Survey Research Methods (2022)
Vol. 16, No. 2, pp. 227-240
doi:10.18148/srm/2022.v16i2.7823
European Survey Research Association

Answer refused: Exploring how item non-response on domestic abuse questions in a social survey affects analysis.

Valeria Skafida
School of Social and Political Science
The University of Edinburgh
Scotland, UK

Fiona Morrison
Centre for Child Wellbeing and Protection
The University of Stirling
Scotland, UK

John Devaney
School of Social and Political Science
The University of Edinburgh
Scotland, UK

We explore the pattern, potential drivers, and implications of item non-response on survey questions about domestic abuse. We draw on a longitudinal representative prospective survey on children and their families in Scotland (N = 3646) and use multivariate regression models to look at non-response on domestic violence questions among mothers of young children. By triangulating data from multiple survey sweeps we hypothesise that item non-response may be due to mothers experiencing violence, and we observe that factors which predict experiencing violence also predict item non-response. We compare conservative and generous dependent variables on domestic abuse prevalence and find that both yield similar results in multivariate models, but that the actual social gradient of domestic violence is likely to be steeper than we can see in survey data. We discuss the ethical implications of imputing missing data and argue that sometimes it is unethical to do so.

Keywords: domestic abuse prevalence; domestic violence; ethics of imputation; item non-response; longitudinal data; missing data

1 Introduction

In this paper we wish to better understand the pattern, potential drivers, and implications of item non-response in questions on domestic abuse and abuse within national social surveys. Establishing domestic violence prevalence is methodologically challenging (Myhill, 2017). Prevalence studies rely on the assumption that when sample respondents are asked very sensitive questions they will answer, and that they will answer truthfully. Yet, there are realistically speaking no other ways to establish prevalence than to ask about this in nationally representative studies, especially if the prevalence estimate is to encompass different forms of abuse, such as coercive control and psychological abuse. These newly criminalised types of abuse may be harder to detect in administrative criminal justice data. Only a small proportion (circa 7%) of the published literature on intimate partner violence are prevalence studies according to a systematic review of the literature (Capaldi, Knoble, Shortt, & Kim, 2012).

The prevalence studies that have been undertaken show that prevalence differs enormously across countries. For example, a review of prevalence studies across countries found that the prevalence of lifetime domestic violence (including emotional, physical and sexual abuse) varied from 1.9% in Washington, US, to 70% in Hispanic Latinas in Southeast US (Alhabib, Nur, & Jones, 2010). A recent review of national prevalence estimates in the Americas reports that lifetime physical and/or sexual abuse rates ranged from 14% to 17% among women in Brazil, Panama, and Uruguay to 58.5% in Bolivia (Bott et al. 2019). A WHO multi-site prevalence study focusing on non-industrialised settings and exploring prevalence of sexual and physical abuse in ten countries found that lifetime prevalence of physical and/or sexual partner violence varied from 15% to 71% (Garcia-Moreno, Jansen, Ellsberg, Heise, & Watts, 2006). A 2014 survey conducted by the European Union Agency for Fundamental Rights (2017) among the 28 European Union Member States showed that the lifetime prevalence of physical and/or sexual violence against women by intimate partners was 22%, with the country rate varying between 13% (Austria, Croatia, Poland, Slovenia, Spain) and 32% (Denmark, Latvia), highlighting what Gracia and Merlo (2016) refer to as the

Contact information: Valeria Skafida, 15a George Square, EH8 9LD Edinburgh, UK (E-Mail: valeria.skafida@ed.ac.uk)

Nordic paradox—whereby the countries with a high level of gender equality are also more likely to have higher estimates of intimate partner violence when surveys are undertaken, and these estimates reflect both differences in incidence and differences in perceptions of violence across countries (Piispa, 2003). There are generally few data sources on domestic abuse among mothers of young children, and a recent systematic review of relevant literature—which focuses on prevalence estimates of child abuse but included surveys measuring exposure to domestic abuse—notes that most surveys fail to adequately measure domestic abuse severity in their questionnaires (Mathews, Pacella, Dunne, Simunovic, & Marston, 2020).

Prevalence estimates also differ significantly across methodologies used (Waltermaurer, 2005). For example, in the review by Alhabib et al. (2010), the most frequently used mode was face-to-face interviews (55%), then self-administered questionnaires (30%), and then telephone interviews (13%). Within this review, face-to-face interviews led to higher rates of disclosures of abuse than when using self-reports or telephone interviews. Similar findings have also been previously reported by others (Hamby, Poindexter, & Gray-Little, 1996) who argue that multiple and open-ended questions increase accurate reporting, whereas written screening may underestimate intimate partner violence prevalence (McFarlane, Christoffel, Bateman, Miller, & Bullock, 1991). Contrary to the above, other related studies on the reporting of crime victimisation (Laaksonen and Heiskanen 2014) and non-consensual sexual behaviour (Testa, Livingston, & VanZile-Tamsen, 2005) suggest the opposite, i.e. that greater disclosure rates can be obtained via self-administered surveys and that the presence of an interviewer can be detrimental. It is not clear why the data collection mode might affect domestic abuse reporting differently. Some hypotheses are that studies conducted by an interviewer may be better able to reassure respondents that data will be kept anonymous, or that respondents may be less fearful than when disclosing information via post/computer to people they have never met. Perhaps skilled interviewers are also better able to build rapport with respondents which in turn facilitates the collection of more truthful responses.

There is a considerable body of work looking at survey attrition and non-response, and this is a topic which has interested researchers for decades (Craig & Mccann, 1978; Smith, 1983). The literature is dominated by work which has focused on overall survey attrition in terms of participants not opting into surveys in a variety of survey settings and countries (Abraham, Maitland, & Bianchi, 2006; Eisner, Murray, Eisner, & Ribeaud, 2019; Fulton, 2018; Robert M Groves, 2006; Lynn & Lugtig, 2017; Matsuo, Billiet, Loosveldt, Berglund, & Kleven, 2010). Longitudinal research on survey non-response also suggests non-response is increasing (J. M. Brick & Williams, 2013), though in some settings there is evidence that the correlates of non-response are also transforming over time (Olson & Witt, 2011). In relative terms, there are fewer studies looking at item non-response, defined here as respondents who do not provide a valid response for specific questions within a survey. For the purposes of this paper, a "non-valid response" is defined as one where the response includes skipping a question or selecting some form of "don't know" response option. There are examples of better researched topics in this respect, such as item non-response on income questions, an area which has been researched for three decades (Bollinger, Hirsch, Hokayem, & Ziliak, 2019; Lillard, Smith, & Welch, 1986). There are also studies looking at item non-response to sensitive questions (other than income). For example, Kays et al find that from a range of sensitive questions on health and wellbeing, the most sensitive questions according to the authors, i.e. those on sexual practices, mental health and substance use, were more likely to be left unanswered (Kays, Gathercoal, & Buhrow, 2012). However, there is a paucity of research which has looked at item non-response in surveys asking about domestic violence.

From the few studies there are, we know that item non-response on domestic violence questions is a problem in other survey contexts, such as the Belgian Health Interview Survey (Drieskens, Demarest, D'Hoker, Ortiz, & Tafforeau, 2017). A highly relevant study looking at a Finnish survey on violence against women, and specifically at partial item non-response on abuse questions (Piispa, 2003) found that willingness to respond to questions about partner violence varied by the age and educational level of the respondent, where older women and those with lower educational qualifications were more likely to have partial item non-response on relevant questions. In a different European multi-national study seeking to measure domestic violence against home-dwelling older women (aged >60 years) in Europe (De Donder et al., 2013), 2880 individuals were interviewed by three different data collection modes (i.e. postal, face-to-face, telephone). The researchers found that there was a pattern to the missing values of 34 indicators of abuse, with principal component analysis indicating that response patterns were associated with different types of data collection. These studies are very much the exception though. Even prevalence studies do not seem to report any findings related to item non-response (Garcia-Moreno et al., 2006), and any response related reporting tends to be about overall survey response (Alhabib et al., 2010) rather than item non-response.

When it comes to item non-response in complex surveys, especially longitudinal cohort surveys, it is possible to fill out many pieces of a missing puzzle, and to learn more about any item non-response by triangulating information from different survey sweeps. With respect to understanding prevalence of domestic violence between intimate partners, we aim to explore if this cross-checking exercise can help us better un-

derstand if there are systematic differences between those choosing not to answer questions about abuse and those who do, and how this may affect analyses which only use valid responses. This is an investigation which cannot be undertaken when looking at domestic violence specific surveys where the survey non-response and the topic specific non-response overlap. To our knowledge, there is no study which has undertaken an analysis of this nature, and we hope the findings, but also the research questions we ask, can be used as a guide by future researchers hoping to better understand the nature of item non-response in sensitive questions, and particularly in relation to domestic abuse questions. Within the Total Survey Error paradigm, we discuss the specific sources of error which may affect our own study and similar surveys exploring domestic abuse, and we hope this work can help inform future survey and question design for social surveys enquiring about domestic abuse and other sensitive information.

2 Aims

In this paper we set out to address four key aims.

1. First, we aim to assess the magnitude and nature of item non-response on a domestic violence questionnaire module within a larger longitudinal and nationally representative social survey on parents and children.

2. Second, we explore whether item non-response on domestic violence is stratified by key factors.

3. Third, we assess how likely it is that item non-response is masking unreported experiences of abuse.

4. Fourth, drawing on the previous points, we compare generous versus conservative estimates of domestic violence between current or former intimate partners and reflect on how item non-response may affect the validity and reliability of analyses based only on valid responses.

3 Dataset description

In this paper we use data from the Growing Up in Scotland (GUS; ScotCen Social Research, 2022) longitudinal cohort study. It is a nationally representative prospective study of young children and their families in Scotland. It is the only longitudinal dataset in the UK to have questions on domestic violence and abuse asked of mothers of young children. Though the study overall has three different cohorts, we focus on the one cohort which asked about intimate partner domestic violence. At the first survey sweep this cohort had 5,217 participating families, for which the "study-children" were born between June 2004 and May 2005. The study children were approximately 10 months old at the first interview which took place in participants' homes with children's mothers. Interviews for the 6th sweep of data took place between April 2010 until June 2011. Children are interviewed at the same stage in their development (as close as possible to the target age of 58.5 months) but children's birthdays span across an entire year to account for seasonal effects on various outcomes collected in the survey. Mothers and partners (if present) were asked a range of questions about themselves and their children's development in one-to-one CAPI interviews, and the median and mean interview length was 62 and 70 minutes respectively. Interviews were administered in English and translation was available but was needed only for a handful of participants. The domestic abuse module featured at the half-way mark. Circa 75% of all interviews were conducted by women. It is possible that interviewer gender may have affected response-rates, for example, by women being deterred from responding in the presence of a male interviewer. However, no interviewer would have been able to see the respondent's responses on the self-complete module. Large social surveys should consider making interviewer gender available as a survey variable to allow for this factor to be taken into account in analysis.

The stratified random sample was drawn from a record of Child Benefit claimants, a State welfare benefit virtually all families were eligible to, and which has an estimated 97% coverage of the eligible population. For a survey of this size and for this target population, attrition rates were low (87% response rate of questionnaires issued, and 70% response rate based on sample at sweep 1). The official user guide for the first sweep of data describes the survey design in further detail (Corbett, Marryat, & Bradshaw, 2007). GUS received ethics approval by the Scotland "A" MREC committee, and the research reported in this paper received ethical approval from the University of Edinburgh's School of Social and Political Science ethics committee.

4 Variable description

4.1 Domestic violence and abuse questions

The domestic violence and abuse questionnaire featured once at the 6th survey sweep when children were c.6 years old. At this stage, 3646 mothers were still part of the survey. While the majority of the survey was interviewer led, the domestic violence and abuse module was a self-complete questionnaire given to mothers to complete privately. The interviewers were aware that for sweep 6, the specific self-complete section included sensitive questions on domestic violence. Interviewers were briefed to remain alert to respondents becoming upset and to remind respondents that they could skip questions if they preferred. GUS has separate mother and partner interviews and in theory it would be unlikely that a mother would be answering the self-complete part of her survey with an onlooking partner. We do not know if and how interviewers may have dealt with any sensitive scenarios such as the one described. As part of standard procedure at every annual survey, interviewers gave participants a leaflet with various helplines of relevant organisations. For the 6th sweep, this leaflet also included details, helplines and

websites of organisations which help women suffering domestic abuse. Leaflets were left with all families.

Mothers were asked to report if they had experienced a range of different types of abuse which were conceptually separated into three subsections: (a) coercive control questions, (b) questions about threats, and (c) questions about physical and sexual violence. The module drew on the Scottish Crime and Justice Survey questions used at the time when the GUS questionnaire was designed. The key difference is that GUS asks about experiences of such abuse in a 6-year period from the birth of the study child to the present day, whereas the Scottish Crime and Justice Survey asks respondents to report on violence occurring in the last 12 months. Table 1 below shows the wording of each of the original 12 yes-or-no questions in the survey and full details including response categories can be found in the data user guide (Bradshaw, Corbett, and Tipping n.d.). There were different forms of response and non-response for this module. Some mothers for example, chose to skip the entire module (resulting in missing data on all questions) while others started the module but selected the "don't remember" or "refuse to answer" option for some or even all 12 questions. Figure 1 summarises the item non-response pattern.

4.2 Relationship "health"

One hypothesis is that the item non-response for these questions could be mainly driven by mothers who have experienced abuse and who do not wish to disclose it. In order to understand to what extent this hypothesis might hold, we use information on the nature of the mother's relationship with her partner which had been captured in survey sweeps undertaken before sweep 6 (when the domestic violence and abuse module featured). In sweeps 2 and 4 of the GUS survey (children aged 2y and 4y respectively), mothers were asked a range of questions on the "health" of their relationship with their then partner (where there was a partner present). Of these questions, the 5 questions listed below explored negative relationship patterns with the mother's partner (responses were on a 5-point Likert scale):

- I sometimes feel lonely even when I am with my husband/partner.
- I suspect we may be on the brink of separation.
- How often do you and your partner argue?
- How often is there anger or hostility between you and your partner?
- How often do you have arguments with your partner that end up with people, pushing, hitting, kicking or shoving?

Overall 3432 mothers had a partner (not necessarily the same one) at both sweeps 2 and 4. Using these repeated questions from both sweeps for these mothers, we created a standardised scale (Cronbach's alpha 0.76) to capture relationship "unhealthiness" and hostility. To facilitate with interpretation, this scale was then dichotomised to differentiate the first fifth capturing those mothers who across both sweeps 2 and 4 made the most negative statements regarding their relationship. We use this dichotomous variable to explore whether mothers who had item non-response for domestic abuse questions at sweep 6 had a higher probability of having previously reported negative relationship patterns. This is a crude measure, considering that we have not accounted for whether the mother is reporting on the same partner at either sweeps 2, 4 or 6, or whether this partner has changed at some stage, or who was "responsible" for these behaviours. However, the vast majority of mothers (73% of mothers in the sample to be precise) were with the same partner throughout the six sweeps. Even with this limitation this crude measure was insightful, as we discuss in subsequent sections. Small sample sizes among those with more complex partner histories, and also the relatively small probability of reported intimate partner abuse did not allow for a more complex variable to be constructed.

4.3 Socioeconomic and family characteristics

It is well established that both survey non-response and item non-response is socially stratified (Berinsky, n.d; De Donder et al., 2013). Thus, we use some key socioeconomic characteristics in the analysis. Maternal education is a derived variable which is based on the highest educational qualifications obtained by the mother. Maternal social class status was based on the National Statistics Socio-Economic Classification (NS-SEC) scheme. This is calculated on the basis of information regarding the individual's working conditions, job security, timing of payments, opportunities for promotion and incremental pay (Rose, O'Reilly, & Martin, 1997). A banded variable indicating the mother's highest educational qualifications was used for the analysis to represent the mother's educational level. We also control for equivalised household income. Income data were obtained by asking the mother to select one of 17 income bands that reflected total household income before tax. Equivalised income was calculated using the Organisation for Economic Co-operation and Development (OECD) modified equivalence scales and procedure (Chanfreau & Burchardt, 2008). The mother's age is presented in banded form. More details on the derivation and operationalisation of all control variables can be found in Supplementary File 1.

4.4 Generous versus conservative prevalence estimates

In Table 4, we compare a conservative and a generous estimate of domestic violence using two binary dependent variables. The conservative binary estimate of experiencing any domestic violence includes mothers who answered "yes" to experiencing any abuse (irrespective of how many types) and also includes mothers who answered "yes" to at least one abuse question, even if they refused/forgot to answer

Table 1
Growing Up in Scotland survey questions on domestic abuse at sweep 6 (children aged 6y)

In the time since child was born, has any partner or ex-partner ever done any of the following things to you? [yes/no]
Coercive control
Stopped you having a fair share of the household money or taken money from you
Repeatedly put you down so that you felt worthless
Behaved in a jealous or controlling way, e.g. restricting what you can do, who you can see, what you can wear
Physical abuse
Pushed you or held you down
Kicked, bitten or hit you
Choked or tried to strangle/smother you
Used a weapon against you, for example and ashtray or a bottle
Forced you or tried to force you to take part in any sexual activity when you did not want to
Threats of abuse
Threatened to hurt you
Threatened to hurt someone close to you, such as your children, family members, friends or pets
Threatened to, attempted to, or actually hurt themselves as a way of making you do something or stopping you from doing something
Threatened to kill you

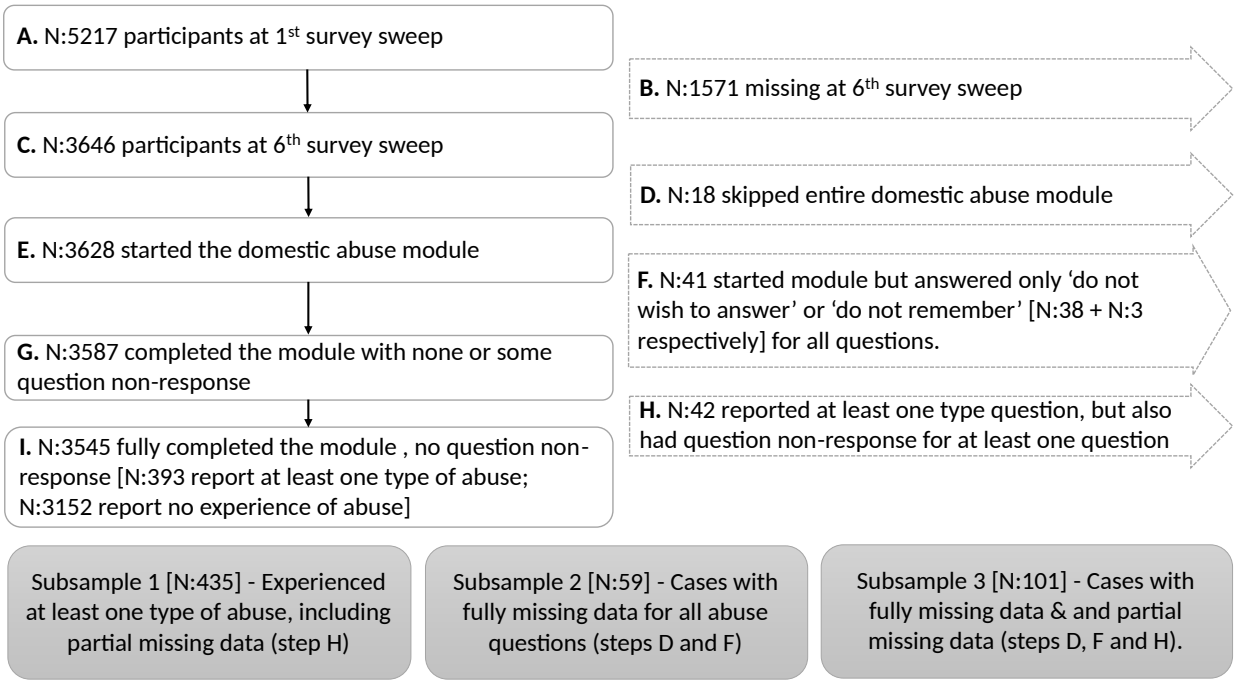

Figure 1. Outline of response and non-response patterns for domestic abuse module in Growing Up in Scotland survey (Sweep 6)

any other of the twelve questions (Subsample 1 from Figure 1). The generous binary variable of having experienced any abuse includes the above 435 mothers as well as the 59 mothers who had complete missing data for all questions relating to abuse (Subsample 1 and Subsample 2 described in Figure 1). The generous measure of abuse made the assumption that all full item non-response for all questions of the domestic abuse module indicated that mothers had experienced abuse, and coded them as such. We discuss later whether and in what way using the conservative versus generous variable affects an analysis of prevalence.

5 Statistical analysis

We carried out weighted bivariate descriptive statistics and proceeded to use a series of multivariate regression models. We used multivariate multinomial regression (Table 4) to estimate Risk Ratios for independent variables predicting different response types to the domestic violence questions and we differentiated between three different outcomes: those who did not experience any domestic violence and did not have item non-response; those who experienced domestic violence and did not have any item non-response; and finally those who had either partial or full item non-response for any or all of the domestic violence questions. We also used binary logit models (Table 5) to look at the relationship between our independent variables and two versions—the conservative and the generous version—of a dependent variable measuring experience of domestic violence and abuse. The conservative measure only coded as "having experienced violence" those mothers who reported experiencing violence in their answers. Since simply comparing odds ratios between different models is not advisable (Mood, 2010), we use the Stata margins command to calculate marginal effects of experiencing violence which allow us to compare results across different binary logit models to see if the relationship between independent variables and experiencing abuse changes depending on which measure is being used (see Table 5). Marginal Effects (ME) express the calculated probability of experiencing violence for each independent variable and category of categorical variables in the model. These show the change in probability when the predictor or independent variable increases by one unit. For categorical variables, the change is from the reference category to a different one. For each ME, the values of other independent variables are held constant at their weighted mean value. For analyses, the appropriate longitudinal survey weights have been applied to adjust for attrition. Stata 15.1 was used for all analyses.

6 Results

6.1 Extent and stratification of item non-response for domestic violence questions

As figure 1 summarises, of the 3646 mothers in the survey at sweep six, there are 393 who have provided a valid answer for all 12 domestic violence questions and who report experiencing at least one type of abuse. There is a further 42 mothers who reported at least one type of abuse, but who did not answer all 12 questions with valid responses. Together, these two groups give us a subsample of 435 mothers whom we know experienced at least one type of abuse. Subsample 2 is focusing on mothers for whom we have no insight at all on whether they experienced abuse. These 59 mothers either skipped the entire self-complete module, or filled in the module by giving a non-valid response to each of the 12 questions. Subsample 3 encompasses both subsample 2—so, the 59 mothers with complete item non-response for all abuse questions, as well as the 42 mothers included in Subsample 1, who reported some form of abuse, but who then also provided non-valid responses for other questions. Subsample 3 is a useful category when we wish to explore the extent to which full or partial item non-response is socially stratified.

We also explored non-response for each of the three categories of abuse questions, and found a statistically significant difference (X^2, $p < 0.05$) between the three categories in terms of the probability of respondents reporting that they didn't remember a certain type of abuse occurring or that they did not wish to answer. This is also evident when comparing ratios of don't remember/don't wish to answer responses from the total reported incidence of abuse for each of the three categories (Table 2). These suggest that respondents were more likely to report "not remembering" when answering the coercive control questions, and more likely to "not wish to respond" when answering questions about physical abuse and threats of abuse questions.

Table 3 shows the patterns of social stratification for each of these subsamples in descriptive bivariate analysis. Whether one looks at maternal education, maternal social class, her age at birth, or equivalised household income, the trend is generally the same: mothers from more disadvantaged circumstances are both more likely to experience abuse, and more likely to have either full item non-response or partial item non-response. To elucidate with some extracted examples, 28% of mothers aged under 20 at the time of birth of the study child experienced any abuse compared to 9% for those aged 30 or older. As for item non-response, 1% of mothers with degree level qualifications or above had full or partial item non-response, compared to 9% of mothers with no qualifications.

In table 4 we see that some correlations between socioeconomic factors and item non-response remain significant even in multivariate multinomial regression models. We also

Table 2
Breakdown of non-response type by violence type

	"I don't remember"	"I don't wish to answer"	Total who reported violence type occuring	Rato: don't remember/reported violence	Ratio: don't wish to answer /reported violence
Coercive control	45	87	336	0.13	0.26
Physical abuse	7	71	224	0.03	0.32
Threats of abuse	13	73	185	0.07	0.39

see that some independent variables which predict item non-response also predict experiencing abuse (with coefficients of similar magnitude and similar direction). For example, mothers who have no item non-response in the lowest income fifth have a risk ratio (RR) of 3.55 to 1 of experiencing abuse, and also a RR of 3.97 of having full or partial item non-response on the domestic abuse survey module. Having missing income data was the variable most strongly predictive of full or partial item-non response (RR 5.80).

6.2 Domestic abuse item non-response and prior "unhealthy" relationship

Another one of our aims was to explore if item non-response at sweep 6 correlated with the mother reporting an unhealthy relationship with a partner at sweeps 2 and 4. This variable is described in detail in the methods section. The dichotomised scale differentiates between the fifth who gave the most negative answers when reflecting on their relationships with their partners at sweep 2 and 4, compared to the rest of the sample. This variable is focusing only on the subset of mothers who had a partner at either or both sweeps 2 and 4, thus it excludes 341 mothers (c.9% of sweep 6 valid sample) who did not have a partner at both sweeps 2 and 4. As a result, we are looking at a reduced subsample, and at outcomes with a small sample size to begin with, so the cell counts for this variable and "full item non-response" are too small (N:6) to be meaningful. As for full and partial item non-response, it seems that mothers who described relationships at sweep 2 and 4 as most hostile were more likely to either not answer domestic abuse questions, or to only answer these partially (4.4% compared to 1.4% among those in less hostile relationships). In multivariate models (Table 4) we see that controlling for several socio-economic confounders, mothers in the most "unhealthy" relationships at sweeps 2 and 4 were far more likely to both experience abuse (RR 4.34) and to not answer some or all abuse questions (RR 3.70).

6.3 Comparing generous and conservative estimates of domestic violence and abuse

We have found that mothers who previously reported being in "poorer" relationships had an increased chance of having full or partial item non-response at sweep 6 on questions about domestic violence compared to mothers who had not reported that. Thus, we create a generous dichotomous variable of experience of "any abuse" which encompasses the sub-sample of mothers for whom we have full item non-response on all domestic abuse questions. We compare this to a more conservative dichotomous measure of abuse which only categorises as experiencing "any abuse" mothers who reported this for at least one question (this includes mothers with partial item non-response). Table 5 shows both Odds Ratios (ORs) for both the conservative and generous dependent variables, as well as Marginal Effects (MEs) to aid comparison of the two models. The reference categories for each categorical variable are assigned an OR of 1, and a ME of 0.00. Negative MEs, such as those for maternal education, indicate that categories of independent variables have a lower predicted probability than the set reference category with a ME of 0.00, and these values also correspond to ORs below 1.

For example, using the conservative measure of violence, the predicted probability of the lowest income fifth experiencing violence is 14 points higher than for the highest income fifth (the reference category) when all other covariates are held constant at the weighted mean. In terms of general trends, experience of domestic violence, regardless of which binary variable is used, is highly socially stratified, as per the bivariate statistics in Table 3. Mothers in more disadvantaged circumstances are more likely to have experienced abuse, with the exception for the variable on maternal education where the trend has been reversed in the multivariate model compared to the bivariate analysis, and we discuss these socioeconomic stratification of experiences of violence in further detail elsewhere (Skafida, Morrison, & Devaney, 2021). In terms of differences between the two models, classifying mothers with full item non-response as having experienced "any abuse" led to the following interesting differences between the conservative and generous variables. Overall the

Table 3
What proportion of mothers responded in what way to the domestic abuse module – bivariate analysis of response outcomes by key independent variables

	No abuse and no item non-response			Any abuse—incl. partial item non-response[a]			Full item non-response on all abuse questions[b]			Full and partial item non-response		
		95% CI			95% CI			95% CI			95% CI	
Weighted data	%	lower	upper	%	lower	upper	%	lower	upper	%	lower	upper
Maternal Education												
Degree or equivalent	90	88	92	10	8	12	1	0	1	1	1	2
Vocational qualifications	83	81	85	16	14	17	2	1	2	4	3	5
Higher grade or equivalent	92	87	95	8	5	12	0	0	2	1	0	4
Standard grade	82	78	86	15	11	19	3	2	6	4	2	7
No qualifications	76	69	82	17	12	23	7	5	12	9	6	14
Equivalised Income												
1st quintile	73	69	76	23	20	27	4	2	5	6	5	8
2nd	84	81	87	14	12	18	2	1	3	3	2	5
3rd	86	83	89	13	10	16	1	1	3	3	2	5
4th	93	91	95	6	5	8	0	0	1	0	0	1
5th	93	91	95	6	5	8	0	0	1	1	0	2
Missing income data	85	78	90	9	6	14	6	3	12	8	4	14
Maternal NS-SEC												
Managerial and professional	91	89	92	8	7	10	0	0	1	1	1	2
Intermediate	85	81	88	14	11	18	1	0	3	2	1	4
Small employers/own account holders	87	82	91	10	7	15	3	1	6	3	1	6
Lower supervisory and technical	83	76	88	15	10	22	2	1	5	4	2	8
Semi-routine and routine	71	68	74	24	21	28	4	3	7	8	6	10
Never worked	67	49	81	22	11	39	11	4	26	11	4	26
Mother's age at childbirth												
Under 20	70	61	78	28	21	37	2	0	8	5	2	11
20–29	81	79	83	16	14	18	3	2	4	4	3	6
30–39	90	88	91	9	8	10	1	1	2	2	1	3
40 or older	90	83	94	9	5	15	1	0	5	2	1	6
"Unhealthy" relationship scale Age 2–4[d]												
Quintile 2–5	92	91	93	7	6	8	1	1	2	1	1	2
Quintile 1 (least healthy)	72	68	76	27	23	31	1	0	2	4	3	7
N	3152			435			59			101[c]		

[a] Mothers reported experiencing at least 1 type of abuse, but for at least 1 question in the module they refused or forgot the answer.
[b] Mothers either skipped the entire domestic abuse module (N=18) or completed the module by refusing (N=38) or responding "don't remember" for all questions (N=3). [c] Columns 1–3 add up to 100%. Column 4 takes the 59 cases from column 3, and the 42 cases from column 2 who had partial non-response and groups them together. [d] Variable valid only for mothers living with partners at sweep 2 or 4 (N=3242).

marginal effects for experiencing abuse were almost identical, and with an under three percentage points difference from each other when comparing the two models.

There were two differences that were of three percentage points or greater. For mothers who had never been in work, the margianal effects of experiencing abuse changed from 3% with the conservative variable, to 10% with the generous variable. Also, the predicted probability of experiencing abuse was 5 percentage points greater with the generous dependent variable for mothers who had missing income data (ME of 7% versus 2%). For all covariates including these two, the point estimates for the "conservative" and the "generous" models were within each other's confidence intervals, which—in the absence of a better metric—is one way of determining that the generous variable is not differing substantially from the conservative one.

Table 4
Multivariate multinomial logistic regression of different response categories for the domestic abuse module (N:3634)

	Any abuse & no item non-response[a]			Full or partial item non-response[a]			Any abuse & no item non-response			Full or partial item non-response		
		95% CI			95% CI			95% CI			95% CI	
Weighted data	RR	low.	upp.	RR	low.	upp.	RR	low.	upp.	RR	low.	upp.
"Unhealthy" relationship scale[b]												
Age 2–4												
Quintile 2–5	-	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	1.00
Quintile 1 (least healthy)	-	-	-	-	-	-	4.34**	3.25	5.79	3.70**	2.02	6.80
Maternal Education												
Degree or equivalent	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Vocational qualifications	0.76	0.54	1.08	1.33	0.63	2.81	0.71*	0.49	1.03	1.34	0.56	3.23
Higher grade or equivalent	0.36**	0.19	0.69	0.53	0.14	1.99	0.44**	0.23	0.85	0.91	0.24	3.40
Standard grade	0.58**	0.37	0.92	1.19	0.51	2.75	0.59**	0.37	0.93	1.09	0.38	3.08
No qualifications	0.58**	0.33	1.00	1.87	0.73	4.79	0.48*	0.22	1.03	2.52	0.67	9.45
Equivalised Income												
1st quintile	3.55**	2.15	5.85	3.97**	1.20	13.15	2.91**	1.54	5.51	2.17	0.55	8.56
2nd	2.05**	1.28	3.29	2.35	0.68	8.17	1.84**	1.10	3.09	1.77	0.47	6.71
3rd	1.82**	1.16	2.88	2.95*	0.84	10.32	1.51	0.91	2.50	2.81	0.77	10.27
4th	1.09	0.70	1.70	0.60	0.13	2.68	1.12	0.73	1.73	0.58	0.13	2.58
5th	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Missing income data	1.35	0.70	2.60	5.80**	1.60	21.08	1.02	0.49	2.09	4.02*	0.91	17.77
Maternal NS-SEC												
Managerial and professional	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Intermediate	1.23	0.86	1.75	1.27	0.54	2.99	1.18	0.78	1.81	1.31	0.51	3.38
Small employers/own account holders	0.95	0.58	1.57	1.53	0.63	3.72	0.87	0.48	1.59	1.85	0.74	4.64
Lower supervisory and technical	1.24	0.75	2.05	2.11*	0.87	5.12	1.28	0.73	2.27	1.34	0.41	4.40
Semi-routine and routine	1.62**	1.09	2.42	3.63**	1.79	7.36	1.55*	0.95	2.52	2.53**	1.18	5.41
Never worked	1.69	0.65	4.41	4.49**	1.25	16.15	0.38	0.04	3.89	0.00**	0.00	0.00
Mother's age at childbirth												
Under 20	2.46**	1.10	5.50	1.35	0.30	6.06	2.37	0.80	7.02	0.33	0.05	2.29
20–29	1.70	0.83	3.51	1.81	0.50	6.50	2.22	0.91	5.39∗	1.06	0.29	3.83
30–39	1.19	0.59	2.43	1.19	0.32	4.41	1.28	0.53	3.13	0.46	0.12	1.80
40 or older	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Mother's ethnicity												
White	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Other ethnic background	0.64	0.30	1.35	0.95	0.33	2.77	0.68	0.28	1.68	1.00	0.25	3.95
Number of children at home												
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.57**	0.43	0.76	0.97	0.57	1.67	0.60**	0.42	0.85	2.00	0.91	4.43∗
3	0.66**	0.46	0.94	0.89	0.47	1.69	0.64*	0.41	1.00	1.21	0.41	3.53
4 or more	0.79	0.51	1.22	1.00	0.40	2.50	0.88	0.53	1.47	0.81	0.17	4.01
Child's gender												
Male	1.20*	0.97	1.49	1.20	0.80	1.81	1.05	0.81	1.36	0.98	0.57	1.67
Female	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
N		390			99			290			57	

[a] The reference category for both models are mothers who experienced no abuse and who had no item non-response.
[b] Excl. mothers with no partner at sweep 2 & 4 (N=3233)
* $p \leq 0.1$ ** $p \leq 0.05$

Table 5
Multivariate multinomial logistic regression of different response categories for the domestic abuse module (N:3634)

	Conservative: Any abuse			Generous: Any abuse + full item non-response[a]			Conservative: Any abuse			Generous: Any abuse + full item non-response[a]		
		95% CI			95% CI			95% CI			95% CI	
Weighted data	OR	low.	upp.	OR	low.	upp.	ME	low.	upp.	ME	low.	upp.
Maternal Education												
Degree or equivalent	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Vocational qualifications	0.81	0.58	1.13	0.83	0.59	1.16	−0.02	−0.06	0.01	−0.02	−0.07	0.02
Higher grade or equivalent	0.39**	0.21	0.74	0.38**	0.20	0.71	−0.08**	−0.13	−0.04	−0.09**	−0.14	−0.04
Standard grade	0.55**	0.35	0.87	0.65**	0.44	0.97	−0.06**	−0.10	−0.02	−0.05**	−0.09	−0.00
No qualifications	0.56**	0.33	0.94	0.75	0.45	1.26	−0.06**	−0.11	−0.01	−0.03	−0.10	0.03
Equivalised Income												
1st quintile	3.55**	2.23	5.67	3.49**	2.17	5.61	0.14**	0.09	0.18	0.15**	0.09	0.20
2nd	2.05**	1.29	3.26	2.04**	1.29	3.20	0.06**	0.02	0.10	0.07**	0.03	0.11
3rd	1.90**	1.22	2.96	1.94**	1.26	3.00	0.05**	0.02	0.09	0.06**	0.02	0.10
4th	1.01	0.65	1.57	1.04	0.68	1.57	0.00	−0.03	0.03	0.00	−0.03	0.03
5th	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Missing income data	1.33	0.72	2.47	2.07**	1.19	3.60	0.02	−0.03**	0.07	0.07	0.01	0.13
Maternal NS-SEC												
Managerial and professional	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Intermediate	1.20	0.86	1.68	1.24	0.90	1.72	0.02	−0.02	0.05	0.02	−0.01	0.06
Small employers/own account holders	0.88	0.54	1.42	1.05	0.68	1.61	−0.01	−0.05	0.03	0.00	−0.04	0.05
Lower supervisory and technical	1.31	0.85	2.03	1.38	0.91	2.09	0.03	−0.02	0.08	0.04	−0.01	0.09
Semi-routine and routine	1.64**	1.15	2.36	1.94**	1.36	2.77	0.05**	0.01	0.10	0.08**	0.04	0.13
Never worked	1.36	0.55	3.35	2.14	0.92	4.97	0.03	−0.07	0.13	0.10	−0.03	0.23
Mother's age at childbirth												
Under 20	2.60**	1.30	5.21	2.17**	1.06	4.45	0.11**	0.04	0.18	0.09**	0.01	0.17
20–29	1.66	0.85	3.23	1.73	0.90	3.31	0.05*	−0.01	0.10	0.06*	−0.00	0.12
30–39	1.15	0.59	2.28	1.20	0.63	2.26	0.01	−0.04	0.06	0.02	−0.04	0.07
40 or older	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Mother's ethnicity												
White	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Other ethnic background	0.64	0.32	1.29	0.72	0.35	1.50	−0.04	−0.09	0.01	−0.03	−0.10	0.03
Number of children at home												
1	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.65**	0.48	0.88	0.63**	0.49	0.81	−0.05**	−0.08	−0.01	−0.06**	−0.09	−0.02
3	0.73	0.51	1.04	0.70**	0.53	0.92	−0.04*	−0.08	0.00	−0.05**	−0.08	−0.01
4 or more	0.91	0.60	1.39	0.82	0.53	1.28	−0.01	−0.06	0.04	−0.03	−0.08	0.03
Child's gender												
Male	1.22**	1.01	1.48	1.20	0.99	1.45	0.02**	0.00	0.04	0.02*	−0.00	0.04
Female	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
N		435			494			435			494	

[a] Fully missing data from item non-response coded as mother experiencing abuse. [b] Marginal effects calculated after logit models.
* $p \leq 0.1$ ** $p \leq 0.05$

7 Discussion

We set out to address four key study aims. Firstly, we hoped to assess the scale and nature of item non-response on domestic abuse questions within the chosen survey setting. What we found was that from a working sample of 3646 mothers 10.8% of mothers experience at least one form of abuse, when we look at mothers who answered all domestic abuse questions. This proportion is 11.9% when including partial item non-response, and 13.6% if assuming that full item non-response is masking experiences of abuse. The most common reason for partial or full item non-response was due to reluctance to disclose this information, such as mothers opting out of the entire domestic violence survey module or mothers responding "do not wish to answer".

The second aim was to understand how item non-response is stratified by key socio-economic and family characteristics. Here we find that measures of socioeconomic disadvantage are strongly correlated with a higher probability of partial and full item non-response. Younger mothers, those with fewer educational qualifications, those in lower paying jobs or not in employment, and those with lower household incomes had a higher probability of not responding to domestic abuse questions. These findings are partially in line with what Piispa reports on a Finnish sample of women, which in line with our findings showed that those with lower educational qualifications were more likely to have partial item non-response on abuse questions. However, while Piispa reports that older mothers were more likely to have item no-response, we find that it is the younger mothers who are most likely to not respond (Piispa, 2003). Piispa notes that the category of older women in the Finish surveys is broad and refers to mothers aged 44–64, whereas the GUS categorisation of the age variable is looking at mothers' age at the time of the birth of the study child, and thus is capturing a different phenomenon. As such, the GUS age ranges are more narrow with most women in the 40+ category aged between 40–45 and only a handful more aged 46–50. It was also evident that a tendency to not respond to one sensitive question was correlated with a tendency to not respond to other sensitive questions, seeing as missing income data was strongly correlated with item non-response for domestic abuse questions, both in bivariate and multivariate analyses.

Thirdly, we sought to explore whether item non-response is likely to be driven by experiences of abuse. We found that in both bivariate and multivariate analysis, having previously reported an "unhealthy" relationship with a partner at previous survey sweeps was very strongly associated with both reporting experiences of abuse at sweep 6 and also with partial or full item non-response for domestic abuse questions. The effect size of prior relationship "unhealthiness" predicting non-response was large despite the numerous covariates in the multivariate models. We hypothesise on the back of this result that having experienced abuse is likely to be a key driver of refusal to answer these questions.

The above brings us to our fourth aim, which is to compare how a generous measure (which includes non-responders) versus a conservative measure (which excludes non-responders) of domestic abuse "perform" in multivariate analysis of domestic abuse experiences. This is done in order to gauge to what extent and in what way using a conservative measure of domestic abuse may misrepresent the pattern of social stratification of domestic abuse experiences. We find from this analysis that there are no great changes between a conservative versus a generous measure of abuse in our sample, and that the two measures yield roughly similar results in terms of coefficient sizes and directions.

7.1 Next steps and ethical considerations

This paper is part of a research programme where a robust indicator of experience of domestic violence and abuse is of key importance. There are potentially two ways forward from this point. One would be to use the generous measure of domestic abuse as the go-to variable. This would not be too dissimilar to implementing the so-called "golden standard" for dealing with missing data, which is what multiple imputation is often considered to be, and which has been used in other research looking at perpetration of intimate partner violence (Emery, 2010). Multiple imputation is a data-led technique which extrapolates what the responses for missing data are likely to be based on what the collected responses are for participants with similar characteristics to those for whom data is missing (Schafer, 1999). The other approach would be to argue that since the results remain largely unchanged it may be preferable to use the conservative variable for substantive analyses on ethical and moral grounds. All survey research remains possible simply due to the willingness and time that participants devote to answering questions. In so doing, when they choose not to disclose information, this choice should be respected. Using the conservative measure avoids having to impose answers onto participants against their will. Yet, having carried out this methodological exercise does help defend the robustness of the conservative variable on domestic abuse against criticisms related to under-reporting, and more importantly it shows us what the pattern of underreporting is likely to be and therefore how this may affect analyses and interpretation.

It is useful to contextualise our case-study analysis of non-response within the broader paradigm of Total Survey Error (TSE). In brief, this is a theoretical framework through which to conceptualise the various sources of error which may ultimately affect data quality (Biemer, 2010). Though there have been many different conceptualisations of the TSE idea, though generally they tend to differentiate in some way between sampling and non-sampling error (Robert M. Groves, 2009; Robert M. Groves & Lyberg, 2010; Lavrakas, 2008). In any prevalence survey on domestic violence and abuse

there are likely to be several sources of error in prevalence estimates, as summarised below:

• Stage 1—*Sampling errors* that occur in generating the sampling frame and obtaining the sample. In the specific case of domestic abuse research underreporting can happen from not being able to include women who are in non-permanent accommodation due to them missing from most existing/accessible sampling frames.

• Stage 2—*Unit non-response error* within longitudinal surveys such as the one used in this study comes both from respondents failing to take part in the initial survey, as well as from survey drop-out after the initial wave, as discussed among those who have sought to implement TSE theory in a longitudinal survey context (Lynn & Lugtig, 2017).

• Stage 3—*Item non-response error* refers to the error which arises from specific questions being left unanswered within a broader survey, as has been the focus of this paper.

• Stage 4—*Measurement error* can arise from different sources, but the sensitivity of domestic abuse as a research topic means that in this context error can arise from participant non-disclosure or from untruthful answers being given to questions on domestic abuse.

From our analysis, we see that the socio-economic variables most strongly linked to experiencing domestic abuse, are also those which are most strongly linked with item non-response on abuse questions (Stage 3). From analysis of GUS attrition carried out elsewhere (Bradshaw, Corbett, & Tipping, n.d.), we know these are the same variables which predict higher survey attrition in the GUS survey (Stage 2) and which survey weights aim to correct for. We also know that some of these characteristics also predict loss of participants through Stage 1, sometimes referred to as noncoverage of a sampling frame (J. Brick & Kalton, 1996) (e.g. not being in sampling frames due to not having a permanent address). What we do not know is whether and in what way Stage 4 loss of information is socially stratified. Nonetheless, we are looking at multiple instances which can lead to misreporting, and in this case most likely underreporting of domestic abuse prevalence. However, there are still interesting insights to be gained by looking at the data we do have, so long as it is not forgotten that any social gradient in domestic abuse prevalence is likely to be much steeper in reality than what is detectable in most prevalence surveys. Walby and Towers (2017) offer a promising alternative questionnaire for prevalence surveys of gendered violence, and a critical analysis of the flaws of existing leading questionnaires already in use in the UK.

7.2 Limitations

Some study limitations are worth highlighting. This is an in-depth study of item non-response in this specific survey and so the extent to which the patterns we observe may be generalised to other surveys and populations is not clear. As mentioned previously, our "unhealthy relationship" measure is a crude variable with some obvious flaws. It does not take into account of who is "responsible" for these behaviours. Nor does it take account if a mother's responses are about the same partners at either sweeps 2, 4 or 6, or whether this partner has changed at some stage. The main weakness of the relationship variable would be that, since it groups together answers from both sweeps 2 and 4, a mother in a very unhealthy relationship at sweep 2 who either re-partnered or separated by sweep 4, would score overall lower on the combined scale capturing relationships at both sweep 2 and 4. Having said that, most mothers (73%) were with the same partner throughout the six sweeps, and small sample sizes among those with more complex partner histories, and also the relatively small probability of reported intimate partner abuse did not allow for a more complex variable to be constructed. Another consideration to note is that some of our analyses involved variables with small sub-sample sizes, and this may have hindered our ability to detect statistically significant results.

8 Conclusion

We set out to study item non-response on domestic violence questions in surveys with four central aims. First, we aimed to assess the magnitude and nature of item non-response on domestic violence questions. We find that the prevalence estimate of domestic violence among mothers of young children in Scotland varies between 10.8% and 13.6% depending on how item non-response is dealt with. Second, we explored whether item non-response on domestic violence questions is stratified by key factors and find that item non-response is socially stratified, and mothers from more disadvantaged backgrounds are more likely to have not responded to some or all domestic violence questions. Third, we assessed how likely it is that item non-response is masking unreported experiences of abuse. We do this by using information on the characteristics of mother' relationships which had been provided in survey sweeps which preceded the domestic violence questions. We find that the same characteristics which predict experiencing domestic violence also predict item non-response for these questions, and we also find that early indicators that the mothers were in unhealthy relationships were strongly associated with not answering domestic violence questions later in time. Fourth, we compared a generous versus a conservative estimate of domestic violence, where the generous estimate assumed that all item non-response reflected undisclosed experiences of violence. We find that in a multivariate model looking at prevalence of violence, we get similar results with either measure. However, the social gradient of violence is likely to be steeper in the population than what we can see in survey data.

We believe that our research may prove of interest to researchers interested in a range of different sensitive topics,

since we show that simply ignoring complete non-response to a module does not appear to bias results. This could be particularly useful for survey contexts where there are not enough variables or survey sweeps to undertake the type of "triangulation" exercise we have done here. We conclude by reflecting on the ethics of filling in the missing pieces when it comes to item non-response in surveys. This relates to data imputation techniques as well as situations where one is triangulating information using multiple survey sweeps to obtain an educated guess of what a missing response may have been. It is important to know, especially for research into highly sensitive topics, the extent to which survey estimates may be valid and representative, and undertaking careful analysis of non-response can help us become better aware of the strengths and weaknesses of our key variables. What each researcher does from then on will depend on context, but we recommend researchers ask themselves: how would the respondent who chose to not answer this question feel about being assigned an answer?

Acknowledgement

This study is supported by the UK's Nuffield Foundation grant number WEL/43875.

References

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American Time Use Survey. *Public Opinion Quarterly*, *70*(5), 676–703. doi:10.1093/poq/nfl037

Alhabib, S., Nur, U., & Jones, R. (2010). Domestic violence against women: Systematic review of prevalence studies. *Journal of Family Violence*, *25*(4), 369–82. doi:10.1007/s10896-009-9298-4

Berinsky, A. J. (n.d). Survey non-response. In W. Donsbach & M. Traugott (Eds.), *The SAGE handbook of public opinion research* (pp. 309–22). London: Sage.

Biemer, P. P. (2010). Total survey error: Design, implementation and evaluation. *Public Opinion Quarterly*, *74*(5), 817–48. doi:10.1093/poq/nfq058

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., & Ziliak, J. P. (2019). Trouble in the tails? What we know about earnings nonresponse 30 years after Lillard, Smith and Welch. *Journal of Political Economy*, *127*(5), 2143–85. doi:10.1086/701807

Bradshaw, P., Corbett, J., & Tipping, S. (n.d.). *Growing up in Scotland sweep 6: 2010–2011 user guide.* Edinburgh: Scottish Centre for Social Research.

Brick, J. M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The Annals of the American Academy of Political and Social Science*, *645*(1), 36–59. doi:10.1177/0002716212456834

Brick, J., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, *5*(3), 215–38. doi:10.1177/096228029600500302

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner Abuse*, *3*(2), 231–80. doi:10.1891/1946-6560.3.2.231

Chanfreau, J., & Burchardt, T. (2008). *Equivalence scales: Rationales, uses and assumptions.* Edinburgh: Scottish Government.

Corbett, J., Marryat, L., & Bradshaw, P. (2007). *Growing up in scotland sweep 1–2005: User guide*. Edinburgh: Scottish Centre for Social Research.

Craig, C. S., & Mccann, J. M. (1978). Item nonresponse in mail surveys: Extent and correlates. *Journal of Marketing Research*, *15*(2), 285–89. doi:10.1177/002224377801500215

De Donder, L., Lang, G., Penhale, B., Ferreira-Alves, J., Tamutiene, I., Verte, D., & Luoma, M. L. (2013). Item non-response when measuring elder abuse: Influence of methodological choices. *The European Journal of Public Health*, *23*(6), 1021–26. doi:10.1093/eurpub/cks172

Drieskens, S., Demarest, S., D'Hoker, N., Ortiz, B., & Tafforeau, J. (2017). Is a health interview survey an appropriate tool to assess domestic violence? *European Journal of Public Health*, *27*(5), 903–9. doi:10.1093/eurpub/ckx078

Eisner, N. L., Murray, A. L., Eisner, M., & Ribeaud, D. (2019). A practical guide to the analysis of non-response and attrition in longitudinal research using a real data example. *International Journal of Behavioral Development*, *42*(1), 24–34. doi:10.1177/0165025418797004

Emery, C. R. (2010). Examining an extension of johnson's hypothesis: Is male perpetrated intimate partner violence more underreported than female violence? Item non-response in the project on human development in Chicago neighborhoods. *Journal of Family Violence*, *25*(2), 173–81. doi:10.1007/s10896-009-9281-0

European Union Agency for Fundamental Rights. (2017). Violence against women: An EU-wide survey. Retrieved from https://fra.europa.eu/en/publication/2014/violence-against-women-eu-wide-survey-main-results-report#

Fulton, B. R. (2018). Organizations and survey research: Implementing response enhancing strategies and conducting nonresponse analyses. *Sociological Methods & Research*, *47*(2), 240–276. doi:10.1177/0049124115626169

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L., & Watts, C. H. (2006). Prevalence of intimate partner violence: Findings from the WHO multi-country study

on women's health and domestic violence. *The Lancet*, *368*(9543), 1260–1269. doi:10.1016/S0140-6736(06)69523-8

Gracia, E., & Merlo, J. (2016). Intimate partner violence against women and the nordic paradox. *Social Science & Medicine*, *157*, 27–30. doi:10.1016/j.socscimed.2016.03.040

Groves, R. M. [Robert M]. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–75. doi:10.1093/poq/nfl033

Groves, R. M. [Robert M.]. (2009). *Survey methodology* (2nd ed.). Hoboken, N.J: Wiley.

Groves, R. M. [Robert M.], & Lyberg, L. (2010). Total survey error: Past, present and future. *Public Opinion Quarterly*, *74*(5), 849–79. doi:10.1093/poq/nfq065

Hamby, S. L., Poindexter, V. C., & Gray-Little, B. (1996). Four measures of partner violence: Construct similarity and classification differences. *Journal of Marriage and the Family*, *58*(1), 127. doi:10.2307/353382

Kays, K., Gathercoal, K., & Buhrow, W. (2012). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior*, *28*(1), 251–256. doi:10.1016/j.chb.2011.09.007

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods.* Thousand Oaks, Calif.: SAGE Publications.

Lillard, L., Smith, J. P., & Welch, F. (1986). What do we really know about wages? The importance of nonreporting and census imputation. *Journal of Political Economy*, *94*(3), 489–506. doi:10.1086/261386

Lynn, P., & Lugtig, P. J. (2017). Total survey error for longitudinal surveys. In P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, … B. T. West (Eds.), *Total survey error in practice* (pp. 279–98). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Mathews, B., Pacella, R., Dunne, M. P., Simunovic, M., & Marston, C. (2020). Improving measurement of child abuse and neglect: A systematic review and analysis of national prevalence studies. *15*(1), e0227884. doi:10.1371/journal.pone.0227884

Matsuo, H., Billiet, J., Loosveldt, G., Berglund, F., & Kleven, Ø. (2010). *Measurement and adjustment of non-response bias based on non-response surveys: The case of Belgium and Norway in the European Social Survey round 3*. doi:10.18148/SRM/2010.V4I3.3774

McFarlane, J., Christoffel, K., Bateman, L., Miller, V., & Bullock, L. (1991). Assessing for abuse: Self-report versus nurse interview. *Public Health Nursing*, *8*(4), 245–50. doi:10.1111/j.1525-1446.1991.tb00664.x

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do and what we can do about it. *European Sociological Review*, *26*(1), 67–82. doi:10.1093/esr/jcp006

Myhill, A. (2017). Measuring domestic violence: Context is everything. *Journal of Gender-Based Violence*, *1*(1), 33–44. doi:10.1332/239868017X14896674831496

Olson, K., & Witt, L. (2011). Are we keeping the people who used to stay? Changes in correlates of panel survey attrition over time. *Social Science Research*, *40*(4), 1037–50. doi:10.1016/j.ssresearch.2011.03.001

Piispa, M. (2003). Violence against women as conveyed by surveys—the Finnish case. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, *3*(2), 173–93. doi:10.1080/14043850310013160

Rose, D., O'Reilly, K., & Martin, J. (1997). The ESRC review of government social classifications. *Population trends*, *89*, 49–89.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, *8*(1), 3–15. doi:10.1177/096228029900800102

ScotCen Social Research. (2022). Growing up in Scotland: Cohort 1, sweeps 1-10, 2005–2020: Special licence access. [data collection]. 19th edition. UK data service. SN 5760. doi:10.5255/UKDA-SN-5760-12

Skafida, V., Morrison, F., & Devaney, J. (2021). Prevalence and social inequality in experiences of domestic abuse among mothers of young children: A study using national survey data from Scotland. *Journal of Interpersonal Violence*, *37*(11-12), 9811–9838. doi:10.1177/0886260520980392

Smith, T. W. (1983). The hidden 25 percent: An analysis of nonresponse on the 1980 General Social Survey. *Public Opinion Quarterly*, *47*(3), 386. doi:10.1086/268797

Testa, M., Livingston, J. A., & VanZile-Tamsen, C. (2005). The impact of questionnaire administration mode on response rate and reporting of consensual and nonconsensual sexual behavior. *Psychology of Women Quarterly*, *29*(4), 345–52. doi:10.1111/j.1471-6402.2005.00234.x

Walby, S., & Towers, J. (2017). Measuring violence to end violence: Mainstreaming gender. *Journal of Gender-Based Violence*, *1*(1), 11–31. doi:10.1332/239868017X14913081639155

Waltermaurer, E. (2005). Measuring intimate partner violence (IPV): You may only get what you ask for. *Journal of Interpersonal Violence*, *20*(4), 501–6. doi:10.1177/0886260504267760