51.65)		$14.28 (-$1.66 - $54.92)		$29.35 ($21.38 - $37.32)		$21.46 ($16.44 - $26.49)	

The last two models nest an error component associated with the non-nested logit model. These error components systematically vary jointly the utility functions of these alternatives, producing correlated utilities that imply that respondents are more likely to trade between these alternatives than between these alternatives and the status quo option. This type of model appears not to be widely used within the health economics literature, rather being used in environmental economics to account for status quo effects in unlabelled DCEs (see Scarpa et al. 2005). Examining the results obtained from the two EC models show that attempting to account for status quo effects need not solve issues related to preference imbalance between alternatives when such preference imbalance exists between the non-status quo alternatives. In both EC models, the error components are statistically significant suggesting that there does indeed exist substation effects between the non-status quo alternatives present within the data. In the first model, with a single status quo constant, the constant is no longer statistically significant, whilst the ASCs of the second model are statistically significant and positive. Of particular interest are the WTP estimates which are markedly different to those obtained from the MNL and MMNL models, suggesting that the error components are accounting for effects not dealt with by either of these models. However, the comparing the WTP estimates obtained by the two EC models suggest differences, which underscores our argument that preference imbalance can exist in unlabelled DCEs between non-status quo alternatives which only the inclusion of ASCs can account for.

7.0 Consequences

In a recent paper, [3] demonstrated the ability of DCEs to predict real world outcomes. Whilst we do not argue against the overall findings of this paper, we do suggest that caution should be given to how this result should be interpreted. As indicated above, econometrically, the constants of discrete choice models ensure that the predicted choice shares match as closely as possible the actual market shares within the data, after accounting for the role of the design attributes. Given that DCEs involve the creation of multiple hypothetical and non-existent markets, it is questionable why the market shares from data collected from such experiments would be similar to those of real markets. To demonstrate, consider three choice tasks contained within the data. Table 7 presents the actual data for the first respondent for choice tasks 1 and 5, and for the second respondent for choice task 5 only. The base of the table, in the absence of a real-world vaccine, we create a single hypothetical vaccine to represent a revealed preference alternative. The final two columns of the table present the choice probabilities derived for each choice task based on the two MNL models presented in Table 7.

It is evident that the individual choice tasks represent markets defined by different product offerings, none of which matches the real-world market and that the estimated choice probabilities do not relate to those obtained from the revealed preference data. Further, even if one choice task in the experiment is designed to match real world market offerings, as per [3], discrete choice models are estimated on all observations within the data, the majority of which will not match real-world market conditions.

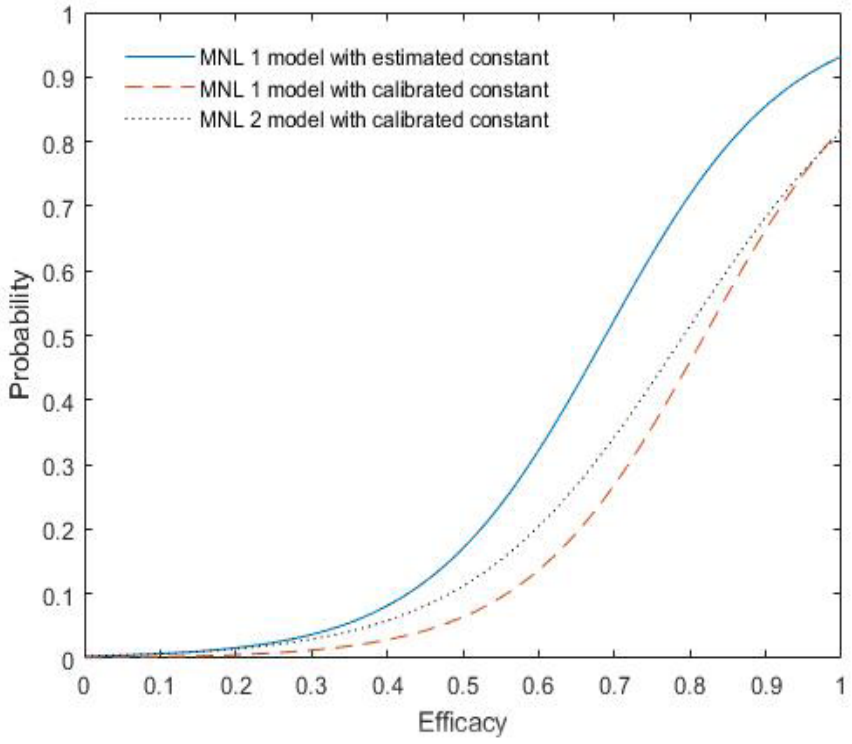
Table 7: Empirical choice data

Resp	Set	Alt	ASC1	ASC2	ASC3	No	Mld	Sev.	Eff.	Ndle	Doc.	Pharm.	Avail.	Pr	Pr (MNL1)	Pr (MNL2)
1	1	1	1	0	0	0	10	2	84	-1	-1	-1	14	100	0.05	0.06
1	1	2	0	1	0	0	20	10	99	-1	1	0	2	0	0.83	0.83
1	1	3	0	0	1	0	200	2	84	1	0	1	10	80	0.07	0.06
1	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0.05
1	5	1	1	0	0	0	20	2	94	-1	1	0	4	40	0.38	0.43
1	5	2	0	1	0	0	10	1	94	-1	0	1	4	40	0.54	0.50
1	5	3	0	0	1	0	100	2	89	1	0	1	12	140	0.05	0.03
1	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.03
2	5	1	1	0	0	0	10	2	99	-1	1	0	2	40	0.84	0.86
2	5	2	0	1	0	0	200	10	89	1	0	1	14	120	0.02	0.02
2	5	3	0	0	1	0	20	1	89	1	1	0	8	60	0.11	0.09
2	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0.03
Pop^n	RP	Vac-cine	0	0	0	0	8	5	97	1	0	1	5	36.3	0.91	0.99
Pop^n	RP	No	0	0	0	1	0	0	0	0	0	0	0	0	0.09	0.01

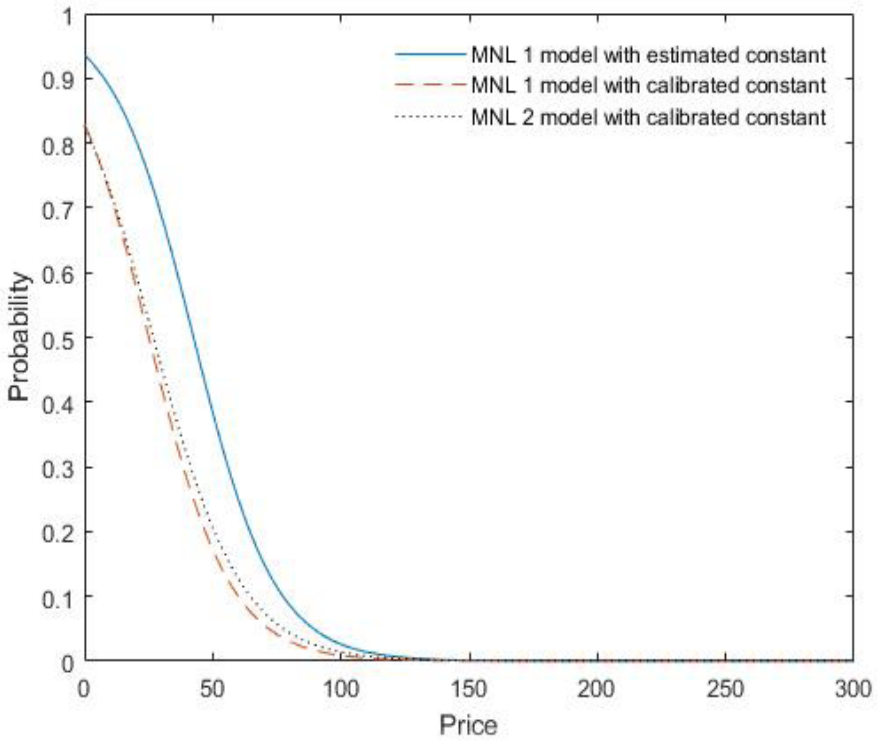
In addition, we argue that if the objective of a study is prediction, it will generally be possible to calibrate a model to reproduce existing market shares, using procedures such as those as outlined by [8]. Putting aside any requirement to also calibrate the non-constant parameters to match those obtained from revealed preference data sources, assume that the real share of vaccine uptake is 0.78. Applying the parameters from the first model to the RP data reported in Table 4 but changing the constant to 5.6901 reproduces this '*real*' market share. Similarly, applying the estimates from the second model but changing the constant to 4.4637 will result in a predicted uptake for the vaccine of 0.78.

As noted, it is generally possible to predict known market shares simply by calibrating the model (possibly including calibrating the non-constant terms of the model at the same time). It is for this reason that we suggest care be taken in interpreting the results reported by [3]. Indeed, we argue that the true test of how well the outputs of DCEs perform is how well the model predicts outcomes given changes to the market attributes.

To demonstrate, consider the two calibrated models applied to our hypothetical real market choice task. Both models predict a 78 percent uptake of the vaccine given the attribute levels assumed. Changing the efficacy attribute from 97 percent to 90 percent however leads to a predicted 66.39 precent vaccine uptake based on the first model, and a 68.30 percent uptake based on the second model. Which model predicts better given the changing market conditions is the more relevant test, given that both models predict the same outcomes given the base scenario. To highlight this point further, Figure 2 presents three sigmoidal curves obtained from the two MNL models reported in Table 5, using the original constant obtained from MNL 1, the first MNL with the new calibrated constant, and MNL 2 with the new calibrated model. Figure 2(a) represents the sigmoidal probability curve computed by changing the Efficacy attribute level, holding all other attributes constant at the RP levels assumed in Table 7. Figure 2(b) represents the sigmoidal probability curve for price derived in the same manner. As can be seen, the sigmoidal curves for the uncalibrated constants are extremely different than for those for the models with calibrated constants suggesting that failure to calibrate the constant terms may result in different predictions to those based on calibrated models. Further, as can be seen in Figure 2(a), despite predicting the same percent of vaccine uptake at the calibrated forecast level, significant differences in predictions can occur as one changes attribute levels away from this initial forecast amount. Again, this highlights the importance of constants in forecasting, and demonstrates that if properly calibrated, the importance of forecasting is not in predicting the initial forecast market share, but rather, in how the model performs given changes from the initially assumed attribute levels.



(a) **Sigmodal curve for efficacy**



(b) **Sigmodal curve for price**

8.0 Discussion and conclusions

In this paper, we argue that many papers within the health economics literature dealing with unlabelled alternatives, particularly those that include a no-choice alternative, assume utility functions with either no constants or only a single constant. We demonstrate that in cases where the choice shares of the non-status quo alternative are non-uniformly spread, such a specification can result in biased parameter estimates, including for the non-constant attributes of the model.

We therefore recommend that all discrete choice models include constants for all but one alternative, even if said constants are not statistically significantly different from one another. This recommendation is not limited to DCEs that include a status quo or no choice alternative, but all models estimated using unlabelled DCEs. Indeed, this recommendation extends to models estimated on both stated preference and revealed preference data, as the issues discussed herein extend to both data types. In making this recommendation, we note that given utility is relative, it should not matter which alternatives these alternative specific constant be associated with. We further recommend that, space permitting, research papers report on the descriptive statistics of the data, and not just discuss the experimental design used to generate the data. In many cases, such as those reported herein, the design properties will not translate to the data set. Given that models are estimated on the final data, and not on the design, it stands to reason that more time should be devoted to reporting on the data, similar to Table 4 presented within this paper. In doing so, we hope that researchers will provide more detailed commentary of the outputs of discrete choice experiments, particularly with respect to the estimated constants.

As evidenced through an empirical data set, the constants obtained from discrete choice models reflect the market shares of the hypothetical choice sets defined from the experiment, which in many cases may be meaningless in terms of any attempt to provide real world behavioural meaning to them, without reference to the descriptive statistics of the data.

We also note that there is an increasing interest in testing whether the results obtained from DCEs are externally valid or not. We argue that how such questions are examined be carefully considered within the future. As noted, the outputs of any discrete choice model can be calibrated to forecast existing market shares, whereas the real test of model performance with respect to prediction, is how well the model predicts, at least in the aggregate, after changes to the market occur. In this regard, understanding how well models forecast is likely to much more difficult, given that it will be necessary to observe real changes to the levels of the attributes of the alternatives existing with real markets over time, and not just observing the attribute levels at one point of time.

Acknowledgements. This project was undertaken as part of an Australian Research Council grant examining issues related to the external validity of discrete choice experiments (ARC DP180103718). We would also like to acknowledge support from Swiss National Foundation.

Author contributions. All of the authors contributed to the conception and planning of the study, the analysis and/or interpretation of the data; drafting and critical revision of the manuscript; and approved the version of the manuscript submitted for publication.

Compliance with Ethical Standards

Funding. The third author was funded to visit the University of Technology Sydney by the Swiss National Foundation (P2TIP1_187933). The overall study was conducted as part of an Australian Research Council Discovery Project grand scheme (ARC DP180103718).

Conflict of interest. There are no conflicts of interest to disclose.

Data availability statement. All data and codes associated with this this study are available from the authors upon request.

References

Allanson, P.F., Brown, E.A., Kopasker, D. and Kwiatkowski, A. (2020) An Investigation of the Stability of Patients' Treatment Preferences Over the Course of a Clinical Trial, Value in Health, 23, 622-629.

Chen, T.-T., Tung, T.-H., Hsueh, Y.-S., Tsai, M.-H., Liang, H.-M., Li, K.-L., Chung, K.-P. and Tang, C.-H. (2015) Measuring Preferences for a Diabetes Pay-for-Performance for Patient (P4P4P) Program using a Discrete Choice Experiment, Value in Health, 18, 578-586.

Doiren, D. and Yoo, H.I. (2017) Temporal Stability of Stated Preferences: The Case of Junior Nursing Jobs, Health Economics, 25, 802-809.

Dong, D., Ozdemir, S., Mong Bee, Y., Toh, S.-A., Bilger, M. and Finkelstein, E. (2016) Measuring High-Risk Patients' Preferences for Pharmacogenetic Testing to Reduce Severe Adverse Drug Reaction: A Discrete Choice Experiment, Value in Health, 19, 767-775.

Esther W de Bekker-Grob, E.W., Swait, J.D., Kassahun, H.T., Bliemer, M.C.J, Jonker, M.F., Veldwijk, J., Cong, K., Rose, J.M., Donkers, B. (2019) Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models, Value in Health, 22(9), 1050-1062.

Flynn, T.N., Bilger, M., Malhotra, C. and Finkelstein, E.A. (2016) Are Efficient Designs Used in Discrete Choice Experiments Too Difficult for Some Respondents? A Case Study Eliciting Preferences for End-of-Life Care, PharmacoEconomics,34, 273-284.

Heidenreich, S., Watson, V., Ryan, M. and Phimister, E. (2018) Decision heuristic or preference? Attribute non-attendance in discrete choice problems, Health Economics, 27, 157-171.

Hole, A.R. (2011) A discrete choice model with endogenous attribute attendance, Economics Letters, 110, 203–205.

Holte, J., Sivey, P. Abelsena, B. and Olsen, J.A. (2016) Modelling Nonlinearities and Reference Dependence in General Practitioners' Income Preferences, Health Economics, 25, 1020-1038.

Krucien, N., PelletierFleury, N. and Gafni, A. (2019a) Measuring Public Preferences for Health Outcomes and Expenditures in a Context of Healthcare Resource Re-Allocation, PharmacoEconomics, 37, 407-417.

Krucien, N., Sicsic, J. and Ryan, M. (2019b) For better or worse? Investigating the validity of best–worst discrete choice experiments in health, Health Economics, 28, 272-586.

Lancsar, E., Fiebig, D.G. and Hole, A.R. (2017) Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software, PharmacoEconomics, 35, 697-716.

Lim, S., Jonker, M.F., Oppe, M., Donkers, B. and Stolk, E. (2018) SeverityStratified Discrete Choice Experiment Designs for Health State Evaluations, PharmacoEconomics, 36, 1377-1389.

Marshall, D.A., Gonzalez, J.M., MacDonald, K.V. and Johnson, F.R. (2017) Estimating Preferences for Complex Health Technologies: Lessons Learned and Implications for Personalized Medicine, Value in Health, 20, 32-39.

Mohammadi, T. Bansback, N., Marra, F., Kahkban, A., Campbell, J.R., FitzGerald, J.M., Lynd, L.D. and Marra, C.A. (2017) Testing the External Validity of a Discrete Choice Experiment Method: An Application to Latent Tuberculosis Infection Treatment, Value in Health, 20, 969-975.

Mühlbacher, A, Johnson, F.R., Yang, J.-H., Happich, M. and Belger, M. (2016) Do You Want to Hear the Bad News? The Value of Diagnostic Tests for Alzheimer's Disease, Value in Health, 19, 66-74.

Mulhern, B., Norman, R., Lorgelly, P., Lancsar, E., Ratcliffe, J., Brazier, J. and Viney, R. (2017) Dimension Order Important when Valuing Health States Using Discrete Choice Experiments Including Duration? PharmacoEconomics, 35, 439-451.

Norm, R. Moorin, R., Maxwell, S., Robinson, S. and Brims, F. (2020) Public Attitudes on Lung Cancer Screening and Radiation Risk: A Best-Worst Experiment, Value in Health, 23(4), 495-505.

Ostermann, J., Mühlbacher, A., Brown, D.S., Regier, D.A., Hobbie, A., Weinhold, A., Alshareef, N., Derrick, C. and Thielman, N.M. (2020) Heterogeneous Patient Preferences for Modern Antiretroviral Therapy: Results of a Discrete Choice Experiment, Value in Health, 23(7), 851-861.

Quaife, M., Vickerman, P. Manian, S., Eakle, R., Cabrera-Escobar, M.A., Delany-Moretlwe, S. and Terris-Prestholt, F. (2018) The effect of HIV prevention products on incentives to supply condomless commercial sex among female sex workers in South Africa, Health Economics, 27, 1550-1556.

Ryan, M., Krucien, N. and Hermens, F. (2018) The eyes have it: Using eye tracking to inform information processing strategies in multi-attributes choices, Health Economics, 27, 709-721.

Ryan, M., Watson, V. and Entwistle, V. (2009) Rationalising the 'irrational': a think aloud study of discrete choice experiment responses, Health Economics, 18, 321–336.

Sándor, Z. and Wedel, M. (2005) Heterogeneous conjoint choice designs, Journal of Marketing Research, 42(2), 210–218.

Sicsic, J., Pelletier-Fleury, N. and Moumji, N. (2018) Women's Benefits and Harms Trade-Offs in Breast Cancer Screening: Results from a Discrete-Choice Experiment, Value in Health, 21, 78-88.

van de Wetering, L., van Exel, J., Bobinac, A. and Brouwer, W.B.F. (2015) Valuing QALYs in Relation to Equity Considerations Using a Discrete Choice Experiment, PharmacoEconomics, 33, 1289-1300.

van Dijk, J.D., Groothuis-Oudshoorn, C.G.M., Marshall, D.A. and IJzerman, M.J. (2016) An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery, Value in Health, 19, 316-322.

Vass, C., Rigby, D. and Payne, K. (2018) Investigating the Heterogeneity in Women's Preferences for Breast Screening: Does the Communication of Risk Matter? Value in Health, 21, 219-228.

Veldwijk, J., Essers, B.A.B., Lambooij, M.S., Dirksen, C.D., Smit, H.A. and, de Wit, G.A. (2016) Survival or Mortality: Does Risk Attribute Framing Influence Decision-Making Behavior in a Discrete Choice Experiment? Value in Health, 19, 202-209.

Wong, C.K.H, Man, K.K.C., Ip, P., Kwan, M. and McGhee, S.M. (2018) Mothers' Preferences and Willingness to Pay for Human Papillomavaris Vaccination for Their Daughters: A Discrete Choice Experiment in Hong Kong, Value in Health, 21, 622-629.

Wright, S. Ulph, F., Dharni, N. and Payne, K. (2017) Eliciting Preferences for Information Provision in Newborn Bloodspot Screening Programs, Value in Health, 20, 651-661.