SURVEY FUTURES

SURVEY DATA COLLECTION METHODS COLLABORATION

Working Paper 19:

When, and how much, might effect estimates be biased when analysing mixed-mode survey data? The roles of mode effects, mode selection, and mode split.

Georgia D Tomova, Richard J Silverwood, Liam Wright

Centre for Longitudinal Studies, UCL

June 2026

www.surveyfutures.net

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (grant ES/X014150/1) aimed at bringing about a step change in survey research to ensure that high quality social survey research can continue in the UK. The initiative brings together social survey researchers, methodologists, commissioners and other stakeholders from across academia, government, private and not-for-profit sectors. Activities include an extensive programme of research, a training and capacity-building (TCB) stream, and dissemination and promotion of good practice. The research programme aims to assess the quality implications of the most important design choices relevant to future UK surveys, with a focus on inclusivity and representativeness, while the TCB stream aims to provide understanding of capacity and skills needs in the survey sector (both interviewers and research professionals), to identify promising ways to improve both, and to take steps towards making those improvements. Survey Futures is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefitting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveyfutures.net.

This paper is a product of the project "Assessing and disseminating methods for handling mode effects", led by Liam Wright (UCL), which forms part of Survey Futures Research Strand 6, "Reducing and evaluating mode effects".

Prior to citing this paper, please check whether a final version has been published in a journal. If so, please cite that version. In the meanwhile, the suggested form of citation for this working paper is:

Tomova GD, Silverwood RJ & Wright L (2026) 'When, and how much, might effect estimates be biased when analysing mixed-mode survey data? The roles of mode effects, mode selection, and mode split, *Survey Futures Working Paper* no. 19. Colchester, UK: University of Essex. Available at https://surveyfutures.net/working-papers/.

Table of Contents

Abstract

Survey data are increasingly collected using mixed-mode designs (e.g. personal interview and web questionnaire). Little is known, however, about the extent to which this introduces bias in subsequent analyses, or whether simply including mode as a covariate addresses it. Using data simulations, we identified the conditions under which mode effects (mode measurement differences) and mode selection (respondent differences) introduce bias, complementing this with empirical illustrations using mixed-mode data from the 1958 National Child Development Study. In simulations, absent mode selection, substantial bias arose only from unusually large mode effects, but was amplified when mode selection was present. Controlling for mode under strong mode selection introduced substantial bias, including sign-reversal of the estimate. Bias was more pronounced when the sample was split equally between modes. The direction and size of bias will depend on the direction of all effects. In our empirical illustration, the results were largely unchanged after controlling for mode, possibly reflecting weaker mode selection and mode effects in these data. These findings highlight that, while possible, substantial bias from mixed-mode designs may be unlikely in many practical settings. However, the risk of bias, and the appropriate strategy to address it, should be considered on an analysis-specific basis.

Key words

survey data, mixed-mode data, mode effects, measurement effects, mode selection, mode preference

Background

Mixed-mode survey designs – where the same data are collected from different respondents using some combination of telephone, face-to-face, or video interview, and web or paper questionnaire – have become increasingly common, driven by rising costs, declining response rates, and shifts in the coverage of different modes (Brown and Calderwood 2020). The specifics of mixed-mode designs can vary substantially between studies, including what modes are offered, whether they are offered sequentially or concurrently, and whether allocation is based on participant preference, allocation by survey managers (e.g. on available contact information) or - as in the sequential mixed-mode design - initial non-response (de Leeuw 2005; Martin 2011).

Regardless, a common challenge in mixed-mode surveys is that participants may respond differently to questions depending on the mode used. This introduces a type of systematic measurement error, commonly referred to as a *mode effect* (d'Ardenne et al. 2025). For instance, participants may be unwilling to reveal socially sensitive or incriminating information truthfully to an interviewer, whereas they might provide this information more readily in an anonymous mode, like a web questionnaire, leading to systematically different responses across modes. Importantly, mode effects do not only introduce differences in *measurement*. They can also introduce spurious associations *between variables* that are subject to mode effects. If one mode elicits more socially desirable responding, variables that are subject to this (e.g. smoking and history of sexually transmitted infections) will appear correlated, even in the absence of any true relationship. Two variables subject to mode effects can therefore be described as containing dependent measurement error, which can also be conceptualised as confounding by mode (Hernan and Cole 2009; Tomova et al. 2026a).

Differences in responses between modes may also arise because different types of participants end up responding by each mode. Depending on the survey design, mode may be determined directly according to known participant characteristics (e.g. individuals without an email address may be automatically assigned to offline modes) or participant preference (e.g. where participants choose mode), or, as in sequential designs, indirectly due to initial non-response. Regardless of the reason, there typically exist systematic differences between individuals in each mode; a phenomenon referred to as *mode selection* (Hox et al. 2017).

Mode selection can introduce spurious associations in two distinct ways. First, it can happen directly. If frequent internet users are more likely to respond via web, and respondents are more likely to disclose poor mental health in a web survey, then internet use will appear associated with mental health outcomes purely as an artefact of the survey design. Second, spurious associations can also be introduced by naively controlling for mode due to a phenomenon known as *collider bias* (Cole et al. 2010; Tomova et al. 2026a). When two variables are independent sources of mode selection, e.g. if loneliness and chronic pain are both reasons why someone may prefer responding by telephone rather than web, then mode is said to be a 'collider' (Arah 2019) of these two causes. Controlling for the collider, mode (e.g. by including as a covariate in a regression model, commonly referred to as the 'indicator method'), introduces spurious associations between its causes: if one knows someone responded via telephone and that they *do not* have chronic pain, then they are more likely to be lonely. This means that controlling for mode would induce an apparent association between loneliness and chronic pain even when there is no true effect and could be sufficient to reverse the direction of the observed estimate. See **Figure 1** for an illustration and Tomova et al. (2026a) for further explanations and examples.

The extent of bias introduced by adopting a mixed-mode survey design is an empirical question and depends on the presence and strength of mode effects and mode selection, what the causes of mode selection are (in particular, whether they involve the exposure and outcome), and the proportion of participants responding by each mode. Whether – and when – these are sufficiently important to materially bias results is a key question for both survey methodologists and analysts of survey data, particularly as mixed-mode surveys become more common. A large literature has assessed the effects of mode on item measurement experimentally (Tomova et al. 2026b), but by themselves these findings cannot be used to indicate whether, and how much, bias would be introduced when estimating relationships between variables. Direct empirical evidence on this is very limited. Although some articles have assessed whether mixed-mode designs alter the observed associations between variables in the data, they predominantly focus on changes in statistical significance (Clarke and Bao 2022; Dolan and Kavetsos 2016; Jäckle et al. 2006; Piccitto et al. 2022; Villar and Fitzgerald 2017), which is a highly discouraged practice (Greenland 2017; McShane et al. 2019), or only examine a few variables (Jäckle et al. 2006; Martin and Lynn 2011; Vannieuwenhuyze 2015), which makes it difficult to generalise beyond them. Most importantly, however, the existing literature does not distinguish between scenarios where controlling for mode may reduce bias (by reducing the impact of mode effects) or introduce bias (by introducing collider bias in the presence of mode selection). As such, it cannot offer any conclusions as to whether any differences in the observed associations are desirable or not, and this is a significant limitation, because it cannot equip researchers with a way to reason about any potential consequences in their own analyses of mixed-mode data. Regardless, examining such questions is not straightforward in real data – it can be difficult to know whether a change in estimate represents more or less bias, without knowing what the true effect is.

Data simulations, however, allow such questions to be investigated. By generating data according to known mechanisms, the true relationships between variables are known by design, making it possible to quantify bias under a range of structural and analytic scenarios (Morris et al. 2019). Although simulations are necessarily simplifications of reality, they are well-suited to methodological questions of this type and can be grounded in plausible structures. They can help demonstrate the extent of bias that is expected to occur under different structural scenarios, but not how commonly such structures may occur in practice. This is why they can also be complemented by accompanying analyses of real data. Together, simulations can help identify the conditions under which bias may arise, whereas real data can help provide an empirical sense of the implications of such bias in practice.

The present study has two overarching aims:

1) In simulated data, identify the conditions under which substantial bias may be introduced in the analyses of mixed-mode data, in particular resulting from:
    a. mode effects
    b. naïve controlling for mode in the presence of mode selection

2) In real data, assess the practical implications of mode effects and mode selection for analyses of mixed-mode data by examining whether, and how much, observed estimates change after controlling for mode across scenarios, where we expect *a priori* that:
    a. in the presence of mode effects without mode selection, a change in estimate would reflect *a decrease* in bias
    b. in the presence of mode selection, a change in the estimate would reflect an *increase* in bias.

Methods

Simulated data

Illustrative example

To ground the simulations, suppose we are interested in estimating the effect of an exposure X on an outcome Y (henceforth the 'focal relationship') in a dataset collected using a (two-mode) mixed-mode design. A common goal in such settings is to obtain estimates in which mode-related measurement differences have been controlled for, usually with the aim to approximate what would have been observed had all respondents answered via the same reference mode. This is commonly attempted by including mode as a covariate in the regression model.

Throughout, we keep concepts and examples as general as possible. Simulation parameters are chosen to be informative of the range of conditions under which bias may or may not arise in practice, rather than attempting to replicate a single particular dataset. We describe variables using general terminology referring to their roles, namely exposure (X), outcome (Y), survey mode (M), and an unobserved common cause of survey mode and the outcome (U). We further distinguish between the latent (true) version of a variable (i.e. X or Y) and its measured version obtained in a survey (i.e. X^* or Y^*).

Data generating mechanisms

To explore and demonstrate the extent of bias likely to be introduced under different scenarios, we considered several data generating mechanisms involving mode effects and mode selection. All scenarios are variations of a common underlying structure, depicted in **Figure 2**, where dashed paths represent structural features that were varied across scenarios: the presence and strength of the focal relationship ($X \to Y$), mode effects on the measured exposure and outcome ($M \to X^*, M \to Y^*$), mode selection driven by the exposure ($X \to M$) and/or outcome ($Y \to M$), and mode selection via an unobserved cause of mode ($U \to M$) and the outcome ($U \to Y$). **Figure 3** summarises all parameters that were varied across the simulations. The only structurally invariant paths were $X \to X^*$ and $Y \to Y^*$, which represent the underlying relationship between the latent exposure or outcome and the structural 'true' part of their measured versions.

We did not simulate scenarios in which U is cause of X. While likely in practice, including it on top of existing paths would only additionally introduce classical confounding, which is a standard and well-understood issue (Greenland et al. 1999) distinct from the mode-related challenges examined here, therefore adding more complexity without additional insight. Although collider bias can similarly be introduced when U is a common cause of X and M, the structural principles generalise from the examples considering collider bias arising when U is a common cause of Y and M. The scenarios considered, while not exhaustive, are sufficient to illustrate the core principles, which extend naturally to a wider range of settings.

Data simulations

Data were simulated according to the illustrative data generating mechanism depicted in **Figure 2** and the simulation parameters in **Figure 3** using the '*dagitty'* package (v0.3-4) (Textor et al. 2017) in R (v4.6.0) (R Core Team 2025) by directly specifying the causal structures of interest.

For simplicity, all variables were simulated as continuous and standardised, and all relationships assumed linear. Path coefficients therefore represent standardised effects expressed in standard deviation units. Survey mode M was the only variable that was dichotomised from its underlying continuous form, using a threshold that results in the desired mode split (50/50 or 80/20, the latter reflecting a more common mixed-mode split (Wright et al. 2024)).

To illustrate the process, consider a single simulation condition with a focal relationship $X \to Y$ of 0.2 SD and an equal 50/50 mode split. For this condition, datasets of 10,000 observations each were simulated for each of the nine mode selection scenarios depicted in **Figure 3**, by varying the path coefficients for $X \to M, Y \to M$, $U \to M,$ and $U \to Y$. M was then dichotomised and within each dataset, all combinations of mode effects ($M \to X^*, M \to Y^*$) were applied in turn to the simulated dataset, by deriving X^* (or Y^*) as a function of X (or Y), M (using the desired mode effect coefficient), and noise. Noise was added because, although mode effects are primarily systematic rather than random error, not all observations (respondents) will always experience the same mode effect. The noise was drawn from a normal distribution with mean 0 and SD equal to 25% of the mode effect coefficient, meaning the amount of noise scaled with the size of the mode effect. The parameters selected for mode effects (i.e. 0 SD, 0.2 SD and 0.5 SD) were chosen to provide an informative overview of the degree of bias that may be introduced. A mode effect of 0.2 SD is relatively large, but not that uncommon, e.g. in a recent systematic review, 37.3% of mode effects between face-to-face and web responses exceeded 0.2 SD (Tomova et al. 2026b), and therefore provides a useful insight for the consequence of material mode effects in practice. Very few mode effects exceed 0.5 SD in reality, providing a useful bound for the bias that may occur under more extreme scenarios. These parameters also provide informative choices for mode selection. To give an intuition of the degree of mode selection represented in terms of SD, consider $Y \to M$, where Y is years of education (mean=13, SD=3), and M is mode (50% web vs telephone). A 0.5 SD effect of educational attainment on mode will result in a respondent with average educational attainment (i.e. 13 years) having approximately a 50% probability of responding via web, and a respondent 1 SD above the mean (i.e. 16 years) having approximately a 70% probability, i.e. a difference of 20 percentage points. For $Y \to M$ of 0.2 SD, the difference will be 8 percentage points. Both are plausible degrees of mode selection, though the exact strength will of course differ in practice across survey designs, variables, and populations. Using the same set of parameter values across mode effects, mode selection, and the focal relationship allows their relative strengths to be compared intuitively across scenarios. To maintain feasibility and clarity, we do not vary the paths $U \to M$ and $U \to Y$, therefore examining the presence (as opposed to absence), but not the strength, of mode selection arising from U.

For all combinations of scenarios depicted in **Figure 3**, 10,000 datasets were simulated, each containing 10,000 observations.

Ground truth and models examined

Throughout, we maintain an illustrative interest in estimating the effect of an exposure (X) on an outcome (Y), where the ground truth effect is 0, 0.2, or 0.5 SD. In practice, X and Y are not directly observed and the measured versions X^* and Y^* are used instead. Across all simulated scenarios, we obtained estimates from the following two models:

1. Unadjusted for M, corresponding to a mixed-mode analysis that does not control for mode:
$$E[Y^*|X^*] = \alpha_0 + \alpha_1 X^*$$
2. Adjusted for M, corresponding to a mixed-mode analysis that controls for mode:
$$E[Y^*|X^*, M] = \beta_0 + \beta_1 X^* + \beta_2 M$$

For each model, we report the median effect estimate and 95% simulation interval (SI), defined as the 2.5th and 97.5th effect estimate centiles across all simulated datasets. Under all simulated conditions, we compared the obtained estimates against the ground truth, as well as against a null effect of zero to assess the risk of sign-reversal. Model 2 was examined to determine when and to what extent the indicator adjustment method reduced or introduced bias.

Real data

Illustrative example

To complement the simulation analyses, we use data from a real mixed-mode survey to examine the practical implications for study results of mode effects and mode selection, including the consequences of controlling for mode, across a range of examples. Since in real data true causal relationships are unknown, a direct empirical replication is not generally possible. However, with *a priori* expectations of which variables may be related to mode selection, as well as empirical evidence on the presence of mode effects obtained from an embedded experiment, we can classify scenarios in terms of the presence and absence of mode selection and/or mode effects, that logically follows the simulation conditions examined. In a range of exposure-outcome pairs across these scenarios, we examine whether, and how much, any observed associations between variables may change after controlling for mode, and based on the assigned classification, whether this therefore may represent an increase or a decrease in bias.

Data

We used data from the 1958 National Child Development Study (NCDS), a multidisciplinary cohort study following over 17,000 individuals in Great Britain born during a single week in 1958, with immigrants born in the same week added during childhood and adolescence, resulting in a total number of 18,558 individuals (Power & Elliott, 2006). Participants have been followed up at multiple time points across the life course, with data collected on a wide range of physical, psychological, and social characteristics. The 2013 follow-up (at age 55) adopted a sequential mixed-mode design in which participants were first offered a web survey, and non-respondents then offered a telephone interview. In this wave, an experiment was embedded in which a random sub-sample of individuals (for which telephone details were available) were allocated directly to telephone (n=1,476), while the rest of the participants were allocated to the sequential mixed-mode arm (n=9,110). For further details, see Brown and Hancock (2015). There were 9,134 respondents in total (n=7,546 in the mixed-mode arm, n=1,588 in the telephone arm). In the present study, we conducted analyses within the mixed-mode arm (i.e. within which mode was *not* randomised), where 5,612 individuals responded by web and 1,934 by telephone, resulting in a web-telephone mode split of 74-26%. We also utilised the experiment between arms to estimate mode effects, as detailed further in the sections below.

Variable selection

We selected variables from the NCDS Age 55 wave to serve as examples in this illustrative analysis, with the aim of obtaining a range of variables that differ in terms of their mode selection and mode effect properties, in order to conceptually match the broad scenarios explored in the simulation. Mode selection status was classified *a priori* on theoretical grounds, i.e. based on whether a plausible mechanism existed by which the variable could directly influence mode selection. Although associations with mode can be estimated empirically, this is challenging in observational mixed-mode data since any observed association between a variable and mode may reflect confounding. Further, when constructing DAGs, such as for the purpose of determining appropriate variable adjustment, this process is based on *a priori* theory anyway (Digitale et al. 2022). Mode effect classification was, however, determined empirically using the randomised experiment embedded within the wave. We obtained estimates of the complier average causal effect (CACE) for each variable of interest using random mode allocation as an instrumental variable. For simplicity and ease of binary classification, a variable was considered to be subject to a mode effect if the 95% confidence interval of the estimate did not contain the null. Further details on the estimation of mode effects are available in the **Supplementary Methods**. For each of the four classification categories, i.e. 1) both mode selection and mode effects, 2) mode selection with no mode effects, 3) no mode selection with mode effects, 4) no mode selection and no mode effects, we selected 10 variables, corresponding to 40 variables in total. A list of these variables, including justifications of their mode selection and mode effects status is available in **Table S1**. Finally, to examine whether any differences in results after controlling for mode may be attenuated or inflated in the presence of additional covariates, we also selected a range of variables from previous waves to serve as illustrative covariate adjustment sets. **Table S2** contains details of all variables from NCDS selected across waves, their purpose, and descriptive statistics.

Models examined

Following the classification of variables by their mode effect and mode selection status, we examined whether controlling for mode alters the observed associations between such variables. Any such change can, in principle, reflect either a reduction in bias (where controlling for mode accounts for mode effects) or an introduction of bias (where controlling for mode in the presence of mode selection induces collider bias). Which of these applies depends on the classification: for example, where a mode effect is present, but mode selection is not, a change would be expected to indicate a decrease in bias, whereas where mode selection is present, it may indicate an increase in bias. Crucially, in real data this cannot be verified directly, as the true exposure-outcome effect is unknown. We therefore interpret all results in light of this theoretical classification and the simulation findings.

Within each classification category, we selected exposure-outcome pairs, such that each variable was used once, resulting in 20 pairs from the 40 variables. While aiming to maintain plausible exposure-outcome choices, this was constrained by the pool of variables available that can be reasonably classified into the defined categories. A list of all exposure-outcome pairs is available in **Table S3**. For each pair, we compared the following models run in the mixed-mode arm:

1. Unadjusted for M and C, corresponding to a mixed-mode analysis that does not control for mode or any other covariates:
$$E[Y^*|X^*] = \gamma_0 + \gamma_1 X^*$$
2. Adjusted for M, but unadjusted for C, corresponding to a mixed-mode analysis that controls for mode but no other covariates:
$$E[Y^*|X^*, M] = \delta_0 + \delta_1 X^* + \delta_2 M$$
3. Unadjusted for M, but adjusted for C, corresponding to a mixed-mode analysis that does not control for mode but controls for other covariates:
$$E[Y^*|X^*, C] = \zeta_0 + \zeta_1 X^* + \zeta_2 C$$
4. Adjusted for M and C, corresponding to a mixed-mode analysis that controls for both mode and other covariates:
$$E[Y^*|X^*, C, M] = \eta_0 + \eta_1 X^* + \eta_2 C + \eta_3 M,$$

where M represents survey mode, and C represents a set of covariates for each exposure-outcome pair (available in **Table S3**). We obtained point estimates and corresponding 95% confidence intervals (CI) from each model and compared estimates that differed in adjustment for M (i.e. $\hat{\gamma}_1$ vs $\hat{\delta}_1$, and $\hat{\zeta}_1$ vs $\hat{\eta}_1$), to assess whether, and how, controlling for mode altered the observed estimates. Estimates were presented as average marginal effects, including standardised effect sizes (in SD units) for continuous variables and probability differences for binary variables, using the '*marginaleffects'* package (Arel-Bundock et al. 2024). Note, these analyses are entirely illustrative for the purpose of assessing the impact of mode, therefore the resulting estimates should not be interpreted substantively.

Finally, we implemented an alternative approach to handling mode effects (other than mode control), in which empirical evidence on mode effects derived from experiments is used to shift values observed in one mode to their expected value in the other. This is conceptually similar to calibration or bias-correction, and is within the wider family of quantitative bias analyses (Fox et al. 2021). The use of this approach is still emerging (Wright et al. 2024), and guidance on application to different variable types is limited. We therefore restricted this analysis to a subset of exposure-outcome pairs in which at least one variable was continuous, where this is most straightforwardly applied. The approach is detailed in the **Supplementary Methods**.

Results

Simulated data

Figures 4 and **5** present results for the simulation with ground truth $X \to Y$ effect of 0.2 SD, presented respectively for 50-50% and 80-20% mode splits. Each panel corresponds to a different mode selection scenario, and within each panel, different mode effect conditions are depicted along the y-axis. Points represent the median estimated effect of X^* on Y^* across 10,000 simulations, with error bars showing the 95% SI, and compared against the ground truth (grey line) and the null (dashed red line). Results for the alternative ground truth scenarios (i.e. other $X \to Y$ effect sizes under each mode split) are available in **Figures S1-S4**. The overview of results below is organised in themes according to mode selection, mode split, and the strength of the focal $X \to Y$ relationship. The implications of mode effects are further discussed within each theme. Throughout, we use **Figure 4** as a starting point for interpretation, and refer to other figures where relevant. The complete set of all simulation results is available in **Table S4**.

No mode selection

In the absence of mode selection (**Figure 4A**), the unadjusted model generally recovered the true effect across most conditions (0.20 SD [95% SI: 0.18, 0.22], compared to ground truth of 0.20 SD). Estimates were only attenuated when X was subject to stronger (0.5 SD) mode effects (0.19 SD [95% SI: 0.17, 0.21]) and inflated when both X and Y were subject to strong mode effects (0.25 SD [95% SI: 0.23, 0.26]). However, adjusting for mode in these scenarios successfully corrected the bias and recovered the true effect (0.20 SD [95% SI: 0.18, 0.22] in both cases).

Mode selection via the exposure and/or outcome

When mode selection was driven by X alone (**Figure 4B**), then mode effects on X attenuated the observed estimate, whereas mode effects on Y or both X and Y inflated it, and the size of the bias increased with both the size of the mode effects themselves, as well as the size of the mode selection from X. In all scenarios, controlling for mode recovered the true effect. For example, under strong mode effects on both X and Y and strong mode selection from X (**Figure 4C**), unadjusted results were: 0.27 SD [95% SI: 0.25, 0.29], adjusted: 20 SD [95% SI: 0.18, 0.22], ground truth 0.20 SD). When mode selection was driven by Y (**Figure 4D**), the presence of mode effects inflated the observed estimate and this inflation increased with the increasing size of the mode effects and mode selection. However, in this scenario, controlling for mode did not recover the true effect and instead attenuated the estimate in the other direction. This was most evident when mode selection from Y was stronger (0.5 SD) (**Figure 4E**) (unadjusted: 0.34 SD [95% SI: 0.32, 0.36], adjusted: 0.017 SD [95% SI: 0.15, 0.09], ground truth 0.20 SD), but weaker when the selection itself was weaker (0.2 SD) (unadjusted: 0.29 SD [95% SI: 0.27, 0.30], adjusted: 0.19 SD [95% SI: 0.17, 0.21], ground truth 0.20 SD). These patterns were considerably more pronounced when both the exposure and outcome were sources of mode selection (**Figure 4F**). Mode effects introduced greater bias in the unadjusted model, and controlling for mode instead biased the estimate in the opposite direction. Most strikingly, when both the exposure and outcome were strong sources of mode selection (0.5 SD) (**Figure 4G**), controlling for mode was sufficient to completely sign-reverse the results (unadjusted: 0.38 SD [95% SI: 0.37, 0.40], adjusted: -0.04 SD [95% SI: -0.06, -0.02], ground truth 0.20 SD).

Mode selection via an unobserved variable

When mode selection was introduced via U (an unobserved common cause of M and Y), under both weaker (0.2 SD) (**Figure 4H**) and stronger (0.5 SD) (**Figure 4I**) mode selection, results in the presence of mode effects were broadly similar to those observed under comparable exposure- and outcome-driven mode selection. However, the implications of controlling for mode were not as pronounced. Under stronger selection paths, adjusting for mode did not sign-reverse the estimates but attenuated them by half (unadjusted: 0.34 SD [95% SI: 0.32, 0.36], adjusted: 0.10 SD [95% SI: 0.08, 0.12], ground truth 0.20 SD).

Mode split

Results differed according to mode split. In general, the consequences of mode effects, mode selection, and inappropriate controlling for mode were all less pronounced under the 80-20% split (**Figure 5, Figure S2, Figure S4**) compared to the 50-50% split (**Figure 4, Figure S1, Figure S3**). Under the 80-20% split, the most extreme scenario of strong mode selection (0.5 SD) from both X and Y did not sign-reverse the estimates, but it attenuated them towards zero (unadjusted: 0.33 SD [95% SI: 0.32, 0.35], adjusted: 0.03 SD [95% SI: 0.01, 0.05], ground truth 0.20 SD). The overall pattern of results was otherwise consistent across mode splits, with bias differing in magnitude but not direction. Uncertainty in estimates was also higher under the unequal split.

Strength of the true focal relationship

The results also differed according to whether the true focal $X \rightarrow Y$ relationship was 0 SD (**Figure S1**, **Figure S2**), 0.2 SD (**Figure 4**, **Figure 5**) or 0.5 SD (**Figure S3**, **Figure S4**). When the true effect was larger (0.5 SD), the consequences of mode selection, mode effects, and inappropriate controlling for mode were relatively less pronounced (**Figure S3**, **Figure S4**). The most extreme scenarios attenuated estimates substantially but did not produce sign-reversal as observed when the ground truth was 0.2 SD (**Figure 4**, **Figure 5**). When the true effect was null (0 SD) (**Figure S1**, **Figure S2**), both mode effects and controlling for mode in the presence of mode selection introduced greater bias compared to when a true effect did exist (**Figure 4**, **Figure 5**). A key difference in the context of a true null effect is that, unlike in all other scenarios, when mode selection is caused by Y alone, controlling for mode does not introduce any bias. The reason for this is that, when Y is the outcome, controlling for a descendant of it (mode) creates a dependency between the structural part of Y and all random determinants of Y. Typically, this results in attenuation of the effect towards the null (e.g. **Figure 4**). However, no attenuation towards the null can occur if the effect is already at the null. In all scenarios, any discussed implications were less severe in the 80-20% mode split compared to 50-50%.

Real data

Figure 6 presents results from the illustrative mixed-mode analysis of NCDS Age 55 data. Across all 20 exposure-outcome pairs, controlling for mode had little if any impact on the observed estimates, regardless of whether the variables involved were subject to mode effect and/or mode selection. When the variables were classified as having no mode selection and no mode effects, the results were identical before and after controlling for mode. When variables were classified as sources of mode selection (top panels in **Figure 6**), controlling for mode resulted in a negligible attenuation of the estimate. These findings were also consistent across comparisons after controlling for an illustrative set of other covariates, though any attenuation was even more negligible. Using quantitative bias analysis to account for the mode effects correctly, rather than controlling for mode as a covariate, also produced largely identical results (**Figure S5**). Overall, in this particular dataset, controlling for mode did not lead to any notable change in the obtained results.

Discussion

Using data simulations and an empirical illustration, we examined the conditions under which mode effects and mode selection are likely to bias estimates of substantive relationships of interest, and the circumstances under which naïvely controlling for mode can introduce, rather than reduce, bias. Our findings demonstrate that the consequences of mode effects and mode selection depend critically on the underlying causal structure, particularly the presence and strength of mode effects, the sources and strength of mode selection, and the mode split proportion. In some structural scenarios, where mode is a 'collider', naïvely controlling for it can introduce substantial bias, including sign-reversal. However, in the empirical illustration, controlling for mode had no practical impact on the observed estimates across a range of exposure-outcome pairs. This means that both in settings where bias may have been expected to exist, it was not observed, and where controlling for mode may have been expected to reduce any existing bias, it had hardly any impact on the observed associations. Overall, this suggests that while the potential for bias exists under certain conditions, the actual magnitude may be negligible in practice depending on the structure of the real data.

Key messages

First, in the absence of mode selection, mode effects generally need to be large and affect both the exposure and outcome to introduce meaningful bias. For example, even mode effects of 0.2 SD (which is likely to exceed most mode effects observed in practice [Tomova et al. 2026b]) did not introduce bias, which is reassuring. Only very large mode effects (0.5 SD) introduced bias, and even then, in the absence of mode selection, controlling for mode was sufficient to remove it. However, in the presence of mode selection, a mode effect on either the exposure or outcome alone can be sufficient to introduce substantial bias because mode effects then constitute a form of differential measurement error (Van Smeden et al. 2020). In other words, the presence of mode selection amplifies any bias from mode effects.

Second, whether naively controlling for mode would introduce bias depends on the underlying causal structure. When only the exposure is a source of mode selection (which may be more likely if the exposure and outcome are measured at different time points), then including mode as a model covariate will generally not introduce bias. Some bias was introduced when mode selection (of 0.2 SD magnitude) was driven by the outcome or both the exposure and outcome, but bias of the magnitude observed in our simulations is likely to fall within the range of uncertainty that is typical of observational studies, where other sources of bias would also operate. Substantial collider bias was, however, introduced by controlling for mode when both the exposure and outcome were sources of stronger mode selection (0.5 SD), and, depending on the strength of the substantive relationship of interest, this then resulted in sign-reversed or heavily attenuated estimates. This highlights the critical importance of carefully considering the potential for mode selection when analysing mixed-mode data, while at the same time providing reassurances that, in practice, substantial collider bias requires very strong mode selection mechanism and is therefore likely to be rare.

Third, the scenarios in which mode effects are large enough to require correcting (i.e. large mode effects amplified by accompanying strong mode selection), are precisely the scenarios in which controlling for mode is most likely to introduce collider bias. Since mode selection can amplify mode effects, their co-existence compounds the problem that needs resolving and increases the risk of inappropriate mode adjustment being made. Such scenarios therefore require careful consideration, and handling mode effects may necessitate other approaches, such as quantitative bias analysis (Fox et al. 2021; Wright et al. 2024).

Fourth, in the illustrative analysis of mixed-mode data from NCDS, controlling for mode had little practical impact on observed estimates across a range of exposure-outcome pairs, regardless of the presence of mode selection or mode effects. This suggests that, at least in this dataset, the conditions required for substantial bias to materially alter study conclusions were not met. Indeed, in the data there was an uneven mode split (75-25%) and most of the mode effects were under 0.1 SD (Goodman et al. 2022). A possible explanation for why mode selection may have been weaker too is that in a longitudinal study with many follow-up waves, the main drivers of mode selection are likely to be individual characteristics collected in earlier waves, rather than the specific variables collected at this particular follow-up. Therefore, the possibility for substantial collider bias which requires direct selection paths is very small. While this is likely to generalise to other cohorts with similar design and provides some reassurance for applied researchers, the results should not be overinterpreted as each mixed-mode study will have its own unique structure and underlying selection mechanisms.

Taken together, these findings suggest that while mixed-mode designs can introduce substantial bias under certain conditions, the risk in many practical settings may be modest as it requires sufficient mode selection, mode effects, and/or a relatively equal mode split. Reassuringly, the conditions that may be commonly encountered in practice, i.e. unequal mode splits with one mode being dominant in the dataset, and lack of strong mode effects and strong mode selection related to the variables being studied, are those in which our simulations suggest bias is unlikely to materially bias the results. This was further demonstrated by our illustration of mixed-mode data analyses in NCDS, where controlling for mode made no notable difference to the observed estimates. However, two important caveats apply. First, the conditions under which bias can be substantial are, although more rare, not implausible: mode effects of 0.5 SD have been reported in various settings (Tomova et al. 2026b), and the magnitude of mode selection can be expected to vary between studies and populations. Second, even modest sources of bias may overwhelm a relationship that is itself small. Different substantive domains tend to work with effects of different standardized sizes (Lortie-Forgues and Inglis 2019; Lovakov and Agadullina 2021; Schäfer and Schwarz 2019), and those typically studying smaller effects would be disproportionately vulnerable to this, even at magnitudes that would be inconsequential for larger effects. The present findings therefore provide cautious reassurance for many applied settings, while serving as an important reminder that the appropriateness of any analytical approach is contingent on the causal structure of the specific study. An assumption made for one mixed-mode setting cannot necessarily be generalised to others.

Implications and recommendations

The results have direct implications for the large body of existing research using mixed-mode data. This includes both studies involving variables subject to mode effects as well as those in which mode has been included as a covariate without explicit consideration of mode selection. Previous publications had already raised concerns that the common practice of naively controlling for mode may introduce bias (Institute for Social and Economic Research 2025; Maslovskaya et al. 2020; Tomova et al. 2026a; Wright et al. 2024). Whether this is the case, however, depends largely on the underlying causal structure in a given setting. It is important to note that this true underlying structure is typically unknown. However, our results indicate that a combination of large mode effects and large mode selection paths are what is necessary to introduce substantial bias (including but not limited to sign-reversal or full attenuation), especially where the substantive exposure-outcome relationship is weak. This is perhaps unlikely to be the case most of the time, especially given recent evidence that mode effects rarely reach magnitudes like 0.5 SD (Tomova et al. 2026b), suggesting that existing findings may not have been impacted substantially. Indeed, in our analysis of real mixed-mode data, the results across 20 illustrative models were essentially the same regardless of whether mode was controlled for or not. However, systematically reviewed empirical evidence on the typical size of mode selection paths does not exist, and it is therefore very difficult to make a sufficiently general statement regarding this. It is worth noting, however, that any consequences will be further minimised by the fact that typical mode splits in mixed-mode surveys include one dominant mode (Wright et al. 2024), which we demonstrated to offer further protection against bias.

Going forward, to translate these findings in practice, applied researchers are encouraged to consider the underlying causal structure of their specific mixed-mode dataset, before proceeding with an analysis if they anticipate mode effects and mode selection are present, and especially if they are considering controlling for mode. Finding information on the potential size of mode effects can be relatively straightforward, given the substantial experimental evidence now available, which has been recently synthesized in a systematic review, with the findings accessible via a searchable database of mode effects (Tomova et al. 2026b). In some cases, where a mode experiment may have been conducted as part of a mixed-mode wave, mode effects can be estimated directly from the data. Typically, however, a plausible assumption based on the existing literature is required. Assessing mode selection is considerably more challenging, as no equivalent experiments are possible to conduct for mode selection. In theory, each path from a variable of interest to mode can be estimated using appropriate data, given a set of confounders, chosen and measured sufficiently well. However, this is clearly only possible for observed variables, which the latent exposure and outcome are, by nature, not. It may therefore be more practical for researchers to reason about the mode allocation process qualitatively based on *a priori* theory. For example, by conceptualising it as a spectrum extending from a fully random process to a fully pre-determined one. Where the mode allocation process is random (e.g. in a mode experiment survey), then mode selection is not generally of concern, unless there is non-compliance or differential non-response, although it is rare to analyse mode experiments to answer substantive questions anyway. A concurrent mixed-mode design is also likely to have weaker mode selection than a sequential design, since initial non-response in the latter introduces an additional selection mechanism. Finally, whenever mode is pre-allocated, mode selection is likely to be largest in size, as it would usually be allocated based on pre-defined characteristics, and may be near-deterministic in nature (e.g. those without access to the internet are automatically allocated to non-web modes). When strong mode selection cannot be ruled out, but mode effects need correcting, then researchers must consider alternative strategies for handling mode effects, such as quantitative bias analysis (Fox et al. 2021; VanderWeele and Li 2019; Wright et al. 2024).

Strengths and limitations

We used data simulations, which allowed the extent of bias to be quantified against a known ground truth – something not possible in real data, where the true relationship between variables is never known. By varying the strength of mode effects, mode selection, the mode split, and the substantiative effect of interest, we were able to examine a broad range of structural scenarios and draw conclusions that offer a more complete understanding of the mechanisms that can lead to bias, beyond what can be achieved in any single dataset or study design. The explicit specification of data generating mechanisms using DAGs adds particular transparency. We also complemented these simulations with an empirical illustration across many models using real mixed-mode data. Although this does not offer the properties of simulated data in terms of knowing the ground truth, it provides a practical sense of whether and how much controlling for mode may actually alter observed estimates, and whether this change is likely to reflect a decrease or an increase in bias, based on *a priori* theoretical grounding and empirical evidence on mode effects.

For simplicity and because we were not aiming to simulate a specific single scenario, all variables were simulated as continuous and normally distributed with linear relationships assumed throughout, which may not reflect the complexity of real survey data. To maintain feasibility and a realistic number of specifications, we only examined scenarios in which all path coefficients were in the same direction (positive). The direction of bias would reverse if all signs were reversed, and its magnitude can differ when signs are mixed, since contributions from paths with opposing signs may partly offset one another and result in effects being cancelled out. The simulation parameters, while plausible, were chosen for illustrative purposes, and real-world values may of course differ substantially. The precise magnitudes of bias reported here may therefore not always translate to all real-world settings. In both simulated and real data, we only considered two-mode designs. We did not consider other sources of measurement error beyond mode effects, nor effect modification, nor did we explicitly aim to simulate and contrast cross-sectional and longitudinal designs. These simplifications are inherent to the simulation approach and do not undermine the qualitative conclusions but suggest that the precise magnitudes of bias reported here should be interpreted as indicative rather than definitive. Although we used 40 different variables from the NCDS age 55 follow-up, the results might have differed if other variables, or other exposure-outcome pairs among those variables, had been selected. However, it is challenging to assign equal numbers of variables into the pre-defined mode selection and mode effect classifications, which limits the number of variables, and therefore models, that can be explored.

Conclusion

Whether mixed-mode survey designs can introduce bias in substantive analyses largely depends on the underlying structure and characteristics of the dataset. Substantial bias can arise under specific conditions, particularly when the variables of substantive interest are sources of mode selection, when the mode effects are large, or when respondents are more equally split between modes. However, these conditions may not be commonly met in practice, as illustrated by our empirical analysis. Naïvely controlling for mode is not a universally appropriate way of handling mode effects and can introduce, rather than reduce, bias, in the presence of strong mode selection. Applied researchers should carefully consider the structure and characteristics of their specific mixed-mode data setting when deciding whether, and how, to account for mode effects, recognising that assumptions made in one context cannot be generalised to others.

Supplementary materials

Available on the Open Science Framework: https://osf.io/q75bn/

Statements

Funding

This research was supported by UKRI ESRC strategic research grant (ES/X014150/1) for "Survey data collection methods collaboration: securing the future of social surveys", known as Survey Futures. The UCL Centre for Longitudinal Studies is also supported by ESRC (ES/W013142/1). The funder had no role in the study conceptualisation, analysis, writing, or decision to publish.

Data availability

All data from the National Child Development Study used in the analyses are freely available from the UK Data Service (http://doi.org/10.5255/UKDA-Series-2000032). The R code used to simulate and analyse all data is available on the Open Science Framework (https://osf.io/q75bn/).

Conflicts of interest

None.

Acknowledgements

None.

Authorship (CRediT)

GDT: Conceptualisation, Investigation, Formal analysis, Visualisation, Writing – original draft; RJS: Conceptualisation, Writing – review & editing, Funding acquisition; LW: Conceptualisation, Writing – review & editing, Funding acquisition.

References

Arah, O. A. (2019), "Analyzing Selection Bias for Credible Causal Inference: When in Doubt, DAG It Out," *Epidemiology*, 30, 517–520. https://doi.org/10.1097/EDE.0000000000001033.

d'Ardenne, J., Bull, R., Das, A., Perera, Z., and Sexton, O. (2025), "Survey Practice Guide 2: How to mitigate against measurement effects when surveys move online."

Arel-Bundock, V., Greifer, N., and Heiss, A. (2024), "How to Interpret Statistical Models Using marginaleffects for *R* and *Python*," *Journal of Statistical Software*, 111. https://doi.org/10.18637/jss.v111.i09.

Brown, M., and Calderwood, L. (2020), "Mixing modes in longitudinal surveys: an overview. CLS Working Paper 2020/3.," London: UCL Centre for Longitudinal Studies.

Brown, M., and Hancock, M. (2015), *National Child Development Survey 2013 Follow-up: A guide to the datasets*, London: Centre for Longitudinal Studies. https://doi.org/https://cls.ucl.ac.uk/wp-content/uploads/2017/07/ncds_2013_follow_up_guide_to_the_datasets.pdf.

Clarke, P. S., and Bao, Y. (2022), "Estimating mode effects from a sequential mixed-mode experiment using structural moment models," *The Annals of Applied Statistics*, 16. https://doi.org/10.1214/21-AOAS1557.

Cole, S. R., Platt, R. W., Schisterman, E. F., Chu, H., Westreich, D., Richardson, D., and Poole, C. (2010), "Illustrating bias due to conditioning on a collider," *International Journal of Epidemiology*, Oxford University Press (OUP), 39, 417–420. https://doi.org/10.1093/ije/dyp334.

Digitale, J. C., Martin, J. N., and Glymour, M. M. (2022), "Tutorial on directed acyclic graphs," *Journal of Clinical Epidemiology*, 142, 264–267. https://doi.org/10.1016/j.jclinepi.2021.08.001.

Dolan, P., and Kavetsos, G. (2016), "Happy Talk: Mode of Administration Effects on Subjective Well-Being," *Journal of Happiness Studies*, 17, 1273–1291. https://doi.org/10.1007/s10902-015-9642-8.

Fox, M. P., MacLehose, R. F., and Lash, T. L. (2021), *Applying Quantitative Bias Analysis to Epidemiologic Data*, Statistics for Biology and Health, Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-82673-4.

Goodman, A., Brown, M., Silverwood, R. J., Sakshaug, J. W., Calderwood, L., Williams, J., and Ploubidis, G. B. (2022), "The impact of using the Web in a mixed-mode follow-up of a longitudinal birth cohort study: Evidence from the National Child Development Study," *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 185, 822–850. https://doi.org/10.1111/rssa.12786.

Greenland, S. (2017), "Invited Commentary: The Need for Cognitive Science in Methodology," *American Journal of Epidemiology*, 186, 639–645. https://doi.org/10.1093/aje/kwx259.

Greenland, S., Pearl, J., and Robins, J. M. (1999), "Confounding and Collapsibility in Causal Inference," *Statistical Science*, 14. https://doi.org/10.1214/ss/1009211805.

Hernan, M. A., and Cole, S. R. (2009), "Invited Commentary: Causal Diagrams and Measurement Bias," *American Journal of Epidemiology*, 170, 959–962. https://doi.org/10.1093/aje/kwp293.

Hox, J., De Leeuw, E., and Klausch, T. (2017), "Mixed-Mode Research: Issues in Design and Analysis," in *Total Survey Error in Practice*, eds. P. P. Biemer, E. Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, and B. T. West, Wiley, pp. 511–530. https://doi.org/10.1002/9781119041702.ch23.

Institute for Social and Economic Research (2025), Understanding Society: Waves 1-15, 2009-2024 and Harmonised BHPS: Waves 1-18, 1991-2009, User Guide, Colchester, UK: University of Essex.

Jäckle, A., Roberts, C., and Lynn, P. (2006), Telephone versus face-to-face interviewing: mode effects on data quality and likely causes: report on phase II of the ESS-Gallup mixed mode methodology project, ISER Working Paper Series.

de Leeuw, E. (2005), "To Mix or Not to Mix Data Collection Modes in Surveys," *Journal of Official Statistics*, 21, 233–255.

Lortie-Forgues, H., and Inglis, M. (2019), "Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?," *Educational Researcher*, 48, 158–166. https://doi.org/10.3102/0013189X19832850.

Lovakov, A., and Agadullina, E. R. (2021), "Empirically derived guidelines for effect size interpretation in social psychology," *European Journal of Social Psychology*, 51, 485–504. https://doi.org/10.1002/ejsp.2752.

Martin, P. (2011), "A Good Mix? Mixed Mode Data Collection and Cross-national Surveys," Research & Methods, 20, 5–26.

Martin, P., and Lynn, P. (2011), "The Effects of Mixed Mode Survey Designs on Simple and Complex Analyses," London: Centre for Comparative Social Surveys.

Maslovskaya, O., Calderwood, L., Ploubidis, G., and Nicolaas, G. (2020), "GenPopWeb2: Adjustments for Mode Effects."

McShane, B. B., Gal, D., Gelman, A., Robert, C., and Tackett, J. L. (2019), "Abandon Statistical Significance," *The American Statistician*, 73, 235–245. https://doi.org/10.1080/00031305.2018.1527253.

Morris, T. P., White, I. R., and Crowther, M. J. (2019), "Using simulation studies to evaluate statistical methods," *Statistics in Medicine*, 38, 2074–2102. https://doi.org/10.1002/sim.8086.

Piccitto, G., Liefbroer, A. C., and Emery, T. (2022), "Does the Survey Mode Affect the Association Between Subjective Well-being and its Determinants? An Experimental Comparison Between Face-to-Face and Web Mode," *Journal of Happiness Studies*, 23, 3441–3461. https://doi.org/10.1007/s10902-022-00553-y.

R Core Team (2025), "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing.

Schäfer, T., and Schwarz, M. A. (2019), “The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases,” *Frontiers in Psychology*, 10, 813. https://doi.org/10.3389/fpsyg.2019.00813.

Textor, J., Van Der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., and Ellison, G. T. H. (2017), “Robust causal inference using directed acyclic graphs: the R package ‘dagitty,’” *International Journal of Epidemiology*, dyw341. https://doi.org/10.1093/ije/dyw341.

Tomova, G. D., Silverwood, R. J., Tennant, P. W. G., and Wright, L. (2026a), “How can the use of different modes of survey data collection introduce bias? An introduction to mode effects using directed acyclic graphs (DAGs),” *American Journal of Epidemiology*, kwag017. https://doi.org/10.1093/aje/kwag017.

Tomova, G. D., Silverwood, R. J., and Wright, L. (2026b), “Mode effects on survey item measurement: A systematic review of the experimental evidence,” SocArXiv. https://doi.org/10.31235/osf.io/bc7qn_v1.

Van Smeden, M., Lash, T. L., and Groenwold, R. H. H. (2020), “Reflection on modern methods: five myths about measurement error in epidemiological research,” *International Journal of Epidemiology*, 49, 338–347. https://doi.org/10.1093/ije/dyz251.

VanderWeele, T. J., and Li, Y. (2019), “Simple Sensitivity Analysis for Differential Measurement Error,” *American Journal of Epidemiology*, 188, 1823–1829. https://doi.org/10.1093/aje/kwz133.

Vannieuwenhuyze, J. T. A. (2015), “Mode effects on variances, covariances, standard deviations, and correlations,” *Journal of Survey Statistics and Methodology*, 3, 296–316. https://doi.org/10.1093/jssam/smv009.

Villar, A., and Fitzgerald, R. (2017), “Using Mixed Modes in Survey Research: Evidence from Six Experiments in the ESS.,” in *Values and Identities in Europe*, Routledge.

Wright, L., Ploubidis, G., and Silverwood, R. (2024), “Handling mode effects in the CLS cohort studies: User guide.”

Figures

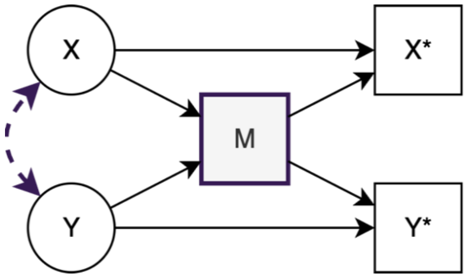

Figure 1. A directed acyclic graph (DAG) depicting collider bias induced by conditioning on survey mode. Ellipses depict latent (true) variables, whereas squares depict their observed (measured) versions. Mode (M) is caused by both the latent exposure (X) and outcome (Y). Conditioning on M induces a spurious association between X and Y (dashed arrow). Although according to this DAG there is no true $X \rightarrow Y$ effect, an association may therefore be observed simply as a result of conditioning on mode.

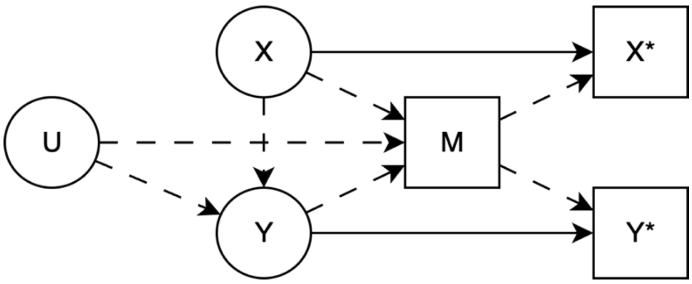

Figure 2. A directed acyclic graph (DAG) showing the data generating mechanism used to simulate data. Ellipses depict latent variables, i.e. the exposure (X), outcome (Y), and an unobserved common cause of mode and the outcome (U); squares depict observed variables, i.e. survey mode (M) and the observed measurements of the exposure (X^*) and outcome (Y^*). Dashed paths, representing the focal relationship $(X \rightarrow Y)$, mode selection $(X \rightarrow M, Y \rightarrow M, U \rightarrow M)$, unobserved causes of the outcome $(U \rightarrow Y)$, and mode effects $(M \rightarrow X^*, M \rightarrow Y^*)$ show path coefficients which were varied in their presence and magnitude across simulations. Solid paths, representing the structural part between a latent variable and its measured form $(X \rightarrow X^*, Y \rightarrow Y^*)$ were not varied.

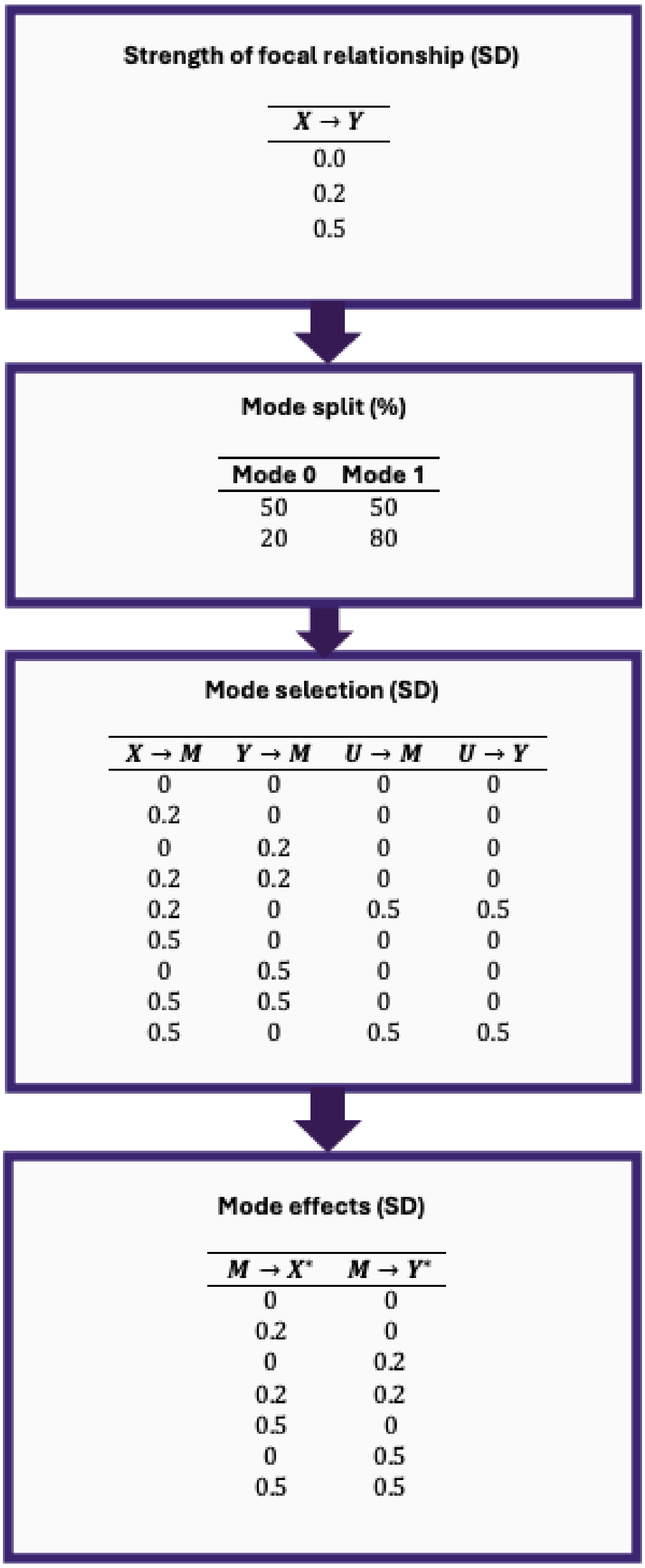

Figure 3. Overview of the simulation flow and parameters, depicting the hierarchical structure of the simulation design, with each box representing a parameter that was varied and the corresponding values or combinations considered. The flow corresponds to the simulation process in which each variation was nested within the previous.

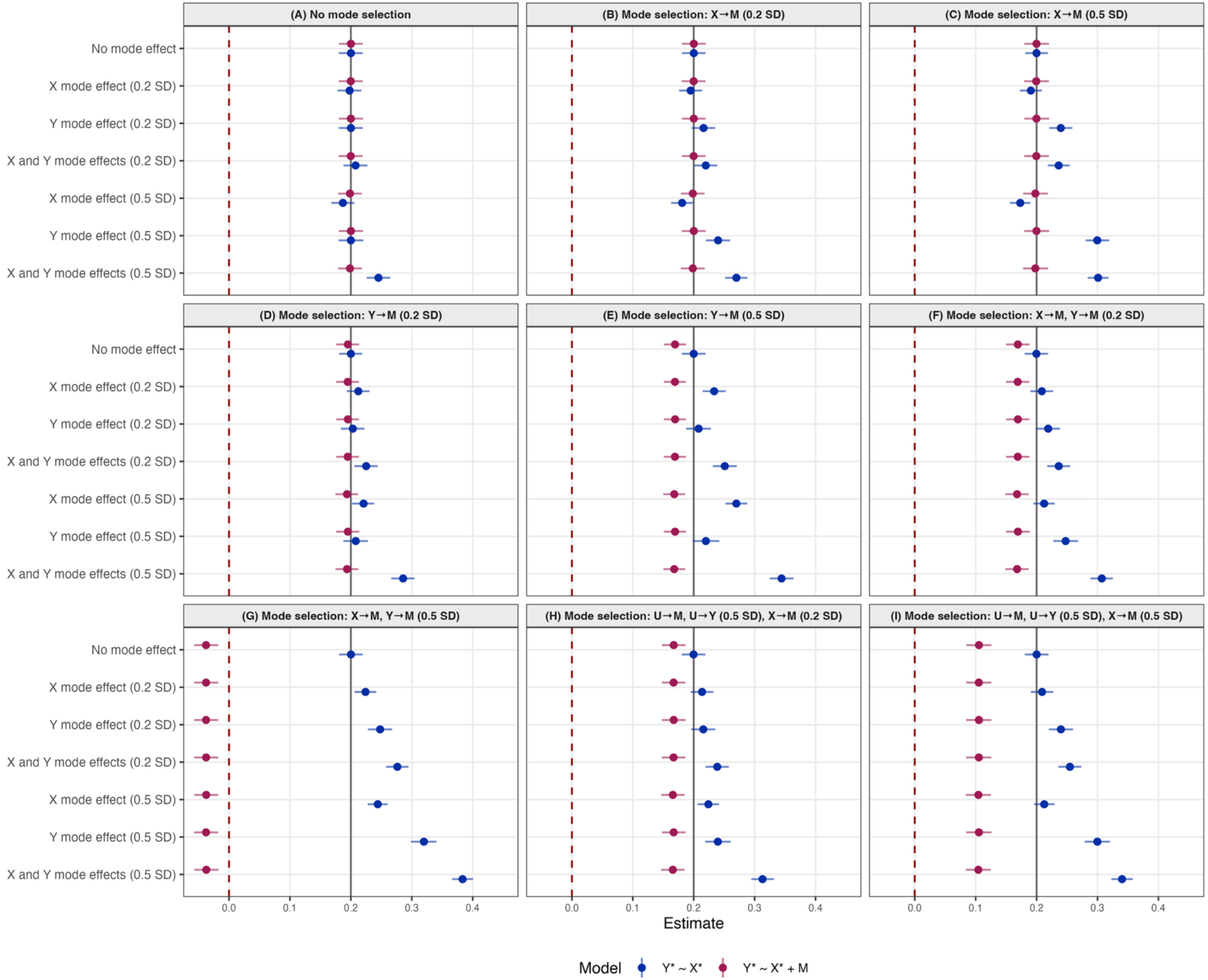

Figure 4. Results from simulated data with true $X \rightarrow Y$ effect of 0.2 SD and 50-50% mode split. Each panel corresponds to a different mode selection scenario (varying the path coefficients $X \rightarrow M,\ Y \rightarrow M,\ U \rightarrow M,\ and\ U \rightarrow Y$), and within each panel, different mode effect conditions are depicted along the y-axis. Points represent the median estimated effect of X^* on Y^* across 10,000 simulations, with error bars showing the 95% SI. The grey vertical line indicates the true effect (0.2 SD), whereas the red dashed line indicates a null effect of zero.

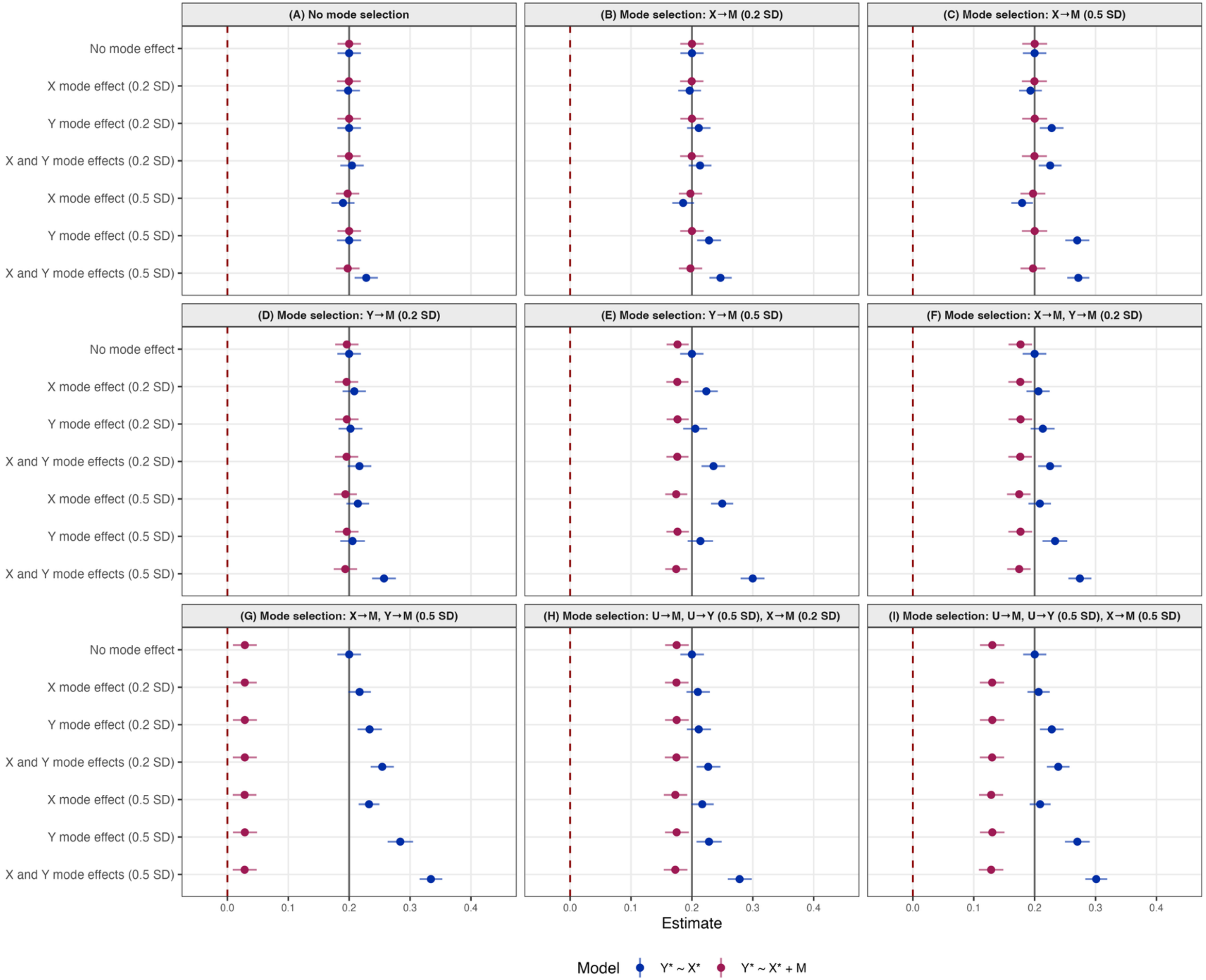

Figure 5. Results from simulated data with true $X \rightarrow Y$ effect of 0.2 SD and 80-20% mode split. Each panel corresponds to a different mode selection scenario (varying the path coefficients $X \rightarrow M$, $Y \rightarrow M$, $U \rightarrow M$, *and* $U \rightarrow Y$), and within each panel, different mode effect conditions are depicted along the y-axis. Points represent the median estimated effect of X^* on Y^* across 10,000 simulations, with error bars showing the 95% SI. The grey vertical line indicates the true effect (0.2 SD), whereas the red dashed line indicates a null effect of zero.

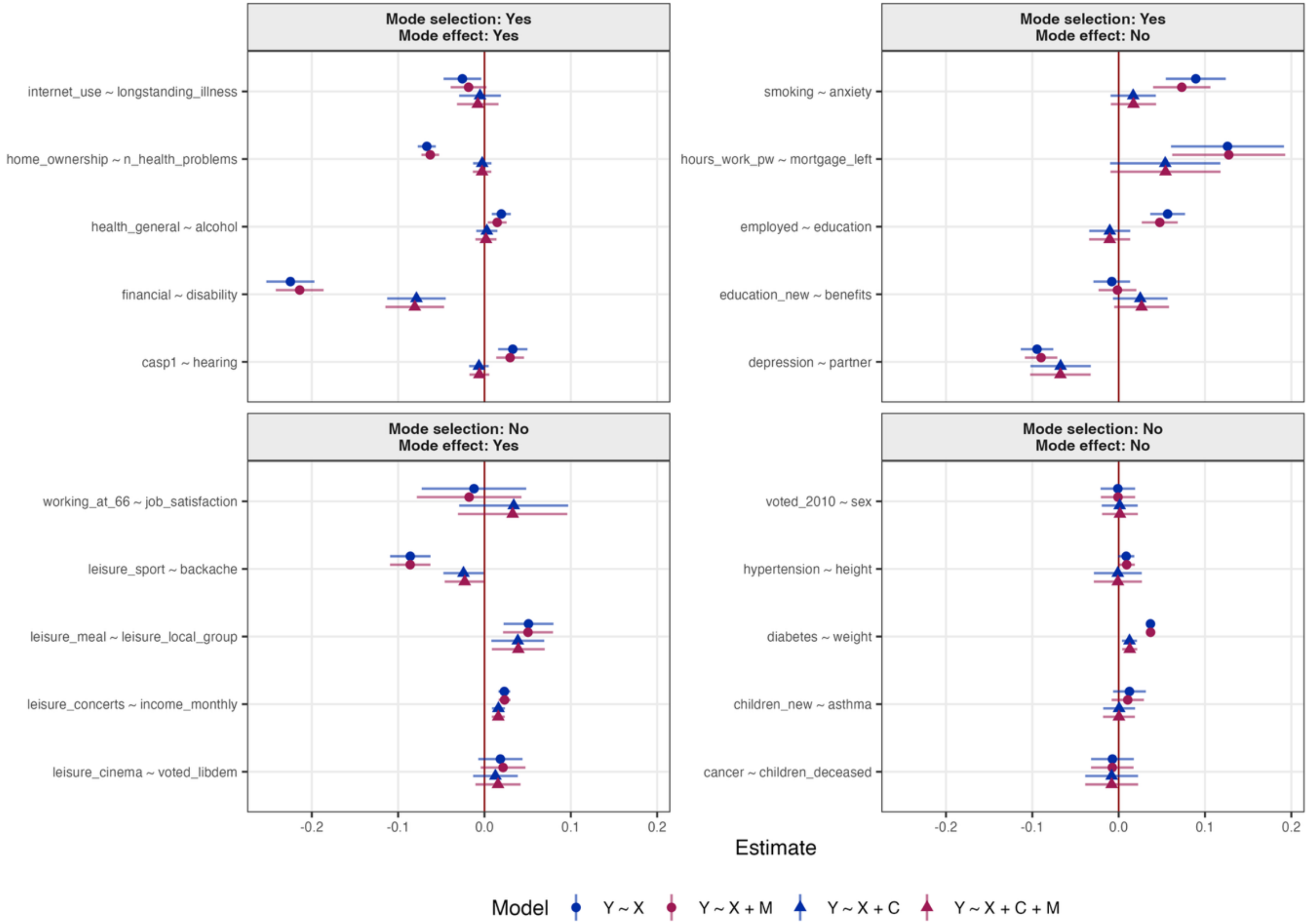

Figure 6. Results for the illustrative analysis of mixed-mode data from NCDS Age 55. Each panel corresponds to a different combination of mode selection and mode effect status (classified as described in the Methods). Within each panel, outcome~exposure pairs are displayed along the y-axis, and point estimates with 95% confidence intervals are shown for four models differing in their adjustment for mode (M) and confounders (C). For a list of confounders, see Table S3. Comparisons between models should be made for models containing the same variables but differing in adjustment for M only. Estimates are presented as average marginal effects: standardised effect sizes (SD units) for continuous variables and probability differences for binary variables.