Received: 23 January 2020 | Revised: 17 December 2020 | Accepted: 22 January 2021

DOI: 10.1002/hec.4246

RESEARCH ARTICLE

Health Economics WILEY

Stated versus revealed preferences: An approach to reduce bias

Kaat de Corte | John Cairns | Richard Grieve

Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

Correspondence
Richard Grieve, Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, WC1H 9SH London, UK.
Email: richard.grieve@lshtm.ac.uk

Funding information
Health Services and Delivery Research Programme, Grant/Award Number: 13/54/62

Abstract

Stated preference (SP) survey responses may not predict actual behavior, leading to hypothetical bias. We developed an approach that harnesses large-scale routine data to help SP surveys provide more accurate estimates of revealed preferences (RPs), within a study which elicited preferences for alternative changes to the blood service in England. The SP survey responses were used to predict the mean number of annual whole blood donations. *Ex ante*, the iterative survey design estimated hypothetical bias by contrasting pilot SP survey responses ($N = 1254$), with individually linked data on RPs, to inform the main SP survey design ($N = 25{,}187$). *Ex post*, the analysis recognized mediation of the relationship between SP and RP when blood donation is deferred. The pilot survey reported that donors' intended donation frequency of 3.2 (men) and 2.6 (women) times per year, exceeded their actual frequency by 41% and 30% respectively. Choice scenario attributes for the main SP survey were then modified, and over-prediction subsequently decreased to 34% for men and 16% for women. The mediating effect of deferrals explained 29% (men) and 86% (women) of the residual discrepancy between SP and RP. Future studies can use this approach to reduce hypothetical bias, and provide more accurate predictions for decision-making.

KEYWORDS
causal inference, choice experiments, external validity, hypothetical bias, revealed preferences, stated preferences

1 | INTRODUCTION

1.1 | Preferences: stated versus revealed

Stated preference (SP) methods are widely used in health economics, for example to estimate the relative value of alternative service changes (DeBekker-Grob, et al., 2012), and to elicit the willingness to pay for health gains. Progress has been made in the development of innovative SP designs (Street, et al., 2001), and analytical methods that respect the structure of SP data (Louviere, 2006). However, a barrier to using SP results in decision-making is that the preferences stated may not predict actual behavior (Quaife et al., 2018; Viney et al., 2002). The general literature comparing stated and revealed preferences (RP) has found that individuals tend to overstate their valuation of a particular good, service or outcome, which can lead to misleading estimates of relative value (Fifer et al., 2014). In the health context, the opportunities for contrasting SP and RP have been limited, with little attention given to understanding the reasons for any differences (Ryan et al., 2010).

Previous comparisons of SP and RP have tried to assess whether SP surveys represent underlying preferences, that is, whether the results have "external validity" (Lancsar & Swait, 2014; Quaife et al., 2018; Ryan & Gerard, 2003) or "predictive validity" (Whitehead, 2005). The bias from individuals overstating their preferences for a particular set of choices has also been termed "hypothetical bias" (Johansson-Stenman & Svedsater, 2012; Loomis, 2011). Good survey design aims to capture the choice context and its inherent restrictions as well as possible, but there is no consensus on how to minimize hypothetical bias, nor on whether some level of bias is inevitable. In the contingent valuation literature, meta-analyses of the discrepancies between SP and RP reported that individuals overstate their preferences in hypothetical settings, and that the magnitude of this discrepancy tends to be higher for publicly funded goods (List & Gallet, 2001; Murphy, et al., 2005). Murphy et al. (2005) found that this discrepancy was smaller for studies which sampled RP and SP for the same individuals (within-subject design), versus across different samples (between-subject design) where the potential for hidden bias is greater due to unobserved differences across individuals.

Outside health economics, several studies have postulated reasons for hypothetical bias. Johnson-Stenman and Svedsater (Johansson-Stenman & Svedsater, 2012) drew on evidence from psychology and behavioral economics, which suggests that people have a positive self-image, and an incentive to overstate their preferences, particularly for what has been termed "moral goods". Other studies have suggested that people derive positive utility from expressing attitudes that show social responsibility (Taylor & Brown, 1994), especially when this does not imply actions that are binding (Kahneman & Knetsch, 1992).

Ryan et al. (2010) and Lancsar and Swait (2014) highlight that little attention has been given to examining hypothetical bias in preference modeling within health economics, and that further research to improve predictions from SP surveys is needed. While previously an important barrier was the lack of RP data in many health settings, the advent of large-scale electronic datasets opens up the possibility of using RP data to inform the design and analysis of SP surveys, so as to reduce hypothetical bias. However, as Lancsar and Swait recognize, access to RP data is insufficient, a conceptual framework is required that draws from adjacent literatures and extends to the design and analysis of choice experiments.

1.2 | Proposed conceptual framework

We propose an approach for reducing hypothetical bias in the prediction of RP from SP. Our causal framework combines the econometric perspective of Morikawa's model of consumer behavior in the transport literature, with the psychological perspective of Ajzen's Theory of Planned Behavior (Ajzen, et al., 2004; Morikawa, 1989). Morikawa built upon the rich history of preference research in transport economics and treats SP and RP as two separate manifestations of a latent concept of "real" preferences. This framework recognizes that market behavior is influenced by situational constraints, in that for SP to provide accurate predictions they should be adjusted for those contextual factors anticipated to influence that behavior. Ajzen's theoretical approach characterizes human behavior, and has been used extensively in the blood donation literature to analyze motivations, barriers, and intentions (France, et al., 2007; France et al., 2014; Masser, et al., 2009; Robinson, et al., 2008; Veldhuizen, et al., 2011). The model approaches the concept of hypothetical bias from a psychological perspective. This theory of planned behavior characterizes the relationship between intention and behavior as mediated by "actual behavioral control" which is analogous to the relationship between SP and RP in Morikawa's framework.

Our approach will draw on these conceptual insights by adjusting for constraints on preference expression when predicting behavior directly from SP responses. These constraints are distinct from the choice formation, evaluation, and selection of attributes that should be captured within the design of a choice experiment. Our approach to the SP survey design makes full use of the RP data, so that the choice of attributes and levels, recognizes the importance of minimizing the discrepancy between SP and RP.

Our proposed approach follows Lancsar and Swait's general advice to consider process rather than just final outcomes but differs in the way constraints are recognized. Lancsar and Swait conceptualize any discrepancy as originating from within the choice process and therefore predictive of SP (Lancsar & Swait, 2014). By contrast our paper draws from both Morikawa and Ajzen's conceptual models in recognizing that first the SP survey should be designed to try and minimize hypothetical bias, and second that the analysis must recognize that constraints can mediate the relationship between SP and RP.

Figure 1 summarizes the approach and shows that from a causal inference perspective, models of SP and RP can be integrated into one directed acyclic graph (DAG). This approach clarifies that the question of hypothetical bias can be reframed as an identification problem. When trying to understand and consequently predict behavior, the policy relevant parameter, and therefore the *estimand* of interest, is the RP but the available data for the *estimator* is the responders' SP. The insight from the DAG (Figure 1), is that the total effects of SP on RP (behavior) consist of both direct and indirect effects, and that constraints are a mediator. Intuitively, behavior can be affected by constraints, but as the DAG reveals, the possible effect of preferences on constraints introduces a pathway of indirect effects, that as the example will show, can be recognized in the analysis to reduce hypothetical bias (see Section 5.5). Therefore, the direct effect of SP on RP cannot be identified without allowing for constraints. The proposed framework addresses this challenge and rather than attempting to predict RP from SP, we allow for the potential mediation effect of constraints to identify hypothetical bias, and estimate the direct effect of SP on RP.

The aim of this paper is to develop an approach for reducing hypothetical bias in the prediction of RP from SP. The proposed approach has two interlinking strands. First, we develop an iterative survey design, whereby responses to a large SP pilot survey are contrasted with RP data, to provide initial estimates of hypothetical bias. These initial estimates of hypothetical bias can be used to modify the choice of attributes and levels in the final SP survey (ex ante). Second, the approach recognizes that there may be constraints, for example clinical regulations or guidance which stop the individual's actual behavior from reflecting their SPs. Our analysis recognizes that it may be inappropriate to incorporate

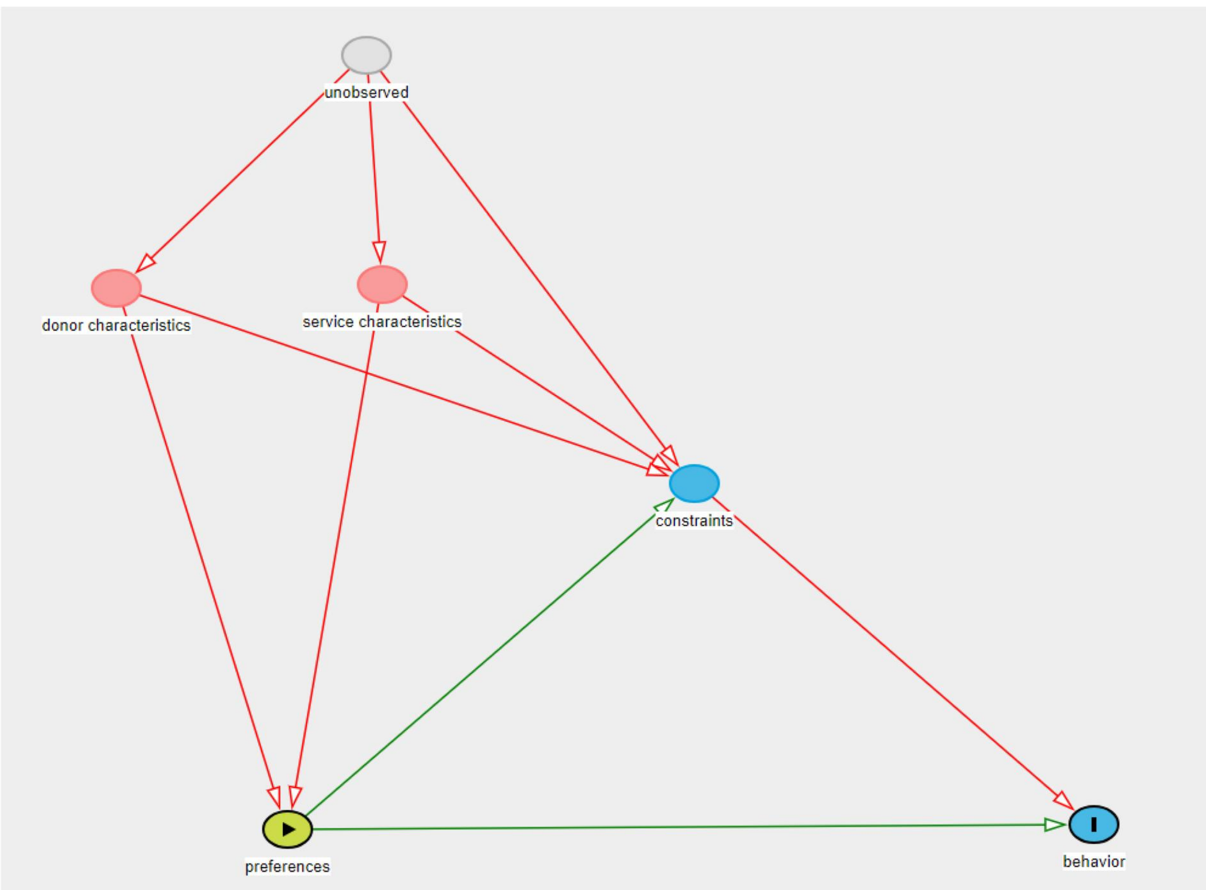

FIGURE 1 Directed acyclic graph (DAG) showing the relationship between preferences, behavior and constraints

such exogenous variables within the choice tasks respondents are required to complete, and we propose instead that they are included as mediators (ex post), when estimating the relationship between SP and RP.

The paper draws on an application of SP modeling from blood donation which is outlined in Section two. Sections three and four describe the SP survey and the source of the RP data. Section five describes the estimation of the discrepancy between SP and RP (hypothetical bias) and is partitioned into a design effect (ex ante) and constraints modeling effect (ex post), with the accompanying results in Section six. Section seven completes the paper with a discussion of the findings, and outlines areas for further research.

2 | THE MOTIVATING EXAMPLE: BLOOD DONATION

2.1 | Overview

The National Health Service Blood and Transplant (NHSBT) organization supplies blood to the English health system at an annual cost of £180 million. NHSBT's strategic aim is to maintain the current supply of whole blood but reduce costs, and this requires evidence about the relative costs and consequences of alternative future service changes. Some of these possible changes, such as providing a health report to donors, have not been implemented in the UK, while others, such as extending the opening hours for blood collection, have been partially adopted. The HEMO (Health Economics MOdeling of blood donation) study recognized the importance of providing timely evidence on the relative preference for alternative service changes including those that had not been implemented even in the research setting, by undertaking a SP survey to estimate the effect of changes to service attributes on the predicted frequency of donation (Grieve et al., 2018). This SP survey was designed to estimate the effect of alternative future changes to the blood service on the frequency with which donors are willing to donate whole blood. For each set of attribute levels, the donor was asked to state the frequency with which they would donate blood ("How many times a year would you give blood?"). The survey included an opt-out ("I would probably not donate"). The form of response variable, and alternative response categories (once per year, twice per year, etc.) was informed by RP data extracted from the PULSE database. More generally, this SP form of survey design was chosen rather than a Discrete Choice Experiment, which could not provide the estimates of donation frequencies required for evaluating the relative value of alternative collection strategies.

The study design recognized that for an altruistic activity such as blood donation, individuals may overstate the frequency with which they are willing to donate blood. This study was therefore designed a priori to contrast preferences for alternative frequencies of blood donation with actual donation activity in the past twelve months so that any predictions can be anchored to RP data. All donors invited to complete the SP survey were randomly sampled from eligible donors within the PULSE database, a registry which includes longitudinal measures about blood donation for all blood donors in England (approximately 1.2 million donors). At any one time approximately 400,000 donors met the eligibility criteria (Grieve et al., 2018). The SP survey design was iterative in that responses to the pilot survey were contrasted with observed donation frequency to provide estimates of hypothetical bias, and the insights from these results, used to modify the design of the main survey. The pilot SP survey had a large sample size ($n = 1211$ survey responders), and the responses contrasted with RP to provide a meaningful estimate of discrepancy (hypothetical bias). These estimates of hypothetical bias were then discussed with donors and NHSBT, and the likely reasons for the bias informed the design of the main SP survey. We now outline the steps in the iterative survey design.

3 | SP SURVEY

3.1 | Survey design: pilot

The choice of attributes for the pilot study was informed by a literature review, preliminary findings from qualitative research with blood donors, and input from policy makers at NHSBT. The literature review found that while blood donation is partly an altruistic act (Rapport & Maggs, 2002, Titmuss, 1970), the frequency with which blood donors donate is likely to depend on the convenience and the opportunity costs of donation (Schreiber et al., 2006), a finding supported by donors and NHSBT.

Drawing on these insights, we included five attributes related to strategies judged to be of policy relevance: travel time, total donation time, opening times of the blood collection venue, provision of a health report, and the maximum number of donations permitted per year (Table 1). The first three attributes influence the cost to the donor of blood donation, whereas the provision of a health report was anticipated to benefit donors (Goette, et al., 2009; Mews, 2013; Mortimer, et al., 2013; Ringwald, 2010). The INTERVAL trial investigated the safety of increasing the maximum frequency of donation (from three to four times per year for women, from four to six for men; Di angelantonio et al., 2017). This attribute is therefore included to understand how donors might respond if the limits were altered. The appropriate levels for each attribute were defined according to summary estimates from the PULSE database, NHSBT market research and consultation with blood donors. The pilot survey included two questions for donors to provide feedback on the survey design, to inform further iterations of the design process, and refinement of the choice tasks.

The survey design was split into two sections to try and capture some aspects of the context not included in the attributes (e.g., familiarity of staff; Lynch & Cohn, 2017). The first section asked donors to think about a donation

TABLE 1 Attributes and levels of the pilot versus main stated preference (SP) surveys

Attribute	Levels	Pilot?	Main survey?
1. Travel time to blood collection venue	10 min shorter than typical travel time	Yes	Yes
	Your typical travel time (only option for LP)		
	15 min longer than typical travel time		
	30 min longer than typical travel time		
2. Health report provided	Yes, after each donation	Yes	Yes
	No		
3. Maximum number of donations per year	3 (females only)	Yes	Yes
	4		
	5 (males only)		
	6 (males only)		
4. Total time to donate	45 min		
	90 min	Yes	No
5. Opening times	9 AM–12 PM and 2–5 PM	No	Yes
	9 AM–5 PM		
	9 AM–8 PM		
	2–8 PM		
6. Appointment availability	Every day: Mon–Sun	No	Yes
	Every weekday: Mon–Fri		
	1 day every 2 months: Mon–Fri		
	1 day every 2 months: Sat or Sun		
7. Opening times and days	Mon–Fri: 9 AM–12 PM and 2–8 PM	Yes	No
	Mon–Fri: 9 AM–5 PM		
	Mon–Fri: 9 AM–12 PM and 2–8 PM		
	Mon–Fri: 9 AM–8 PM		
	Mon–Sun: 9 AM–12 PM and 2–5 PM		
	Mon–Sun: 9 AM–5 PM		
	Mon–Sun: 9 AM–12 PM and 2–8 PM		
	Mon–Sun: 9 AM–8 PM		

TABLE 2 Characteristics of the responders to the pilot and main stated preference (SP) surveys, each compared to the target population in PULSE, each stratified by gender

	N = 774 Males Pilot survey	*N* = 353,763 PULSE Mar 2016	*N* = 15,652 Main survey	*N* = 437 Females Pilot survey	*N* = 427,265 PULSE Mar 2016	*N* = 8329 Main survey
Blood group						
High demand	80 (10.34)	46,998 (13.29)	1551 (9.91)	54 (12.36)	64,950 (15.20)	921 (11.06)
Standard demand	694 (89.66)	306,765 (86.71)	14,101 (90.09)	383 (87.64)	362,315 (84.80)	7408 (88.94)
Ethnicity						
White	735 (94.96)	323,912 (91.56)	14,639 (93.53)	418 (95.65)	400,968 (93.85)	7700 (92.45)
Black/mixed black	3 (0.39)	3518 (0.99)	98 (0.63)	2 (0.46)	4797 (1.12)	103 (1.24)
Asian/mixed Asian	17 (2.20)	12,677 (3.58)	367 (2.34)	12 (2.75)	9050 (2.12)	195 (2.34)
Other or not stated	19 (2.46)	13,656 (3.86)	548 (3.50)	5 (1.14)	12,450 (2.91)	331 (3.97)
Session type						
Static center	51 (6.59)	52,808 (14.93)	1307 (8.35)	31 (7.09)	55,003 (12.87)	746 (9.97)
Mobile session	723 (93.41)	300,955 (85.07)	14,345 (91.65)	406 (92.91)	372,262 (87.13)	7583 (91.04)

Note: *N* (%) unless stated.

opportunity in the same context, defined as: “the last place you gave blood” (LP), and the second section, asked donors to consider donation at a “different place” (DP).

Six SP questions were included in each survey (2 LP and 4 DP questions). The survey design, and all results were stratified by gender to recognize that current guidelines specify different maximum annual donations for men and women.

We considered adopting a full factorial design for the pilot study, but this would have involved 96 possible LP ($1^1 \times 2^2 \times 3^1 \times 8^1$) and 384 DP scenarios ($2^2 \times 3^1 \times 4^1 \times 8^1$), for men, and 64 LP ($1^2 \times 2^3 \times 8^1$) and 256 DP scenarios ($2^3 \times 4^1 \times 8^1$) for women as calculated based on the number of levels and attributes in Table 1. Instead, we adopted an efficient design, using Ngene™ and considering each section of the survey as one choice set compared to an “opt out”, resulting in 24 LP and 24 DP scenarios for men (72 versions of the survey), with 8 LP and 12 DP scenarios for women (12 versions). Two men were invited to complete the survey for every woman, to reflect the greater number of requisite scenarios.

The pilot survey was administered online to 5016 donors. Selected donors were sent an email invitation from NHSBT with a link to the online survey built using FluidSurveys™. Donors who did not complete the survey (except those who refused consent) were re-contacted by email five days later. The survey closed seven days after the reminder email.

The responses to the pilot survey were used in a model that predicted donation frequency at an individual level (see Section 5 and Appendix). These predictions were contrasted with the individual donor's RP, according to their donations reported in the PULSE donor register. We then estimated the magnitude of the discrepancy between SP and RP, overall and according to pre-specified subgroups, and used this as a measure of hypothetical bias. We hypothesized that a large discrepancy would suggest that the choice of attributes and/or levels should be modified for the main survey. We presented the estimates of hypothetical bias to policy-makers at NHSBT and used insights from these, and the open text responses from blood donors to refine the design.

3.2 | Survey design: main survey

Following the results of the pilot study and discussion with NHSBT, the donation time attribute was judged as least relevant for future blood donation policy and was excluded from the main survey. Factor analysis of the discrepancy from the pilot study, as well as thematic analysis of the free text question responses suggested that an appointment availability attribute would be worth including in the main survey (See Table 1). For the main survey, it was judged important to capture interaction effects, and so a full factorial design was adopted. 100,000 donors were invited that met the eligibility criteria.

4 | PULSE: SOURCE OF REVEALED PREFERENCE DATA

4.1 | Extraction and analysis of observational data from PULSE

The PULSE database contains information on the actual frequency of blood donations for all 1.2 million registered blood donors in England, with a median period of follow-up of five years. In line with the study's objective of providing timely evidence to NHSBT, the RP data on the frequency of blood donation was only available up until the time the final SP survey was administered. The database also contains details on the characteristics of the blood donors and of the blood collection centers. Subgroups were defined a priori according to: blood type, ethnicity, mobile or static venue, number of previous donations, and age (see Table 2). In addition, the service configuration experienced by an individual (henceforth described as the "baseline service characteristics") was characterized using variables from PULSE.

This "mapping" of PULSE variables to attributes of the survey facilitates comparison between the observed donation frequency given the baseline characteristics, and the predictions from the SP survey for those who responded.

5 | SP ANALYSIS AND COMPARISON TO RP

5.1 | Empirical model (see also Appendix 1)

The responses to both the pilot and the main survey were used in a general empirical model that estimated the probability of donating at each frequency (once, twice a year, etc.) at the individual level. The model recognized that a donor can express their level of preference for the alternative donation choices, given the attributes and their levels, by choosing the donation frequency to maximize utility. The models included main terms pertaining to the attributes in each SP survey. The model also included terms representing the above donor characteristics. The increased sample size of the main survey and its corresponding higher power, allowed for the inclusion of second order interactions to allow for the potential effect modification of alternative attributes with one another, and of the individual donor's characteristics with each attribute.

Separate models were specified for each gender. The uncertainty around the estimated model coefficients were reported with robust standard errors, to allow for the panel nature of the response data, that is, the potential correlation of the survey responses for each individual.

5.2 | Model estimation: pilot survey

The response data from the pilot SP survey was analyzed using a multinomial logit regression (MNL) model, accepting some loss of information, but maintaining flexibility in handling the ordered categorical responses ("I would probably not donate," "I would donate once per year," "two times per year," etc.; Jones, 2007; Kirkwood & Sterne, 2003). The mlogit command was used to implement the model in Stata™ (Rabe-Hesketh & Skrondal, 2012). Predicted probabilities were obtained for different service configurations for all stated frequencies (see 5.3 for details).

5.3 | Model estimation: main survey

Our approach to model choice followed general guidance for SP models in initially considering alternative models according to a priori reasoning and the form of response data, and then comparing measures of model fit (AIC and BIC) across the range of plausible models (Hauber et al., 2016; Lancsar et al., 2017). The corresponding RP data on actual donation frequency for donors responding to the main survey, was not used to inform model choice. Instead, these data were "held back" to then be used to assess the accuracy of the predictions from the SP survey.

TABLE 3 Main survey measures of model fit used for model selection: Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC)

	AIC			BIC		
	Ordered logistic	**Negative binomial**	**Gamma**	**Ordered logistic**	**Negative binomial**	**Gamma**
Male	316,744	377,799	402,963	318,594	379,602	404,766
Female	133,047	169,529	180,964	134,579	171,034	182,469

While a priori it was anticipated that the response to the SP questions could consider donation frequency as a count variable, and so the negative binomial model would be a potential choice of analytical model we also recognized that the SP question required patients to choose a category of donation frequency (one per year, twice per year etc). We drew on qualitative insights and discussions with donors and NHSBT to postulate that donation frequency would be chosen by donors so as to maximize utility, recognizing potential feelings of altruism and donor identity associated with the categories of donation frequency available to the donors and this encouraged consideration of the ordered logistic regression model (Lynch & Cohn, 2017). For completeness, we also considered a model (Gamma) that regarded the dependent variable as continuous, but bounded at zero, and allowing for a right-skewed distribution. As shown in Table 3, we found that the ordered logistic provided the best fit to the response data from the main survey, and therefore used this model to predict SP. The ologit command was used to implement the model in StataTM.

5.4 | Predicted donation frequency and estimation of hypothetical bias (both surveys)

For the pilot and main surveys, we predicted annual donation frequency, Prob $(Y = \mathrm{k} \mid X_{NHSBT}) = \hat{p_k} \quad = f$ $(X_{NHSBT}\hat{\beta}_{Survey})$, by combining the coefficients estimated by each model ($\hat{\beta}_{Survey}$), with information from the blood donor register about the levels of each blood service attribute.

(X_{NHSBT}). This information pertained to each donor at their most recent blood donation appointment in the 0-12 months preceding receipt of the respective survey. We then combined the predicted probability of each category of donation frequency, with the corresponding level (once, twice a year, etc.), to predict the expected annual frequency of blood donation for that individual as $E(frequency) = \hat{F}_i = \sum \hat{p}_k Q_{k)}$.

For both surveys, we compared each individual's predicted annual donation frequency (predictions from the SP model), to each person's donation frequency observed over the preceding 12-month period (the RP data). We reported the magnitude of the discrepancy between each individual's predicted and observed donation frequency as the estimated hypothetical bias, $\mathrm{HB} = (F_i - \hat{F}_i)$.

For both surveys, we estimated the mean hypothetical bias overall, and according to pre-specified subgroups. For both surveys, we estimated the mean hypothetical bias overall, and according to pre-specified subgroups. We then reported uncertainty intervals around the predicted mean levels of hypothetical bias, reporting 95% Confidence Intervals (CI) to reflect the sampling variation in these predictions.

5.5 | Estimates of hypothetical bias during iterative design (ex ante)

We reported the mean (95% CI) discrepancy between predictions from the pilot SP survey versus the observed data. We did not calibrate the predictions of the pilot study, but instead we discussed with NHSBT and blood donors the potential reasons for the estimated discrepancy and modified the design of the main SP survey accordingly.

5.6 | Estimates of hypothetical bias after allowing for constraints (ex post)

For the main SP survey, we reported the mean (95% CI) of the predictions versus the observed data. Informed by the DAG (Figure 1), we recognized constraints (mediators) in estimating the magnitude of the discrepancy between RP and SP. Within the context of blood collection, this implies recognizing that attending a blood donation appointment does not always lead to a successful donation. Following a health screening that precedes every intended donation, the donor may have their blood donation deferred, most commonly because of low hemoglobin levels in their blood. Other reasons could include travel to certain countries, a recent tattoo, or a health condition. These constraints are not captured in the choice scenario attributes, and so any discrepancy between SP and RP could be due to intended donations that were deferred rather than hypothetical bias. Our approach adjusts for the effects of this exogenous factor[1]. A model of deferrals was used to adjust the probability of donation as predicted by the SP model. The deferral model was based on the INTERVAL trial data as this included deferred appointments and adjusted for inter-donation time. The probability of deferral was predicted based on logistic regression for grouped data using data on the number of deferrals and attendances, adjusting for personal characteristics (age, demand blood group, ethnicity, donor experience, and fixed or mobile center). Separate models were estimated for men and women using the 12-week inter-donation interval group as baseline for men, and 16 weeks for women as is current practice. We found that, once we adjusted for whether or not the donor had to defer their blood donation, the magnitude of the difference between the predicted versus observed levels of blood donation was small, and that no further model calibration was required.

6 | RESULTS

6.1 | Survey responses

Of the 5016 invitees to the pilot survey, 1211 (774 males, 437 females) completed at least one SP question. For the main survey, 23,981 (15,652 males, 8329 females) of the 100,000 sampled responded. The sociodemographic characteristics of the responders were similar across the two surveys. For both surveys the donors responding tended to be older, and to donate more frequently compared to the overall group of eligible donors sampled from the PULSE database (Table 2). The estimated coefficients from the main survey are presented in the appendices (Table A1). The results show that the estimated effect of the alternative levels of each attribute on the probability of donors choosing each category of donation frequency, are in line with our expectations as to which attributes and levels would increase versus decrease the utility of the service configuration to the donors.

6.2 | Ex ante – iterative design

The RP data showed that the average number of observed donations was similar between the pilot and the main surveys; for men the average donation frequency was 2.30 (95% CI from 2.2 to 2.37) per year in the pilot versus 2.24 (95%

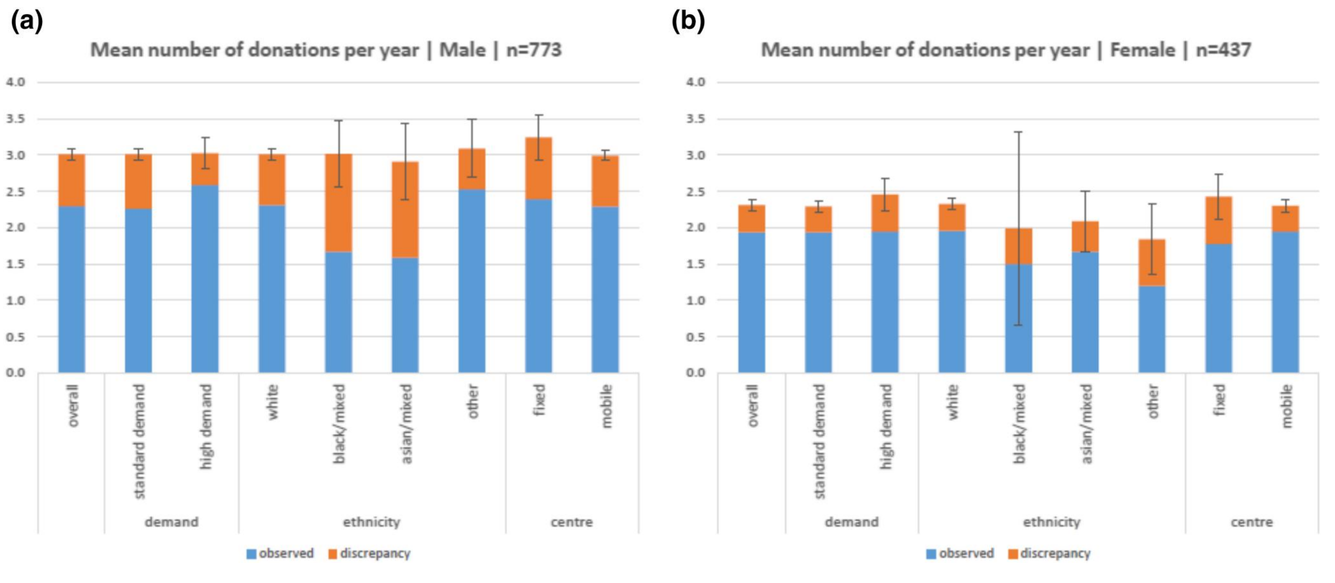

FIGURE 2 Estimated mean (95% CI) discrepancy, mean stated preference (SP) and revealed preference (RP) for the pilot SP survey, overall and by subgroup, with all results stratified by gender

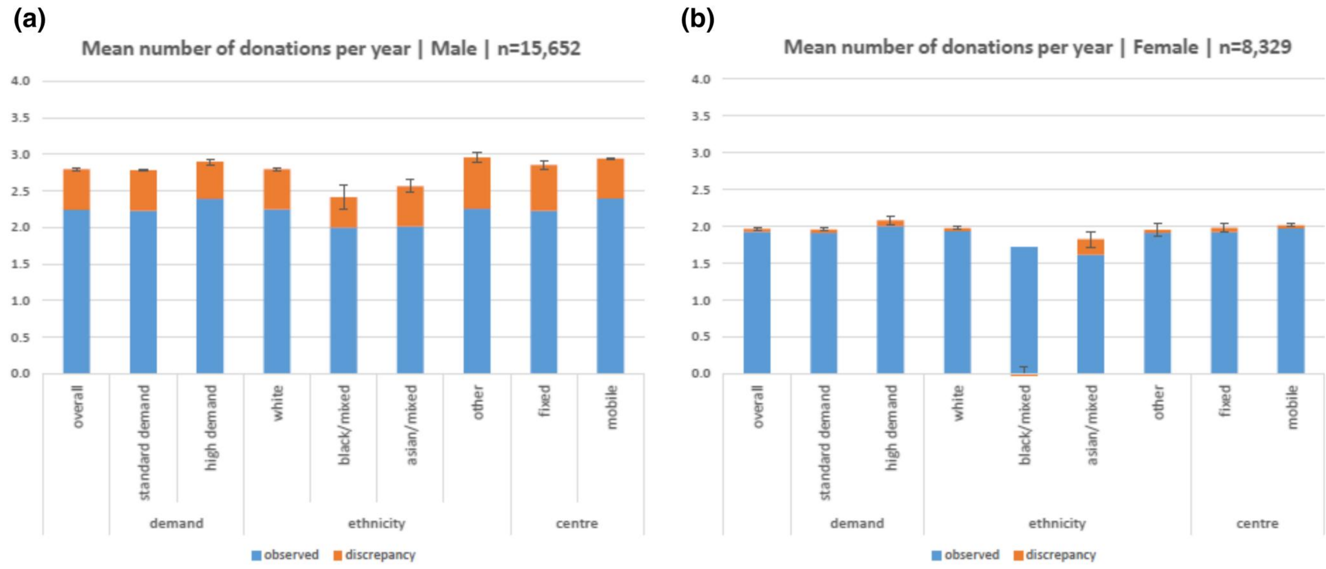

FIGURE 3 Estimated mean (95% CI) discrepancy, mean stated preference (SP) and revealed preference (RP) for the main SP survey, overall and by subgroup, with all results stratified by gender. Results are *before* adjustment for the mediator (deferral rate).

CI from 2.23 to 2.26) in the main survey, and for women the corresponding frequencies were 1.94 (95% CI from 1.85 to 2.02) and 1.93 (95% CI 1.91 to 1.94) (Figures 2 and 3).

The comparison of predictions from the SP survey with the RP data revealed that following the iterative survey design, the discrepancy for men decreased from 0.71 (95% CI from 0.64 to 0.78) with the pilot survey to 0.55 (95% CI from 0.54 to 0.56) for the main survey, a decrease of 0.16 (95% CI from 0.11 to 0.22, $p < 0.00005$). For women, the discrepancy decreased from 0.37 (95% CI from 0.30 to 0.45) in the pilot, to 0.04 (95% CI from 0.03 to 0.06) in the main survey [difference = 0.33 (95% CI from 0.25 to 0.40), $p < 0.00005$]. For male donors whose blood type was defined as being in "standard demand" the discrepancy decreased from 0.74 (95% CI from 0.67 to 0.82) to 0.55 (95% CI from 0.54 to 0.57) [difference = 0.19 (95% CI from 0.13 to 0.25, $p < 0.00005$)], whereas for male donors whose blood type was in "high demand" the discrepancy stayed the same (0.43, 0.22–0.64, compared to 0.51, 0.47-0.54) [difference = 0.07 (95% CI from −0.10 to 0.25, $p = 0.42$)]. For women, the reduction in discrepancy from the pilot to the main survey was similar for all blood types.

For all subgroups defined by ethnicity, the average discrepancy following the main survey was substantially smaller than for the pilot survey except for men in the "other" ethnicity category, where the sample size was small and the confidence intervals comparing surveys overlap. In other words, the ex ante hypothetical bias mitigation strategy of iterative choice task design resulted in a significant decrease in discrepancy.

6.3 | Ex post – constraint modeling

Once the analysis was extended to allow for the constraint of deferral, the estimated levels of hypothetical bias were reduced further. For women, the majority of the hypothetical bias was explained by deferrals, but for men the impact was less pronounced. Figure 4 shows the proportion of the discrepancy due to deferrals. The deferral model explains 86% of the discrepancy between SP and RP for women however there is some heterogeneity amongst subgroups. When the deferral model is applied to men, it only explains 29% of the discrepancy overall. Again, there is some heterogeneity amongst ethnicity but not by blood type.

7 | DISCUSSION

This paper develops an approach for reducing hypothetical bias when predicting RP from SP surveys. This approach combines an iterative survey design which exploits large-scale RP data (ex ante), with allowance for mediators in the

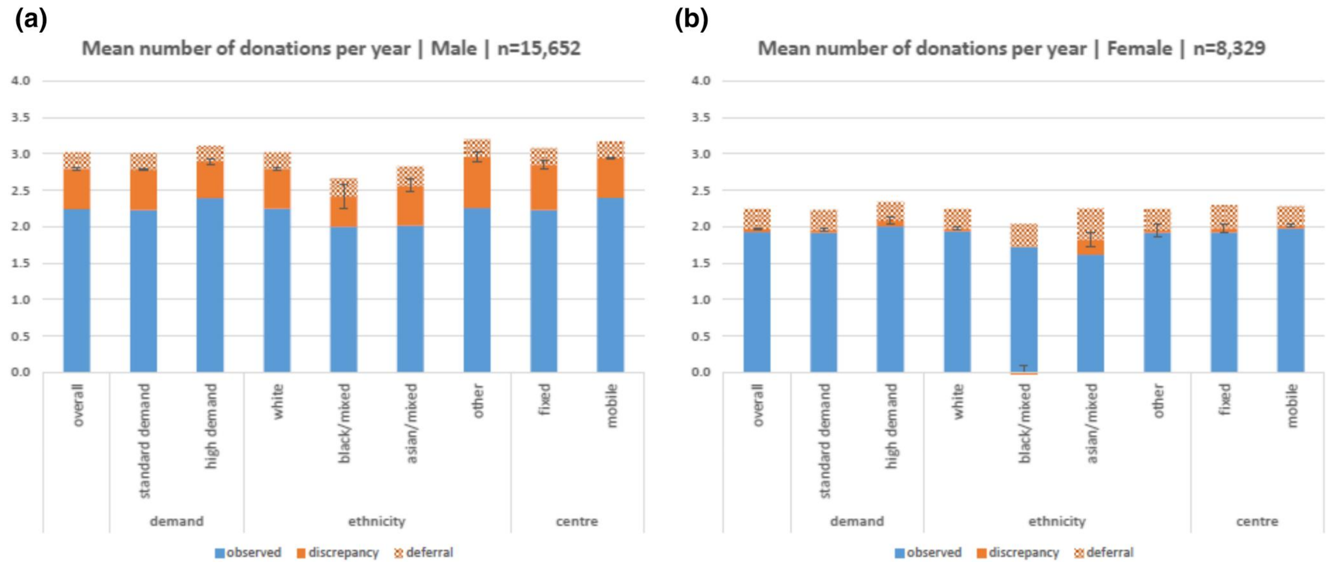

FIGURE 4 Estimated mean (95% CI) discrepancy, mean stated preference (SP) and revealed preference (RP) for the main SP survey, overall and by subgroup, with all results stratified by gender. Results are *after* adjustment for the mediator (deferral rate).

relationship between SP and RP (ex post), to reduce the level of hypothetical bias. The approach is exemplified within an evaluation of alterative policies to increase the frequency of whole-blood donation, and illustrates how the design, and analysis, of SP surveys can be improved to provide predictions that are more accurate and suitable for informing decision-making. Here, the use of causal diagrams can help address identification problems, and in this study, adjustment for a single mediator substantially reduces hypothetical bias.

The paper extends previous approaches to address the problem of hypothetical bias when using responses to SP surveys to predict RP (Broadbent, 2014; Lancsar & Swait, 2014; Loomis, 2014; Ryan & Gerard, 2003; Ryan et al., 2016; Norwood, 2005; Norwood et al., 2008; Whitehead et al., 2011; Whitehead, 2005; Whitehead & Cherry, 2007; Whitehead et al., 2008; Whitehead et al., 2016). Previous research has suggested that SP survey responses do not accurately predict actual behavior, and that these designs are unsuitable for providing the requisite evidence for evaluating changes to health services. Our paper finds that the discrepancies between RP and SP can be reduced by extending the design and analysis of SP surveys to make better use of available RP data.

While methodological guidelines for the design of SP surveys emphasize the importance of piloting, and drawing on qualitative insights to guide the design of the main survey, the emphasis is on the choice of priors, attributes and levels (Coast et al., 2012; Carlsson, 2010; Caussade et al, 2005; Johnson et al., 2013; List & Gallet, 2001; Louviere, 2006). Our approach extends these design principles by revising the main survey design according to insights gained about the magnitude of discrepancy between the SP from the pilot study with RP data. These insights can come by drawing on qualitative insights about the likely reasons for the discrepancy with policy-makers (in our case NHSBT) and survey respondents (active blood donors), and modifying the study design accordingly, by changing the attributes and levels selected.

Our approach emphasizes using RP data to assess hypothetical bias at the pilot stage, to allow these estimates to inform the design of the main survey. The "raw" RP data pertaining to individuals included in the final survey is "held back" to provide a "true" assessment of hypothetical bias. Hence, while the form of RP data, in this case the definition and range of response categories for the donation frequency variable, is used in defining the SP response variable, the RP data does not directly inform the choice for the SP response model. The final predictions are also adjusted to allow for potential mediators of the relationship between SP and RP. The approach is therefore more in keeping with recommendations by Lancsar and Swait (2014) who emphasize the importance of conceptualizing the underlying reasons for hypothetical bias, and incorporating RP at the design stage, rather than relying on RP to calibrate the predictions (Buckell & Hess, 2019; Ghijben et al., 2014; Mark & Swait, 2004).

We also follow recommendations from Lancsar et al. (2017) in choosing the range and final models for analyzing the SP data. We initially select a range of models that recognize the form of response data (e.g., count, ordered categorical), but in making a final choice also consider assessment of model fit which in this context can be used to contrast non-nested models. Nonetheless we acknowledge an alternative approach would be to make fuller use of the RP data within the model development. First, the RP data could be used to calibrate the predictions. Second, if the RP data are of quite different form to the SP data (e.g., categorical vs. continuous), then this can be recognized as part of the process of model choice for the SP data. It may be especially helpful to consider the RP data when choosing the models for the SP response variable if estimates of hypothetical bias remain high.

The proposal to recognize the role of mediators in the relationship between SP and RP goes beyond recent developments in analytical models for SP data (Hensher et al., 2015). While mediation approaches have been considered elsewhere in health economics (Hawkins et al., 2012; Lamu & Olsen, 2018), they have not been previously applied in SP studies. We show how the causal inference framework captures the influence of constraints on the pathway from preferences to behavior. While it is common practice to include constraints (e.g., travel time) as attributes within SP surveys, we distinguish between constraints which have a direct effect on SP and warrant inclusion as attributes in the survey, and those which do not effect SP per se, but which modify the relationship between SP and RP (e.g., having a blood donation appointment postponed for medical reasons).

The approach was considered in the context of blood donation, but the principles for improving the conduct of SP surveys is widely applicable. For example, the design of SP surveys on compliance for alternative drugs prescribed in primary care (e.g., antidiabetics) could use RP data from large-scale prescribing databases to modify the survey design (ex ante). The analysis could then recognize demand and supply side factors that could modify the relationship between SP and RP. In the diabetes example, side effects could mediate the relationship between intended and actual compliance. Similarly, vaccination uptake, could be constrained by allergies or illness, receipt of laparoscopic rather than open surgery may be influenced by the patient's suitability for local anesthesia. These constraints also occur on the supply side. A dialysis patient may have a strong preference for kidney transplant but the availability of organs prevents the realization of this preference, the uptake of weight-loss interventions may reflect food availability.

The proposed approach to reducing the discrepancy between SP and RP uses a "within sample" design to minimize hidden bias that arise in "across sample" comparisons from unobserved differences between the samples. The paper also illustrates how a national large-scale longitudinal registry can provide RP data for the same individuals included in large, representative SP surveys. Here, we were able to examine whether there was heterogeneity in the estimates of hypothetical bias according to subgroups. The results showed that while following the proposed approach hypothetical bias was generally low, there was residual bias for some policy-relevant subgroups – notably men – with blood types that were in high demand.

The paper has the following limitations, which provoke areas for further research. First, the study was required to deliver timely evidence on SP, and only had access to RP data for the period before, rather than after survey administration. While it is conceivable that donors' preferences could have changed subsequently, previous studies have shown that the donation frequency for existing blood donors is stable over time (Di angelantonio et al., 2017; Kaptoge et al., 2019; Transplant, 2015). Second, we compared SP and RP of the status quo, not after a policy change. Third, the reduction in the magnitude of the discrepancy between the pilot and main survey may have partly been due to differences in the sample size or overall design (full factorial vs. efficient design). While in the base case the analysis, models differed between the pilot and main surveys, the sensitivity analyses reported that when the same models were used for both surveys, results were similar to the base case. Fourth, the CI for the predicted mean levels of hypothetical bias did not reflect the uncertainty in the estimates of the underlying model coefficients, and may have underestimated the overall uncertainty surrounding the estimates of hypothetical bias.

Future research is required that considers how in practice this approach can be implemented across a range of settings. Future rapid evaluations of service change or new treatments could follow the structure and principles presented here. It would be possible to quickly administer large-scale pilot surveys by email, and extract accompanying RP data from electronic health records (e.g., hospital episode statistics, clinical practice research database). The proposed approach of estimating the magnitude and reasons for hypothetical bias at the pilot stage, can then inform changes to the design of the SP survey to reduce this bias, combined with using insights from causal diagrams to adjust the predictions to allow for reasons why stated and actual behavior may differ. Studies should build the additional costs and time associated with the acquisition of RP at the proposal stage. We argue that this investment is worthwhile, as if SP predictions are shown to accurately predict RP, then this can help ensure they are useful for informing future policy and clinical decision-making.

This research also provokes further methodological considerations. In particular, it would be useful to consider how insights from the causal inference framework outlined, could be applied to settings where there are multiple constraints (mediators) to an individual's behavior. Further studies could assess the extent to which the proposed approach minimizes not just the *absolute* magnitude of the discrepancy, but also the *relative* effect of a service change on individual's SPs versus behavior. Finally, future work could explore developments in machine learning for prediction and consider more flexible data adaptive approaches to the problem of predicting RPs from SPs.

ACKNOWLEDGEMENTS

We appreciate the help of our colleagues who worked on the Health Economics Modeling of Blood Donation Study (HEMO): Sarah Willis, Zia Sadique, Silvia Perra, Mark Pennington, Neil Hawkins, David Roberts, Carmel Moore, Gail Miflin, Crispen Wickenden, Cathy Koppitz and Jen Turner. We also thank Stephen O'Neill for his detailed comments on an earlier version of the paper. And we are also grateful to Professor Anthony Scott for his helpful discussion of an earlier version of this paper at the international Health Economics Association conference, 2019. ESRC PhD studentship (KDC), NIHR HS&DR 13/54/62 (JC and RG). This research was funded by an ESRC PhD studentship (ESRC), a research grant from the NIHR Health Services and Delivery Research programme (JC), and an NIHR senior research fellowship (RG).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study were made available by NHSBT. Restrictions apply to the availability of these data, which were used under a data sharing agreement for this study.

ORCID

Richard Grieve https://orcid.org/0000-0001-8899-1301

ENDNOTE

[1] Hemoglobin level is considered outside the direct control of the donor. In addition, the assumption is made that it is not influenced by donation history (time since last donation) as all frequencies considered are deemed safe by National Health Service Blood and Transplant with long enough intervals for hemoglobin to recover between donations.

REFERENCES

Ajzen, I., Brown, T. C., & Carvajal, F. (2004). Explaining the discrepancy between intentions and actions: The case of hypothetical bias in contingent valuation. *Personality and Social Psychology Bulletin*, *30*, 1108–1121.

Broadbent, C. D. (2014). Evaluating mitigation and calibration techniques for hypothetical bias in choice experiments. *Journal of Environmental Planning and Management*, *57*, 1831–1848.

Buckell, J., & Hess, S. (2019). Stubbing out hypothetical bias: Improving tobacco market predictions by combining stated and revealed preference data. *Journal of Health Economics*, *65*, 93–102.

Carlsson, F. (2010). Design of stated preference surveys: Is there more to learn from behavioral economics? *Environmental and Resource Economics*, *46*, 167–177.

Caussade, S., De Dios Ortuzar, J., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B*, *39*, 621–640.

Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., & Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: Issues and recommendations. *Health Economics*, *21*, 730–741.

De Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: A review of the literature. *Journal of Health Economics*, *21*, 145–172.

Di angelantonio, E., Thompson, S. G., Kaptoge, S., Moore, C., Walker, M., Armitage, J., Ouwehand, W. H., Roberts, D. J., & Danesh, J. (2017). Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): A randomised trial of 45 000 donors. *Lancet*, *390*, 2360–2371.

Fifer, S., Rose, J., & Greaves, S. (2014). Hypothetical bias in stated choice experiments: Is it a problem? And if so, how do we deal with it? *Transportation Research: Part A: Policy and Practice*, *61*, 164–177.

France, J. L., France, C. R., & Himawan, L. K. (2007). A path analysis of intention to redonate among experienced blood donors: An extension of the theory of planned behavior. *Transfusion*, *47*, 1006–1012.

France, J. L., Kowalsky, J. M., France, C. R., Mcglone, S. T., Himawan, L. K., Kessler, D. A., & Shaz, B. H. (2014). Development of common metrics for donation attitude, subjective norm, perceived behaviorla control, and intention for the blood donation context. *Transfusion*, *54*, 839–847.

Ghijben, P., Lancsar, E., & Zavaresk, S. (2014). Preference for oral anticoagulants in atrial fibrillation: A best-best discrete choice experiment. *PharmacoEconomics*, *32*, 1115–1127.

Goette, L., Stutzer, A., Yavuzcan, G., & Frey, B. M. (2009). Free cholesterol testing as a motivation device in blood donations: Evidence from field experiments. *Transfusion*, *49*, 524–531.

Grieve, R., Willis, S., De Corte, K., Sadique, M. Z., Hawkins, N., Perra, S., et al. (2018). Options for possible changes to the blood donation service: Health economics modelling. *Health Services and Delivery Research*, *6*(40).

Hauber, A. B., Gonzalez, J. M., Groothuis-Oudshoorn, C. G. M., Prior, T., Marshall, D. A., Cunningham, C., Ijzerman, M. J., & Bridges, J. F. P. (2016). Statistical methods for the analysis of discrete choice experiments: A report of the ispor conjoint analysis good research practices task force. *Value in Health*, *19*, 300–315.

Hawkins, N., Richardson, G., Sutton, A. J., Cooper, N. J., Griffiths, C., Rogers, A., & Bower, P. (2012). Surrogates, meta-analysis and cost-effectiveness modelling: A combined analytic approach. *Health Economics*, *21*, 742–756.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied choice analysis*. Cambridge University Press.

Johansson-Stenman, O., & Svedsater, H. (2012). Self-image and valuation of moral goods: Stated versus actual willingness to pay. *Journal of Economic Behavior & Organization*, *84*, 879–891.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Muhlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B., & Bridges, J. F. P. (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experiemntal design good research practices task force. *Value in Health*, *16*, 3–13.

Jones, A. (2007). *Applied econometrics for health economists*. Radcliffe Publishing Ltd.

Kahneman, D., & Knetsch, J. L. (1992). Valueing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, *22*, 57–70.

Kaptoge, S., Di Angelantonio, E., Moore, C., Walker, M., Armitage, J., Ouwehand, W. H., Roberts, D. J., Danesh, J., Thompson, S. G.; INTERVAL Trial Group. (2019). Longer-term efficiency and safety of increasing the frequency of whole blood donation (INTERVAL): Extension study of a randomised trial of 20757 blood donors. *Lancet Haematology*, *6*(10), e510–e520. https://doi.org/10.1016/S2352-3026(19)30106-1

Kirkwood, B. R., & Sterne, J. A. (2003). *Essential medical statistics*. Blackwell Science Ltd.

Lamu, A. N., & Olsen, J. A. (2018). Yes, health is important, but as much for its importance via social life: The direct and indirect effects of health on subjective well-being in chronically ill individuals. *Health Economics*, *27*, 209–222.

Lancsar, E., Fiebig, D. G., & Hole, A. R. (2017). Discrete choice experiments: A guide to model specification, estimation and software. *PharmacoEconomics*, *35*, 697–716.

Lancsar, E., & Swait, J. (2014). Reconceptualising the external validity of discrete choice experiments. *PharmacoEconomics*, *32*, 951–965.

List, J. A., & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values?. *Environmental and Resource Economics*, *20*, 241–254.

Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys*, *25*, 363–370.

Loomis, J. B. (2014). Strategies for overcoming hypothetical bias in stated preference surveys. *Journal of Agricultural and Resource Economics*, *39*, 34–46.

Louviere, J. (2006). What you don't know might hurt you: Some unresolved issues in the design and analysis fo discrete choice experiments. In W. Adamowicz, & J. R. Deshazo (Eds.), *Special issue: Frontiers on stated choice methods*. Environmental and resource economics. Springer.

Lynch, R., & Cohn, S. (2017). Beyond the person: The construction and transformation of blood as a resource. *Critical Public Health*, *27*, 362–372.

Mark, T. L., & Swait, J. (2004). Using stated preference and revealed preference modeling to evaluate prescribing decisions. *Health Economics*, *13*, 563–573.

Masser, B. M., White, K. M., Hyde, M. K., Terry, D. J., & Robinson, N. G. (2009). Predicting blood donation intentions and behavior among Australian blood donors: Testing an extended theory of planned behavior model. *Transfusion*, *49*, 320–329.

Mcfadden, D. (1974). *Conditional logit analysis of qualitative choice behavior*. Academic Press.

Mews, M. (2013). To pay or not to pay - the role of monetary incentives in an optimal blood doatnion service bundle. *International Journal of Nonprofit and Voluntary Sector Marketing*, *18*, 192–202.

Morikawa, T. (1989). *Incorporating stated preference data in travel demand analysis*. Doctor of Philosophy. Massachusetts Institute of Technology.

Mortimer, D., Ghijben, P., Harris, A., & Hollingsworth, B. (2013). *Incentive-based and non-incentive-based interventions for increasing blood donation*. The Cochrane Library.

Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, *30*, 313–325.

Norwood, F. B. (2005). Can calibration reconcile stated and observed preferences?. *Journal of Agricultural & Applied Economics*, *37*, 237–248.

Norwood, F. B., Lusk, J. L., & Boyer, T. (2008). Forecasting hypothetical bias: A tale of two calibrations. *Environmental economics, experimental methods. Cherry*. In T. L. Kroll, S. Shogren, & F. Jason (Eds.), *Routledge explorations in environmental economics*. Taylor and Francis, Routledge.

Quaife, M., Terris-Prestholt, F., Di Tanna, G. L., & Vickerman, P. (2018). How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *The European Journal of Health Economics*, *19*, 1053–1066.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using stata volume II: Categorical responses, counts, and survival*. State Press.

Rapport, F. L., & Maggs, C. J. (2002). Titmuss and the gift relationship: Altruism revisited. *Journal of Advanced Nursing*, *40*, 495–503.
Ringwald, J. (2010). Established ways to keep donor's interest alive. *ISBT Science Series*, *5*(n1), 17–23. https://doi.org/10.1111/j.1751-2824.2010.01389.x
Robinson, N. G., Masser, B. M., White, K. M., Hyde, M. K., & Terry, D. J. (2008). Predicting intentions to donate blood among nondonors in Australia: An extended theory of planned behavior. *Transfusion*, *48*, 2559–2567.
Ryan, M., & Gerard, K. (2003). Using discrete choice experiments to value health care programmes: Current practice and future research reflections. *Applied Health Economics and Health Policy*, *2*, 55–64.
Ryan, M., Gerard, K., & Amaya-Amayak, M. (2010). *Using Discrete choice experiments to value health and health care*. Springer.
Ryan, M., Mentzakis, E., Jareinpituk, S., & Cairns, J. (2016). External validity of contingent valuation: Comparing hypothetical and actual payments. *Health Economics*, *26*(11), 1467–1473. https://doi.org/10.1002/hec.3436
Schreiber, G. B., Schlumpf, K. S., Glynn, S. A., Wright, D. J., Tu, Y., King, M. R., Higgins, M. J., Kessler, D., Gilcher, R., Nass, C. C., Guiltinan, A. M., & Study, N. H. L. B. I. R. E. D. (2006). Convenience, the bane of our existence, and other barriers to donating. *Transfusion*, *46*, 545–553.
Street, D. J., Bunch, D. S., & Moore, B. J. (2001). Optimal designs for 2k paired comparison experiments. *Communications in Statistics, Theory and Methods*, *30*, 2149–2171.
Taylor, S. E., & Brown, J. E. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Science*, *10*, 80–83.
Titmuss, R. M. (1970). *The gift relationship: From human blood to social policy*. Allen & Unwin.
Transplant, N. B. A. (2015). *Blood 2020: A strategy for the blood supply in England and north wales*. NHS Blood and Transplant.
Veldhuizen, I., Ferguson, E., De Kort, W., Donders, R., & Atsma, F. (2011). Exploring the dynamics of the theory of planned behavior in the context of blood donation: Does donation experience make a difference? *Transfusion*, *51*, 2425–2437.
Viney, R., Lancsar, E., & Louviere, J. (2002). Discrete choice experiments to measure consumer preferences for health and health care. *Expert Review of Pharmacoeconomics & Outcomes Research*, *2*, 319–326.
Whitehead, J., Haab, T., & Huang, J.-C. E. (2011). *Preference data for environmental valuation: Combining revealed and stated approaches, Routledge explorations in environmental economics*. Routledge: Taylor and Francis.
Whitehead, J. C. (2005). Environmental risk and averting behavior: Predictive validity of jointly estimated revealed and stated behavior data. *Environmental and Resource Economics*, *32*, 301–316.
Whitehead, J. C., & Cherry, T. L. (2007). Willingness to pay for a green energy program: A comparison of ex-ante and ex-post hypothetical bias mitigation approaches. *Resource and Energy Economics*, *29*, 247–261.
Whitehead, J. C., Dumas, C. F., Herstine, J., Hill, J., & Buerger, B. (2008). Valuing beach access and width with revealed and stated preference data. *Marine Resource Economics*, *23*, 119–135.
Whitehead, J. C., Weddell, M. S., & Groothuis, P. A. (2016). Mitigating hypothetical bias in stated preference data: Evidence from sports tourism. *Economic Inquiry*, *54*, 605–611.

How to cite this article: de Corte K, Cairns J, Grieve R. Stated versus revealed preferences: An approach to reduce bias. *Health Economics*. 2021;1–29. https://doi.org/10.1002/hec.4246

APPENDIX

SP model for main survey

We postulate a general model for estimating the probability of blood donation at each possible frequency defined by the donor's survey response, defined according to an ordered categorical variable (once a year, two times a year etc) (see supporting table 2). We assume that an individual blood donor, i faced with a choice amongst M categories of annual donation frequency (indexed j = 1, 2, . . . , M) chooses so as to maximise their utility (McFadden, 1974). Alternative j is chosen if $U_{ij} = \max(U_{i1,\ldots,}U_{iM})$, assuming that the utility from choosing a particular alternative can be decomposed into the linear combination of the attributes levels of this alternative, and an error term:

$$U_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \in_{ij}$$

Where vector $\boldsymbol{X}_{ij}$ comprises indicators of the levels of each of the blood service attributes, and respondent characteristics.A donor can express their level of preference for the alternative donation choices, given the attributes and their levels, by choosing the donation frequency to maximise utility.The probability that frequency j is chosen is given by:

$$P\{y_i = j\} = P\{U_{ij} = \max(U_{i1}, \ldots . U_{iM})\} = P\left\{\boldsymbol{X}_{ij}\boldsymbol{\beta} + \in_{ij} > \max_{k=1,\ldots J,\ k\neq j}\{\boldsymbol{X}_{ik}\boldsymbol{\beta} + \in_{ik}\}\right\}$$

The ordered-logistic model recognised that the response variable is categorical, with a natural ordering (three times per year, two times per year etc), and we assume that there is a ranking of utilities for each frequency,

$$P\{y_i = j\} = \left[\mathrm{F}\left(\lambda_j - \boldsymbol{X}_{ij}\boldsymbol{\beta}\right) - \mathrm{F}\left(\lambda_{j-i} - \boldsymbol{X}_{ij}\boldsymbol{\beta}\right)\right]$$

Where F is the logistic cumulative distribution function (CDF), and λ_{j} ($\lambda_{\mathrm{j\text{-}1}}$) is the threshold at which the individual would cease to choose frequency *j* (*j*-1). Given the attributes and levels within the main SP survey, the analytical model is specified as:

$$\begin{aligned}\boldsymbol{X}_{ij}\boldsymbol{\beta} = {} & \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{Traveltime_j} + \beta_2 HealthReport_i + \beta_3\boldsymbol{MaxDonations_j} + \boldsymbol{\beta}_4\boldsymbol{AppAvailability}_j + \boldsymbol{\beta}_5\boldsymbol{OpeningTimes_j} \\ & + \beta_6 Place_j + \beta_7 Venue_i + \boldsymbol{\beta}_8\boldsymbol{Ethnicity_i} + \beta_9 Age_i + \beta_{10} ExperiencedDonor_i + \beta_{11} NoDonationsLastYear_i \\ & + \beta_{12} HighDemand_i + \sum\gamma\boldsymbol{W}_j + \sum\delta_{i}\boldsymbol{Z}_{ij}\end{aligned}$$

Where parameters or variables are in **bold**, they are vectors as the variable has multiple levels. The coefficients $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_5$ show effects on stated donation frequencies for levels of attributes pertaining to different durations of travel time, provision of a health report, maximum number of donations per year, availability of blood donation appointments, and blood donation venue opening times, $\boldsymbol{\beta}_6$ to whether or not the proposed blood service change pertained to the venue where the donor last donated, and $\boldsymbol{\beta}_7$ is an indicator for whether the donors last donation was at a static versus a mobile venue. The effect of individual donor's characteristics on stated donation frequency are represented by coefficients $\boldsymbol{\beta}_8$ to $\boldsymbol{\beta}_{12}$ which represent ethnicity, age, whether the donor is experienced or not, whether the donor's blood group is in high demand or not, and the number of donations in the year preceding the survey. The model allowed for potential effect modification of alternative attributes with one another, through the inclusion of W_j, a vector of 15 different interaction terms each with 4-10 levels. Z_{ij} represents a vector of 36 further interactions to allow for potential effect modification of the individual donor's characteristics with each attribute.

Separate models were specified for each gender. Robust standard errors were reported to allow for the panel nature of the response data, that is the potential correlation of the survey responses for each individual.

To estimate the predicted number of donations for each donor from the estimated model, we combined the coefficients with the levels of each blood service attribute the donor experienced at their last donation based on information in the blood donor register to obtain the predicted probability for each frequency, $\hat{P}\{y_i = j\} = [\mathrm{F}(\hat{\lambda}_{\mathrm{j}} - \boldsymbol{X}_{ij}\hat{\boldsymbol{\beta}}) - \mathrm{F}(\hat{\lambda}_{\mathrm{j\text{-}1}} - \boldsymbol{X}_{ij}\hat{\boldsymbol{\beta}})]$. The expected annual frequency of blood donation per individual was calculated as a weighted average to obtain $\widehat{freq} = \sum(j \times \hat{P}\{y_i = j\})$

TABLE A1 Model coding

Variable	Level	Coding	
Travel time	10 min shorter than typical travel time	1	
	Your typical travel time (only option for LP)	2	(Reference)
	15 min longer than typical travel time	3	
	30 min longer than typical travel time	4	
Health report	No	1	(Reference)
	Yes, after each donation	2	
		Female	**Male**
Maximum number of donations per year	3 (female reference)	1	
	4 (male reference)	2	1
	5		2
	6		3
Appointment availability	Every day: Mon-Sun	1	
	Every weekday: Mon-Fri	2	(Reference)
	1 day every 2 months: Mon-Fri	3	
	1 day every 2 months: Sat or Sun	4	
Opening times	9 AM–12 PM and 2–5 PM	1	(Reference)
	9 AM–5 PM	2	
	9 AM–8 PM	3	
	2–8 PM	4	
Place	Last	LP	(Reference)
	Different	DP	
Venue	Static		(Reference)
	Mobile		
Ethnicity	White		(Reference)
	Black/mixed		
	Asian/mixed		
	Other		
Age		Mean	(Reference)
Donor experience	Nursery		(Reference)
	Experienced		
Number of donations in the past year	ndon	Mean	(Reference)
Blood type demand	Standard	0	(Reference)
	High	1	

TABLE A2 ordered logistic model estimates for women

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
Main effects							
Travel time	1	0.494	0.158	0.002	**	0.184	0.805
	2	Reference					
	3	−0.660	0.156	0.000	***	−0.965	−0.355
	4	−0.815	0.173	0.000	***	−1.154	−0.476
Availability	1	1.187	0.126	0.000	***	0.940	1.434
	2	Reference					
	3	−0.885	0.133	0.000	***	−1.147	−0.624
	4	−0.089	0.144	0.534		−0.371	0.192
Opening times	1	Reference					
	2	0.441	0.140	0.002	**	0.167	0.714
	3	1.745	0.132	0.000	***	1.486	2.004
	4	1.424	0.138	0.000	***	1.153	1.695
Health report	No	Reference					
	Yes	0.034	0.096	0.724		−0.154	0.222
Place	Last	Reference					
	Different	−0.222	0.127	0.082		−0.471	0.028
Maximum number of donations	Three	Reference					
	Four	0.656	0.110	0.000	***	0.440	0.873
Age	per year	0.022	0.002	0.000	***	0.017	0.026
Blood type demand	Standard	Reference					
	High	0.076	0.098	0.439		−0.116	0.267
Ethnicity	White	Reference					
	Black/mixed	−0.431	0.321	0.179		−1.061	0.198
	Asian/mixed	−0.118	0.212	0.578		−0.534	0.298
	Other	−0.250	0.169	0.140		−0.582	0.082
Venue	Static	Reference					
	Mobile	0.366	0.110	0.001	**	0.150	0.581
Number of donations	per donation	0.106	0.040	0.008	**	0.028	0.184
Donor experience	Nursery	Reference					
	Experienced	−0.052	0.078	0.502		−0.204	0.100
INTERACTIONS							
traveltime#availability	2 1	−0.178	0.078	0.022	*	−0.330	−0.026
	2 3	−0.020	0.080	0.799		−0.176	0.136
	2 4	−0.046	0.085	0.590		−0.213	0.121
	3 1	0.136	0.080	0.089		−0.021	0.292
	3 3	0.200	0.083	0.016	*	0.038	0.362
	3 4	0.288	0.082	0.000	***	0.126	0.449
	4 1	−0.177	0.080	0.027	*	−0.334	−0.021
	4 3	0.214	0.086	0.013	*	0.045	0.383
	4 4	0.388	0.086	0.000	***	0.220	0.556

TABLE A2 (Continued)

	Level	Coefficient	Robust SE	p value		95% CI Bounds	
traveltime#openingtimes	2 2	0.098	0.084	0.244		−0.067	0.263
	2 3	−0.066	0.072	0.362		−0.207	0.076
	2 4	0.132	0.077	0.088		−0.019	0.283
	3 2	0.069	0.086	0.425		−0.100	0.239
	3 3	−0.042	0.076	0.577		−0.191	0.107
	3 4	−0.007	0.078	0.933		−0.160	0.147
	4 2	0.118	0.084	0.157		−0.046	0.282
	4 3	−0.399	0.077	0.000	***	−0.550	−0.247
	4 4	−0.186	0.083	0.025	*	−0.349	−0.023
traveltime#healthreport	2 2	−0.208	0.054	0.000	***	−0.314	−0.101
	3 2	−0.027	0.059	0.642		−0.143	0.088
	4 2	−0.210	0.060	0.000	***	−0.326	−0.093
traveltime#place	2#LP	0.000	(Empty)				
	2#DP	0.000	(Omitted)				
	3#LP	0.000	(Empty)				
	3#DP	0.000	(Omitted)				
	4#LP	0.000	(Empty)				
	4#DP	0.000	(Omitted)				
traveltime#maxdonations	2 2	0.210	0.062	0.001	**	0.088	0.333
	3 2	−0.446	0.059	0.000	***	−0.562	−0.329
	4 2	−0.723	0.065	0.000	***	−0.849	−0.596
availability#openingtimes	1 2	−0.070	0.065	0.279		−0.197	0.057
	1 3	−0.493	0.064	0.000	***	−0.619	−0.366
	1 4	−0.574	0.067	0.000	***	−0.705	−0.443
	3 2	−0.219	0.069	0.002	**	−0.355	−0.082
	3 3	−0.281	0.064	0.000	***	−0.406	−0.156
	3 4	−0.276	0.065	0.000	***	−0.404	−0.148
	4 2	−0.175	0.069	0.011	*	−0.310	−0.040
	4 3	−0.805	0.070	0.000	***	−0.943	−0.668
	4 4	−0.800	0.072	0.000	***	−0.941	−0.658
availability#healthreport	1 2	0.017	0.045	0.699		−0.071	0.106
	3 2	0.090	0.049	0.069		−0.007	0.186
	4 2	0.092	0.048	0.056		−0.002	0.187
availability#place	1#DP	0.033	0.066	0.614		−0.096	0.162
	3#DP	0.044	0.065	0.503		−0.085	0.172
	4#DP	−0.013	0.068	0.854		−0.146	0.121

(Continues)

TABLE A2 (Continued)

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
availability#maxdonations	1 2	0.311	0.047	0.000	***	0.220	0.402
	3 2	−0.245	0.049	0.000	***	−0.342	−0.148
	4 2	−0.099	0.052	0.057		−0.201	0.003
openingtimes#healthreport	2 2	−0.081	0.048	0.093		−0.175	0.014
	3 2	−0.015	0.045	0.748		−0.103	0.074
	4 2	−0.055	0.045	0.221		−0.144	0.033
openingtimes#place	2#DP	−0.112	0.070	0.110		−0.250	0.025
	3#DP	0.140	0.061	0.022	*	0.020	0.260
	4#DP	−0.023	0.066	0.731		−0.152	0.107
openingtimes#maxdonations	2 2	−0.056	0.051	0.272		−0.157	0.044
	3 2	0.217	0.052	0.000	***	0.116	0.318
	4 2	0.247	0.052	0.000	***	0.146	0.349
healthreport#place	2#DP	0.179	0.050	0.000	***	0.081	0.277
healthreport#maxdonations	2 2	0.145	0.036	0.000	***	0.075	0.215
place#maxdonations	DP#2	−0.178	0.051	0.001	**	−0.279	−0.077
traveltime#c.age	2	−0.001	0.002	0.650		−0.005	0.003
	3	0.002	0.002	0.305		−0.002	0.007
	4	0.001	0.003	0.769		−0.004	0.006
availability#c.age	1	−0.012	0.002	0.000	***	−0.016	−0.008
	3	0.005	0.002	0.006	**	0.002	0.009
	4	−0.009	0.002	0.000	***	−0.013	−0.005
openingtimes#c.age	2	−0.002	0.002	0.447		−0.005	0.002
	3	−0.017	0.002	0.000	***	−0.020	−0.013
	4	−0.016	0.002	0.000	***	−0.020	−0.012
healthreport#c.age	2	0.000	0.001	0.730		−0.002	0.003
place#c.age	DP	−0.004	0.002	0.027	*	−0.007	0.000
maxdonations#c.age	2	−0.010	0.002	0.000	***	−0.013	−0.007
traveltime#demand	2 1	−0.093	0.089	0.296		−0.269	0.082
	3 1	−0.062	0.096	0.519		−0.251	0.127
	4 1	−0.108	0.108	0.318		−0.318	0.103
availability#demand	1 1	−0.033	0.073	0.645		−0.176	0.109
	3 1	0.061	0.080	0.448		−0.096	0.218
	4 1	0.050	0.091	0.585		−0.129	0.229
openingtimes#demand	2 1	−0.003	0.081	0.973		−0.161	0.156
	3 1	−0.090	0.080	0.260		−0.246	0.066
	4 1	−0.028	0.082	0.728		−0.189	0.132
healthreport#demand	2 1	−0.021	0.056	0.701		−0.130	0.088
place#demand	DP#1	0.079	0.071	0.267		−0.060	0.217
maxdonations#demand	2 1	0.049	0.065	0.445		−0.077	0.176

TABLE A2 (Continued)

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
traveltime#ethnicity	2#black/mixed	0.036	0.283	0.898		−0.518	0.591
	2#asian/mixed	0.181	0.189	0.338		−0.189	0.551
	2#other	0.140	0.144	0.331		−0.142	0.422
	3#black/mixed	0.102	0.262	0.696		−0.411	0.615
	3#asian/mixed	0.269	0.170	0.113		−0.063	0.601
	3#other	0.264	0.144	0.068		−0.020	0.547
	4#black/mixed	0.212	0.287	0.459		−0.350	0.774
	4#asian/mixed	0.291	0.198	0.141		−0.097	0.679
	4#other	0.194	0.164	0.238		−0.128	0.516
availability#ethnicity	1#black/mixed	0.173	0.240	0.470		−0.298	0.644
	1#asian/mixed	0.160	0.158	0.309		−0.149	0.469
	1#other	−0.068	0.120	0.572		−0.302	0.167
	3#black/mixed	0.317	0.244	0.193		−0.161	0.795
	3#asian/mixed	0.408	0.160	0.011	*	0.095	0.721
	3#other	0.139	0.128	0.277		−0.111	0.389
	4#black/mixed	0.195	0.273	0.475		−0.340	0.731
	4#asian/mixed	0.462	0.194	0.017	*	0.082	0.843
	4#other	0.042	0.152	0.781		−0.256	0.340
openingtimes#ethnicity	2#black/mixed	0.229	0.215	0.288		−0.193	0.650
	2#asian/mixed	0.307	0.185	0.097		−0.056	0.669
	2#other	0.231	0.122	0.058		−0.008	0.471
	3#black/mixed	−0.281	0.269	0.295		−0.808	0.245
	3#asian/mixed	−0.134	0.182	0.462		−0.490	0.223
	3#other	0.169	0.122	0.168		−0.071	0.409
	4#black/mixed	0.022	0.224	0.921		−0.416	0.461
	4#asian/mixed	−0.153	0.179	0.392		−0.503	0.197
	4#other	0.193	0.135	0.153		−0.071	0.457
healthreport#ethnicity	2#black/mixed	0.329	0.149	0.027	*	0.038	0.621
	2#asian/mixed	0.015	0.113	0.891		−0.206	0.236
	2#other	0.070	0.086	0.414		−0.098	0.239
place#ethnicity	DP#black/mixed	0.198	0.220	0.369		−0.234	0.630
	DP#asian/mixed	−0.122	0.151	0.417		−0.418	0.173
	DP#other	−0.179	0.117	0.126		−0.407	0.050
maxdonations#ethnicity	2#black/mixed	0.071	0.191	0.708		−0.302	0.445
	2#asian/mixed	0.004	0.133	0.978		−0.257	0.264
	2#other	−0.099	0.104	0.340		−0.302	0.104

(Continues)

TABLE A2 (Continued)

	Level	Coefficient	Robust SE	p value		95% CI Bounds	
traveltime#site	2#mobile	−0.039	0.107	0.712		−0.249	0.170
	3#mobile	−0.263	0.101	0.010	*	−0.462	−0.064
	4#mobile	−0.426	0.114	0.000	***	−0.649	−0.203
availability#site	1#mobile	−0.198	0.086	0.020	*	−0.366	−0.031
	3#mobile	0.417	0.095	0.000	***	0.230	0.604
	4#mobile	0.374	0.101	0.000	***	0.176	0.572
openingtimes#site	2#mobile	−0.107	0.091	0.240		−0.286	0.072
	3#mobile	−0.133	0.090	0.137		−0.309	0.042
	4#mobile	0.021	0.096	0.830		−0.167	0.208
healthreport#site	2#mobile	−0.083	0.061	0.174		−0.203	0.037
place#site	DP#mobile	−0.007	0.082	0.935		−0.168	0.155
maxdonations#site	2#mobile	0.133	0.071	0.061		−0.006	0.273
traveltime#ndon	2	−0.091	0.036	0.012	*	−0.162	−0.020
	3	0.042	0.039	0.279		−0.034	0.118
	4	−0.064	0.044	0.146		−0.149	0.022
availability#ndon	1	0.011	0.030	0.722		−0.048	0.070
	3	0.048	0.032	0.131		−0.014	0.110
	4	0.059	0.037	0.105		−0.012	0.131
openingtimes#ndon	2	0.004	0.033	0.898		−0.060	0.069
	3	0.017	0.032	0.587		−0.045	0.079
	4	0.033	0.033	0.324		−0.032	0.097
healthreport#ndon	2	0.018	0.022	0.398		−0.024	0.061
place#ndon	DP	0.042	0.030	0.154		−0.016	0.100
maxdonations#ndon	2	0.246	0.027	0.000	***	0.194	0.298
traveltime#experience	2 1	−0.024	0.068	0.723		−0.158	0.110
	3 1	−0.167	0.073	0.022	*	−0.310	−0.025
	4 1	−0.135	0.081	0.096		−0.295	0.024
availability#experience	1 1	0.140	0.058	0.015	*	0.027	0.253
	3 1	−0.011	0.061	0.856		−0.130	0.108
	4 1	0.017	0.070	0.808		−0.120	0.154
openingtimes#experience	2 1	−0.013	0.061	0.838		−0.133	0.108
	3 1	0.085	0.060	0.158		−0.033	0.204
	4 1	0.090	0.063	0.154		−0.034	0.215
healthreport#experience	2 1	0.003	0.041	0.946		−0.078	0.084
place#experience	DP#1	0.050	0.056	0.374		−0.060	0.159
maxdonations#experience	2 1	0.139	0.050	0.005	**	0.042	0.237
Measures of fit							
Log pseudolikelihood	−66349.5						
Number of free parameters	174						

TABLE A2 (Continued)

	Level	Coefficient	Robust SE	*p* value	95% CI Bounds
AIC	133046.9				
BIC	134579.1				
Number of responses	49,298				
Number of respondents	8329				

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; robust standard errors adjusted for clustering by individual respondent.

Abbreviations: AIC, akaike information criterion; BIC, Bayesian information criterion.

TABLE A3 Ordered logistic model estimates for men

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
Main effects							
Travel time	1	0.083	0.127	0.513		−0.166	0.332
	2	Reference					
	3	−0.892	0.138	0.000	***	−1.162	−0.623
	4	−1.526	0.137	0.000	***	−1.795	−1.257
Availability	1	1.051	0.105	0.000	***	0.845	1.257
	2	Reference					
	3	−0.839	0.113	0.000	***	−1.061	−0.617
	4	−0.081	0.121	0.504		−0.319	0.157
Opening times	1	Reference					
	2	0.110	0.105	0.295		−0.096	0.316
	3	1.464	0.106	0.000	***	1.257	1.671
	4	1.353	0.112	0.000	***	1.133	1.573
Health report	No	Reference					
	Yes	0.159	0.074	0.031	*	0.014	0.304
Place	Last	Reference					
	Different	−0.145	0.097	0.134		−0.334	0.045
Maximum number of donations	Four	Reference					
	Five	0.567	0.079	0.000	***	0.412	0.721
	Six	1.185	0.107	0.000	***	0.975	1.395
Age	per year	0.012	0.002	0.000	***	0.008	0.016
Blood type demand	Standard	Reference					
	High	0.067	0.072	0.353		−0.075	0.209
Ethnicity	White	Reference					
	Black/mixed	−0.651	0.262	0.013	*	−1.164	−0.138
	Asian/mixed	−0.178	0.141	0.207		−0.455	0.098
	Other	−0.015	0.113	0.895		−0.236	0.206
Venue	Static	Reference					
	Mobile	0.117	0.080	0.142		−0.039	0.273
Number of donations	per donation	0.204	0.025	0.000	***	0.156	0.253
Donor experience	Nursery	Reference					
	Experienced	−0.139	0.061	0.022	*	−0.259	−0.020

TABLE A3 (Continued)

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
INTERACTIONS							
traveltime#availability	2 1	−0.133	0.055	0.015	*	−0.241	−0.025
	2 3	0.036	0.055	0.505		−0.071	0.143
	2 4	0.042	0.059	0.477		−0.073	0.157
	3 1	−0.158	0.054	0.003	**	−0.264	−0.052
	3 3	−0.031	0.059	0.594		−0.146	0.084
	3 4	0.030	0.062	0.632		−0.092	0.152
	4 1	−0.068	0.057	0.228		−0.179	0.043
	4 3	0.169	0.059	0.004	**	0.053	0.284
	4 4	0.452	0.061	0.000	***	0.332	0.572
traveltime#openingtimes	2 2	0.046	0.064	0.467		−0.078	0.171
	2 3	0.073	0.058	0.212		−0.041	0.187
	2 4	0.103	0.059	0.079		−0.012	0.219
	3 2	0.088	0.066	0.179		−0.040	0.217
	3 3	0.011	0.059	0.848		−0.105	0.127
	3 4	0.112	0.059	0.057		−0.003	0.227
	4 2	0.035	0.062	0.579		−0.087	0.156
	4 3	−0.156	0.061	0.011	*	−0.276	−0.036
	4 4	−0.036	0.064	0.575		−0.160	0.089
traveltime#healthreport	2 2	0.018	0.040	0.652		−0.060	0.096
	3 2	0.010	0.041	0.808		−0.070	0.090
	4 2	−0.009	0.044	0.832		−0.095	0.076
traveltime#place	2#LP	0.000	(Empty)				
	2#DP	0.000	(Omitted)				
	3#LP	0.000	(Empty)				
	3#DP	0.000	(Omitted)				
	4#LP	0.000	(Empty)				
	4#DP	0.000	(Omitted)				
traveltime#maxdonations	2 2	−0.040	0.042	0.340		−0.121	0.042
	2 3	0.052	0.055	0.343		−0.055	0.159
	3 2	−0.263	0.047	0.000	***	−0.354	−0.171
	3 3	−0.404	0.053	0.000	***	−0.508	−0.299
	4 2	−0.365	0.045	0.000	***	−0.454	−0.277
	4 3	−0.747	0.054	0.000	***	−0.852	−0.642

(Continues)

TABLE A3 (Continued)

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
availability#openingtimes	1 2	0.098	0.054	0.070		−0.008	0.204
	1 3	−0.333	0.046	0.000	***	−0.423	−0.243
	1 4	−0.290	0.047	0.000	***	−0.381	−0.198
	3 2	0.007	0.051	0.896		−0.093	0.107
	3 3	−0.026	0.046	0.571		−0.116	0.064
	3 4	−0.002	0.047	0.959		−0.095	0.091
	4 2	0.008	0.056	0.886		−0.102	0.118
	4 3	−0.509	0.048	0.000	***	−0.603	−0.415
	4 4	−0.520	0.052	0.000	***	−0.621	−0.419
availability#healthreport	1 2	0.005	0.030	0.873		−0.053	0.063
	3 2	−0.002	0.031	0.954		−0.062	0.058
	4 2	−0.043	0.033	0.197		−0.108	0.022
availability#place	1#DP	0.039	0.044	0.376		−0.048	0.126
	3#DP	−0.001	0.046	0.988		−0.091	0.089
	4#DP	−0.021	0.049	0.666		−0.116	0.074
availability#maxdonations	1 2	0.024	0.034	0.480		−0.043	0.091
	1 3	0.199	0.044	0.000	***	0.112	0.285
	3 2	−0.109	0.036	0.003	**	−0.179	−0.038
	3 3	−0.064	0.043	0.137		−0.148	0.020
	4 2	−0.005	0.038	0.887		−0.081	0.070
	4 3	−0.161	0.049	0.001	**	−0.256	−0.066
openingtimes#healthreport	2 2	−0.062	0.033	0.062		−0.127	0.003
	3 2	−0.057	0.034	0.094		−0.123	0.010
	4 2	−0.068	0.035	0.052		−0.136	0.001
openingtimes#place	2#DP	0.079	0.055	0.149		−0.028	0.186
	3#DP	0.105	0.046	0.022	*	0.015	0.196
	4#DP	0.015	0.048	0.761		−0.079	0.108
openingtimes#maxdonations	2 2	0.000	0.038	0.990		−0.074	0.073
	2 3	0.104	0.048	0.030	*	0.010	0.199
	3 2	0.111	0.032	0.000	***	0.049	0.173
	3 3	0.311	0.045	0.000	***	0.224	0.399
	4 2	−0.005	0.034	0.875		−0.072	0.061
	4 3	0.251	0.046	0.000	***	0.161	0.341
healthreport#place	2#DP	−0.023	0.032	0.467		−0.087	0.040
healthreport#maxdonations	2 2	−0.085	0.025	0.001	**	−0.135	−0.035
	2 3	0.012	0.030	0.687		−0.047	0.071
place#maxdonations	DP#2	0.015	0.035	0.661		−0.052	0.083
	DP#3	−0.052	0.044	0.233		−0.138	0.034

TABLE A3 (Continued)

	Level	Coefficient	Robust SE	p value		95% CI Bounds	
traveltime#c.age	2	−0.003	0.002	0.111		−0.006	0.001
	3	0.005	0.002	0.008	**	0.001	0.009
	4	0.010	0.002	0.000	***	0.006	0.014
availability#c.age	1	−0.010	0.001	0.000	***	−0.013	−0.007
	3	0.008	0.002	0.000	***	0.005	0.011
	4	−0.004	0.002	0.026	*	−0.007	0.000
openingtimes#c.age	2	−0.003	0.001	0.086		−0.005	0.000
	3	−0.021	0.002	0.000	***	−0.024	−0.018
	4	−0.020	0.002	0.000	***	−0.023	−0.017
healthreport#c.age	2	0.001	0.001	0.365		−0.001	0.003
place#c.age	DP	0.001	0.001	0.264		−0.001	0.004
maxdonations#c.age	2	−0.008	0.001	0.000	***	−0.010	−0.006
	3	−0.025	0.001	0.000	***	−0.028	−0.022
traveltime#demand	2 1	0.053	0.067	0.431		−0.078	0.184
	3 1	−0.078	0.080	0.328		−0.234	0.078
	4 1	0.061	0.080	0.442		−0.095	0.217
availability#demand	1 1	−0.022	0.057	0.704		−0.134	0.091
	3 1	0.006	0.060	0.923		−0.112	0.124
	4 1	−0.075	0.069	0.277		−0.210	0.060
openingtimes#demand	2 1	−0.003	0.057	0.953		−0.115	0.108
	3 1	−0.063	0.058	0.276		−0.176	0.050
	4 1	−0.010	0.061	0.873		−0.129	0.110
healthreport#demand	2 1	−0.003	0.040	0.939		−0.080	0.074
place#demand	DP#1	−0.059	0.052	0.251		−0.161	0.042
maxdonations#demand	2 1	−0.041	0.043	0.347		−0.126	0.044
	3 1	0.062	0.056	0.270		−0.048	0.172
traveltime#ethnicity	2#black/mixed	0.484	0.245	0.048	*	0.004	0.965
	2#asian/mixed	0.179	0.129	0.165		−0.074	0.433
	2#other	0.207	0.103	0.045	*	0.005	0.409
	3#black/mixed	0.385	0.287	0.179		−0.177	0.948
	3#asian/mixed	0.183	0.141	0.194		−0.093	0.459
	3#other	0.070	0.121	0.563		−0.168	0.308
	4#black/mixed	0.572	0.244	0.019	*	0.094	1.051
	4#asian/mixed	0.574	0.147	0.000	***	0.286	0.863
	4#other	0.009	0.130	0.944		−0.245	0.263

(Continues)

TABLE A3 (Continued)

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
availability#ethnicity	1#black/mixed	0.282	0.221	0.202		−0.151	0.714
	1#asian/mixed	−0.069	0.121	0.567		−0.305	0.167
	1#other	0.050	0.087	0.567		−0.121	0.221
	3#black/mixed	0.587	0.249	0.018	*	0.100	1.074
	3#asian/mixed	0.180	0.113	0.110		−0.041	0.402
	3#other	0.148	0.093	0.110		−0.034	0.329
	4#black/mixed	0.106	0.263	0.688		−0.410	0.621
	4#asian/mixed	0.193	0.117	0.098		−0.036	0.423
	4#other	−0.035	0.104	0.740		−0.239	0.170
openingtimes#ethnicity	2#black/mixed	0.149	0.206	0.470		−0.255	0.553
	2#asian/mixed	−0.069	0.108	0.521		−0.280	0.142
	2#other	0.003	0.088	0.976		−0.170	0.175
	3#black/mixed	0.109	0.223	0.625		−0.327	0.545
	3#asian/mixed	−0.107	0.120	0.373		−0.342	0.128
	3#other	−0.205	0.093	0.028	*	−0.388	−0.023
	4#black/mixed	0.112	0.241	0.643		−0.361	0.585
	4#asian/mixed	−0.137	0.126	0.277		−0.385	0.110
	4#other	−0.046	0.095	0.630		−0.232	0.140
healthreport#ethnicity	2#black/mixed	0.260	0.157	0.097		−0.047	0.567
	2#asian/mixed	0.090	0.076	0.240		−0.060	0.240
	2#other	0.034	0.064	0.599		−0.092	0.160
place#ethnicity	DP#black/mixed	−0.085	0.145	0.557		−0.369	0.199
	DP#asian/mixed	−0.114	0.103	0.267		−0.315	0.087
	DP#other	−0.010	0.083	0.907		−0.172	0.152
maxdonations#ethnicity	2#black/mixed	0.018	0.156	0.910		−0.289	0.324
	2#asian/mixed	−0.230	0.084	0.006	**	−0.394	−0.065
	2#other	−0.036	0.069	0.604		−0.172	0.100
	3#black/mixed	−0.438	0.183	0.017	*	−0.796	−0.079
	3#asian/mixed	−0.405	0.106	0.000	***	−0.613	−0.197
	3#other	−0.098	0.090	0.277		−0.274	0.078
traveltime#site	2#mobile	0.140	0.077	0.070		−0.012	0.292
	3#mobile	0.048	0.086	0.576		−0.120	0.216
	4#mobile	−0.175	0.084	0.037	*	−0.340	−0.011
availability#site	1#mobile	−0.141	0.064	0.028	*	−0.266	−0.015
	3#mobile	0.185	0.069	0.008	**	0.049	0.321
	4#mobile	0.208	0.077	0.007	**	0.057	0.358
openingtimes#site	2#mobile	−0.056	0.064	0.378		−0.182	0.069
	3#mobile	0.031	0.066	0.641		−0.099	0.160
	4#mobile	0.132	0.071	0.064		−0.008	0.272

TABLE A3 (Continued)

	Level	Coefficient	Robust SE	*p* value		95% CI Bounds	
healthreport#site	2#mobile	0.006	0.046	0.889		−0.083	0.096
place#site	DP#mobile	−0.154	0.059	0.009	**	−0.270	−0.039
maxdonations#site	2#mobile	0.148	0.049	0.002	**	0.053	0.244
	3#mobile	0.232	0.064	0.000	***	0.106	0.357
traveltime#ndon	2	−0.017	0.023	0.444		−0.061	0.027
	3	0.046	0.026	0.080		−0.005	0.097
	4	−0.033	0.027	0.226		−0.086	0.020
availability#ndon	1	−0.035	0.019	0.073		−0.072	0.003
	3	0.036	0.020	0.078		−0.004	0.076
	4	−0.033	0.023	0.153		−0.079	0.012
openingtimes#ndon	2	−0.011	0.020	0.570		−0.050	0.027
	3	−0.024	0.021	0.247		−0.064	0.016
	4	−0.018	0.021	0.386		−0.059	0.023
healthreport#ndon	2	0.009	0.014	0.525		−0.018	0.035
place#ndon	DP	0.019	0.017	0.267		−0.015	0.054
maxdonations#ndon	2	0.104	0.014	0.000	***	0.076	0.132
	3	0.254	0.020	0.000	***	0.215	0.294
traveltime#experience	2 1	−0.010	0.056	0.859		−0.120	0.100
	3 1	−0.019	0.065	0.768		−0.146	0.108
	4 1	−0.131	0.064	0.041	*	−0.257	−0.006
availability#experience	1 1	0.073	0.047	0.123		−0.020	0.166
	3 1	0.008	0.049	0.866		−0.088	0.105
	4 1	0.067	0.057	0.238		−0.044	0.179
openingtimes#experience	2 1	0.051	0.048	0.293		−0.044	0.145
	3 1	0.301	0.051	0.000	***	0.201	0.400
	4 1	0.185	0.051	0.000	***	0.085	0.286
healthreport#experience	2 1	−0.030	0.034	0.378		−0.095	0.036
place#experience	DP#1	−0.020	0.044	0.654		−0.105	0.066
maxdonations#experience	2 1	0.022	0.036	0.536		−0.048	0.092
Measures of fit							
Log pseudolikelihood	−158176						
Number of free parameters	196						
AIC	316743.9						
BIC	318593.8						
Number of responses	92,782						
Number of respondents	15,652						

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; robust standard errors adjusted for clustering by individual respondent.

Abbreviations: AIC, akaike information criterion; BIC, Bayesian information criterion.