**Abstract**

A paradox confronting survey researchers is that the persistent decline in response rates witnessed over the decades is not accompanied by an increase in nonresponse bias. Yet it is a common conviction that maximising the response rate will mitigate nonresponse bias in a survey and considerable expenditure in time and budget has been dedicated to do so. Empirical evidence of a link between response rates and nonresponse bias is limited or even contradictive. Our study implements a network modelling approach to clarify what mechanisms – if any – lead to nonresponse bias. Our data arise from an original and systematic review of the literature published between 1978 and 2020, featuring a diverse set of survey design characteristics across 187 unique studies making it one of the most extensive meta-analyses on the topic. Furthermore, we present the first application of network modelling to the issue of nonresponse rates and nonresponse bias, discussing the merits and limitations of the method. Our findings are mostly in agreement with prior literature: nonresponse bias is generally independent of the response rate and survey design. We note one exception: a conditional dependence between nonresponse rate and nonresponse bias in surveys administered by mail/paper. Even so, these relationships are complex and nuanced, meriting careful reflection from experts in the field.

**Methods**

Literature Search

Our study is a meta-analysis of estimates of nonresponse rate and nonresponse bias from 187 studies. Studies were identified from a systematic search of the literature that was based on the strategy implemented by @grovesImpactNonresponseRates2008. Articles were identified from five electronic databases: Google Scholar, Web of Science, Scopus, Web Survey Methodology and PsychInfo. Additional sources included conference proceedings (e.g., the American Statistical Association Survey Research Methods section)., articles cited in book chapters (e.g., Wiley Series in Survey Methodology), and internet searches of the key terms to obtain a wide range of articles and grey literature. The primary search criteria were that the study contained a survey and recorded nonresponse rates and nonresponse bias for that survey. Example of the search terms are *response bias, nonresponse bias, selection bias, survey bias, participation rates.* Further inclusion criteria are:

1. Article contains an empirical study about nonresponse rates.
2. A probability-sample is drawn from a population.
3. A survey is conducted.
4. Sample frame information is available for respondents *and* nonrespondents.

Screening of the initial search results was aided by the automatic screening tool ASReview (v0.19.1). Since all articles considered in the meta-analysis by @grovesImpactNonresponseRates2008 were granted automatic inclusion, we used them as the training set for ASReview. ASReview returns the articles classed as relevant or irrelevant according to the training set. Five independent researchers quality checked the output from ASReview by manually coding their relevance, stopping when 25 consecutive articles were coded as irrelevant. Articles that were incorrectly classed by ASReview were amended. Articles about panel surveys were excluded as the pattern of nonresponse is more complicated than for cross-sectional surveys. The process resulted in 187 studies as described in Figure 3.

The same five independent researchers undertook the task of coding each survey’s design characteristics. Most of the characteristics are related to the design of the study (e.g., nonresponse rate, nonresponse bias, survey mode, survey topic, survey incentivisation, survey sponsorship, type of population, use of reminders). Some characteristics are related to the type of outcome statistics (e.g., mean, median, total, or percentage) and the type of question (e.g., attitudinal, demographic, behavioural). Table 1 describes the characteristics that were selected for the current analysis. A full overview of all characteristics can be found in the appendix.

Nonresponse Rate and Nonresponse Bias

Nonresponse rates and nonresponse bias were calculated for each survey following the same procedure as @grovesImpactNonresponseRates2008. This was possible because all studies included had information on the respondents, nonrespondents, and the population or “truth”. Nonresponse rate is defined as a percentage of all the potentially eligible units that do not respond to all items in the survey (i.e., unit nonresponse). The nonresponse rate is found by subtracting the response rate from 1. The response rate is calculated as:

\begin{align\*}

RR &= 100\*\frac{n\_R}{n\_E} \\

\textrm{where}, \\

n\_R &= \textrm{the number of eligible units in the responding sample}, \\

n\_E &= \textrm{is the total number of eligible units invited to respond}. \\

\end{align\*}

Nonresponse bias consists of two components. The first is the nonresponse rate. The second is the *difference* between the respondents and the nonrespondents on the estimate or survey outcome. Here, we calculate the *absolute relative bias* as:

\begin{align\*}

Bias(\hat{y}\_r) &= \left\lvert\frac{100\*(\hat{y\_r} - \hat{y\_n})}{\bar{y\_n}}\right\rvert \\

\; \textrm{where}, \\

\hat{y\_r} &= \textrm{the survey estimate for respondents}, \\

\hat{y\_n} &= \textrm{the survey estimate for nonrespondents}, \\

p\_n &= \textrm{the nonresponse rate.} \\

\end{align\*}

Statistical Analysis

Network modelling estimates the (conditional) statistical dependencies among nonresponse rates, nonresponse bias, and the survey design characteristics described in Table 1. Network models are families of probability distributions that satisfy several statements of conditional (in)-dependence that are represented in graphical form [@haslbeckStructureEstimationMixed2015]. Networks consist of nodes (variables) and edges (statistical relations between variables). In our case nodes represent the survey design characteristics and are from different probability distributions (i.e., categorical and continuous variables).

The Mixed Graphical Model (MGM; [@haslbeckStructureEstimationMixed2015; @chenSelectionEstimationMixed2015; @yangMixedGraphicalModels2014]) caters to this by modelling the joint probability density of the appropriate distribution for a given node. A pairwise interaction between two continuous nodes, denoted $\mathit{s}$ and $\mathit{r}$, is represented by a single parameter, $\theta\_{s,r}$, when the interaction is non-zero. A pairwise interaction between two categorical variables, denoted $\mathit{m}$ and $\mathit{u}$, is represented by $\mathit{R = (m-1)\*(u-1)}$ parameters corresponding to indicator functions (dummy variables) for $\mathit{R}$ states (categorical levels). A pairwise interaction between a continuous and a categorical variable is represented by by $\mathit{R = 1\*(m-1)}$ parameters corresponding to $\mathit{(m-1)}$ corresponding indicator functions for all $\mathit{R}$ states [@haslbeckMgmEstimatingTimeVarying2020]. Estimation results in an undirected graph, denoted $G = (V, E)$, that has nodes $\mathit{V = \{1, 2,..., p\}}$ and edges $\mathit{E \subseteq V \* V}$.

The MGM are estimated in two steps: First, the conditional distribution of a node is estimated separately holding all other nodes constant, using the generalised linear regression framework. This results in two estimations per node going in opposite directions (i.e. N1 -- N2 and N2 -- N1). Second, the average of these two estimates is computed and it is this estimate that is represented in the network structure. An edge is represented in the network if \*either\* nodewise regression estimate is non-zero (i.e. we specify an \*\*"OR"\*\* rule instead of an \*\*"AND"\*\* rule that requires both estimates to be non-zero). The MGM uses regularization to avoid estimating spurious relations among nodes (i.e., false positives). The regularization method used is the Least Absolute Shrinkage and Selection Operator (LASSO) and is also known as the \*beta-min-condition\*, shrinking very small edges to zero and returning a conservative network with good specificity (at the cost of sensitivity). A final model is selected by optimizing the Extended Bayesian Information Criterion (EBIC). The EBIC relies on tuning the hyperparameter, $\gamma$, that controls the amount of extra penalization for model complexity and is set to 0.25 here.

Our study takes an exploratory approach to investigating the association between nonresponse rates, nonresponse bias, and survey design characteristics. We first estimate a network containing all survey design characteristics from Table 1. However, previous research has shown that nonresponse rates and nonresponse bias are different for different survey modes. Consequently, we estimate and compare a network for face-to-face, mail, web, and telephone surveys to investigate how survey design characteristics influence nonresponse rates and nonresponse bias for each mode.

As we have no confirmatory hypothesis we also do not explicitly perform statistical significance testing. Instead, we estimate the accuracy and stability of our network models by bootstrapping the edge weights and centrality measures and construct confidence intervals. Our primary aim is to explore the (conditional) statistical dependencies in the data to uncover potential drivers of nonresponse rates and nonresponse bias.

**Software:** All analysis is conducted in RStudio (v4.1.2). Network analysis is performed by the following package: `mgm` [@haslbeckMgmEstimatingTimeVarying2020] for network estimation, `bootnet` [@epskampEstimatingPsychologicalNetworks2018] for network stability analysis, and `qgraph` [@epskampQgraphNetworkVisualizations2012] for network visualisation.

**Results**

Descriptive Statistics

Our first step was to conduct descriptive analysis similarly to @grovesImpactNonresponseRates2008 to get a sense of how the nonresponse rate and nonresponse bias in surveys has changed since their early meta-analysis. A comparison of their original 49 studies to the new studies included in our analysis is presented in Table 2. Nonresponse rates are higher in the new studies at 44% compared to 36% in the earlier studies. This is consistent with the consensus that survey response rates have shown consistent decline over the last decades (i.e., nonresponse is increasing). A slight increase in nonresponse bias from 8% to 10% is also present in the new studies. Nonresponse rate and nonresponse bias is also broken down by survey mode in Table 2. As expected, there is variation in nonresponse rates and nonresponse bias across modes. Face-to-face surveys have the lowest nonresponse rate at 26.5% while drop-off paper surveys have the highest nonresponse rate at 55.23%. Face-to-face surveys also have the lowest rate of nonresponse bias at 8.44% which is similar to mail surveys at 8.98%. Mail and drop-off paper surveys have similar rates of nonresponse bias at 11.14% and 11.13% respectively, while bias in web surveys sits at 12.67%. Notice that the survey mode with the highest nonresponse rate is not the survey mode that has the largest nonresponse bias (e.g., mail surveys), indicating that nonresponse bias is driven by other factors (e.g., survey design characteristics).

Correlational analysis returned results similar to @grovesImpactNonresponseRates2008 and can be seen in Figure 4. @grovesImpactNonresponseRates2008 report a weighted correlation between nonresponse rates and nonresponse bias of 0.20 which is comparable to the weighted correlation of 0.23 that we find in the newly included studies. Our analysis found a between-study correlation of 0.47 in the studies from @grovesImpactNonresponseRates2008 and 0.40 in the new studies. A substantial study-level component to the relationship between nonresponse rate and nonresponse bias was also reported by [@brickResponsiveSurveyDesigns2017] but is here based on a much larger sample size. The increasing trend of nonresponse rates and relative stability of nonresponse bias from Table 2 is also seen in Figure 5.

Network Models

In this section we present five MGMs that are estimated for the different survey modes. Attention should be paid to several features to adequately interpret these models. Nodes are connected by edges when a conditional statistical dependency is estimated from the data. Blue edges indicate a positive association, red edges indicate a negative association, and grey edges indicate that no sign can be determined for the association (this is the case when an edge represents the associations of a variable with > 2 categories). To ease the visual comparison of the network models, the thickness of the edges is scaled to the maximum edge-weight found over all five models.

Interactions between continuous variables can be interpreted as conditional covariances. Interactions involving categorical variables is more intricate as the edge-weights represent a summary of several parameters. In the case where we have an interaction between a categorical variable and a continuous variable, we have two sets of parameters to interpret. One set represents the increase in the continuous variable associated with each level of the categorical variable. The other set represents the probability of every level of the categorical variable – which can be positively or negatively associated with an increase in the continuous variable. Parameters are estimated for the categorical variable regressed on the continuous variable and vice versa. Interactions between two categorical variables can be interpreted as probabilities compared to the reference (first) category.

Lastly, the pie-charts surrounding the nodes indicate the within-sample predictability of that node. The $R^2$ or explained variance is shown for the continuous nodes (nonresponse rate, nonresponse bias, and year) and the normalised correct classifications or accuracy is shown for categorical nodes.

Network including all survey modes

In the first network model there are 11 nodes including nonresponse rates, nonresponse bias, all survey modes, and other survey design characteristics (Figure 6). The nodewise weighted adjacency matrix is presented in Table 3. A full list of parameter interactions can be found in the Appendix.

After accounting for all other nodes in the network and regularizing out edges that are too weak, there is a direct association between nonresponse rates and nonresponse bias. Although it is difficult to see, there are additional direct associations between nonresponse bias to reminders and survey mode. As the edge to reminders is positive, this indicates that nonresponse bias increase slightly when reminders are used. In survey research, reminders are typically distributed in an effort to boost response rates. Indeed, in the network reminders are negatively associated with nonresponse meaning nonresponse decreases when reminders are used). The lingering association with nonresponse bias may occur if reminders only boost response rates from a particular group of nonrespondents associated with bias.

Several design characteristics are associated with nonresponse rates. A weak positive association with the publication year is evident and aligns with the descriptive statistics. Other positive associations are with salient survey topics, incentives being given, and surveys in special populations. It is rather surprising that nonresponse increases when incentives are given and when the survey topic is salient to the respondent. Regarding incentives, one explanation is if incentives are offered when response rates are already trailing it may be too late for a positive effect on response rates to take effect. We observe in the network that surveys with salient topics are less likely to be incentivised. Perhaps the topic is not salient enough to the respondent to increase responses and the lack of incentivisation is what matters more. Sponsored surveys and surveys about miscellaneous topics also lead to decreasing nonresponse rates.

The associations with nonresponse rate and nonresponse bias are all rather weak. The most central (densely connected) node in the network appears to be survey mode. The associations between different modes and other nodes are estimated in reference to drop-off paper surveys, finding that nonresponse rate decreases in face-to-face, mail, and telephone surveys. The effect of web surveys on nonresponse rate is estimated to be zero. Given how densely connected survey mode is to the other design characteristics, we estimate a network for each survey mode in turn.

Networks for the different survey modes

A network model was estimated for each survey mode to explore how the statistical relations change from the overall network model presented before. In each of these mode-specific network models there are 10 nodes (the same as the overall network excluding the node for mode). The nodewise weighted adjacency matrices for each network is presented in Tables 4 to 8. A full list of the nodewise parameter interactions for each network can be found in the Appendix.

Network for face-to-face surveys

The statistical associations in the face-to-face network are noticeably weaker than in the overall network. In face-to-face surveys, the use of reminders, health-related surveys, and surveys in special populations are associated with increasing nonresponse rates. The reason for reminders increasing nonresponse is the same as described before – reminders are not leading to better response rates in face-to-face surveys. The increase in nonresponse with special populations could be because special populations can be difficult to reach. Similarly, if we assume that health surveys are primarily targeted to people receiving some form of medical care (e.g., hospital patients, GP patients) then we would expect higher nonresponse rates as illness is likely to interfere with their ability to respond. Another notable difference is the absence of a relationship between the publication year and nonresponse rate. We know from prior research that nonresponse rates are increasing year-on-year generally, but face-to-face surveys specifically do not share in this trend.

The association between nonresponse rate and nonresponse bias is absent in face-to-face surveys, indicating that the design characteristics that lead to an increase in nonresponse rates are not in turn causing an increase in nonresponse bias. Although the edges are too weak to be observed in the network, nonresponse bias does decrease in face-to-face surveys with salient topics and when incentives are used. Recall that face-to-face surveys had the lowest nonresponse rates and lower nonresponse bias compared to other modes (Table 2), which could be the reason for absence of direct association.

Network for telephone surveys

The network for telephone surveys bears similarities to the face-to-face network. Namely, there are weak direct associations between some design characteristics and nonresponse rates. Consistent with the overall trend, nonresponse rates in telephone surveys are increasing year-on-year. However, in telephone surveys it appears that nonresponse rates decrease when the topic is salient to the nonrespondent, when reminders are given, when surveys are sponsored and when the survey is health related. In contrast to the overall and face-to-face network, the mentioned design characteristics have a more desired effect on nonresponse rates. Response rates for health surveys conducted over the telephone could be higher because most patients will have access to their phone even in hospital.

Despite the high nonresponse rate and nonresponse bias in telephone surveys (Table 2) there are few design characteristics driving this apart from year (see Table 5). Telephone surveys are unique from other modes in that they require instant reciprocity from the respondent (arguably also for face-to-face surveys, but interviewers can offer to drop off the questionnaire). Several transient influences can interfere with this, for instance, if the telephone call comes at a busy time or the mood of the respondent makes them unwilling to participate. Such circumstances are not captured here but correspond to the random propensities model [@brickResponsiveSurveyDesigns2017] and can explain why we do not find many effects on nonresponse rates and nonresponse bias in telephone surveys.

Network for mail surveys

The network for mail surveys shows several direct associations between nonresponse rates and survey design characteristics. Consistent with the general trend, nonresponse rates in mail surveys is also increasing year-on-year. Interestingly, nonresponse rate is also increasing when the survey topic is salient to the respondent, when incentives are used, and when the survey is in a special population. However, the network also shows that web surveys are less likely to use reminders when the survey topic is salient to the respondent and when special populations are targeted. This could mean that response rates to mail surveys are more dependent on reminders than other survey modes and this indirectly influences the relationship between nonresponse rates and other survey features. This is intuitive, as mail surveys can become hidden or lost when delivered alongside other important mail. Consider the scenario where the recipient received several letters in the mail. These could be household bills, important appointments, or updates from an insurance company. If the mail survey is delivered alongside these letters, it may go unnoticed or be low priority. Under these circumstances reminders can be very effective if they are delivered at a better time.

Nonresponse rate decreases when mail surveys issue reminders, are sponsored, and are health related. This implies that reminders and survey sponsorship are having the intended effect on nonresponse rates. Health surveys conducted by mail are also more accessible to patients than face-to-face surveys for instance as they can be completed at a convenient time for the respondent. Here, health surveys are also strongly associated with high topic saliency and use of reminders. A likely explanation is that mail health surveys are specific and targeted to a particular group, such as check-ins or monitoring of a treatment group. Together these circumstances lead to the reduction in response rates observed (Figure 7) and lower nonresponse bias (Table 2).

Despite high nonresponse rates in mail surveys the rate of nonresponse bias is low compared to the other modes (Table 2), and yet a direct positive association is present between them in the network model. This eloquently demonstrates the complexity of the relationship between nonresponse rates and nonresponse bias and the importance of survey design in making sense of it.

Network for web surveys

In our sample of web surveys, the nonresponse rate and nonresponse bias are notably greater than the other survey modes (Table 2). Yet in the estimated network model there are fewer direct associations between these nodes with the other design characteristics. Nonresponse rate increases slightly when reminders and incentives are used and when the survey is health related. Nonresponse bias also increases very slightly when the survey question is about an observable characteristic of the respondent. One reason for this is the anonymity of web surveys allow respondents to give inaccurate answers to certain questions.

A key difference about the network for web surveys is the strong negative associations between key design characteristics like topic saliency, incentivisation, reminders, and survey sponsorship that mean these are less likely to occur in web surveys. These characteristics usually encourage more responses if used and the lack of such features is an explanation for the very high nonresponse rate but *not* for the large amount of nonresponse bias.

**Robustness of the Network**

Assessing the robustness of statistical network models is important as all parameters in the network are estimated and not observed. The estimation of the parameters is vulnerable to uneven variation in the sample and low overall sample size. These are particular issues for the mode-specific network models, which are estimated from a reduced sample. In our study we assessed robustness in two ways: by bootstrapping the edge-weights (parameters) and centrality measures (node strength).

Figure 9 shows the bootstrapped edge-weights in each of the presented network models. Edge-weights are most stable for the overall network that utilised the full sample and on the mail survey network which has the highest sample size of the mode-specific networks. In the face-to-face, web, and telephone survey networks many of the estimates are zero (indicating the absence of an edge). The estimates that are non-zero have considerably wide confidence intervals (shaded region) and the bootstrap mean often diverges from the sample estimate. Although the sample sizes for each network are sufficient for estimation the number of cases per level of each categorical variable can become extremely low. In part, this is a side effect of the frequency of survey design characteristics across modes. It is also due to the case-wise deletion that occurs in network estimation. Increasing the sample size of the face-to-face, telephone, and web surveys is the only solution to this issue.

Figure 10 shows the correlation between the sample and bootstrapped centrality measures in each network. We only estimated node strength which tells us which node is the most densely and strongly connected in the network. Overall, the stability of node strength is excellent even as a higher percentage of cases are dropped from the bootstrap sample. A similar effect of sample size occurs in face-to-face, telephone, and web networks which have wider confidence intervals around the stability of node strength. Nonetheless, the mean stability does not drop below the recommended benchmark of 0.50.

**Discussion**

Our primary objective is to evaluate if – and how – survey design features contribute to the relationship between nonresponse rates and nonresponse bias. The rise in nonresponse rate has triggered several changes in how we design, conduct, and analyse survey data. Choices over survey design are some of the most crucial and challenging decisions that practitioners face. A compromise is often made between what we believe is the optimal survey design and budget or feasibility constraints. A common belief is that more expensive survey designs are necessarily better, leading to the use of reminders, incentives, longer field periods and preference for certain survey modes. It is sensical that the higher the response rate the more accurate and unbiased our survey estimates will be. Yet in our study, as in previous studies, we find no strong evidence that nonresponse rate is consequential for nonresponse bias.

In the first network model the most central node was the survey mode. Survey mode is expected to have direct implications for other aspects of survey design. Interviews who have can directly speak with respondents have the opportunity to make the purpose of the survey more explicit and salient as is the case for face-to-face and telephone surveys. Reminders are more common in mail and web surveys as having a (physical or electronic) address for the respondent makes delivery easier. What is more interesting is that the survey mode is directly associated with nonresponse rates *and* nonresponse bias, making it a common cause. According to the causal models presented at the beginning of this paper, a common cause of two variables can introduce a (spurious) connection between them. Indeed, we see that nonresponse rate and nonresponse bias are directly as positively associated in this network model meaning that as nonresponse increases so does nonresponse bias. The fact that this direct association between nonresponse rate and nonresponse bias is then absent in all but one of the mode-specific networks is further evidence that this association could be spurious. In the network including only mail surveys the association remains intact and the effect size is similar to the effect size in the first network model (0.18 and 0.20). There are two possible explanations here. One that we have already mentioned is that the connection between nonresponse rates and bias is spurious. They share a common cause in the network including only mail surveys too – reminders decrease the nonresponse rate but (very slightly) increase nonresponse bias. A second explanation is that the effect of a high response rate on nonresponse bias is specific to mail-mode surveys. Nonresponse rate is sizeable in mail surveys (41.45%) yet nonresponse bias is low (8.98%), so this is a curious finding. A key question is if the design features of mail surveys that increase the response rate are doing so only for a particular group of people. In this case, the increase in response rates would be biased to that group. I can think of three possible contributors to such a scenario based on the mail-only network. The only other factors that decrease nonresponse rate are the use of reminders, sponsorship of the survey, and health-related surveys. It has been noted in other studies that reminders can potentially increase bias if they are effective in the group already overrepresented in the survey. Sponsor bias can occur if respondents hold preconceived opinions about the sponsor that cause them to respond in a certain manner. As such, the overall response rate can be high, but those responses may be positively or negatively biased based on the respondents feelings about the sponsor. Health surveys are typically important to the respondent (e.g., a care quality survey or post-treatment follow-up) leading to a higher response rate. Health surveys can be biased when the health of the respondent is a barrier to participation, leading to a survey that has a high response rate but only for healthy participants. In these scenarios the reduction in the nonresponse rate leads to an increase in nonresponse bias and explains the connections seen in the mail network.

Most of the other characteristics of a survey consistently relate to nonresponse rates, but in a less consistent manner across survey modes. In telephone and mail surveys for example, the use of reminders and survey sponsorship are associated with lower nonresponse. Yet in web and face-to-face surveys they are associated with an increase in nonresponse rate. Another surprising effect is that the saliency of a survey topic is sometimes associated with rising nonresponse (as in mail and web surveys) and at other times is associated with a decrease (in face-to-face and telephone surveys). This highlights the complex and multidimensional nature of how survey design operates distinctly across different modes and how this can have surprising effects on nonresponse rates. The value of the network model is in revealing these dynamics at once. Choices about survey design are important if we want to control the nonresponse rate. If the goal is to reduce nonresponse bias the additional cost in resources may not be worth the increase in responses.

Limitations

Some limitations to our study warrant discussion. The strategy for coding survey design characteristics is subject to the information that is provided in the published text or other publicly available information (e.g., supplementary information, protocols). In some cases it was rather ambiguous if a certain design feature was present – just because a survey does not mention incentivisation does not mean that participants did not receive any form of compensation. We decided that surveys would only be coded as offering incentives, reminders, or as being sponsored if it was explicitly mentioned. Otherwise, the survey was coded as not having these features. Additionally, the coding of topic saliency was a (subjective) decision made by the coding team based on the stated topic and target population (i.e., given the target population and the purpose of the survey, is it likely the topic is salient to the respondents?). Moreover, the process of coding is ongoing and ass more data is being collected the inter-rater reliability has not yet been calculated. However, the coders followed the same instruction sheet, and any disagreements or discrepancies were resolved in discussion. Nevertheless, the coding of the survey design features is subject to some amount of human error or variation that has not been quantified in this study.

The parameters of the network model are estimated and not observed. As this is the first application of network modelling to survey design research, we have no benchmark of what the expected results should look like. Our opinion is that there are reasonable explanations for all of the edges in the network models. Still, the robustness analysis revealed considerable variability (large confidence intervals) and some discrepancies between the original estimates and the bootstrapped estimates. The likely reason is the variation in sample size across the survey modes and the observed frequencies of the survey design characteristics. In some cases, there are very few observations for a given feature (e.g., only 6.7% of telephone surveys mention incentivisation) and estimates based on such few observations are understandably less stable.

Class imbalance also reduces the power to detect statistical effects and it is possible that many of the (weak) edges in the networks models are a result of this. Moreover, the estimation procedure of the `mgm` package uses (LASSO) regularization *and* (EBIC) model selection which results in a highly specific but low sensitivity network. It is therefore possible that survey design features are more consequential for nonresponse bias than our implied by our results. We hope that as we continue to collect more data from surveys that the issues caused by class imbalance across the design features are minimized. For now, we encourage you to consider the relationships in these network models with caution.

**Conclusion**

In the present study we implemented a novel technique to an antiquated problem and demonstrated the utility of network modelling in generating causal theories about the mechanisms of nonresponse rates and nonresponse bias in surveys. Most of our findings are in accordance with the literature – the contribution of nonresponse rate to nonresponse bias is small but nuanced. Network visualisations expose us to this nuance in a way that is not possible with more traditional modelling frameworks like meta-regression. In a single glance we can see the intricacies of how survey design features are related to nonresponse rate and nonresponse bias and are too numerous to discuss in a single paper. Our hope is to inspire critical thought about these network structures among survey researchers and to open a debate about the use of causal inference in survey research.

A myriad of factors is relevant to the outcome of a survey and the design characteristics chosen here are just a few selected within the constraints of our sample. Our results are therefore contingent on what is featured in the network models. There is scope for other survey design characteristics to be included, such as survey length, the survey country, and the fieldwork period. We also believe that future work should expand the set of explanatory variables to include more transient influences on survey climate and participation such as the personality of the respondent and their physical environment (e.g., time constraints, urbanicity, demographic information). The ubiquity of surveys in modern society itself may be setting the threshold to participate higher and researchers continue to be challenged with selecting the optimal set of covariates to understanding survey participation.