

Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial

Ian R. White,^{a,*†} Eleftheria Kalaitzaki^b and Simon G. Thompson^a

Missing outcome data and incomplete uptake of randomised interventions are common problems, which complicate the analysis and interpretation of randomised controlled trials, and are rarely addressed well in practice. To promote the implementation of recent methodological developments, we describe sequences of randomisation-based analyses that can be used to explore both issues. We illustrate these in an Internet-based trial evaluating the use of a new interactive website for those seeking help to reduce their alcohol consumption, in which the primary outcome was available for less than half of the participants and uptake of the intervention was limited.

For missing outcome data, we first employ data on intermediate outcomes and intervention use to make a missing at random assumption more plausible, with analyses based on general estimating equations, mixed models and multiple imputation. We then use data on the ease of obtaining outcome data and sensitivity analyses to explore departures from the missing at random assumption. For incomplete uptake of randomised interventions, we estimate structural mean models by using instrumental variable methods.

In the alcohol trial, there is no evidence of benefit unless rather extreme assumptions are made about the missing data nor an important benefit in more extensive users of the intervention. These findings considerably aid the interpretation of the trial's results. More generally, the analyses proposed are applicable to many trials with missing outcome data or incomplete intervention uptake. To facilitate use by others, Stata code is provided for all methods. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: missing data; missing not at random; non-compliance; randomised trials

1. Introduction

Missing outcome data and incomplete uptake of trial interventions are common problems in randomised controlled trials. A key consideration in handling both issues is the intention-to-treat (ITT) principle [1], which states that all individuals randomised in a clinical trial should be included in the analysis, in the groups to which they were randomised, regardless of any departures from randomised treatment. Following this principle preserves the benefit of randomisation, that the treatment groups cannot differ systematically on any factors except those assigned in the trial, and avoids selection bias. However, it is not universally agreed how the ITT principle applies when some outcomes are missing [2]. Further, the ITT principle does not tell us how to estimate the effects that might have been observed with better uptake of trial interventions.

Missing outcome data are problematic because they cause a loss of power and can lead to biased estimates of intervention effects. Once data are missing, the loss of power cannot be reversed, but it can be minimised by appropriate analysis choices, in particular by including all observed data in the analysis [3]. Estimates of intervention effects are typically biased if the analysis makes the wrong assumption about the missing data. However, any analysis with missing data must make partly or completely untestable

^aMedical Research Council Biostatistics Unit, Cambridge, U.K.

^bMedical Research Council General Practice Research Framework, London, U.K.

*Correspondence to: Ian R. White, Medical Research Council Biostatistics Unit, Cambridge, U.K.

†E-mail: ian.white@mrc-bsu.cam.ac.uk

assumptions, so we can rarely be sure that we have the correct analysis. For this reason, sensitivity analysis is recommended [4–6].

The assumptions of many (but not all) statistical methods for handling missing data can be expressed using the framework of Little and Rubin [7]. Data are missing completely at random (MCAR) if the probability of data being missing does not depend on any missing or observed values. Data are missing at random (MAR) if the probability of a particular set of values being missing for an individual does not depend on the values themselves, conditional on the observed values of other variables. Otherwise, data are missing not at random (MNAR).

Incomplete uptake of trial interventions often means that randomised groups have more similar experience than the investigators had intended, which usually causes the difference in outcomes to be smaller than it would have been with better uptake [8]. However, bias in the estimated intervention effect is not always towards zero: incomplete uptake in equivalence or non-inferiority trials, or in trials where non-trial interventions are available, can inflate differences between randomised groups [9]. Estimating the effect of *allocating* an intervention does not require adjustment for incomplete uptake, in contrast to estimating the effect of a particular level of intervention uptake. The latter is commonly carried out by per-protocol analysis, which excludes data observed when participants had poor intervention uptake. However, per-protocol analysis is undesirable because it is subject to selection bias. Randomisation-respecting alternatives achieve the same aim by using only comparisons of groups as randomised [8]. One such method is principal stratification [10], which leads to estimation of the complier-average causal effect (CACE) [11] in problems where intervention uptake is dichotomous. An alternative, suitable for quantitative intervention uptake, is the structural mean model (SMM) [12].

This paper aims to promote the implementation of recent methodological developments by describing a sequence of analyses that explores both issues and to illustrate the methods using data from an Internet-based trial. This trial is a good example because the issues are particularly acute, but they arise in a wide range of other trials. The Internet-based trial is described in Section 2. Methods for tackling missing data are described in Section 3, with results in Section 4. Methods for tackling incomplete uptake of interventions are described in Section 5, with results in Section 6. We conclude with a discussion in Section 7.

2. The Down Your Drink trial

Hazardous drinking in the general population is an important public health problem [13]. Brief interventions are effective [14] but hard to implement. The Internet is increasingly used to deliver behaviour change interventions [15], and a new ‘Down Your Drink’ (DYD) website was developed, building on psychological theories and aiming to engage users by providing interactive tools [16].

The DYD trial was a randomised evaluation of the DYD website compared with a non-interactive control website providing information only [17]. All stages of the trial—recruitment, randomisation, intervention and data collection—were conducted online. This presented a number of challenges [18]. The key challenge relevant to the present paper was whether the numbers of participants using the intervention website and providing follow-up data would be sufficient.

The primary trial outcome was alcohol consumption in the previous week, which was recorded by the TOT-AL, a specially developed online questionnaire [19]. When an outcome assessment was due, participants received an email with a link to the trial website where they could complete the outcome questionnaires. Alcohol consumption was transformed in all analyses to $\log(\text{number of units in the last week plus } 1)$.

This paper uses data, summarised in Table I, from the pilot trial, which recruited 3746 individuals from 16 February to 16 October 2007. Outcome data were collected at 1 and 3 months; we focus on estimating the intervention effect at 3 months. The correlation between baseline and 3-month alcohol consumption was 0.41 (0.45 for baseline and 1 month; 0.53 for 1 month and 3 months).

Although baseline data were complete, poor follow-up response rates were anticipated because there was no personal contact with participants. To increase response rates, all participants who did not complete the outcome questionnaires within 7 days of the first email invitation received second and (if necessary) third invitations at weekly intervals. A fourth email inviting participants to provide their outcome data directly by email to the investigators yielded no further responses. Offline follow-up was attempted for users who had provided a telephone number or address, but was not successful [18]. Incentives were also trialled [20]. Participants were additionally randomised to complete only one of four secondary outcome measures in order to reduce the assessment burden and improve response rates

Table I. Down Your Drink trial: data description.

	Intervention (<i>n</i> = 1880)		Control (<i>n</i> = 1866)	
Baseline variables				
Age (years)	37.5	(10.9)	37.5	(10.9)
Male	45%		44%	
Has degree	51%		50%	
AUDIT-C score (0 to 12)	8.4	(2.1)	8.4	(2.1)
EQ-5D score (−0.6 to 1)	0.85	(0.18)	0.85	(0.18)
Confidence score (1 to 5)	2.8	(1.2)	2.8	(1.2)
TOT-AL (units/week)	56.0	(36.8)	54.7	(37.3)
log(TOT-AL + 1)	3.80	(0.82)	3.77	(0.84)
Compliance variables				
No. of logins in the first month:	0	5%	5%	
	1	59%	72%	
	≥2	36%	23%	
No. of pages hit in the first month	64.9	(78.6)	12.7	(12.8)
Complier at 1 month	78%		—	
No. of logins in the first 3 months:	0	4%	5%	
	1	51%	60%	
	≥2	45%	35%	
No. of pages hit in the first 3 months	70.4	(89.7)	13.7	(13.9)
1-month outcome variables				
Responded	50%		60%	
TOT-AL (units/week)	39.8	(34.0)	39.5	(32.8)
log(TOT-AL + 1)	3.30	(1.10)	3.30	(1.10)
3-month outcome variables				
Responded	38%		46%	
TOT-AL (units/week)	38.6	(32.6)	37.0	(32.5)
log(TOT-AL + 1)	3.25	(1.12)	3.18	(1.18)

Values are arithmetic mean (SD) or %. AUDIT-C, Alcohol Use Disorders Identification Test-C; EQ-5D, EuroQol Five Dimensional.

Table II. Down Your Drink trial: outcome data at 3 months by number of emails sent.

	Intervention				Control			
	<i>n</i>	%	log(TOT-AL + 1)		<i>n</i>	%	log(TOT-AL + 1)	
Responded after email			Mean	SD			Mean	SD
1	348	19	3.16	1.16	441	24	3.17	1.19
2	194	10	3.28	1.04	236	13	3.18	1.19
3	174	9	3.41	1.11	178	10	3.22	1.17
Never	1164	62	—	—	1011	54	—	—
Total	1880	100	3.25	1.12	1866	100	3.18	1.18

[21]. The number of emails sent to each participant is summarised in Table II and is used in the analysis in Section 4. In the intervention arm, the mean log(TOT-AL + 1) is larger in later respondents than earlier respondents, suggesting a MNAR mechanism, with non-respondents perhaps having an even higher mean log(TOT-AL + 1).

Similarly, low use of the website was a concern. It is hard to define and measure website use [22]; in particular, although each page download was recorded, the length of time that participants spent actually using the website is unknown. We summarised website use by the number of login sessions and the total number of pages downloaded in the first month; in calculating the latter, multiple downloads of the same page were counted only if they occurred in different login sessions. Individuals were automatically logged in to the intervention or control websites after randomisation, but a few who immediately left the trial website had no logins.

The main findings of the trial were that alcohol consumption in responders dropped substantially from baseline to 1 month, and again slightly from 1 to 3 months, but that the drops were very similar across

randomised groups [23]. The ratio of (geometric mean) 3-month alcohol consumption in the intervention group compared with the control group was 1.04 (95% confidence interval 0.94 to 1.16). However, missing data were substantial and more common in the intervention group (Table I). Use of the intervention website was greater than use of the control website, but the majority of participants in both arms had only one login session.

The outstanding questions that this paper aims to answer are whether the results are robust to different assumptions about the missing data and whether interpretation is affected by incomplete use of the website.

3. Missing outcome data: methods

We propose a modelling strategy that starts with simple data on baseline and outcome and then progressively adds in intermediate outcomes, website use and ease-of-contact data.

For the i th participant, let z_i denote their randomised group, \mathbf{x}_i a vector of baseline covariates (assumed complete), y_{i1} , y_{i2} the outcomes at two follow-up times and r_{i1} , r_{i2} whether each outcome was observed (1) or missing (0). Our methods generalise easily to more than two follow-up times. If the data were complete, then an adjusted analysis for the outcome at follow-up time 2 would estimate β in the model

$$y_{i2}|z_i, \mathbf{x}_i \sim N(\alpha + \beta z_i + \boldsymbol{\gamma}'\mathbf{x}_i, \sigma^2). \quad (1)$$

An unadjusted analysis is the same without the $\boldsymbol{\gamma}'\mathbf{x}_i$ term.

3.1. Complete cases

A first analysis fits model (1) in the subset with $r_{i2} = 1$, the ‘complete cases’. This analysis is inefficient because it does not make use of individuals with y_{i2} missing but y_{i1} observed. It is valid if the model is correctly specified and the data are ‘covariate-dependent missing completely at random’ [24]: that is, if the missing data mechanism depends only on the baseline covariates included in the model. If the model is incorrectly specified (e.g. if it should contain a nonlinear function of \mathbf{x}_i), then the analysis is in general valid only if the data are MCAR within randomised groups [25].

3.2. Using repeated outcome measures

We now consider three methods that jointly model both y_{i1} and y_{i2} . These are valid if the data (y_{i1}, y_{i2}) are MAR given (z_i, \mathbf{x}_i) : in particular, for participants with y_{i1} observed, dropout at follow-up time 2 is now allowed to depend on y_{i1} .

A generalised estimating equations (GEE) approach [26] fits the model:

$$E \left[\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} | z_i, \mathbf{x}_i \right] = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix} = \begin{pmatrix} \alpha_1 + \beta_1 z_i + \boldsymbol{\gamma}'_1 \mathbf{x}_i \\ \alpha_2 + \beta_2 z_i + \boldsymbol{\gamma}'_2 \mathbf{x}_i \end{pmatrix} \quad (2)$$

$$\text{var} \left(\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} | z_i, \mathbf{x}_i \right) = \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \quad (3)$$

Normality is not assumed, but the residual variance σ^2 is assumed to be equal at the two times. Estimation uses the standard estimating equations [26]; the parameter of main interest is β_2 . If the model is misspecified (in particular, if the residual variance is different at the two times), then valid standard errors can still be obtained by the robust (sandwich) method. With incomplete data, point estimates for a correctly specified model are valid if the data are MAR, whereas point estimates for an incorrectly specified model are valid if the data are MCAR; weighted estimating equations can relax the latter condition to MAR [27]. We allow the coefficients in the two components of (2) to be different: this amounts to allowing interactions between time and the baseline variables \mathbf{x} and z . In general, it is best to use an unstructured working correlation matrix; with only two time points, this is the same as an exchangeable working correlation matrix.

A *mixed models* approach [28] modifies the model defined by (2) and (3) by adding the distributional assumption

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} | z_i, \mathbf{x}_i \sim N \left(\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix}, \Sigma \right) \quad (4)$$

and replacing (3) with an unconstrained variance–covariance matrix Σ . The model is estimated using restricted maximum likelihood. It may be appropriate to allow Σ to differ by randomised group.

In *multiple imputation* (MI), several completed data sets are produced by drawing the missing values from their posterior predictive distribution thus acknowledging the uncertainty due to missing data under a MAR assumption [29, 30]. This can be carried out using model (4). It is often sensible to draw imputations separately for each trial arm, because interactions between randomised group and baseline covariates may be of interest [31]. It is sometimes considered that MI offers a way to include all randomised individuals in the analysis (e.g. [32]). However, if the imputation model is the same as the analysis model, then MI is expected to give approximately the same results as a mixed model analysis [33].

3.3. Using compliance

One way to make the MAR assumption more plausible is to introduce other post-randomisation variables \mathbf{v}_i into the analysis. Specifically, we now assume that $(y_{i1}, y_{i2}, \mathbf{v}_i)$ are MAR given (z_i, \mathbf{x}_i) , or if \mathbf{v}_i is complete that (y_{i1}, y_{i2}) are MAR given $(z_i, \mathbf{x}_i, \mathbf{v}_i)$, so that observed values of \mathbf{v}_i are allowed to explain missingness of (y_{i1}, y_{i2}) . Here, we take \mathbf{v}_i as the amount of intervention received (compliance), because this is likely to predict both outcomes (y_{i1}, y_{i2}) and responses (r_{i1}, r_{i2}) , but \mathbf{v}_i could also include trial outcomes that are more observed than y_{i2} .

In the mixed model approach, \mathbf{v}_i can be included using the extended model [5]

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \mathbf{v}_i \end{pmatrix} | z_i = z, \mathbf{x}_i \sim N \left(\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \boldsymbol{\mu}_{i3} \end{pmatrix}, \Sigma \right) \quad (5)$$

where μ_{i1} and μ_{i2} are still as defined in (2) and $\boldsymbol{\mu}_{i3} = \boldsymbol{\alpha}_3 + \boldsymbol{\beta}_3 z_i + \boldsymbol{\gamma}'_3 \mathbf{x}_i$. Σ is modelled completely flexibly. The GEE approach is similar but without the normality assumption; Σ is modelled with an unstructured working correlation and equal variances, so it is advisable to scale the compliance variables to have variances similar to the outcome variables. Including \mathbf{v}_i is easiest under the MI approach, because it can simply be included in the imputation model and excluded from the analysis model.

In many trials, compliance has a very different distribution across the two arms and may have different meaning. In this case, it is important to allow the association between \mathbf{v}_i and (y_{i1}, y_{i2}) to vary by randomised group. This is most conveniently carried out in MI, by imputing separately by arm; it cannot be carried out using standard GEE implementations, but a mixed model could allow Σ in (5) to depend on z_i .

3.4. Sensitivity analyses

The aforementioned models attempt to make a MAR assumption more plausible by including more data in the analysis [34]. However, MAR often remains at least questionable, if not implausible [35]. We now consider sensitivity analyses to departures from MAR. Following Kenward *et al.* [4], we embed the MAR model in a wider family of MNAR models indexed by one or more ‘informative missing parameters’ that express the magnitude of departures from MAR. We then use subject-matter knowledge to specify possible values of the informative missing parameters and re-estimate the intervention effect in each case.

We use a pattern-mixture model [36] that extends Equation (1) by allowing a term δ^Y that controls departures from MAR:

$$y_{i2} | z_i, \mathbf{x}_i, r_{i2} \sim N(\alpha_{CC} + \beta_{CC} z_i + \boldsymbol{\gamma}'_{CC} \mathbf{x}_i + \delta^Y (1 - r_{i2}), \sigma^2). \quad (6)$$

The regression parameters subscripted CC can be estimated by fitting Equation (1) to the complete cases ($r_{i2} = 1$), but δ^Y is not identified by the data. An important extension allows the informative missing parameter δ^Y to differ between randomised groups:

$$y_{i2} | z_i, \mathbf{x}_i, r_{i2} \sim N(\alpha_{CC} + \beta_{CC} z_i + \boldsymbol{\gamma}'_{CC} \mathbf{x}_i + \delta^Y_{z_i} (1 - r_{i2}), \sigma^2) \quad (7)$$

where $\delta_{z_i}^Y(1 - r_{i2})$ can also be written as $\delta_1^Y z_i(1 - r_{i2}) + \delta_0^Y(1 - z_i)(1 - r_{i2})$. This model is plausible because, for example, missing data may well be more informative among individuals who have been encouraged to change their behaviour than among controls, and is important because treatment effects are most affected when departures from MAR behave differently in the two arms [37]. Parameters (δ_0^Y, δ_1^Y) are not identified by the data: δ_0^Y is the mean difference between unobserved and observed outcomes in the control arm, adjusted for \mathbf{x} , and δ_1^Y is the corresponding difference in the intervention arm.

This model has previously been used with an informative prior distribution for (δ_0^Y, δ_1^Y) that was elicited from investigators [37]. In the present paper, investigators' views are used to define plausible values of (δ_0^Y, δ_1^Y) for sensitivity analysis, rather than tackling a fully Bayesian analysis. It is useful to consider three sensitivity analyses: first, to values of $\delta_0^Y = \delta_1^Y$, then to values of δ_0^Y fixing $\delta_1^Y = 0$ and, finally, to values of δ_1^Y fixing $\delta_0^Y = 0$.

Once (δ_0^Y, δ_1^Y) has been specified, estimation is straightforward. Write $\mathbf{w}'_i = (1, z_i, \mathbf{x}'_i)$ and $\boldsymbol{\theta}'_{CC} = (\alpha_{CC}, \beta_{CC}, \boldsymbol{\gamma}'_{CC})$ so the mean part of model (6) is $E[y_{i2}|\mathbf{w}_i, r_{i2}] = \boldsymbol{\theta}'_{CC} \mathbf{w}_i + \delta_{z_i}^Y(1 - r_{i2})$. The model of interest is $E[y_{i2}|\mathbf{w}_i] = \boldsymbol{\theta}' \mathbf{w}_i$ where $\boldsymbol{\theta}' = (\alpha, \beta, \boldsymbol{\gamma}')$. It follows that $\boldsymbol{\theta} = \boldsymbol{\theta}_{CC} + \boldsymbol{\theta}_{ADD}$ where $\boldsymbol{\theta}_{ADD}$ are the coefficients from a regression of $\delta_{z_i}^Y(1 - r_{i2})$ on \mathbf{w}_i . It is easy to estimate $\boldsymbol{\theta}_{CC}$ and $\boldsymbol{\theta}_{ADD}$. Finally, the estimated parameters are independent, so we can estimate $\text{var}(\hat{\boldsymbol{\theta}})$ as $\widehat{\text{var}}(\hat{\boldsymbol{\theta}}_{CC}) + \widehat{\text{var}}(\hat{\boldsymbol{\theta}}_{ADD})$.

3.5. Using the number of attempts

Instead of specifying the informative missing parameter(s) based on subject-matter knowledge, it may be possible to estimate them using data on the number of attempts made to observe outcome y_{i2} . Let r_{i2k} be the outcome of the k th attempt to observe the primary outcome y_{i2} , where $r_{i2k} = 1$ indicates that the outcome was observed, $r_{i2k} = 0$ indicates that it was not observed and $r_{i2k} = \cdot$ indicates that the k th attempt was not made (either because a previous attempt was successful or because the participant had refused or withdrawn from the trial). The association between r_{i2k} and y_{i2} can be identified using Alho's model, which assumes that this association is the same for all k [38]:

$$\text{logit } P(r_{i2k} = 1 | r_{i2k} \neq \cdot, z_i, \mathbf{x}_i, y_{i2}) = \alpha_k^R + \beta^R z_i + \boldsymbol{\gamma}^{R'} \mathbf{x}_i + \delta^R y_{i2}. \quad (8)$$

Here, we allow the probability of responding to vary between attempts, but we assume that the association between fully observed covariates and responding is the same at all attempts, although the latter assumption could easily be relaxed. δ^R is an informative missing parameter, and $\delta^R = 0$ corresponds to MAR.

Estimation of model (8) uses data on individuals with observed outcomes together with the numbers and baseline covariates of individuals with unobserved outcomes. The model may be fitted using a conditional likelihood supplemented by a set of estimating equations [38], but this algorithm is not guaranteed to converge. Alternative estimation methods are based on the full likelihood for model (8) jointly with model (4). A Bayesian approach has been used [39], and a likelihood-based approach is also possible; the likelihood involves integrating out the unobserved values of y_{i2} . Fitting the model by using the full likelihood directly estimates the parameters of (4); an alternative is to use the inverse of the response probability as a weight for analysis of complete cases [38], but care must be taken to obtain standard errors that allow for the often large uncertainty in the weights [39].

As in the previous section, an important extension to model (8) allows the informative missing parameter δ^R to differ between randomised groups:

$$\text{logit } P(r_{i2k} = 1 | r_{i2k} \neq \cdot, z_i, \mathbf{x}_i, y_{i2}) = \alpha_k^R + \beta^R z_i + \boldsymbol{\gamma}^{R'} \mathbf{x}_i + \delta_{z_i}^R y_{i2} \quad (9)$$

where $\delta_{z_i}^R y_{i2}$ can also be written as $\delta_1^R z_i y_{i2} + \delta_0^R(1 - z_i)y_{i2}$.

4. Missing outcome data: analysis of the Down Your Drink trial

4.1. Implementation

In the DYD trial, $z_i = 1$ for individuals randomised to the interactive website and 0 for the control website, and (y_{i1}, y_{i2}) are the 1-month and 3-month alcohol consumption outcomes ($\log(\text{TOT-AL} + 1)$). All analyses were performed both unadjusted and adjusted for \mathbf{x}_i , which comprises the baseline variables

listed in Table I; in general, we prefer the analysis adjusted for baseline covariates, especially the baseline value of the outcome. For GEEs, robust standard errors were used. For mixed models, the unstructured variance–covariance matrix was allowed to differ between randomised groups, although results with a common variance–covariance matrix (not shown) were very similar.

Multiple imputations were drawn separately for each arm by using the chained equations approach [40] implemented in Stata [41, 42]. For method MI1, the imputation model for the outcome at each time was a linear regression including the outcome at the other time and the baseline variables. For method MI2, the imputation models additionally included the log of one plus the numbers of pages hit at 1 and 3 months and the log of one plus the numbers of login sessions at 1 and 3 months. The distribution of the incomplete variables was not fully Normal, even after log transformation, so predictive mean matching was used to improve the imputations [31]. In each case, model (1) was fitted to each of 50 imputed data sets, and the results were combined using Rubin’s rules [29]. Covariates were used in the imputation model even when unadjusted analyses were performed.

For sensitivity analyses, the views of five DYD investigators were quantified before the trial results were known, and these views were used to choose values of the informative missing parameters (δ_0^Y, δ_1^Y) in Equation (7). When the informative missing parameters were assumed the same in both arms, the investigators believed that the mean of the unobserved responses for alcohol consumption at 3 months could be as much as 75% more or 50% less than the mean of the observed responses: these suggest the sensitivity analyses $\delta_0^Y = \delta_1^Y = \log 1.75$ and $\delta_0^Y = \delta_1^Y = \log 0.5$. When the data were assumed to be informatively missing only in the control arm, the investigators believed that the mean of the unobserved responses could be as much as 50% more or 50% less than the mean of the observed responses: these suggest the sensitivity analyses $(\delta_0^Y, \delta_1^Y) = (\log 1.5, 0)$ and $(\log 0.5, 0)$. We also choose the corresponding cases with the data informatively missing only in the intervention arm: $(\delta_0^Y, \delta_1^Y) = (0, \log 1.5)$ and $(0, \log 0.5)$. In addition to these rather extreme sensitivity analyses, we also used more moderate sensitivity analyses with $(\delta_0^Y, \delta_1^Y) = (\log 1.5, \log 1.5)$, $(\log 1.25, 0)$ and $(0, \log 1.25)$.

Analysis of number of attempts used the number of email reminders that were sent to each participant. The conditional likelihood algorithm diverged in some cases, so we used the maximum likelihood approach. ‘Alho 1’ and ‘Alho 2’ refer to models (8) and (9), respectively.

Stata code for these analyses is given in Appendix A.1.

4.2. Results

We summarise the results in Table III and display the covariate-adjusted results in Figure 1. The intervention effect is expressed as the ratio of the geometric mean alcohol consumption (plus 1 unit/week) at 3 months in the intervention group to the corresponding geometric mean in the control group. In the following text, we interpret these figures as percentage increases or decreases.

All methods based on MAR, as well as complete-cases analysis, give very similar results: the point estimate represents a non-significant increase of between 4% and 12% due to the intervention, with a 95% confidence interval that does not extend below a 6% reduction.

Sensitivity analyses show that the estimated intervention effect is not very sensitive to departures from MAR when the informative missing parameter is assumed to be equal across randomised groups, but is very sensitive to departures from MAR that occur differently in the randomised groups. Moderate sensitivity analyses (indicated by * in Table III and Figure 1) yield estimates ranging from an 8% reduction to a 23% increase in alcohol consumption, whereas more extreme sensitivity analyses range from a 32% reduction to a 56% increase. This suggests that the trial’s results are only robust to departures from MAR that are similar in both randomised groups.

Using the number of email reminders and the MNAR models (8) and (9) gives the estimates of the informative missing parameters in Table IV. For model (8), where the informative missing parameter is assumed equal across the two groups, the negative estimate of the informative missing parameter δ^R suggests that heavier drinkers are more likely to be non-responders. However, the informative missing parameter is not significantly different from zero so that the data are consistent with a MAR assumption. For model (9), where the informative missing parameter is allowed to differ between groups, both estimates are again negative and that for the intervention group is larger in magnitude, suggesting that the tendency for heavier drinkers to be non-responders may be greater in the intervention group. Although the informative missing parameter in the intervention group is significantly different from zero ($P = 0.03$), a test for a difference between the arm-specific informative missing parameters is not significant ($P = 0.26$) nor is a test on 2 degrees of freedom for departure from MAR ($P = 0.09$).

Table III. Down Your Drink trial: analysis of alcohol consumption at 3 months, using various assumptions and methods to handle missing outcome data.

Approach	Assumption	Method	Unadjusted	Adjusted
Complete cases	CD-MCAR ¹		1.073 (0.956 to 1.203)	1.043 (0.939 to 1.157)
Using repeated outcome measures	MAR ²	GEE	1.095 (0.982 to 1.222)	1.058 (0.957 to 1.171)
		Mixed model	1.097 (0.983 to 1.223)	1.063 (0.960 to 1.176)
		MI1	1.072 (0.960 to 1.196)	1.050 (0.945 to 1.167)
Using website use	MAR ³	MI2	1.115 (0.991 to 1.255)	1.092 (0.974 to 1.225)
Sensitivity analysis using (7)		exp(δ)		
	MNAR, $\delta_0^Y = \delta_1^Y = \delta$	0.5	1.017 (0.905 to 1.142)	0.990 (0.890 to 1.101)
		1.5*	1.107 (0.986 to 1.242)	1.074 (0.967 to 1.193)
		1.75	1.120 (0.997 to 1.258)	1.087 (0.978 to 1.208)
	MNAR, $\delta_0^Y = 0, \delta_1^Y = \delta$	0.5	0.698 (0.622 to 0.784)	0.680 (0.612 to 0.755)
		1.25*	1.232 (1.098 to 1.381)	1.197 (1.078 to 1.328)
		1.5	1.379 (1.229 to 1.547)	1.339 (1.206 to 1.487)
	MNAR, $\delta_0^Y = \delta, \delta_1^Y = 0$	0.5	1.562 (1.391 to 1.753)	1.519 (1.367 to 1.688)
		1.25*	0.951 (0.847 to 1.066)	0.924 (0.832 to 1.025)
		1.5	0.861 (0.768 to 0.966)	0.837 (0.753 to 0.929)
Using the number of attempts and (9)				
	MNAR, $\delta_0^R = \delta_1^R$	Alho 1	1.086 (0.967 to 1.220)	1.050 (0.945 to 1.167)
	MNAR, $\delta_0^R \neq \delta_1^R$	Alho 2	1.328 (0.926 to 1.904)	1.057 (0.872 to 1.281)

Figures are ratio of geometric means, intervention/control, with 95% confidence interval.

*Moderate sensitivity analyses.

¹ y_{i2} is covariate-dependent MCAR given (z_i, x_i) .

² (y_{i1}, y_{i2}) are MAR given (z_i, x_i) .

³ (y_{i1}, y_{i2}) are MAR given (z_i, x_i, v_i) .

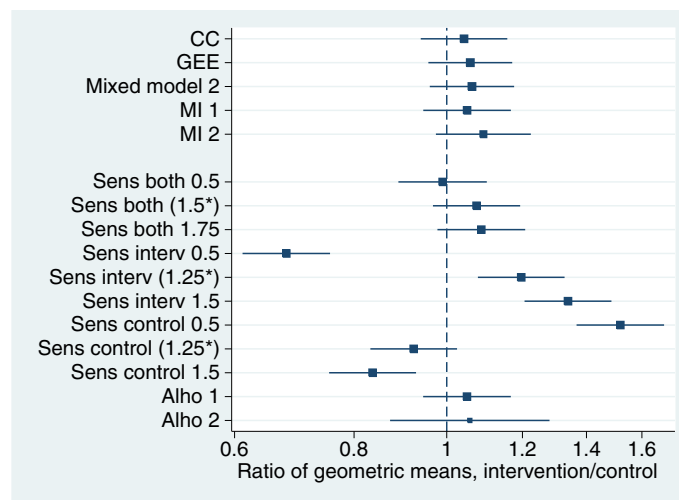


Figure 1. Down Your Drink trial: estimates (95% confidence intervals) of the intervention effect on weekly alcohol consumption, adjusted for baseline covariates, using different methods for handling the missing data.

*Denotes moderate sensitivity analyses.

Table IV. Down Your Drink trial: informative missing parameters δ^R (standard errors), defined as the log odds ratio for response per 1-unit increase in 3-month log(TOTAL + 1), estimated using the Alho model and maximum likelihood.

Model	Arm	Unadjusted	Adjusted
Alho 1	Both	-0.15 (0.10)	-0.10 (0.13)
Alho 2	Intervention	-0.28 (0.13)	-0.11 (0.16)
	Control	-0.06 (0.15)	-0.10 (0.13)

Alho 1: MNAR, common δ^R across attempts and arms (model (8)).

Alho 2: MNAR, common δ^R across attempts (model (9)).

These results do not provide good evidence against a MAR assumption, but they change the estimated intervention effects in Table III when the informative missing parameter is allowed to differ across randomised groups as in model (9). This MNAR analysis indicates a much larger increase due to intervention in the unadjusted analysis, and much wider confidence intervals in both unadjusted and adjusted analyses. We attribute these findings to the great sensitivity of estimated intervention effects to differences in informative missing parameter δ^R between randomised groups, along with the difficulty of estimating this parameter.

5. Incomplete uptake of interventions: methods

Intervention receipt in some randomised trials can be summarised as a binary variable [43], whereas other trials have complex intervention receipt that may be summarised as one or more quantitative variables. We present a SMM that is applicable to both binary and quantitative cases, provided that intervention receipt is univariate.

5.1. Structural mean model

Structural mean models describe the relationship between the observed data and the counterfactual data that would have been observed with a different random allocation [12, 44]. For the i th individual, we define $y_i(1)$ as the outcome that would be observed if they were randomised to intervention and $y_i(0)$ as the corresponding outcome if they were randomised to control. Exactly one of these potential outcomes is observed for each individual. Define $d_i(1)$ as the i th individual's compliance (binary or quantitative) with the intervention, if they were allocated to intervention. We initially ignore compliance with the control. We now assume that the causal effect of the intervention is proportional to the compliance. This implies that individuals who would be complete non-compliers if allocated to intervention have no effect of allocation, the 'exclusion restriction' assumption. We then have the SMM

$$y_i(1) - y_i(0) = \psi d_i(1) + e_i \tag{10}$$

where e_i is a zero-mean error term whose presence allows treatment effects to vary between individuals. Model (10) implies that

$$E[y_i(1) - \psi d_i(1)] = E[y_i(0)]. \tag{11}$$

Estimation proceeds by noting that randomised group z_i is independent of the potential outcomes $y_i(1)$, $y_i(0)$ and $d_i(1)$, so each expectation in (11) can be computed in one arm of the trial. This leads to the estimating equation $\sum_i (z_i - \bar{z})(y_i - \psi d_i) = 0$ where $d_i = d_i(1)$ or 0 for individuals randomised to intervention or control, respectively.

Baseline covariates \mathbf{x}_i that are uncorrelated with z_i and e_i may be used in two ways to improve the efficiency of the estimation procedure. First, we can condition on \mathbf{x}_i in (11) and model $E[y_i(0)|\mathbf{x}_i] = \alpha + \boldsymbol{\gamma}'\mathbf{x}_i$, yielding the alternative estimating equation $\sum_i (z_i - \bar{z})(y_i - \psi d_i - \alpha - \boldsymbol{\gamma}'\mathbf{x}_i) = 0$, to which we add standard estimating equations for α and $\boldsymbol{\gamma}$ [45], giving

$$\sum_i \begin{pmatrix} 1 \\ \mathbf{x}_i \\ z_i \end{pmatrix} (y_i - \psi d_i - \alpha - \boldsymbol{\gamma}'\mathbf{x}_i) = \mathbf{0}. \tag{12}$$

Equation (12) is easy to estimate because it is a standard instrumental variables (IV) model [46], in which z_i is the instrument, d_i is the ‘endogenous’ variable and x_i is the ‘exogenous’ variable. A second approach, not adopted here, makes use of baseline covariates w_i that predict d_i in the intervention group. In this case, precision can be gained by using the interactions $z_i w_i$ as additional instruments, but at the cost of further assumptions [47]. Again, this can be fitted using standard IV software.

5.2. Interpretation

Interpreting the estimated parameter ψ is easiest when compliance is binary. In the aforementioned model, this means that $d_i(1)$ is 0 or 1. In the statistical literature, the two groups formed are often known as ‘compliers’ ($d_i(1) = 1$) and ‘non-compliers’ ($d_i(1) = 0$), referring to an individual’s compliance status if they were randomised to the intervention. In this setting, the ‘exclusion restriction’ assumption, which identifies the model, states that randomised allocation has no effect on non-compliers. The parameter ψ can be interpreted without further assumptions as the CACE, the average of $y_i(1) - y_i(0)$ over the subgroup with $d_i(1) = 1$ [48].

If compliance is not naturally binary, it is tempting to dichotomise it. Clearly, a good definition of ‘compliers’ is needed. It is not necessary to assume that compliers all receive the same benefit of intervention, because the CACE represents an average overall compliers. However, it is essential to assume that non-compliers receive no benefit from intervention. It is therefore typically necessary to use a restrictive definition of non-compliance, classing any individual whose moderate compliance could have brought him or her benefit with the compliers.

An alternative approach is to express compliance $d_i(1)$ quantitatively. This typically makes the exclusion restriction more plausible, because the zero level can be chosen to represent no use of the intervention. However, without further assumptions, no simple interpretation of ψ generalises the CACE. The further assumption usually made is $E[e_i|d_i] = 0$ so that model (10) is correctly specified. In this case, ψd can be interpreted as the average causal effect of allocation to intervention in the subgroup who would comply to an extent d : that is, $E[y_i(1) - y_i(0)|d_i(1) = d] = \psi d$.

5.3. Using control group compliance

When the control group also receives some intervention, such as a placebo or a standard treatment, control-group compliance $d_i(0)$ is also available. The SMM could then be extended as

$$y_i(1) - y_i(0) = \psi_1 d_i(1) - \psi_0 d_i(0) + e_i \quad (13)$$

which allows for a causal effect of the control intervention. The original approach to this problem assumed that $d_i(1)$ is a monotonic function of $d_i(0)$ [49], but the method is very sensitive to departures from this assumption [50]. More recent causal estimation methods for models such as (13) use either an assumption that $d_i(0) \leq d_i(1)$ for all i and Bayesian modelling with slightly informative priors [51], or informative priors for one of the treatment effects [52], or covariates that predict $d_i(0)$ and $d_i(1)$ differently but that do not modify the causal effect of treatment [45]. Because of the complexities of all these approaches, we would prefer to ignore $d_i(0)$ when it is plausible that the control intervention has no causal effect (i.e. that $\psi_0 = 0$ in (13)).

5.4. Missing data

Missing outcome data complicate estimation of the IV model. Standard implementations of IV are restricted to using complete cases only and are thus valid only under MCAR. Three approaches can be used to make them valid under MAR.

First, inverse probability weighting (IPW) can be used [53, 54]. Models are constructed for $p(r_i|x_i, z_i = 1, d_i)$ and $p(r_i|x_i, z_i = 1)$, and the ‘stabilised weights’ [55] in group 1 are formed as the ratio of the fitted values, $\hat{p}(r_i|x_i, z_i = 1)/\hat{p}(r_i|x_i, z_i = 1, d_i)$. The stabilised weights in group 0 are all 1, because d_i does not vary. Weighted IV regression is then performed with robust standard errors.

Second, the ‘adjusted treatment received’ (ATR) method [56, 57] is equivalent to IV regression for complete data and is valid when outcomes are MAR [54]. In this method, a linear regression model is first constructed for actual treatment receipt on randomised group and covariates (using all observations including those with missing y_i), and the residuals are estimated. The causal effect of actual treatment receipt is then estimated by linear regression of y_i on actual treatment receipt, adjusting for

the previously estimated residuals and the covariates. The standard errors from this second stage may be underestimated because they ignore uncertainty in the residuals [54].

Third, MI can be used.

6. Incomplete uptake of interventions: analysis of the Down Your Drink trial

6.1. Implementation

The DYD trial has complex intervention receipt: individuals could use the website on different numbers of occasions, for different lengths of time, and in different ways. Any attempt to estimate the effect of intervention receipt in such data relies on a plausible causal model describing how intervention receipt may affect outcomes. We describe one approach with dichotomised compliance and one with quantitative compliance.

For dichotomised compliance, a non-zero cut-off was chosen, because almost all randomised individuals had at least one login (Table I). Section 5.2 argues for a relatively low cut-off, and we defined compliers as individuals who logged in more than once or accessed more than 10 pages of the website within the first 1 month from randomisation. Our analyses therefore rest on the assumption that an individual who accessed fewer than 10 pages on only one occasion received no benefit, and they estimate the average benefit of the intervention website over a wide range of use.

For quantitative compliance, we defined $d_i(1)$ in Equation (10) as the number of pages downloaded over the first month of the trial, but with an upper limit of 300 pages because we did not believe that use above this level would have further benefit. We did not use website uptake in the control group in the model, because the control website is unlikely to be effective. For the MI approach, we used the imputations constructed using compliance variables as in MI2 of Section 4.

Stata code for these analyses is given in Appendix A.2.

6.2. Results

Of 1880 individuals allocated to intervention, 1461 (78%) were classed as compliers. As a result, the estimated CACE (Table V) was not very different from the ITT MAR estimate (Table III). IPW and ATR methods behaved very similarly. MI gave somewhat different results: although this is unexpected, it is consistent with the differences between MI1 and MI2 in Table III.

Estimates of the causal effect per 100 pages downloaded were somewhat larger than the ITT estimates. This appears to be because the mean number of pages downloaded in the intervention group was 65, so the estimated effect of downloading 100 pages was approximately one and a half (100/65) times the ITT effect. The confidence interval for the intervention effect in these analyses does not extend below an 11% reduction.

7. Discussion

7.1. Conclusions for the Down Your Drink trial

A concern with the DYD trial, and many other online trials, is that high rates of non-response and low intervention uptake makes it hard to draw conclusions about the intervention's effectiveness. Our

Table V. Down Your Drink trial: analysis of alcohol consumption at 3 months, allowing for incomplete use of website.

Missing data method	Unadjusted	Adjusted
Binary compliance: compliers versus non-compliers		
IPW	1.100 (0.950 to 1.273)	1.053 (0.921 to 1.203)
ATR	1.100 (0.949 to 1.275)	1.052 (0.920 to 1.204)
Multiple imputation	1.150 (0.986 to 1.342)	1.121 (0.965 to 1.301)
Continuous compliance: per 100 pages downloaded		
IPW	1.124 (0.937 to 1.349)	1.079 (0.917 to 1.269)
ATR	1.123 (0.935 to 1.348)	1.082 (0.916 to 1.277)
Multiple imputation	1.189 (0.982 to 1.440)	1.151 (0.956 to 1.386)

Figures are ratio of geometric means, intervention/control, with 95% confidence interval. ATR, adjusted treatment received; IPW, inverse probability weighting.

analyses in this paper show that the conclusions were not substantially affected under a range of assumptions about the missing data mechanism, except when we assumed that the informative missing parameter differed between randomised groups. To the extent that the latter assumption may be implausible, our results appear reasonably robust. Similarly, conclusions were not substantially affected when we used causal models to consider the impact of downloading 100 website pages. The latter conclusion depends on a judgement that 100 website pages was a reasonable target for moderately conscientious website use.

Our results therefore provide some support for the use of online trials in general, with two cautions: it is essential to consider the informative missingness parameters differing between randomised groups and to consider how much the observed intervention uptake falls short of what might be hoped for. Analyses allowing for non-response and low intervention uptake are best specified in advance and included in the analysis plan.

7.2. Methodological conclusions

These methods are of potential use in all trials and not just online trials. When rates of missing data are low, sensitivity analysis may be enough to demonstrate that missing data are not a problem. In other cases, including intermediate or other outcomes and/or compliance variables in MAR analyses is a useful strategy, although treatment effect estimates may only be changed when the auxiliary variables are strongly associated with outcome [58]. Sensitivity analyses are always helpful but depend on expert consideration of the plausible degree of departure from MAR. Data on number of attempts to obtain data, or more generally ease of contact, are often recorded and should be more widely used in analysis: results from the Alho model (8) or (9) can be a useful way to allow for extra uncertainty due to the possibility of MNAR data without the need to rely on expert opinion. With pressure on journal space, it may be convenient for all these alternative analyses to be included in web appendices. In the primary publication of the DYD trial [23], which was based on more data than those used here, the primary analysis was the adjusted complete-cases analysis, and web appendices presented alternative analyses for the missing data—a partial last observation carried forward (LOCF; see in the next section), MI and sensitivity analyses using (7)—and analyses adjusting for non-compliance.

A particularly relevant question in a trial with a ‘negative’ result is whether this negative result is attributable to incomplete intervention uptake. In this context, it is important to formulate the causal question carefully, defining a parameter such as the CACE or the causal effect of a particular amount of intervention, and then consider the limits of the confidence interval for the parameter.

7.3. Other methods

We have not reported here an analysis using LOCF, one of the most widely used techniques [3], which simply replaces missing outcomes with the last observed value. LOCF rests on an assumption that outcomes do not change (on average in each arm) after participants drop out of the study, which is often implausible. In the DYD trial, with its large change in outcome after baseline, LOCF would yield implausibly different imputations for individuals with no post-baseline measurement and those with a 1-month measurement. Because LOCF is widely used, the primary DYD trial publication [23] reported a partial LOCF analysis that carried only post-baseline measurements forward. Unfortunately, usual justifications for LOCF rest not on the plausibility of its assumption but on approximate constancy of observed outcomes, or on an appeal to the ITT principle, or on conservatism: none of these are valid [2].

Our sensitivity analyses were based on data from baseline and follow-up time 2 only. Basing the sensitivity analyses on the mixed model (4) might be preferable but is technically more complicated and is unlikely to make much difference in view of the small differences between complete-cases and MAR-based analyses (Table III). The Alho method could also be extended to allow for the repeated measures: for example, better estimating the informative missing parameter δ^R by assuming it to be constant across follow-up times. Another possible assumption about the missing data is that they are ‘latent ignorable’, meaning that they would be MAR if the potential compliance $d_i(1)$ were observed for everyone [59].

7.4. Extensions

For binary outcomes, mixed models become more complex, and GEE or MI methods might be preferred. The MNAR methods can be applied equally well, and the exp (δ) parameters can be interpreted

as informatively missing odds ratios [60,61]. The SMMs described may still be used to estimate causal risk differences, but if causal odds ratios are wanted then generalised SMMs are needed [62].

Methods used for survival outcomes are typically very different from those that we have described. Here, missing data take the form of censoring, and the non-informative censoring assumption takes the place of the MAR assumption: departures from the non-informative censoring assumption are rarely considered but should be. SMMs are not suitable for survival outcomes, but the structural accelerated failure time model is a general alternative for handling incomplete intervention uptake [63, 64], and hazard-based methods are available for handling all-or-nothing uptake [65].

In the DYD trial, all baseline covariates were complete. Incomplete baseline covariates are simply and efficiently handled by single imputation methods such as imputing the overall or centre-specific mean of the covariate [31]. Such simple methods would be inappropriate for missing outcomes: they are appropriate for missing baselines because baseline covariates are independent of randomised group, and adjustment for baseline covariates is not required for unbiased estimation [66].

Stata do-files to implement the analyses presented in this paper are given in Appendix A.

APPENDIX A. Stata code

A.1. Missing data

The succeeding code uses four user-written commands. `ice` [67] and `mim` [68] implement MI and are available from the Statistical Software Components (SSC) archive. `alho` and `rctmiss` implement the ‘number of attempts’ model and sensitivity analyses, respectively, and are available from the first author’s website by typing `net from http://www.mrc-bsu.cam.ac.uk/IW_Stata/` in Stata.

The code shows adjusted analyses; for unadjusted analyses, delete the `global xvvars` and `global time_xvvars` commands. The data are assumed to be in a file `DYDwide.dta` with one record per randomised individual.

```
// Define global macros for different sets of variables
global yvars logtotal1 logtotal3 // Outcome variables
global dvars lpages1 lpages3 lsession1 lsession3
// Intervention uptake variables (logged)
global xvvars logtotal_bs age gender educ_cat confidence_0 AUDITC EQ5D_0
// Baseline covariates
global time_xvvars i.time#c.logtotal_bs i.time#c.age i.time#i.gender ///
i.time#i.educ_cat i.time#c.confidence_0 i.time#c.AUDITC i.time#c.EQ5D_0
// Interaction terms

// Complete Case Analysis
use DYDwide, clear
regress logtotal3 group $xvvars, eform(ratio)

// Generalised Estimating Equations (GEE)
reshape long logtotal, i(userid) j(time) // One record per time per individual
xtset userid time
gen group1 = group*(time==1)
gen group3 = group*(time==3)
xtgee logtotal i.time group1 group3 $xvvars $time_xvvars, corr(uns) robust eform

// Mixed Model, different variances by arm
xtmixed logtotal i.time group1 group3 $xvvars $time_xvvars || userid:,nocons || , ///
residuals(unstructured,t(time) by(group))

// Multiple Imputation
use DYDwide, clear
ice $yvars $xvvars, // Without website use variables
seed(1234) m(50) by(group) match($yvars) saving(imputed1, replace)
ice $yvars $xvvars $dvars, // With website use variables
seed(1234) m(50) by(group) match($yvars) saving(imputed2, replace)
```

```

use imputed1, clear                                // MI1

mim: regress logtotal3 group $xvars, eform(Ratio)
mim, merror eform

use imputed2, clear                                // MI2
mim: regress logtotal3 group $xvars, eform(Ratio)
mim, merror eform

// Sensitivity Analyses
use DYDwide, clear
rctmiss, pmmdelta(ln(0.5))                         eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(1.5))                         eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(1.75))                       eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(0.50)*(group==1))            eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(1.25)*(group==1))            eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(1.50)*(group==1))            eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(0.50)*(group==0))            eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(1.25)*(group==0))            eform(Ratio): reg logtotal3 group $xvars
rctmiss, pmmdelta(ln(1.50)*(group==0))            eform(Ratio): reg logtotal3 group $xvars

// Using number of email reminders
* Alho1: common IMP across attempts and arms
alho logtotal3 group $xvars, repeats(nreminders)

* Alho2: common IMP across attempts but differs across arms
alho logtotal3 group $xvars, repeats(nreminders) imeq(group)
lincom [respond]groupXlogtotal3 + [respond]logtotal3 // IMP in intervention arm

```

A.2. Allowing for incomplete intervention uptake

The following code is for quantitative compliance: results with binary compliance are obtained by redefining `treat`.

```

use DYDwide, clear
gen treat = min(pages1, 300) * (group==1) / 100

// Handling missing data by IPW
gen obs = !missing(logtotal3)
logit obs treat $xvars if group==1
predict pobs1
logit obs $xvars if group==1
predict pobs2
gen weight = pobs2/pobs1 if group==1
replace weight = 1 if group==0
ivregress 2sls logtotal3 $xvars (treat=group) [pw=weight], eform(Ratio)

// Handling missing data by ATR (Nagelkerke)
reg treat group $xvars

predict resid, resid
regress logtotal3 treat $xvars resid, eform(Ratio)

// Handling missing data by MI
use imputed2, clear
gen treat = min(pages1, 300) * (group==1) / 100
mim, category(fit): ivregress 2sls logtotal3 $xvars (treat=group), eform(Ratio)
mim, merror eform

```

Acknowledgements

The National Prevention Research Initiative (<http://www.npri.org.uk>) funded the DYD trial, with support from the following organisations: British Heart Foundation; Cancer Research UK; Chief Scientist Office, Scottish Government Health Directorate; Department of Health; Diabetes UK; Economic and Social Research Council; Health & Social Care Research & Development Office for Northern Ireland; Medical Research Council (MRC); Welsh Assembly Government; and World Cancer Research Fund.

I. R. W. was supported by MRC grant U.1052.00.006, and S. G. T. was supported by MRC grant U.1052.00.001.

References

- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal* 1999; **319**:670–674.
- White IR, Horton N, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *British Medical Journal* 2011; **342**:d40.
- Wood A, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials* 2004; **1**:368–376.
- Kenward MG, Goetghebeur EJT, Molenberghs G. Sensitivity analysis for incomplete categorical tables. *Statistical Modelling* 2001; **1**:31–48.
- Carpenter JR, Kenward MG. *Missing Data in Clinical Trials—A Practical Guide*. National Institute for Health Research, Publication RM03/JH17/MK: Birmingham, 2008. Available at <http://www.pcpoh.bham.ac.uk/publichealth/methodology/projects/RM03JH17MK.shtml>.
- National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The National Academies Press: Washington, DC, 2010.
- Little RJA, Rubin DB. *Statistical Analysis with Missing Data*, 2nd edn. Wiley: Hoboken, N. J., 2002.
- White IR. Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research* 2005; **14**:327–347.
- Sheng D, Kim M. The effects of non-compliance on intent-to-treat analysis of equivalence trials. *Statistics in Medicine* 2006; **25**:1183.
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**:21–29.
- Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* 1997; **25**:305–327.
- Goetghebeur E, Lapp K. The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Journal of the Royal Statistical Society (C)* 1997; **46**:351–364.
- Prime Minister's Strategy Unit. Alcohol Harm Reduction Strategy for England. Cabinet Office, 2004. Available from <http://www.referrersguide.com/Guidelines/Alcohol%20Harm%20Reduction%20Strategy.pdf>.
- Moyer A, Finney J, Swearingen C, Vergun P. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction* 2002; **97**:279–292.
- Murray E. Internet-delivered treatments for long-term conditions: strategies, efficiency and cost-effectiveness. *Expert Review of Pharmacoeconomics and Outcomes Research* 2008; **8**:261–272.
- Linke S, McCambridge J, Khadjesari Z, Wallace P, Murray E. Development of a psychologically enhanced interactive online intervention for hazardous drinking. *Alcohol and Alcoholism* 2008; **43**:669–674.
- Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, Godfrey C, Linke S, Wallace P. The DYD-RCT protocol: an on-line randomised controlled trial of an interactive computer-based intervention compared with a standard information website to reduce alcohol consumption among hazardous drinkers. *BMC Public Health* 2007; **7**:306.
- Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, Thompson S, Wallace P. Methodological challenges in on-line trials. *Journal of Medical Internet Research* 2009; **11**:e9.
- Khadjesari Z, Murray E, Kalaitzaki E, White I, McCambridge J, Godfrey C, Wallace P. Test–retest reliability of an online measure of past week alcohol consumption (the TOT-AL), and comparison with face-to-face interview. *Addictive Behaviors* 2009; **34**:337–342.
- Khadjesari Z, Murray E, Kalaitzaki E, White IR, McCambridge J, Thompson SG, Wallace P, Godfrey C. Impact and costs of incentives to reduce attrition in online trials: Two randomised controlled trials. *Journal of Medical Internet Research* 2011; **13**:e26.
- McCambridge J, Kalaitzaki E, White IR, Khadjesari Z, Murray E, Linke S, Thompson SG, Godfrey C, Wallace P. Can differences in the length or relevance of questionnaires impact upon attrition in trials? A randomised controlled trial. *Journal of Medical Internet Research* in press.
- Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in web-based health behavior change programs. *Journal of Medical Internet Research* 2006; **8**:e15.
- Wallace P, Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, Kalaitzaki E, Godfrey C, Linke S. On-line randomized controlled trial of an Internet based interactive intervention for members of the general public with alcohol use disorders. *PLoS ONE* 2011; **6**:e14740.
- Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 1995; **90**:1112–1121.
- Seaman S, White IR. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research* DOI: 10.1177/0962280210395740.
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
- Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; **90**:106–121.
- Brown H, Prescott R. *Applied Mixed Models in Medicine*. Wiley: Chichester, 1999.
- Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons: New York, 1987.
- Schafer JL. *Analysis of Incomplete Multivariate Data*. Chapman and Hall: London, 1997.
- White IR, Wood A, Royston P. Tutorial in biostatistics: multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* 2011; **30**:377–399.

32. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. *Addiction* 2008; **103**:218–227.
33. Meng XL. Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* 1994; **9**:538–558.
34. Rubin DB, Stern HS, Vehovar V. Handling “don’t know” survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association* 1995; **90**:822–828.
35. Potthoff RF, Tudor GE, Pieper KS, Hasselblad V. Can one assess whether missing data are missing at random in medical studies? *Statistical Methods in Medical Research* 2006; **15**:213–234.
36. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; **88**:125–134.
37. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomised controlled trials. *Clinical Trials* 2007; **4**:125–139.
38. Alho JM. Adjusting for non-response bias using logistic regression. *Biometrika* 1990; **77**:617–624.
39. Wood A, White IR, Hotopf M. Using number of failed contact attempts to adjust for non-ignorable non-response. *Journal of the Royal Statistical Society: Series A* 2006; **169**:525–542.
40. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999; **18**:681–694.
41. Royston P. Multiple imputation of missing values. *Stata Journal* 2004; **4**:227–241.
42. Royston P. Multiple imputation of missing values: update. *Stata Journal* 2005; **5**:188–201.
43. Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Statistics in Medicine* 1991; **10**:45–52.
44. Fischer-Lapp K, Goetghebeur E. Practical properties of some structural mean analyses of the effect of compliance in randomized trials. *Controlled Clinical Trials* 1999; **20**:531–546.
45. Fischer K, Goetghebeur E, Vrijens B, White IR. A structural mean model to analyze the effect of compliance when comparing two active treatments. *Biostatistics* 2011; **12**:247–257.
46. Bowden RJ, Turkington DA. *Instrumental Variables*. Cambridge University Press: Cambridge, 1990.
47. Bond SJ, White IR, Walker AS. Instrumental variables and interactions in the causal analysis of a complex clinical trial. *Statistics in Medicine* 2007; **26**:1473–1496.
48. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**:444–455.
49. Efron B, Feldman D. Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association* 1991; **86**:9–17.
50. Albert JM, Demets DL. On a model-based approach to estimating efficacy in clinical trials. *Statistics in Medicine* 1994; **13**:2323–2335.
51. Jin H, Rubin D. Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association* 2008; **103**:101–111.
52. Bond SJ, White IR. Using expert opinion on non-protocol treatments for causal analysis of randomised controlled trials. *Clinical Trials* 2010; **7**:664–676.
53. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research* 2005; **14**(4):369–395.
54. Maracy M, Dunn G. Estimating dose–response effects in psychological treatment trials: the role of instrumental variables. *Statistical Methods in Medical Research* 2008; **00**:1–25.
55. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**:550–560.
56. Hausman JA. Specification tests in econometrics. *Econometrica* 1978; **46**:1251–1271.
57. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Statistics in Medicine* 2000; **19**:1849–1864.
58. White IR, Moodie E, Thompson SG, Croudace T. A modelling strategy for the analysis of clinical trials with partly missing longitudinal data. *International Journal of Methods in Psychiatric Research* 2003; **12**:139–150.
59. Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999; **86**:365–379.
60. Higgins JPT, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials* 2008; **5**:225–239.
61. White IR, Wood AM, Higgins JPT. Allowing for uncertainty due to missing data in meta-analysis—Part 1: two-stage methods. *Statistics in Medicine* 2008; **27**:711–727.
62. Vansteelandt S, Goetghebeur E. Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society: Series B* 2003; **65**:817–835.
63. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics – Theory and Methods* 1991; **20**(8):2609–2631.
64. White IR, Walker S, Babiker AG, Darbyshire JH. Impact of treatment changes on the interpretation of the Concorde trial. *AIDS* 1997; **11**:999–1006.
65. Loeyts T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics* 2003; **59**:100–105.
66. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomised trials. *Statistics in Medicine* 2005; **24**:993–1007.
67. Royston P. Multiple imputation of missing values: update. *Stata Journal* 2005; **5**:527–536.
68. Carlin JB, Galati JC, Royston P. A new framework for managing and analysing multiply imputed data sets in Stata. *Stata Journal* 2008; **8**:49–67.