

Selection on the Outcome: The Fraction of Nonresponse Bias That Survey Weighting Removes, with an Application to Subjective Well-Being

Shay Tsaban

shay.tsaban@gmail.com

Peres Academic Center

Research Article

Keywords: nonresponse bias, survey weighting, selection on the outcome, missing not at random, subjective well-being, research practice

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10555816/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Selection on the Outcome: The Fraction of Nonresponse Bias That Survey Weighting Removes, with an Application to Subjective Well-Being

Abstract

Weighting on socio-demographic auxiliaries is the standard remedy for unit nonresponse in social surveys. This paper asks what that remedy accomplishes when participation depends on the survey outcome itself rather than only on the auxiliaries. Under such a mechanism, weighting removes a share of the resulting bias exactly equal to the proportion of variance in the outcome that the auxiliaries explain, and leaves the complementary share in place. The residual share depends on neither the strength of the selection nor the power of the auxiliaries to predict participation; it is governed by a single estimable quantity, and the result generalises beyond the linear case. The severity of the problem therefore varies across research fields according to how far their outcomes are demographically patterned. Subjective well-being research is shown to sit at the unfavourable extreme. In 69,719 General Social Survey respondents the variables conventional weighting schemes use explain 1.7 per cent of the variance in reported happiness, so weighting removes under two per cent of any outcome-dependent bias; the estimate is stable across eight specifications and has fallen since the 1970s. A contrasting case supports the mechanism: where national registers supplied auxiliaries that predict the outcome, weighting recovered most of a documented selection bias. An audit of 1,618 articles finds that 72 per cent rest on survey samples while 1.1 per cent mention nonresponse, attrition or coverage. Diagnostics drawn from health-related quality-of-life research are set out.

Keywords: nonresponse bias; survey weighting; selection on the outcome; missing not at random; subjective well-being; research practice

1 Introduction

Unit nonresponse is the central practical threat to inference from social surveys, and response rates have been falling for three decades (Luiten et al. 2020; Meyer et al. 2015). The standard remedy is weighting. Sampled units are reweighted using auxiliary variables for which population benchmarks exist, typically age, sex, education and region, on the understanding that this restores representativeness. The procedure is near-universal, and in most published work it is applied without comment.

Its logic is well understood. Weighting reduces bias to the extent that the auxiliary variables are associated with both the propensity to respond and the survey outcome (Bollen et al. 2016; Brick 2013; Kreuter et al. 2010). Where missingness depends on the unobserved value of the outcome itself, adjustment on observed covariates does not in general resolve it (Little and Rubin 2019; Rubin 1976). These principles are not in dispute. What they do not supply is a magnitude. They establish that adjustment on observables is inadequate under outcome-dependent missingness without saying how inadequate, because the answer depends on the outcome variable and therefore differs across research fields.

This paper supplies that magnitude, and shows that it is governed by a single quantity that any analyst can compute. Under a response mechanism in which the probability of participating depends on the outcome directly as well as on the auxiliaries, weighting removes in full the bias arising from the auxiliaries and removes a fraction of the outcome-driven bias equal to the proportion of variance in the outcome that those auxiliaries explain. The residual share is the complement of that proportion. It does not depend on how strong the selection is, so the result does not require the bias to be large, only to exist. It does not depend on how well the auxiliaries predict participation, so it cannot be dissolved by a better propensity model. The result is derived

first in a transparent linear benchmark and then generalised, and Section 10 examines what happens when each assumption fails.

A corollary follows immediately. Severity varies across fields according to how far their outcomes are demographically patterned. For income or educational attainment, which are strongly patterned, standard weighting performs comparatively well. For outcomes that demographics predict weakly, it performs badly, and the worse the demographic prediction the less weighting repairs.

The rest of the paper applies this to the field that sits at the unfavourable extreme. Subjective well-being has become an institutional statistic: national statistical offices collect it, international organisations rank countries by it, and governments are urged to treat it as an objective of policy (Adler and Seligman 2016; Diener and Seligman 2004; Diener et al. 2015; Frijters and Krekel 2021; Frijters et al. 2020; OECD 2013, 2025). Almost all of that evidence comes from sample surveys in which respondents are asked how satisfied they are with their lives, and the resulting averages are treated as estimates of population well-being (Clark 2018; Diener et al. 2018; Powdthavee 2015).

Two things make this field the sharpest case. Participation is not independent of the quantity being measured. Analyses of interviewer contact data in the German Socio-Economic Panel report a link between reported happiness and the effort required to obtain an interview (Chadi 2019); in a Norwegian population health study with prior-wave data on most non-participants, high mental distress predicted dropping out with an odds ratio of 1.84 (Torvik et al. 2012); and populations with plausibly low well-being, including people who are homeless, have been studied precisely because they sit outside the household frames the field otherwise depends on (Biswas-Diener and Diener 2006). And the governing parameter is unusually small,

because the modest explanatory power of socio-demographic characteristics for well-being is itself one of the better-established findings in that literature (Brown and Rohrer 2020; Dolan et al. 2008; Kaiser and Oswald 2022).

Those two facts have not been put together. Estimated in 69,719 General Social Survey respondents, the variables conventional weighting schemes use explain 1.7 per cent of the variance in reported happiness. Weighting therefore removes under two per cent of any outcome-dependent nonresponse bias and leaves more than ninety-eight. The same statistic that is reported as a substantive result about the limited role of circumstances is also the repair rate of the field's standard correction.

The direct evidence on how strongly well-being affects participation is genuinely mixed. Reported effects range from slight (Moum 1994) through moderate (Torvik et al. 2012) to a clear link with contact effort (Chadi 2019), and one panel study found no resulting bias at all (Herrera et al. 2021), a finding Section 10 engages at length. Because the residual share is invariant to the strength of selection, a claim that depended on settling that dispute would be premature. This one does not.

The argument rests on four procedures rather than on assertion, and each is set out so that a reader can check it. The first is a derivation, in Section 4, generalised beyond the linear benchmark and with its assumptions stated individually in Table 5 alongside the direction in which the conclusion moves if each fails. The second is a direct estimate of the governing parameter, in Section 5, computed in 69,719 respondents and stable across eight specifications. It is paired with a contrasting case in which weighting demonstrably did repair most of a documented selection bias, which tests whether the parameter explains the difference. The third is a calibration, in Section 6, spanning the full plausible range so that no single value carries the

conclusion. The fourth is a systematic audit of 1,618 articles published since 2023 across five well-being and quality-of-life journals, reported in Section 2 and Figure 1. It finds that 72 per cent present survey-based evidence while 1.1 per cent mention nonresponse, attrition, or coverage. Among survey-based articles the figure is 1.4 per cent, flat across four publication years, and a deliberately generous keyword list adjudicated article by article raises it only to about 1.5 per cent. Code accompanies the article.

Five specific contributions follow. First, a closed-form expression for the share of outcome-dependent nonresponse bias that survives standard weighting, with a general form that does not require linearity. Second, the identification of the single parameter that governs that share, and the observation that this literature already estimates it for unrelated reasons. Table 6 states which prior results the argument uses and which claims are made here, each written so that it can be refuted. Third, an estimate of that parameter, which converts the result from a conditional statement into a magnitude. Fourth, evidence on the scale of the omission, from an audit whose denominator is the field's own recent output. Fifth, a transferable remedy set: Section 9 imports five procedures from health-related quality-of-life research, where outcome-dependent missingness has been treated as a first-order problem for two decades, and Section 8 with Table 7 works out the direction of the induced error in six standard inferences.

The scope of what is affected is wide. Because roughly seven in ten articles in this field rest on survey samples, the result bears on the majority of what is published in it, from cross-national comparison to well-being inequality.

Section 2 contrasts the intensity with which this field has audited its measurement scale against the attention it has given its sampling frame, and reports the audit. Section 3 sets out the mechanisms through which well-being enters the participation decision. Section 4 derives the

main result. Section 5 estimates the parameter that governs it and examines a contrasting case. Section 6 calibrates the result across the full parameter range. Section 7 states precisely what is and is not new. Section 8 works through the consequences for six standard inferences. Section 9 sets out diagnostics and reporting practices. Section 10 considers objections, and Section 11 concludes.

2 Two audits, unequally pursued

Well-being research has subjected its measurement instrument to sustained scrutiny. A substantial literature asks whether responses on an ordered scale can support the comparisons routinely drawn from them. T. N. Bond and Lang (2019) show that the conditions for non-parametric identification of two groups' relative average happiness are strong and generally unmet, and that ordered probit results in nine prominent research areas can be reversed under plausible monotone transformations. Kaiser and Vendrik (2020) assess how threatening such transformations are in practice. Kaiser and Oswald (2022) approach the question from the other direction, showing that a single feelings integer predicts later consequential behaviour and carries more predictive power than a collection of socio-economic influences. Ferrer-i-Carbonell and Frijters (2004) examine how much estimated determinants depend on methodological choices, including the treatment of ordinality and fixed effects.

The same care is visible in work on the interpretation of estimated effects. Hirata and Groß (2025) argue that relative effects on life satisfaction are routinely attributed to social comparison even where the data cannot identify the mechanism, and they propose a scheme that separates effects operating through aspirations from those operating through capabilities. Their objection is that a conclusion is being drawn which the design does not license. The argument of this paper has the same form, applied to a different link in the inferential chain.

The scrutiny extends well beyond ordinality. Reliability has been estimated directly (Krueger and Schkade 2008). The dependence of reports on transient context, question order, and framing has been examined for decades (Diener et al. 1999). Different measures behave differently against the same covariate, most famously income (Deaton and Stone 2014; Kahneman and Deaton 2010). The salience of a covariate at the moment of answering can itself distort reports (Kahneman et al. 2006), and whether a single item suffices remains argued (Helliwell et al. 2026). Whether cross-national differences reflect culture or cultural measurement bias has been posed as a distinct question (Veenhoven 2012). A national panel was convened to assess what these measures can support (National Research Council 2013), and international guidance codifies much of the accumulated caution (OECD 2013, 2025).

The response to one comparability threat is especially instructive. Respondents in different countries may use the same verbal categories differently. A dedicated instrument was built to correct that variation: anchoring vignettes, which fix a common referent across groups (King et al. 2003) and have been applied to life satisfaction across European countries (Angelini et al. 2014). The field recognised a threat to comparability, formalised it, and built a tool for it. No comparable instrument exists for the possibility that the samples themselves stand in different relations to their populations, although that threat operates on the same comparisons and in the same direction.

This is a healthy record, and it has been directed almost entirely at one component of the inference: the mapping from an internal state to a reported number. The other component has received far less attention. Whatever the scale means, the average is computed over the people who answered, and if those people are not a random draw with respect to the quantity being

measured, the reported average estimates something other than the population parameter. No refinement of the scale repairs that.

To gauge how much attention this receives, all articles published from January 2023 onwards in five well-being and quality-of-life journals were retrieved through OpenAlex. These are the principal outlets for quantitative work in the field. Abstracts were searched for terms indicating survey-based evidence and terms indicating engagement with nonresponse, attrition, coverage, or weighting.

Table 1 Salience of survey-based evidence and of nonresponse in five well-being and quality-of-life journals, articles published from January 2023

Journal	Articles with abstracts	Survey-based (%)	Nonresponse, attrition or coverage (%)	Survey weights (%)	Of survey-based articles, mention nonresponse (%)
Applied Research in Quality of Life	204	79	1.0	0.0	1.2
Journal of Happiness Studies	197	69	0.0	0.0	0.0
Journal of Positive Psychology	298	60	0.7	0.0	1.1
Quality of Life Research	508	82	2.4	0.6	2.6
Social Indicators Research	411	66	0.5	0.5	0.7
All five	1,618	72	1.1	0.3	1.4

Note. Counts are of abstracts containing at least one term in the relevant set. The final column restricts the denominator to the 1,165 survey-based articles, the population for which the question arises. The measure captures salience, not complete neglect.

Seventy-two per cent present survey-based evidence, just over one per cent mention nonresponse, attrition, or coverage anywhere in the abstract, and three in a thousand mention weighting. Restricting the denominator to the 1,165 articles that actually rest on survey samples

raises the figure only to 1.4 per cent. A check on titles across the full back catalogues returns none at all containing nonresponse, attrition, selection bias, or sampling frame in the two general well-being outlets.

The pattern is stable rather than residual. Across the four publication years covered, the share of survey-based articles rises slightly, from 69 per cent in 2023 to 73 per cent in 2026, while the share mentioning nonresponse moves within a band from 0.9 to 1.3 per cent and shows no trend. Figure 1 displays both cuts. This is not a legacy problem that current practice is correcting.

The measure is deliberately conservative. An abstract is a compressed genre, and its silences are evidence about salience rather than proof of neglect. Salience is nonetheless the relevant quantity: issues bearing on whether a headline estimate is credible normally reach the abstract when authors regard them as consequential.

A keyword classifier can err in both directions, so it was checked in both. The search above uses a narrow term list. Repeating it with a deliberately generous list, adding dropout, loss to follow-up, retention rate, self-selection, volunteer bias, participation rate, non-participant, missing data, representativeness and generalisability, raises the count from 18 to 47 articles, or from 1.1 to 2.9 per cent. Each of the 29 additional articles was then read. Most use the vocabulary for a different construct: generalisability of a scale across cultures or age groups, missing data as an indicator of instrument acceptability, self-selection into a treatment rather than into a survey, and in one case labour force participation. Six engage genuinely with missingness that could depend on the outcome, and all six are in the health-related journal. Adjudicating the broad list therefore yields roughly 24 articles, or 1.5 per cent, which is the figure the narrow and

broad measures bracket. The conclusion does not turn on where within that interval the true value lies.

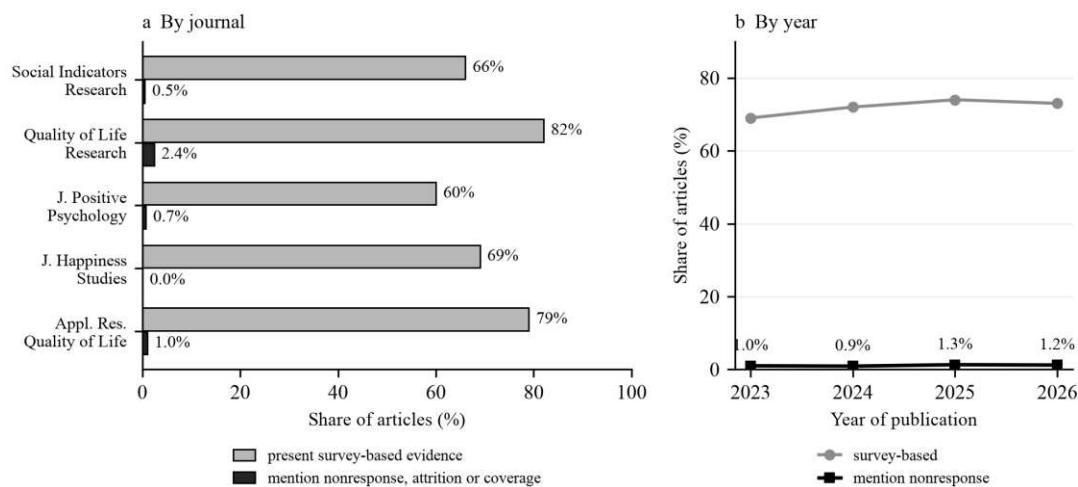


Fig. 1 Share of articles presenting survey-based evidence and share mentioning nonresponse, attrition or coverage, by journal (a) and by year of publication (b), for 1,618 articles published from January 2023 in five well-being and quality-of-life journals

One journal in Table 1 stands apart. Quality of Life Research, whose remit is health-related quality of life, records engagement roughly four times that of the others, and of a different character. Its articles describe outcome data as missing not at random because patients who deteriorate stop completing questionnaires, ask whether non-responders to screening instruments face worse survival, and apply inverse probability weighting and worst-case imputation to bound the consequences. That field met the same structural problem earlier and built machinery for it. Section 9 returns to what can be borrowed.

3 How well-being enters the participation decision

The concern of this paper is a response mechanism in which well-being itself, rather than only its observable correlates, affects the probability of appearing in the data. Four channels are documented in existing evidence. Table 2 summarises them.

Table 2 Channels through which well-being enters the probability of appearing in survey data

Channel	What it is	Evidence	Expected direction
Frame noncoverage	Groups outside the household sampling frame are never eligible	Well-being studied directly in homeless populations, which sit outside household frames (Biswas-Diener and Diener 2006; Johnstone et al. 2015; Panadero et al. 2015) and in prisons (van der Kaap-Deeder et al. 2017)	Excludes low well-being
Unit nonresponse	Sampled persons decline or are not contacted	Reported happiness associated with contact effort and participation motivation (Chadi 2019); a slight direct effect of poor self-assessed health on non-cooperation (Moum 1994)	Under-represents low well-being
Attrition	Panel members leave between waves	Mental distress predicts dropout, moderately (Torvik et al. 2012); attrition follows distinct latent patterns rather than a single exit (Lugtig 2014); general determinants of longitudinal response reviewed by Watson and Wooden (2009)	Retains higher well-being
Reporting distortion	Participation occurs but the report is displaced	Happiness overstated and depression understated in non-anonymous modes (Reisinger 2022)	Inflates observed well-being

The first three are selection: they determine who is in the data. The fourth is measurement: it determines what a present respondent reports. They are kept distinct here because weighting addresses only the first three, and the argument below concerns selection. The

fourth is noted because its expected direction coincides, so that treating it as an independent reassurance would be a mistake.

Both features of these channels matter for what follows. They point in a common direction, so their effects accumulate rather than offset. And they operate on well-being directly rather than only through demographic characteristics: a person in poor health drops out of a panel because of the condition itself, not because of the age cell to which the weighting scheme assigns them.

The health literature adds further evidence of the same shape. Non-response bias has been evaluated directly against mental health outcomes in a large adolescent cohort (de Winter et al. 2005), and in a Swiss health survey where second-stage non-response was traced to determinable characteristics (Volken 2013). Where participation is voluntary rather than compulsory, respondents who are hardest to recruit report more risk behaviour, so prevalence is understated (Cheung et al. 2017). Attrition over fifteen years in a population cohort was concentrated among those with lower education and poorer outcomes (Gustavson et al. 2012). Nonresponse has been analysed in the American Time Use Survey, a source used for well-being research through the day reconstruction approach (Abraham et al. 2006), and response rates in the principal European comparative survey have been documented at length (Stoop et al. 2010).

Evidence from the wider survey literature counsels against assuming either that nonresponse must bias estimates or that it cannot. A telephone survey conducted at much greater effort than its standard counterpart produced very few significant differences across a wide range of attitudinal measures (Keeter et al. 2000). Experimental work shows that nonresponse bias can be induced, and that it varies by the topic measured (Groves et al. 2006). Bias is not predictable from the response rate in either direction (Meterko et al. 2015). Whether it arises depends on the

relation between participation and the particular variable being estimated, which is why the question has to be settled separately for each outcome rather than inherited from studies of other ones.

The general survey-methodology literature has long established the relevant principles. Nonresponse bias in a respondent mean equals the covariance between response propensity and the survey variable, divided by the mean response propensity (Brick 2013; Groves 2006; Groves and Couper 1998). Its consequences depend on which estimate is at issue rather than on the rate alone (Peytchev 2012). Response rates are poor predictors of nonresponse bias, because bias depends on that covariance rather than on the rate itself (Groves and Peytcheva 2008). Weighting adjustments reduce bias effectively only where the auxiliary variable is strongly associated with both the response indicator and the survey outcome, a requirement the authors of that result describe as often overlooked in practice (Kreuter et al. 2010; see also Haziza and Lesage 2016). Missingness that depends on the unobserved value of the variable itself is not ignorable, and adjustment on observed covariates does not in general resolve it (Little and Rubin 2019; Rubin 1976).

Two related phenomena are set aside. One is panel conditioning, where repeated interviewing changes what respondents report (Warren and Halpern-Manners 2012). The other is variation in reporting with the mode and privacy of administration, and with the cognitive process of answering (Kreuter et al. 2008; Reisinger 2022; Tourangeau et al. 2000). Both matter and neither is addressed by weighting, but neither is the subject of this paper.

These principles are not in dispute. What has not been established is what they imply quantitatively for this field, given what this field already knows about its own outcome variable. That is the question of the next section.

The context makes it pressing. Response rates in the surveys underpinning well-being research have fallen substantially, with fieldwork effort rising to partially offset the decline (Luiten et al. 2020; Meyer et al. 2015). Online data collection has introduced fraudulent and automated responses at a scale that has forced explicit screening protocols (Goodrich et al. 2023). The share of a target population that a survey actually reaches has been shrinking for two decades.

4 What weighting removes and what it leaves

4.1 Setup

Let Y denote well-being in a target population, with mean $\mu = E[Y]$ and variance σ_Y^2 . Let R be a response indicator and ρ the response propensity. The estimand is μ . The respondent mean is $E[Y | R = 1]$, and its nonresponse bias is

$$B = E[Y | R = 1] - \mu = \frac{\text{Cov}(\rho, Y)}{\bar{\rho}},$$

where $\bar{\rho} = E[\rho]$. This expression is standard (Brick 2013; Groves 2006).

Let X be the auxiliary variables used in the weighting adjustment, typically age, sex, education, region, and household composition. Weighting on X , by post-stratification or response-propensity adjustment, yields an estimator consistent for $E_X[E[Y | X, R = 1]]$, with bias equal to the average within-cell covariance,

$$B_w = E_X \left[\frac{\text{Cov}(\rho, Y | X)}{\bar{\rho}(X)} \right].$$

Weighting therefore removes the component of the propensity-outcome covariance that runs between values of X and retains the component that runs within them.

4.2 The response mechanism

Consider a response propensity that depends on well-being directly as well as on the auxiliaries,

$$\rho = \alpha + \beta Y + \gamma' X,$$

with $\beta > 0$ representing the tendency of people with higher well-being to participate.

This is a linear probability specification, adopted for transparency. The results below require only that the dependence be approximately linear over the relevant range.

Two auxiliary assumptions are made explicit. First, $\text{Var}(Y | X)$ does not vary systematically with X . Second, realised response rates do not vary so sharply across cells of X that $\bar{\rho}(X)$ departs materially from $\bar{\rho}$. Both are approximations, and Section 10 considers their failure.

4.3 The result

Conditioning on X fixes the term $\gamma' X$, so within a cell the only source of covariance between propensity and outcome is the direct dependence on Y :

$$\text{Cov}(\rho, Y | X) = \beta \text{Var}(Y | X) = \beta \sigma_Y^2 (1 - R_{Y.X}^2),$$

where $R_{Y.X}^2$ is the proportion of variance in well-being explained by a regression of Y on the auxiliaries X . Unconditionally, $\text{Cov}(\rho, Y) = \beta \sigma_Y^2 + \text{Cov}(\gamma' X, Y)$. Taking the ratio of the two bias expressions gives the central result.

Proposition 1. Under the response mechanism above, weighting on X removes in full the bias arising from the dependence of participation on the auxiliaries, and removes the fraction $R_{Y.X}^2$ of the bias arising from the direct dependence of participation on well-being. The residual share of outcome-dependent bias is

$$\frac{B_w}{B_{\text{direct}}} = 1 - R_{Y \cdot X}^2.$$

The result specialises established principles rather than challenging them. Its value lies in identifying which single quantity governs the outcome, and in showing that the quantity is one this literature already estimates.

The linear specification is a benchmark rather than a restriction, and it is worth stating what survives without it. Let the propensity take the general form $\rho = g(X) + h(Y)$, with h increasing but otherwise unrestricted, and drop the homoskedasticity assumption. Conditioning on X again annihilates $g(X)$, so the residual bias is governed by $E_X[\text{Cov}(h(Y), Y \mid X)]$ while the unconditional direct component is $\text{Cov}(h(Y), Y)$.

Proposition 2 (general case). For any increasing h , the share of outcome-dependent bias remaining after weighting on X is

$$\frac{E_X[\text{Cov}(h(Y), Y \mid X)]}{\text{Cov}(h(Y), Y)},$$

which equals $1 - R_{Y \cdot X}^2$ when h is linear and $\text{Var}(Y \mid X)$ is constant, and equals unity whenever X is independent of Y .

The general expression is a ratio of a within- X average to an unconditional quantity, so it is bounded above by one and falls only to the extent that conditioning on X reduces the covariation between well-being and the part of participation that responds to it. Nothing about the shape of h changes that structure. What the linear case adds is an interpretable value for the ratio, namely a statistic the field already reports. Where the conditional covariance is larger in cells with lower response rates, the ratio exceeds $1 - R_{Y \cdot X}^2$, so the linear benchmark is conservative in the direction that matters.

Two features deserve emphasis. First, the residual share does not depend on β . The strength of selection determines how large the bias is, not what proportion of it weighting removes. A field cannot reassure itself that selection is probably mild and therefore adequately handled: mildness affects the magnitude, not the repair rate.

Second, the residual share does not depend on how well X predicts response. Elaborating the propensity model with better predictors of participation improves the treatment of the $\gamma'X$ component, which weighting already handled, and does nothing for the βY component, which is the one at issue. This runs against a natural intuition that better nonresponse modelling is the remedy. It is the remedy only for selection on observables. Figure 2 sets out the two routes and their different fates.

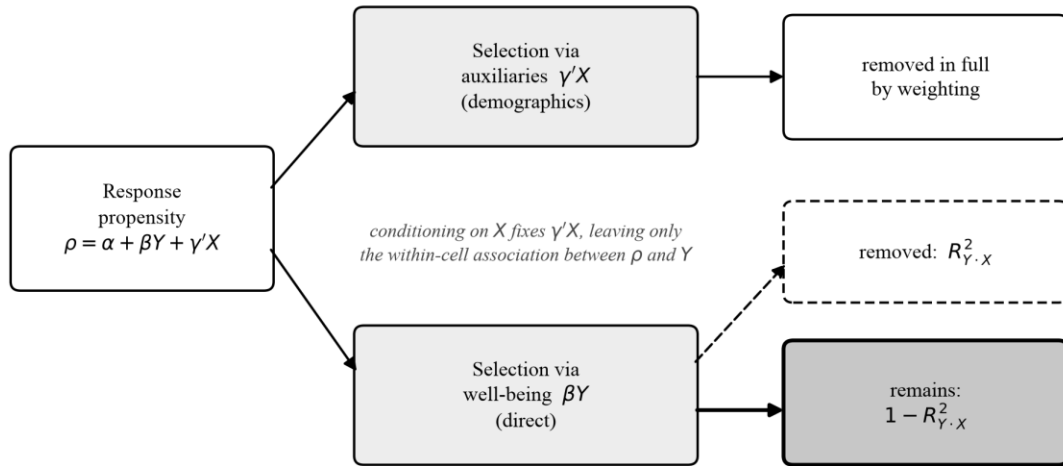


Fig. 2 Decomposition of selection into the component operating through the weighting auxiliaries and the component operating directly through well-being, with the share of each that weighting removes

Proposition 3. The quantity $R_{Y.X}^2$ that governs the repair rate is the coefficient of determination from a regression of well-being on socio-demographic characteristics. It therefore serves two distinct purposes simultaneously: as a substantive statistic describing how weakly circumstances determine well-being, and as the fraction of outcome-dependent nonresponse bias that weighting on those circumstances removes.

The third proposition is where the argument bites. The modest explanatory power of demographic and socio-economic variables is among the more robust and more frequently repeated observations in this literature. Kaiser and Oswald (2022) state it directly, reporting that a single feelings integer carries more predictive power than a collection of socio-economic influences. Reviews of the economic determinants of well-being reach compatible conclusions about how much remains unexplained (Dolan et al. 2008). The best-known version of the claim assigns roughly a tenth of the variance in happiness to life circumstances (Lyubomirsky et al. 2005). That decomposition has been criticised on identification grounds (Brown and Rohrer 2020), and its authors have since accepted much of the criticism while defending the model’s central premise (Sheldon and Lyubomirsky 2021). The criticism is accepted here in full. Proposition 3 does not require any causal interpretation of the decomposition. It requires only the descriptive, predictive quantity: the share of variance in measured well-being that the weighting variables account for.

That distinction matters for the engagement with Brown and Rohrer (2020). Their critique targets reading a variance decomposition as a statement about what can be changed. The present use is different in kind. No claim is made that circumstances cause a given share of well-being, only that a regression of well-being on the weighting variables has a low R^2 . That predictive fact sets the repair rate whatever its causal interpretation. Their argument leaves this

use intact, and their central caution reinforces it. A quantity poorly suited to bearing causal weight can still be the right quantity for a question about prediction.

A further point strengthens the case. Weighting auxiliaries are typically a narrow set, restricted to variables for which reliable population benchmarks exist, and usually a strict subset of the circumstance variables considered in the determinants literature. The relevant $R^2_{Y.X}$ is therefore at most, and generally below, the figures cited there.

5 Evidence on the governing parameter

Proposition 1 makes the residual share depend on one estimable quantity. This section estimates it, and then examines a case in which weighting demonstrably did repair most of a selection problem, in order to test whether the parameter explains the difference.

5.1 An estimate from a large national survey

The General Social Survey provides a long series of a standard happiness item alongside the demographic variables that post-stratification uses. Pooling all rounds from 1972 to 2024 gives 69,719 respondents with valid data. Reported happiness was coded so that higher values indicate greater happiness, and regressed by ordinary least squares on the variables that conventional weighting schemes employ: age and its square, sex, race, educational attainment, and region. Table 3 reports the resulting coefficient of determination and the repair rate that Proposition 1 attaches to it.

Table 3 Variance in reported happiness explained by weighting variables, and the implied repair rate, General Social Survey 1972 to 2024

Specification	$R^2_{Y.X}$	Bias removed	Bias remaining	n
Age and sex only	0.001	0.1%	99.9%	69,719
Age, sex, race, education	0.016	1.6%	98.4%	69,719

Specification	$R^2_{Y.X}$	Bias removed	Bias remaining	n
Core set: age, sex, race, education, region	0.017	1.7%	98.3%	69,719
Core set, survey-weighted	0.025	2.5%	97.5%	69,719
Core set plus year fixed effects	0.037	3.7%	96.3%	69,719
Core set, 2010 to 2024 only	0.013	1.3%	98.7%	21,700
Core set, binary outcome (very happy)	0.010	1.0%	99.0%	69,719
Core set plus marital status	0.072	7.2%	92.8%	69,719

Note. Estimates from the General Social Survey cumulative file. Marital status is included for comparison only; it is not normally a post-stratification variable, since population benchmarks are less readily available. Retrieval and estimation code accompanies this article.

Three features matter. The estimate is small and stable: across every specification using variables a weighting scheme would actually contain, the explained share lies between one and four per cent. It is not an artefact of pooling five decades, since restricting to recent rounds lowers rather than raises it. And it is declining, from 0.031 for 1972 to 1989 to 0.014 for 2020 to 2024, so the repair rate has fallen as the demographic patterning of happiness has weakened.

One correction works against this conclusion and should be applied. Reported well-being contains measurement error, which attenuates the estimated R^2 . Test-retest reliability of a standard life satisfaction item over two weeks has been estimated at 0.59 (Krueger and Schkade 2008). Dividing the core estimate by that figure gives an implied value of approximately 0.029 for the latent quantity, so the repair rate rises from 1.7 to roughly 2.9 per cent. The order of magnitude is unchanged.

Two limitations should be stated. The happiness item is a three-category ordinal measure, so the coefficient of determination describes linear predictable variance in the reported scale rather than in any underlying continuum; the binary specification in Table 3 is reported for that

reason. And these are data from one country, so the estimate anchors the parameter rather than establishing it universally. The code is provided so the same quantity can be computed wherever a survey carries both a well-being item and its weighting variables.

5.2 A contrasting case, and what distinguishes it

If the parameter governs the repair rate, then weighting should succeed where the auxiliaries are informative about the outcome. A study designed to test representativeness directly provides that case. Momen et al. (2022) linked a Danish regional health survey of 32,417 participants to national registers covering the 1,063,082 people eligible to take part, so that the outcome was observed for respondents and non-respondents alike. In the register population, 10.4 per cent had a recorded diagnosis of any mental disorder. Among unweighted survey participants the figure was 8.2 per cent, and after non-response weights were applied it was 9.5 per cent. Weighting closed most of the gap, and the authors conclude that the selection bias was minimised once weights were applied.

That result is not a counterexample. The weights were constructed by the national statistical agency through model-based calibration using register information available for participants and non-participants alike, comprising sex, age, municipality, social background, and healthcare utilisation in the form of visits to general practitioners and somatic hospital admissions. Healthcare utilisation is strongly related to psychiatric morbidity. The auxiliary set therefore had substantial explanatory power for the outcome, which is precisely the condition under which Proposition 1 predicts that weighting will recover most of the bias.

The two cases bracket the parameter rather than contradicting one another. Where the weighting variables are the demographic set ordinarily available, the explained share is on the order of two per cent and almost none of the bias is removed. Where the weighting variables

include administrative measures that predict the outcome, most of it is. The difference between the two situations is not the quality of the statistical adjustment but the informativeness of what the adjustment is allowed to use, which is the quantity this paper identifies. It also indicates that the remedy is attainable rather than hypothetical: the Danish case is what Section 8 recommends, already implemented.

One further detail of that study bears on Section 10. Momen et al. also report that estimated associations between mental disorders and general medical conditions were similar in weighted samples, unweighted samples, and the full register population, even though the prevalence itself was displaced. Conditional associations proved more robust than the level, which is the distinction drawn below in response to the strongest objection to this paper.

6 Calibration and sensitivity

Proposition 1 converts into a table once a value for $R_{Y.X}^2$ is supplied. Table 4 reports the residual share across the full range, from values typical of narrow demographic schemes to optimistic ones assuming rich auxiliaries, so the conclusion does not rest on the estimate of Section 5 alone.

Table 4 Share of outcome-dependent nonresponse bias remaining after weighting

Variance in well-being explained by weighting variables ($R_{Y.X}^2$)	Share of bias removed	Share remaining	Residual gap if raw respondent–nonrespondent difference is 0.30 scale points
0.02	2%	98%	0.294
0.05	5%	95%	0.285
0.10	10%	90%	0.270
0.15	15%	85%	0.255
0.20	20%	80%	0.240
0.30	30%	70%	0.210

Note. The final column applies the residual share to an illustrative raw difference of 0.30 points on a 0–10 scale. It is illustrative, not an estimate for any survey.

Across the entire range, weighting leaves between seventy and ninety-eight per cent of the bias in place. The estimate in Table 3 sits at the bottom of that range, so the row representing operational practice is the first, not the fourth.

Because the result is a derivation, its reliability rests on its assumptions rather than on sampling variability, and each can be examined separately. Table 5 states them with the direction in which the conclusion moves if each fails. The assumptions are conservative: those that can fail either way have no first-order effect, and those that bind push the true residual above the figures in Table 4 rather than below.

Table 5 Assumptions of Proposition 1 and the direction of error if each fails

Assumption	If it fails	Effect on the residual share	Discussed in
Participation depends on well-being linearly over the relevant range	Dependence is non-linear	Exact fraction no longer holds; residual still governed by the within-cell association weighting cannot reach	Section 10
$\text{Var}(Y X)$ does not vary systematically with X	Residual variance is larger in low-response cells	Residual share rises above Table 4	Section 10
Response rates do not vary sharply across cells of X	Rates vary sharply	Second-order; weighting still cannot reach the within-cell term	Section 10
Selection on well-being is present at all	It is absent	No bias to repair; the result is silent rather than wrong	Section 10
Weighting auxiliaries are the usual narrow demographic set	Richer auxiliaries are used	Residual share falls, but only to the extent the additions predict well-being, not response	Section 4

Note. The residual share is invariant to the strength of selection and to the auxiliaries' power to predict participation, so neither appears here.

These magnitudes are not negligible relative to what the field interprets. Differences in mean life satisfaction between neighbouring countries, and shifts in national means across years, are frequently of the order of one or two tenths of a scale point. A residual of comparable size

does not make such comparisons meaningless, but it is not small relative to the signal, which is the condition under which a bias can be set aside.

Figure 3 shows the relationship. The residual share declines linearly in $R_{Y.X}^2$, and the region occupied by realistic weighting schemes sits at the left of the plot, where the repair is smallest.

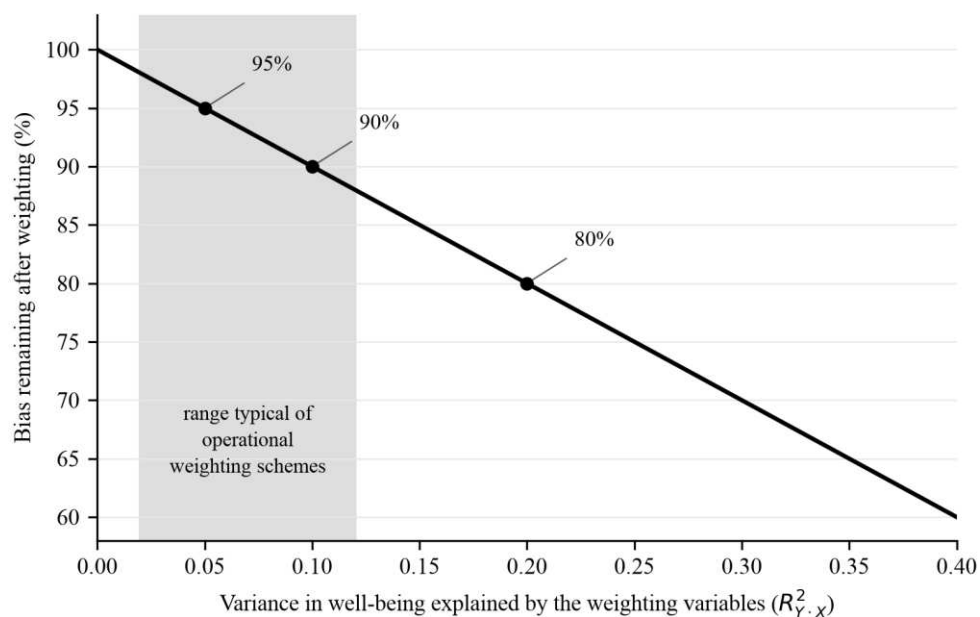


Fig. 3 Share of outcome-dependent nonresponse bias remaining after weighting, as a function of the variance in well-being explained by the weighting variables. The shaded band marks the range of values typical of operational weighting schemes

7 What is new here, and what is not

A conceptual argument should state plainly what it adds, and a reader is entitled to ask whether the preceding sections restate a known result in new vocabulary. The question is taken directly.

Table 6 separates what is drawn from existing literature from what is established here.

The distinction is set out so that a reader can check it rather than accept it.

Table 6 Position of the argument relative to the results it builds on

Established result	Source	What it leaves open	Established here
Nonresponse bias equals the covariance of response propensity with the survey variable, divided by mean propensity	Groves (2006); Brick (2013)	The size of that covariance in any given field	Its decomposition into a component weighting removes and a component weighting cannot reach (Fig. 2)
Weighting reduces bias only where auxiliaries relate to both response and outcome	Kreuter et al. (2010); Haziza and Lesage (2016)	By how much, when the relation to the outcome is weak	The exact residual share $1 - R_{Y.X}^2$ under a stated mechanism (Proposition 1)
Missingness dependent on the unobserved outcome is not ignorable	Rubin (1976)	Which measurable parameter governs severity in a particular literature	Identification of that parameter, its generalisation beyond the linear case (Proposition 2), and a direct estimate of it (Proposition 3, Table 3)
Socio-demographic variables explain a small share of variance in well-being	Kaiser and Oswald (2022); Dolan et al. (2008)	Consequences beyond the substantive reading	That the same statistic is the repair rate of the field's standard correction, estimated here at 0.017
Non-response weights can restore representativeness where registers supply the auxiliaries	Momen et al. (2022)	Why weighting succeeds there and not elsewhere	That the difference is the explanatory power of the auxiliaries for the outcome, which the two cases bracket

Three claims follow, and each is falsifiable. The first is that the residual share is exactly $1 - R_{Y.X}^2$ under the stated mechanism; this is a derivation, checkable line by line, and refuted by showing an assumption fails in a way that changes the result. The second is that the governing parameter is one this field already estimates for unrelated reasons, and that the two roles have not been connected; this is refuted by producing prior work that connects them. The third concerns the direction of the implication, which is the reverse of the intuitive one. The repair rate rises with the demographic predictability of the outcome. Fields whose outcomes are strongly patterned by demographics can weight away outcome-dependent selection; fields whose

outcomes are not, cannot. Well-being research sits at the unfavourable end, and it sits there because of the finding it reports most often. This is refuted by showing that $R^2_{Y.X}$ for standard weighting variables is in fact high.

What changes for a reader is specific. A weighted well-being mean should not be read as though weighting had addressed selection on the outcome. The number that determines how much to discount it is identified and estimable. The measurement that would improve matters is auxiliaries that predict well-being, not auxiliaries that predict response.

8 Consequences for standard inferences

Cross-national comparison. International rankings draw largely on a single multi-country survey programme (Deaton 2008), and measurement invariance is treated as the condition for valid comparison. Röck et al. (2026) apply multi-group confirmatory factor analysis to European data with a bifactor model, report full scalar invariance in thirty of thirty-two countries, and conclude that this supports accurate cross-country comparison. The analysis is careful and invariance is a genuine precondition. The conclusion nonetheless requires a second condition that invariance testing cannot supply, because invariance evaluates whether the measurement model operates identically among those who answered. It is silent on whether the samples stand in the same relation to their populations. Response rates differ substantially across national surveys, and Proposition 1 implies that the associated residual biases differ correspondingly. Two countries can satisfy scalar invariance exactly and still yield means that are not comparable, because the instrument behaves identically while the samples were assembled differently. The amendment is narrow: invariance and comparable coverage are distinct conditions, and a claim of comparability requires both.

National level estimates. M. Bond et al. (2026) report the first Personal Wellbeing Index estimates for the United Kingdom on a weighted probability sample, obtaining a mean below other Western industrialised countries and outside the range predicted by homeostatic theory, which they attribute largely to one low-scoring domain. The domain analysis is persuasive on its own terms. An anomalously low national mean is also, however, the signature that outcome-dependent nonresponse would produce, and the weighting applied is demographic. Table 4 indicates that a residual of similar order to the anomaly cannot be excluded by that weighting. This does not contradict their interpretation and no alternative estimate is offered here. It identifies a competing explanation the design cannot currently rule out.

Trends and group differences. Rohrer et al. (2026) examine why the gender gap in adolescent life satisfaction widened in repeated school surveys, proceeding by testing alternatives rather than settling on a narrative. They rule out a scaling artefact, trace part of the change to compositional shifts in the student population, and identify an effect of the move to tablet administration. Their procedure is the one this paper recommends. The alternatives they consider are nonetheless confined to the scale and to who is in the population, and do not include differential participation among those sampled. School surveys have a specific coverage property here: adolescents absent through exclusion, truancy, or illness are disproportionately those with lower well-being, and their absence can vary across rounds. Adding participation would extend their framework rather than displace it.

Age profiles and adaptation. Estimates of how well-being varies across the life course rest on samples in which the oldest respondents have survived both mortality and attrition, and in which institutional populations sit outside the frame. Attrition among older panel members has been examined for exactly this reason (Herrera et al. 2021). The induced bias runs upward at

older ages. That is where the shape of the age profile is most contested, where the evidence has long been read as showing a rise in later life (Blanchflower 2021; Blanchflower and Oswald 2004), and where the choice of controls has been shown to drive the estimated shape (Bartram 2021). Studies of older adults that are also cross-national compound both exposures at once, since the survivorship structure and the response climate differ by country (Kim and Lee 2026). The same logic applies to adaptation. Estimates of recovery from adverse events are identified from people observed both before and after the event (Lucas et al. 2003, 2004; Luhmann et al. 2012), and those who do not recover are, by the mechanisms in Table 2, disproportionately likely to leave. Adaptation estimated on retained respondents will exceed adaptation in the population, and the gap grows with the number of post-event waves a design requires.

Group comparisons in synthesised evidence. Elsas et al. (2026) review findings on single parenthood and happiness across fifty-four publications, concluding that single parents report lower happiness than other parents in nearly all countries covered. Synthesis inherits the selection properties of what is synthesised. Where a group faces both lower well-being and higher barriers to participation, as is plausible for single parents, the pooled contrast understates the difference rather than overstating it, so the direction works in favour of their conclusion. A findings archive aggregates estimates without aggregating away their common selection problem.

Well-being inequality. Dispersion is more exposed than the mean. Selection that removes observations from the lower tail compresses the observed distribution, so estimated well-being inequality is biased downward and trends in it are biased toward improvement. This matters for a literature that has documented long-run movements in the dispersion of happiness and read them substantively (Clark et al. 2014; Stevenson and Wolfers 2008), and that has

developed dedicated statistics for inequality in an ordinal happiness variable (Kalmijn and Veenhoven 2005). Weighting to demographic margins restores the shape of the demographic distribution, not the shape of the well-being distribution, which is the object of interest.

Table 7 collects these cases. In most of them the direction of the error is predictable, and it runs toward the conclusion the field has tended to draw. The single-parenthood case shows the exception, and it matters: the argument is not that every estimate is inflated, but that the direction is determinate once the selection mechanism is specified, and is therefore something to be worked out rather than assumed away.

Table 7 Direction of the induced error in six standard inferences

Inference	Selection mechanism that bites	Direction of error	Consequence for the usual conclusion
Cross-national comparison	Response rates differ across national surveys, so residual biases differ	Varies by country	Country ranks not identified by invariance alone
National level estimates	Outcome-dependent nonresponse within one frame	Upward on the mean	An anomalous national mean may be partly artefactual
Trends and group gaps	Differential participation changing across rounds	Either, depending on the shift	A change in composition of respondents mimics a change in well-being
Age profiles	Mortality and attrition select survivors; institutions outside frame	Upward at older ages	Overstates the late-life rise
Adaptation to adverse events	Non-recoverers leave the panel	Upward on recovery	Overstates adaptation, worsening with each post-event wave
Well-being inequality	Lower tail is thinned by all three selection channels	Downward on dispersion	Understates inequality and overstates its improvement

9 What follows for practice

The argument implies that outcome-dependent nonresponse cannot be adjusted away with the auxiliaries normally available, and therefore that it should be reported and bounded rather than assumed absent. Health-related quality-of-life research reached this position earlier and has developed the relevant machinery. Table 8 sets out five transferable procedures.

Table 8 Procedures for outcome-dependent missingness available from adjacent fields

Procedure	What it delivers	Established use
Sensitivity analysis under departures from ignorability	A range of estimates across assumed values of β , rather than a single number resting on an untested assumption	Recommended for clinical trials by the National Research Council (2010) and Little et al. (2012); methods in Scharfstein et al. (1999), Molenberghs et al. (2003) and Carpenter and Kenward (2012)
Pattern-mixture and worst-case bounding	Explicit bounds obtained by assigning unfavourable values to the missing, establishing whether a conclusion survives	Standard in quality-of-life endpoint analysis (Coens et al. 2020) and in economic evaluation alongside trials (Faria et al. 2014); the general logic is partial identification (Lee 2009; Manski 1989)
Representativeness indicators	A summary of how far respondents depart from the sampled population on measured characteristics, reportable alongside the response rate	R-indicators developed for official statistics (Schouten et al. 2012)
Auxiliary measurement on nonrespondents	Brief proxy or follow-up measurement that makes the selection partly observable, raising the effective $R^2_{Y,X}$	Proxy correlates used in nonresponse adjustment (Kreuter et al. 2010); register-based auxiliaries in the Danish case (Momen et al. 2022)
Frame extension to excluded populations	Direct measurement in populations absent from household frames	Well-being surveys of homeless populations (Biswas-Diener and Diener 2006; Panadero et al. 2015)

Where missing values are to be filled rather than bounded, multiple imputation is the established apparatus (Rubin 1987; Schafer and Graham 2002; van Buuren and Groothuis-Oudshoorn 2011). Inverse probability weighting is its counterpart for unit-level absence (Seaman and White 2013). Both rest on the conditional-independence assumption examined here, so neither substitutes for the diagnostics below.

A further option operates at the design stage: responsive and adaptive designs target fieldwork effort at underrepresented cases rather than maximising the response rate as such (Brick and Tourangeau 2017). Their value here depends on whether the targeting criteria relate to well-being rather than only to demographics.

Three reporting practices follow. Studies presenting population estimates should report the response or retention rate and the auxiliaries used in weighting, since Proposition 1 shows the

second determines what the weighting achieved. The rate alone is a weak signal. It does not identify the covariance that produces the bias (Davern 2013; Groves and Peytcheva 2008). Studies whose conclusion turns on a difference of the order of a tenth of a scale point should present a bound of the kind in Table 4. Studies of populations with elevated attrition risk should report attrition by baseline well-being, which any panel can already compute and which is the most informative single diagnostic available.

The fourth procedure in Table 8 deserves emphasis as the only one that changes the governing parameter rather than bounding its consequences. Raising $R_{Y.X}^2$ requires auxiliaries that predict well-being, not auxiliaries that predict response. A short well-being item at recruitment, or an administrative correlate such as a health indicator, would do more than any refinement of demographic post-stratification.

10 Objections and limitations

The derivation rests on assumptions that may not hold. Each is listed in Table 5 with the direction in which the conclusion moves if it fails. Two points are worth restating. Under non-linear dependence of participation on well-being the exact fraction no longer holds, although the residual continues to be governed by the within-cell association that weighting cannot reach; the linear case is a transparent benchmark rather than a description of any particular survey. And if residual variance in well-being is larger in cells with lower response rates, the true residual exceeds the figures in Table 4 rather than falling below them.

Selection may be weaker than assumed. This is possible and is not resolved here. It affects the magnitude of the bias and not the share that weighting removes, which is the claim being made. The appropriate response is measurement rather than assumption, which is the point of Table 8.

A selection model would solve this. The standard econometric response to selection on unobservables is to model participation jointly with the outcome (Heckman 1979). That approach identifies the outcome equation only under a restriction: some variable must shift the probability of participating without entering well-being directly. Well-being surveys rarely supply one. Interviewer identity, contact attempts and fieldwork effort are the natural candidates, and they are plausible instruments precisely to the extent that they do not affect how satisfied a person is with their life, which is an assumption the paradata cannot test. Where such an instrument does exist, the correction is available and should be used, and the result derived here then describes what happens in its absence. That absence is the ordinary case in this literature. A related diagnostic asks how far results move as observed controls are added, and uses that movement to bound the influence of what remains unobserved (Oster 2019). It is also why the constructive proposals in Section 9 emphasise bounding the quantity rather than point-identifying it, in the tradition of partial identification (Lee 2009; Manski 1989): a credible interval is preferable to a point estimate resting on an exclusion restriction nobody can defend.

Weighting might not be the right benchmark. Whether to weight at all is itself contested, and the case for weighting depends on the estimand and on what the weights are doing (Bollen et al. 2016; Solon et al. 2015). The argument here does not recommend weighting or oppose it. It takes weighting as the field’s prevailing practice for population estimates and establishes what that practice accomplishes against the specific threat of outcome-dependent participation. An analyst who declines to weight faces the same underlying problem without even the partial repair.

A direct test found no bias. This objection has the most force and deserves a direct answer. Herrera et al. (2021) examined selective attrition in a panel of older adults and report

that although several attrition predictors were associated with subjective well-being, attrition did not produce biased estimates in multivariate models of life satisfaction or depressive symptoms. Their finding is evidence against the practical importance of the problem in that setting, and it is not set aside here.

Two things limit how far it generalises. First, their test concerns estimated associations in multivariate models, whereas the estimand in this paper is a population mean. Selection on an outcome shifts a mean directly, while its effect on a conditional association depends on whether selection also relates to the covariates, and the two need not move together. A design can be robust for regression coefficients while remaining exposed for a national average, and it is averages that are used for ranking and appraisal. Second, a single study in one country and one age group establishes the magnitude in that setting rather than in general.

The broader point is that this is close to the only direct test available. A literature able to cite a dozen such studies, some finding bias and some not, could calibrate Table 4 rather than reason about its range, and the recommendations in Section 9 are aimed at producing exactly that.

The audit measures abstracts, not articles. Table 1 records salience, so the figures understate total engagement; they are informative about how consequential authors judge the issue to be.

Some surveys already do better. Several national statistical producers document coverage and response extensively, and some panels publish attrition analyses. The claim is not that no one attends to this, but that attention is concentrated among data producers and largely absent from the research literature analysing their data, which is what Table 1 measures.

The argument applies to any survey outcome. It does, in the general form given by Rubin (1976). Its severity, however, is governed by $R_{Y.X}^2$, and that parameter is unusually low for subjective well-being. For outcomes that are strongly demographically patterned, such as income or educational attainment, standard weighting performs substantially better. The problem is acute here because of a property of this outcome variable.

11 Conclusion

Subjective well-being research has audited its scale with considerable care and its sampling frame hardly at all. This paper has argued that the imbalance is costly, and it has quantified the cost. Where participation depends on well-being itself, weighting on demographic auxiliaries removes a share of the bias equal to the variance in well-being those auxiliaries explain, and leaves the rest. Estimated in the largest and longest-running survey of its kind, that share is under two per cent, so weighting removes about one fiftieth of the problem and retains the remainder.

The parameter that governs this is not new to the field. It is the familiar observation that circumstances predict feelings weakly, a finding usually reported as a statement about the limits of material determinism. The same number also measures how little of a selection problem the field's standard correction can repair. Both readings are correct, and they have not previously been put side by side.

The implication is not that well-being estimates should be discarded. It is that a weighted mean should not be treated as though the weighting had addressed selection on the outcome, because it has not. The remedies are in routine use one field away and consist mainly of bounding a quantity rather than assuming it away. The first step is the diagnostic any panel can already produce: how the people who left differed in well-being from those who stayed.

Statements and Declarations

Competing Interests: The author has no competing interests to declare that are relevant to the content of this article.

Funding: No funding was received for conducting this study.

Ethics approval: This article is a methodological and conceptual contribution. It does not report any study involving human participants or animals performed by the author. It uses only publicly available, de-identified secondary data released for research use, and article metadata. No ethics approval was required.

References

- Abraham, K.G., Maitland, A., Bianchi, S.M.: Nonresponse in the American Time Use Survey. *Public Opinion Quarterly* 70(5), 676–703 (2006). <https://doi.org/10.1093/poq/nfl037>
- Adler, A., Seligman, M.E.P.: Using wellbeing for public policy: Theory, measurement, and recommendations. *Intl. J. Wellbeing* 6(1), 1–35 (2016). <https://doi.org/10.5502/ijw.v6i1.429>
- Angelini, V., Cavapozzi, D., Corazzini, L., Paccagnella, O.: Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases. *Oxf Bull Econ Stat* 76(5), 643–666 (2014). <https://doi.org/10.1111/obes.12039>
- Bartram, D.: Age and Life Satisfaction: Getting Control Variables under Control. *Sociology* 55(2), 421–437 (2021). <https://doi.org/10.1177/0038038520926871>
- Biswas-Diener, R., Diener, E.: The Subjective Well-Being of the Homeless, and Lessons for Happiness. *Soc Indic Res* 76(2), 185–205 (2006). <https://doi.org/10.1007/s11205-005-8671-9>

- Blanchflower, D.G.: Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries. *J Popul Econ* 34(2), 575–624 (2021). <https://doi.org/10.1007/s00148-020-00797-z>
- Blanchflower, D.G., Oswald, A.J.: Well-being over time in Britain and the USA. *Journal of Public Economics* 88(7-8), 1359–1386 (2004). [https://doi.org/10.1016/S0047-2727\(02\)00168-8](https://doi.org/10.1016/S0047-2727(02)00168-8)
- Bollen, K.A., Biemer, P.P., Karr, A.F., Tueller, S., Berzofsky, M.E.: Are Survey Weights Needed? A Review of Diagnostic Tests in Regression Analysis. *Annu. Rev. Stat. Appl.* 3(1), 375–392 (2016). <https://doi.org/10.1146/annurev-statistics-011516-012958>
- Bond, M., Frings, D., Stenning, A., Morton, J.: National Estimates of Subjective Wellbeing in the UK Using the Personal Wellbeing Index. *J Happiness Stud* 27(2) (2026). <https://doi.org/10.1007/s10902-026-01017-3>
- Bond, T.N., Lang, K.: The Sad Truth about Happiness Scales. *Journal of Political Economy* 127(4), 1629–1640 (2019). <https://doi.org/10.1086/701679>
- Brick, J.M.: Unit Nonresponse and Weighting Adjustments: A Critical Review. *Journal of Official Statistics* 29(3), 329–353 (2013). <https://doi.org/10.2478/jos-2013-0026>
- Brick, J.M., Tourangeau, R.: Responsive Survey Designs for Reducing Nonresponse Bias. *Journal of Official Statistics* 33(3), 735–752 (2017). <https://doi.org/10.1515/jos-2017-0034>
- Brown, N.J.L., Rohrer, J.M.: Easy as (Happiness) Pie? A Critical Evaluation of a Popular Model of the Determinants of Well-Being. *J Happiness Stud* 21(4), 1285–1301 (2020). <https://doi.org/10.1007/s10902-019-00128-4>

- Carpenter, J.R., Kenward, M.G.: Multiple Imputation and its Application. Wiley, Chichester (2012). <https://doi.org/10.1002/9781119942283>
- Chadi, A.: Dissatisfied with life or with being interviewed? Happiness and the motivation to participate in a survey. *Soc Choice Welf* 53(3), 519–553 (2019). <https://doi.org/10.1007/s00355-019-01195-5>
- Cheung, K.L., ten Klooster, P.M., Smit, C., de Vries, H., Pieterse, M.E.: The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health* 17(1) (2017). <https://doi.org/10.1186/s12889-017-4189-8>
- Clark, A.E.: Four Decades of the Economics of Happiness: Where Next? *Review of Income and Wealth* 64(2), 245–269 (2018). <https://doi.org/10.1111/roiw.12369>
- Clark, A.E., Flèche, S., Senik, C.: The great happiness moderation. In: Clark, A.E., Senik, C. (eds.) *Happiness and Economic Growth: Lessons from Developing Countries*, pp. 32–139. Oxford University Press, Oxford (2014). <https://doi.org/10.1093/acprof:oso/9780198723653.003.0003>
- Coens, C., Pe, M., Dueck, A.C., Sloan, J., Basch, E., Calvert, M., Campbell, A., Cleeland, C., Cocks, K., Collette, L., Devlin, N., Dorme, L., Flechtner, H.H., Gotay, C., Griebisch, I., Groenvold, M., King, M., Kluetz, P.G., Koller, M., Malone, D.C., Martinelli, F., Mitchell, S.A., Musoro, J.Z., O'Connor, D., Oliver, K., Piau-Louis, E., Piccart, M., Quinten, C., Reijneveld, J.C., Schürmann, C., Smith, A.W., Soltys, K.M., Taphoorn, M.J.B., Velikova, G., Bottomley, A.: International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials:

- recommendations of the SISAQOL Consortium. *The Lancet Oncology* 21(2), e83–e96 (2020). [https://doi.org/10.1016/S1470-2045\(19\)30790-9](https://doi.org/10.1016/S1470-2045(19)30790-9)
- Davern, M.: Nonresponse Rates are a Problematic Indicator of Nonresponse Bias in Survey Research. *Health Services Research* 48(3), 905–912 (2013). <https://doi.org/10.1111/1475-6773.12070>
- de Winter, A.F., Oldehinkel, A.J., Veenstra, R., Brunnekreef, J.A., Verhulst, F.C., Ormel, J.: Evaluation of non-response bias in mental health determinants and outcomes in a large sample of pre-adolescents. *Eur J Epidemiol* 20(2), 173–181 (2005). <https://doi.org/10.1007/s10654-004-4948-6>
- Deaton, A.: Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll. *Journal of Economic Perspectives* 22(2), 53–72 (2008). <https://doi.org/10.1257/jep.22.2.53>
- Deaton, A., Stone, A.A.: Evaluative and hedonic wellbeing among those with and without children at home. *Proc. Natl. Acad. Sci. U.S.A.* 111(4), 1328–1333 (2014). <https://doi.org/10.1073/pnas.1311600111>
- Diener, E., Oishi, S., Lucas, R.E.: National accounts of subjective well-being. *American Psychologist* 70(3), 234–242 (2015). <https://doi.org/10.1037/a0038899>
- Diener, E., Oishi, S., Tay, L.: Advances in subjective well-being research. *Nat Hum Behav* 2(4), 253–260 (2018). <https://doi.org/10.1038/s41562-018-0307-6>
- Diener, E., Seligman, M.E.P.: Beyond Money. *Psychol Sci Public Interest* 5(1), 1–31 (2004). <https://doi.org/10.1111/j.0963-7214.2004.00501001.x>

- Diener, E., Suh, E.M., Lucas, R.E., Smith, H.L.: Subjective well-being: Three decades of progress. *Psychological Bulletin* 125(2), 276–302 (1999). <https://doi.org/10.1037/0033-2909.125.2.276>
- Dolan, P., Peasgood, T., White, M.: Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology* 29(1), 94–122 (2008). <https://doi.org/10.1016/j.joep.2007.09.001>
- Elsas, S., Möhrle, T., Veenhoven, R.: Happiness and Single Parenthood: A Literature Review Using an Online Findings Archive. *J Happiness Stud* 27(5) (2026). <https://doi.org/10.1007/s10902-026-01030-6>
- Faria, R., Gomes, M., Epstein, D., White, I.R.: A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. *PharmacoEconomics* 32(12), 1157–1170 (2014). <https://doi.org/10.1007/s40273-014-0193-3>
- Ferrer-i-Carbonell, A., Frijters, P.: How Important is Methodology for the Estimates of the Determinants of Happiness? *The Economic Journal* 114(497), 641–659 (2004). <https://doi.org/10.1111/j.1468-0297.2004.00235.x>
- Frijters, P., Clark, A.E., Krekel, C., Layard, R.: A happy choice: wellbeing as the goal of government. *Behav. Public Policy* 4(2), 126–165 (2020). <https://doi.org/10.1017/bpp.2019.39>
- Frijters, P., Krekel, C.: *A Handbook for Wellbeing Policy-Making*. Oxford University Press, Oxford (2021). <https://doi.org/10.1093/oso/9780192896803.001.0001>

- Goodrich, B., Fenton, M., Penn, J., Bovay, J., Mountain, T.: Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. *Applied Eco Perspectives Pol* 45(2), 762–784 (2023). <https://doi.org/10.1002/aepp.13353>
- Groves, R.M.: Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70(5), 646–675 (2006). <https://doi.org/10.1093/poq/nfl033>
- Groves, R.M., Couper, M.P.: *Nonresponse in Household Interview Surveys*. Wiley, New York (1998). <https://doi.org/10.1002/9781118490082>
- Groves, R.M., Couper, M.P., Presser, S., Singer, E., Tourangeau, R., Acosta, G.P., Nelson, L.: Experiments in Producing Nonresponse Bias. *Public Opinion Quarterly* 70(5), 720–736 (2006). <https://doi.org/10.1093/poq/nfl036>
- Groves, R.M., Peytcheva, E.: The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly* 72(2), 167–189 (2008). <https://doi.org/10.1093/poq/nfn011>
- Gustavson, K., von Soest, T., Karevold, E., Røysamb, E.: Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health* 12(1) (2012). <https://doi.org/10.1186/1471-2458-12-918>
- Haziza, D., Lesage, É.: A Discussion of Weighting Procedures for Unit Nonresponse. *Journal of Official Statistics* 32(1), 129–145 (2016). <https://doi.org/10.1515/jos-2016-0006>
- Heckman, J.J.: Sample selection bias as a specification error. *Econometrica* 47(1), 153–161 (1979). <https://doi.org/10.2307/1912352>

- Helliwell, J.F., Layard, R., Sachs, J.D., De Neve, J.E., Aknin, L.B., Wang, S.: Why single-item measures of wellbeing are best. *Nat Hum Behav* 10(3), 437–438 (2026).
<https://doi.org/10.1038/s41562-026-02401-y>
- Herrera, M.S., Devilat, D., Fernández, M.B., Elgueta, R.: Does the selective attrition of a panel survey of older people affect the multivariate estimations of subjective well-being? *Qual Life Res* 30(1), 41–54 (2021). <https://doi.org/10.1007/s11136-020-02612-4>
- Hirata, J., Groß, P.: Relative Effects on Life Satisfaction Revisited: Social Comparison is Only Half the Story. *J Happiness Stud* 26(3) (2025). <https://doi.org/10.1007/s10902-024-00848-2>
- Johnstone, M., Jetten, J., Dingle, G.A., Parsell, C., Walter, Z.C.: Discrimination and well-being amongst the homeless: the role of multiple group membership. *Front. Psychol.* 6 (2015).
<https://doi.org/10.3389/fpsyg.2015.00739>
- Kahneman, D., Deaton, A.: High income improves evaluation of life but not emotional well-being. *Proc. Natl. Acad. Sci. U.S.A.* 107(38), 16489–16493 (2010).
<https://doi.org/10.1073/pnas.1011492107>
- Kahneman, D., Krueger, A.B., Schkade, D., Schwarz, N., Stone, A.A.: Would You Be Happier If You Were Richer? A Focusing Illusion. *Science* 312(5782), 1908–1910 (2006).
<https://doi.org/10.1126/science.1129688>
- Kaiser, C., Oswald, A.J.: The scientific value of numerical measures of human feelings. *Proc. Natl. Acad. Sci. U.S.A.* 119(42) (2022). <https://doi.org/10.1073/pnas.2210412119>
- Kaiser, C., Vendrik, M.C.M.: How threatening are transformations of happiness scales to subjective wellbeing research? GSBE Research Memorandum 2020/032. Maastricht University (2020). <https://doi.org/10.26481/umagsb.2020032>

- Kalmijn, W., Veenhoven, R.: Measuring Inequality of Happiness in Nations: In Search for Proper Statistics. *J Happiness Stud* 6(4), 357–396 (2005). <https://doi.org/10.1007/s10902-005-8855-7>
- Keeter, S., Miller, C., Kohut, A., Groves, R.M., Presser, S.: Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly* 64(2), 125–148 (2000). <https://doi.org/10.1086/317759>
- Kim, H., Lee, B.: Exploring the Determinants of Subjective Well-Being in Older Adults: A Cross-National Analysis of Social, Political, Economic, and Demographic Factors. *J Happiness Stud* 27(2) (2026). <https://doi.org/10.1007/s10902-026-01010-w>
- King, G., Murray, C.J.L., Salomon, J.A., Tandon, A.: Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *Am Polit Sci Rev* 97(4), 567–583 (2003). <https://doi.org/10.1017/S0003055403000881>
- Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M., Raghunathan, T.E.: Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-Response: Examples from Multiple Surveys. *Journal of the Royal Statistical Society Series A: Statistics in Society* 173(2), 389–407 (2010). <https://doi.org/10.1111/j.1467-985X.2009.00621.x>
- Kreuter, F., Presser, S., Tourangeau, R.: Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly* 72(5), 847–865 (2008). <https://doi.org/10.1093/poq/nfn063>
- Krueger, A.B., Schkade, D.A.: The reliability of subjective well-being measures. *Journal of Public Economics* 92(8-9), 1833–1845 (2008). <https://doi.org/10.1016/j.jpubeco.2007.12.015>

- Lee, D.S.: Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies* 76(3), 1071–1102 (2009).
<https://doi.org/10.1111/j.1467-937X.2009.00536.x>
- Little, R.J., D’Agostino, R., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., Neaton, J.D., Rotnitzky, A., Scharfstein, D., Shih, W.J., Siegel, J.P., Stern, H.: The Prevention and Treatment of Missing Data in Clinical Trials. *N Engl J Med* 367(14), 1355–1360 (2012).
<https://doi.org/10.1056/NEJMsrl203730>
- Little, R.J.A., Rubin, D.B.: *Statistical Analysis with Missing Data*, 3rd edn. Wiley, Hoboken (2019). <https://doi.org/10.1002/9781119482260>
- Lucas, R.E., Clark, A.E., Georgellis, Y., Diener, E.: Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology* 84(3), 527–539 (2003). <https://doi.org/10.1037/0022-3514.84.3.527>
- Lucas, R.E., Clark, A.E., Georgellis, Y., Diener, E.: Unemployment Alters the Set Point for Life Satisfaction. *Psychol Sci* 15(1), 8–13 (2004). <https://doi.org/10.1111/j.0963-7214.2004.01501002.x>
- Lugtig, P.: Panel Attrition. *Sociological Methods & Research* 43(4), 699–723 (2014).
<https://doi.org/10.1177/0049124113520305>
- Luhmann, M., Hofmann, W., Eid, M., Lucas, R.E.: Subjective well-being and adaptation to life events: A meta-analysis. *Journal of Personality and Social Psychology* 102(3), 592–615 (2012). <https://doi.org/10.1037/a0025948>

- Luiten, A., Hox, J., de Leeuw, E.: Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys. *Journal of Official Statistics* 36(3), 469–487 (2020). <https://doi.org/10.2478/jos-2020-0025>
- Lyubomirsky, S., Sheldon, K.M., Schkade, D.: Pursuing Happiness: The Architecture of Sustainable Change. *Review of General Psychology* 9(2), 111–131 (2005). <https://doi.org/10.1037/1089-2680.9.2.111>
- Manski, C.F.: Anatomy of the selection problem. *J. Hum. Resour.* 24(3), 343–360 (1989). <https://doi.org/10.2307/145818>
- Meterko, M., Restuccia, J.D., Stolzmann, K., Mohr, D., Brennan, C., Glasgow, J., Kaboli, P.: Response Rates, Nonresponse Bias, and Data Quality. *Public Opinion Quarterly* 79(1), 130–144 (2015). <https://doi.org/10.1093/poq/nfu052>
- Meyer, B.D., Mok, W.K.C., Sullivan, J.X.: Household Surveys in Crisis. *Journal of Economic Perspectives* 29(4), 199–226 (2015). <https://doi.org/10.1257/jep.29.4.199>
- Molenberghs, G., Thijs, H., Kenward, M.G., Verbeke, G.: Sensitivity Analysis of Continuous Incomplete Longitudinal Outcomes. *Statistica Neerlandica* 57(1), 112–135 (2003). <https://doi.org/10.1111/1467-9574.00224>
- Momen, N.C., Lasgaard, M., Weye, N., Edwards, J., McGrath, J.J., Plana-Ripoll, O.: Representativeness of survey participants in relation to mental disorders: a linkage between national registers and a population-representative survey. *IJPDS* 7(4) (2022). <https://doi.org/10.23889/ijpds.v7i4.1759>
- Moum, T.: Is subjective well-being a predictor of nonresponse in broad population surveys? *Soc Indic Res* 32(1), 1–20 (1994). <https://doi.org/10.1007/BF01078463>

- National Research Council: The Prevention and Treatment of Missing Data in Clinical Trials. National Academies Press, Washington, DC (2010). <https://doi.org/10.17226/12955>
- National Research Council: Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience. National Academies Press, Washington, DC (2013). <https://doi.org/10.17226/18548>
- OECD: OECD Guidelines on Measuring Subjective Well-being. OECD Publishing, Paris (2013). <https://doi.org/10.1787/9789264191655-en>
- OECD: OECD Guidelines on Measuring Subjective Well-being (2025 Update). OECD Publishing, Paris (2025). <https://doi.org/10.1787/9203632a-en>
- Oster, E.: Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics* 37(2), 187–204 (2019). <https://doi.org/10.1080/07350015.2016.1227711>
- Panadero, S., Guillén, A.I., Vázquez, J.J.: Happiness on the street: Overall happiness among homeless people in Madrid (Spain). *American Journal of Orthopsychiatry* 85(4), 324–330 (2015). <https://doi.org/10.1037/ort0000080>
- Peytchev, A.: Consequences of Survey Nonresponse. *The ANNALS of the American Academy of Political and Social Science* 645(1), 88–111 (2012). <https://doi.org/10.1177/0002716212461748>
- Powdthavee, N.: Would You Like to Know What Makes People Happy? An Overview of the Datasets on Subjective Well-Being. *Australian Economic Review* 48(3), 314–320 (2015). <https://doi.org/10.1111/1467-8462.12105>
- Reisinger, J.: Subjective well-being and social desirability. *Journal of Public Economics* 214, 104745 (2022). <https://doi.org/10.1016/j.jpubeco.2022.104745>

- Röck, S., Kleinheinz, L., Walde, J.: Assessing Measurement Invariance of Subjective Well-Being Across European Countries. *J Happiness Stud* 27(2) (2026).
<https://doi.org/10.1007/s10902-026-01011-9>
- Rohrer, J.M., McElreath, R., Kachel, G.: Why Did the Gender Gap in Adolescent Life Satisfaction Grow? Evaluating Methodological and Demographic Explanations. *J Happiness Stud* 27(6) (2026). <https://doi.org/10.1007/s10902-026-01068-6>
- Rubin, D.B.: Inference and missing data. *Biometrika* 63(3), 581–592 (1976).
<https://doi.org/10.1093/biomet/63.3.581>
- Rubin, D.B.: Multiple Imputation for Nonresponse in Surveys. Wiley, New York (1987).
<https://doi.org/10.1002/9780470316696>
- Schafer, J.L., Graham, J.W.: Missing data: Our view of the state of the art. *Psychological Methods* 7(2), 147–177 (2002). <https://doi.org/10.1037/1082-989X.7.2.147>
- Scharfstein, D.O., Rotnitzky, A., Robins, J.M.: Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *Journal of the American Statistical Association* 94(448), 1096–1120 (1999). <https://doi.org/10.1080/01621459.1999.10473862>
- Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N., Skinner, C.: Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators. *Int Statistical Rev* 80(3), 382–399 (2012). <https://doi.org/10.1111/j.1751-5823.2012.00189.x>
- Seaman, S.R., White, I.R.: Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res* 22(3), 278–295 (2013).
<https://doi.org/10.1177/0962280210395740>

- Sheldon, K.M., Lyubomirsky, S.: Revisiting the Sustainable Happiness Model and Pie Chart: Can Happiness Be Successfully Pursued? *The Journal of Positive Psychology* 16(2), 145–154 (2021). <https://doi.org/10.1080/17439760.2019.1689421>
- Solon, G., Haider, S.J., Wooldridge, J.M.: What Are We Weighting For? *J. Human Resources* 50(2), 301–316 (2015). <https://doi.org/10.3368/jhr.50.2.301>
- Stevenson, B., Wolfers, J.: Happiness Inequality in the United States. *The Journal of Legal Studies* 37(S2), S33–S79 (2008). <https://doi.org/10.1086/592004>
- Stoop, I., Billiet, J., Koch, A., Fitzgerald, R.: Improving Survey Response: Lessons Learned from the European Social Survey. Wiley, Chichester (2010). <https://doi.org/10.1002/9780470688335>
- Torvik, F.A., Rognmo, K., Tambs, K.: Alcohol use and mental distress as predictors of non-response in a general population health survey: the HUNT study. *Soc Psychiatry Psychiatr Epidemiol* 47(5), 805–816 (2012). <https://doi.org/10.1007/s00127-011-0387-3>
- Tourangeau, R., Rips, L.J., Rasinski, K.: The Psychology of Survey Response. Cambridge University Press, Cambridge (2000). <https://doi.org/10.1017/CBO9780511819322>
- van Buuren, S., Groothuis-Oudshoorn, K.: mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* 45(3), 1–67 (2011). <https://doi.org/10.18637/jss.v045.i03>
- van der Kaap-Deeder, J., Audenaert, E., Vandavelde, S., Soenens, B., Van Mastrigt, S., Mabbe, E., Vansteenkiste, M.: Choosing when choices are limited: The role of perceived afforded choice and autonomy in prisoners' well-being. *Law and Human Behavior* 41(6), 567–578 (2017). <https://doi.org/10.1037/lhb0000259>
- Veenhoven, R.: Cross-national differences in happiness: Cultural measurement bias or effect of culture? *Intl. J. Wellbeing*, 333–353 (2012). <https://doi.org/10.5502/ijw.v2.i4.4>

- Volken, T.: Second-stage non-response in the Swiss health survey: determinants and bias in outcomes. *BMC Public Health* 13(1) (2013). <https://doi.org/10.1186/1471-2458-13-167>
- Warren, J.R., Halpern-Manners, A.: Panel Conditioning in Longitudinal Social Science Surveys. *Sociological Methods & Research* 41(4), 491–534 (2012).
<https://doi.org/10.1177/0049124112460374>
- Watson, N., Wooden, M.: Identifying factors affecting longitudinal survey response. In: Lynn, P. (ed.) *Methodology of Longitudinal Surveys*, pp. 157–181. Wiley, Chichester (2009).
<https://doi.org/10.1002/9780470743874.ch10>