Journal of Clinical Epidemiology 182 (2025) 111753

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

# Exploring the effect of COVID-19 restrictions on the social functioning scale in a clinical trial of antipsychotic reduction: using multiple imputation to target a hypothetical estimand

Louise Marston[a,b,*], Joanna Moncrieff[c,d], Stefan Priebe[e], Suzie Cro[f], Victoria R. Cornelius[f]

[a]Department of Primary Care and Population Health, University College London, London NW3 2PF, UK
[b]Priment Clinical Trials Unit, University College London, London, NW3 2PF, UK
[c]Division of Psychiatry, University College London, London, W1T 7NF, UK
[d]Centre for Mental Health Research, North East London Foundation NHS Trust, London, UK
[e]Centre for Mental Health Research, City, University of London, London, UK
[f]Imperial Clinical Trials Unit, Imperial College London, London, UK

Accepted 28 February 2025; Published online 7 March 2025

## Abstract

**Objectives:** Many trials are affected by unforeseen events after recruitment has commenced. The aim of this study is to explore a hypothetical strategy for dealing with an intercurrent event that occurred during trial follow-up; COVID-19 restrictions.

**Study Design and Setting:** Secondary analysis of a randomized controlled trial (RCT) in schizophrenia, comparing antipsychotic reduction vs maintenance medication on the social functioning scale (SFS) score at 12 months' follow-up. A hypothetical analysis strategy was used to estimate the treatment effect in a COVID-19 restriction-free world. Outcome data were set to missing, and multiple imputation was used to replace values affected by COVID-19.

**Results:** The trial randomized 253 participants, 187 participants had an SFS score at 12 months, and 75 of those were collected during COVID-19 restrictions. In the original complete case regression analysis, targeting a treatment policy estimand, the treatment effect was estimated to be 0.51 (95% CI −1.33, 2.35) points higher in the reduction group. After multiple imputation, targeting the hypothetical estimand, the mean SFS score was −3.01 (95% CI −7.22, 1.20) points lower in the reduction group, but varied with different assumptions about the timing of events and in sensitivity analyses to increase the size of difference between randomized groups.

**Conclusion:** We demonstrated how the intervention effect can change when estimating the intervention effect in a pandemic world (treatment policy estimand) vs a pandemic restriction-free world (hypothetical estimand), and that estimates are sensitive to imputation and input assumptions. Trialists should be aware of potential intercurrent events and plan the analysis to take them into account. 

*Keywords:* Missing data; COVID-19; Estimands; Social functioning; Multiple imputation; Randomized controlled trial

## 1. Introduction

During the COVID-19 pandemic, ongoing randomized controlled trials (RCTs) were disrupted in many ways. For example, some had to stop recruiting, while others could continue follow-up but were unable to do this in person, instead of moving to online or telephone contact and data collection. This overnight step change in the way trials were conducted had a profound impact on ongoing trials. For affected trials to provide useful and reliable estimates of the intervention effect, many statistical challenges needed to be overcome [1,2].

Funding: RADAR-independent research funded by the National Institute for Health Research (Programme Grants for Applied Research, Research into Antipsychotic Discontinuation And Reduction (RADAR), RP-PG-0514-20,004). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health and Social Care. SC is funded by an NIHR advanced research fellowship (NIHR300593). There was no specific funding for this paper.

* Corresponding author. Department of Primary Care and Population Health, University College London, Rowland Hill Street, London NW3 2PF, UK.

*E-mail address:* l.marston@ucl.ac.uk (L. Marston).

https://doi.org/10.1016/j.jclinepi.2025.111753

**Plain Language Summary**

Many medical research studies that enable us to find out how well things work had to change due to COVID-19 restrictions. This may have altered the results. We used data from a randomized controlled trial (RCT) to examine whether there was evidence for this. The main outcome included questions on how often participants went to the cinema, swimming, to church or saw friends or relatives. Many of these activities were not possible during COVID-19 restrictions and became possible again over time. We used statistical methods to replace data that were collected during COVID-19 restrictions with the best possible estimate if COVID-19 had not happened. We found that the trial results were likely to have been different if the effect of COVID-19 restrictions were taken away. It is likely that most studies will be interested in results that do not include data collected during COVID-19.

There was some guidance regarding mitigating the effects of COVID-19 to ongoing trials from regulatory bodies, including the Medicines and Healthcare Products Regulatory Agency (MHRA), European Medicines Agency (EMA), and Food and Drug Administration (FDA) [3–5]. Their main message was to collect as much data as possible and develop an analysis plan for mitigating any problems before database lock. These were practical guidance but did not set out how such trials should be analyzed.

The motivating example for this paper is Research into Antipsychotic Discontinuation and Reduction (RADAR) [6,7], an RCT of people with schizophrenia who were assigned to either reduction or maintenance of antipsychotics over 2 years' follow-up. This trial completed recruitment shortly before COVID-19 restrictions in England (16/03/2020), but the follow-up continued remotely under COVID-19 restrictions, with some participants returning to in-person appointments as restrictions eased.

The trial's primary outcome was the social functioning scale (SFS), a 79-item, seven-section measure (social engagement/withdrawal, interpersonal communication, independence − performance, recreation, prosocial, independence − competence, occupation or employment), which asks about the previous 3 months [8] (Table A1). This trial aimed to identify whether there was a difference in changes in social functioning between randomized groups measured at 6, 12, and 24 months, with primary endpoint at 24 months. However, many SFS items were not relevant during lockdown, as many were related to activities that were impossible to do [1]. For example, the independence−performance section asks about the frequency of using the bus or trains in the previous 3 months, and the prosocial section enquires about activities such as going to the theater, cinema, pub, playing sports, or church activities − all of which were not permitted during restrictions and only slowly reopened as restrictions eased (Appendix A).

The aim of this paper is to compare the trial's original planned treatment estimand and analysis approach, which ignored COVID-19, to a hypothetical estimand which removed the impact of COVID-19 restrictions.

## 2. Methods

### 2.1. The RADAR trial

The trial design is described fully in the protocol paper and main results paper [6,7]. Briefly, RADAR was a two-arm, 1:1 individually RCT examining antipsychotic reduction vs antipsychotic maintenance. The trial included adults aged over 18 years with schizophrenia, schizoaffective disorder, delusional disorder, or nonaffective psychosis, who had at least one previous psychotic episode or a single episode lasting at least a year and were being prescribed antipsychotic medication.

### 2.2. Estimands

The planned primary analysis for RADAR, as detailed in the statistical analysis plan, was written after the pandemic began, with the estimand aligning with the primary analysis not stated. The original approach taken was to use all data for the main analysis without mitigation for the COVID-19 restrictions [9]. Although not stated explicitly in RADAR, this aligns with a treatment policy strategy for handling the intercurrent event of COVID-19 restrictions by ignoring the occurrence of COVID-19, that is, addressed the treatment effect regardless of COVID-19 restrictions [10,11]. A sensitivity analysis, which used a three-level categorical variable in a linear regression model, was originally used to take account of the national level of restrictions (no COVID-19 restriction; partial restriction− nonessential shops being open; full restrictions− nonessential shops being closed) when the data were collected (Appendix B). A second sensitivity analysis where the model included an interaction between the COVID-19 restriction variable and randomized group was conducted. These mitigations did not alter the differences between the randomized groups [6].

However, alternative methods have been proposed to explore the effect of COVID-19 on an outcome using the estimand framework given in The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)-E9 (R1), as described by Cro et al [9,12]. The aim of this research is to explore these methods using the data collected at 12 months from RADAR. The SFS data collected at 12 months was used as

**What is new?**

**Key findings**

- The intervention effect can change when estimating it in a pandemic world (treatment policy estimand) compared to a pandemic-free world (hypothetical estimand).
- Estimates are sensitive to imputation techniques and assumptions made; both should be considered in detail before an analysis is carried out.

**What this adds to what is known?**

- Intercurrent events may not be observed at the beginning of a trial. The estimand framework enables the research question to be reframed once new intercurrent events become clear.
- Changes in intercurrent events are not unique to pandemics. They may occur due to other events that disrupt studies while in progress; for example, flooding or earthquakes. The methods used in this paper could be used in those situations too.

**What is the implication and what should change now?**

- Trialists should be aware of potential intercurrent events and their potential implications as early as possible, adding sensitivity analyses to the statistical analysis plan to explore the effects of intercurrent events.

the outcome in this paper, as a substantial portion was collected during COVID-19 restrictions. Cro et al explain how a hypothetical strategy can be employed to handle the intercurrent event of COVID-19 to understand the treatment effect in the absence of COVID-19 restrictions. A hypothetical strategy is one that assumes that the intercurrent event has not happened [9]. We decided on the additional estimand we were aiming to achieve with this analysis, with respect to the study population, treatment condition of interest, the outcome to answer the research question, how to account for intercurrent events, and the population-level summary to give evidence of a difference between the treatment groups (Fig). Specifically, we targeted the hypothetical estimand of the mean difference in SFS at 12 months for reduction compared to maintenance of antipsychotics for the population of eligible participants recruited to RADAR in a COVID-19 restriction-free world.

### 2.3. Statistical methods

Details on how to calculate the SFS score and how missing items were handled are in Appendix C.

For all analyses, those who died before their 12-month follow-up was due were omitted from analysis. This aligned with the analysis of the trial [6] and the principal stratum intercurrent event strategy [10], which target the intervention effect for only those who would not experience the intercurrent event of death, assuming death was independent of randomized treatment [11].

Initially summary statistics for the SFS at 12 months, including mean and standard deviation (SD), were calculated by the level of restrictions. We performed linear regression analysis, with robust standard errors, including the randomization variable and baseline SFS score in the model. This analysis is the same as the sensitivity analysis in Moncrieff et al [6] (shown in the supplementary material) using the treatment policy estimand. This was repeated including the COVID-19 restriction variable at 12 months.

To estimate the hypothetical estimand, SFS scores at 6 and 12 months collected during any COVID-19 restrictions were set to missing [13]. We then first fitted a similar model to the main model of interest without the data that were collected when there were any COVID-19 restrictions. This analysis assumed that the missing outcomes of those individuals set missing due to COVID-19 restrictions were similar to those with the same value of covariates included in the analysis model who were observed during times without COVID-19 restrictions (ie, missing at random), hence suitable for estimating the hypothetical estimand of interest. We tested which variables were related to the outcome and missingness; more details in Appendix D.

Next, we used multiple imputation by chained equations [14] as an alternative analytical method to estimate the hypothetical estimand in this study. This is suitable for handling the hypothetical intercurrent event because multiple imputation assumes those whose SFS scores were set to missing were similar to those participants with the same covariates included in the imputation model (missing at random). We included SFS at baseline, which would be part of the substantive models, factors associated with missingness of the outcome, and factors associated with the outcome in the imputation model. Additionally, a small number of factors related to demographics and psychiatric history were included in the imputation model. We did not include the dichotomous variable related to COVID-19 timing as we did not want that to influence imputed values. We imputed the two randomized groups separately and appended them together before analysis. For imputation, we utilized linear regression for continuous variables, ordered logistic regression for educational attainment, and alcohol consumption and Poisson regression for the number of mental health admissions at baseline. We could not include drug use (yes or no) or employment (yes or no) in the final imputation model, despite them being predictors of missingness of the outcome, because the imputation models would not converge with them included, even after utilizing bootstrap methods or augmenting the data as suggested by White et al [15]. We chose 100 multiple

| Description in relation to RADAR original main analysis | **Estimand** | Description in relation to RADAR after COVID-19 |
|---|---|---|
| People with schizophrenia or other psychosis who meet the RADAR eligibility criteria | **Population** | People with schizophrenia or other psychosis who meet the RADAR eligibility criteria |
| Discontinuation or maintenance of antipsychotics | **Treatment condition** | Discontinuation or maintenance of antipsychotics |
| Social Functioning Scale at 12 months | **Outcome** | Social Functioning Scale at 12 months |
| Covid pandemic - Treatment policy<br><br>Death – Principal stratum | **Intercurrent event- strategy** | Covid pandemic – hypothetical policy (pandemic free world)<br><br>Death – Principal stratum |
| Difference in SFS means between randomised groups | **Population Level summary measure** | Difference in SFS means between randomised groups |

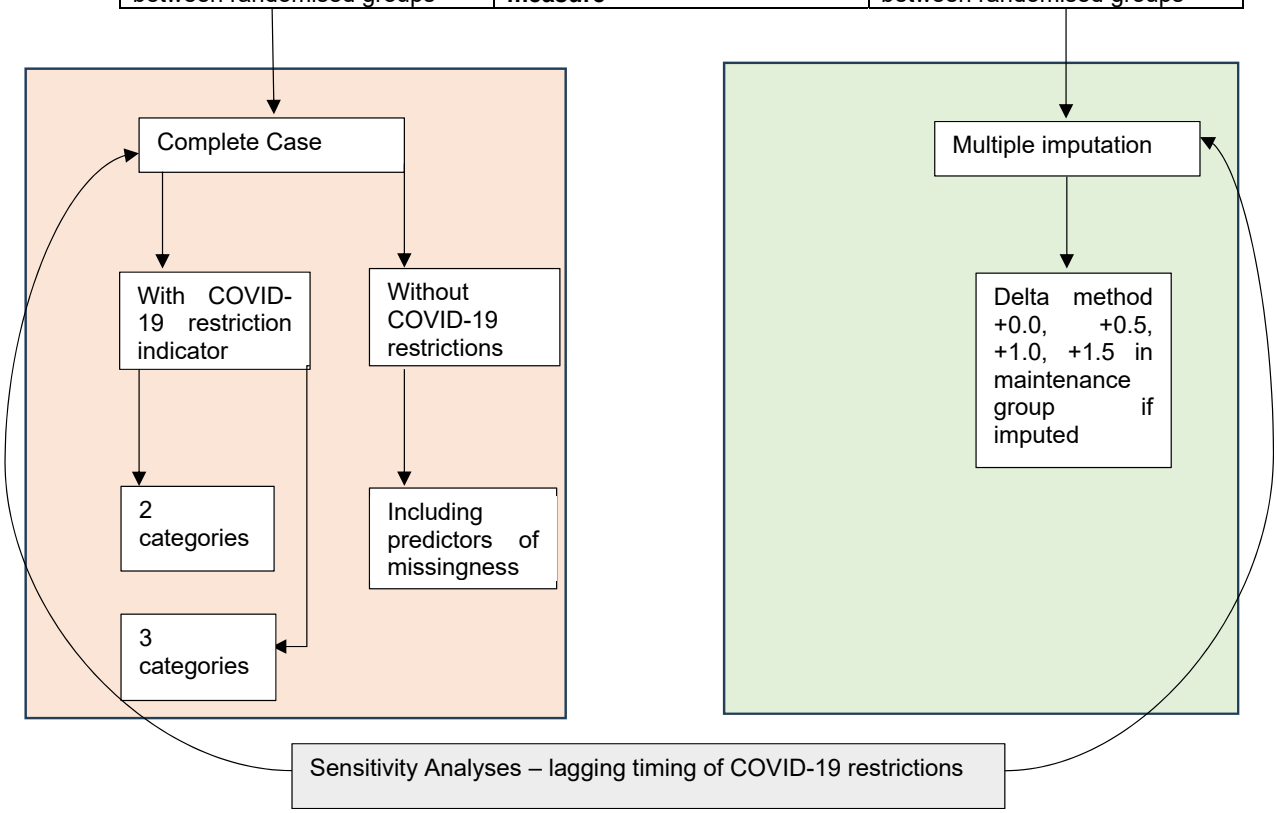


**Figure.** Flow diagram of estimands and analyses undertaken.

imputations for the main and sensitivity imputation models, because this was above the number needed to satisfy the rule of thumb of the Monte Carlo error of an estimate being less than or equal to 10% of its standard error [14]. We did not include data from those who died before a 12-month follow-up, as it would be impossible for them to have 12-month follow-up data. Data missing for other reasons (all data at 12 months missing before, during, or after COVID-19 restrictions and missing baseline data) were similarly imputed. The same linear regression model as the complete case models was used to analyze each imputed dataset, and Rubin's rules [16] were used to obtain the overall treatment effect (95% CI) across imputed datasets. Sensitivity analyses explored the impact of varying restriction start dates for both estimands, and for the hypothetical estimand, the impact of a better outcome than predicted under multiple imputation using a tipping point approach to vary delta until the significance was changed (Appendix E).

The results of the predictors of missingness and association with the outcome analyses for the main and sensitivity analyses can be seen in Tables A2-A3 and A5-A6 in the appendix. Descriptive statistics for complete case and after both multiple imputations are in Table A7 in the appendix.

All analyses were carried out using Stata version 17.

## 3. Results

At 12 months, 193/253 participants provided responses to some of the SFS items, 187 provided sufficient data to calculate an overall SFS score, 112 (60%) of these were when there were no COVID-19 restrictions, 40 (21%) when there were partial restrictions, and 35 (19%) when there were full restrictions (75 (40%) with the data collected at 12 months under any restrictions). The only difference in baseline characteristics between those who did ($n = 62$) and did not have missing SFS data ($n = 187$) at 12 months was educational attainment; a greater percentage of those who did not have missing data had tertiary education compared with those who had missing SFS at 12 months (42% vs 27%, respectively) (Table A8 in appendix). Four participants died before 12 months, three in the reduction group and one in the maintenance group, giving 249 participants who could have provided data at 12 months. Baseline characteristics by COVID-19 restrictions at 12 months are shown in Table 1, and mean (SD) SFS scores under the scenarios used in this paper are in Table 2.

For the hypothetical estimand, when only including data when there were no restrictions, the result changed with a lower score in the reduction group although it remained not significant. After multiple imputation, which included additional variables to predict the data affected by COVID-19 for the hypothetical estimand, there was also a lower mean SFS score for those in the reduction group in comparison to the maintenance group (Table 3).

Under sensitivity analysis conditions altering the COVID-19 restriction start and end dates, there were fewer 12-month SFS scores potentially affected by COVID-19, with 154 SFS completed without COVID-19 restrictions and 33 completed with restrictions. There was a greater percentage of male participants providing 12-month SFS data during restrictions (77%) than not during restrictions (63%), and the mean age was higher in the group providing 12-month SFS data during restrictions (Table A4). At 12 months, those affected by COVID-19 restrictions in the reduction group had a mean SFS score of 106 (SD 8) and 103 (SD 8) in the maintenance group. There were a number of factors associated with 12-month SFS and missingness of the 12-month SFS (Tables A5-A6). Including predictors of missingness in modeling using the treatment policy estimand gave a lower SFS in the reduction than maintenance group. This was further lowered after multiple imputation using the hypothetical estimand.

Both sets of delta-based imputation models required an addition of 1.5 to the imputed SFS score in the maintenance group to reach the tipping point of the coefficient for the randomized group to be statistically significant (Table 3).

**Table 1.** Baseline descriptive statistics by COVID-19 restrictions at 12 months

| | No restrictions | | Any restrictions | |
|---|---|---|---|---|
| **Characteristic** | **n/N** | **%** | **n/N** | **%** |
| Male | 69/112 | 62 | 53/75 | 71 |
| Female | 43/112 | 38 | 19/75 | 25 |
| Transgender | 0/112 | 0 | 3/75 | 4 |
| Age mean (SD) | 47 | (11) | 47 | (12) |
| White | 69/111 | 62 | 57/75 | 76 |
| Black | 26/111 | 23 | 12/75 | 16 |
| Asian | 9/111 | 8 | 4/75 | 5 |
| Other | 7/111 | 6 | 2/75 | 3 |
| SFS score mean (SD) | 109 | (9) | 107 | (10) |

SFS, social functioning scale.

## 4. Discussion

This paper documents the potential problems of using the social function scale during COVID-19 restrictions. The measure contained some items that were inaccessible during the pandemic, which could invalidate SFS as a measure. We found that scores were lower during the pandemic, which supports this. We suggest that the measure is not an ideal outcome to use during a pandemic, but as the primary outcome of a trial cannot be changed, it would be important to undertake adjustments for this in the analysis such as the ones we demonstrate in this paper. Otherwise, the overall score would be too low, risking failure to show a difference between the randomized groups at the outcome time point. The results of modeling the data using the treatment policy estimand, ignoring COVID-19 restrictions, show that social functioning is higher in the reduction group.

In contrast, the hypothetical estimand results that considered if COVID-19 restrictions did not happen were in favor of the maintenance group. This suggests that COVID-19 had an impact on the treatment effect on the SFS score. However, for both estimands the estimated treatment effects were nonsignificant, indicating that this impact was not large enough to change the overall original conclusions of the study. [6]. The SFS has not been used by other published studies that have collected data during the COVID-19 restrictions of 2020 and 2021, so it is difficult to see whether the patterns shown in the data are unique to RADAR.

It may not be reasonable to assume that the effect of the restrictions was immediate on participants. The results of the sensitivity analysis to examine the effect of a delayed impact of the restrictions showed a smaller difference between the randomized groups. This is likely to be because the COVID-19 missing data were for a smaller group of participants who were truly affected by COVID-19. This possibly better represents those affected by COVID-19 than the crude delineation used in the main RADAR paper and the initial analyses in this paper. The delta-based multiple imputation sensitivity analysis widened the gap in SFS

**Table 2.** Descriptive statistics for SFS scores at 12 months under a number of scenarios related to COVID-19 by randomized group

| | Reduction group | Maintenance group |
|---|---|---|
| **Scenario** | **Mean (SD)** | **Mean(SD)** |
| Using all available data ($n$ = 187) | 106 (9) | 107 (10) |
| No restrictions ($n$ = 112) | 106 (9) | 108 (11) |
| Any restrictions ($n$ = 75) | 106 (8) | 105 (10) |
| Hypothetical scenario with data imputed (100 imputations) ($n$ = 248) randomized groups imputed separately | 105 (14)[a] | 108 (14)[a] |
| Sensitivity (MI δ = 0.5) | 105 (14)[a] | 109 (14)[a] |
| Sensitivity (MI δ = 1.0) | 105 (14)[a] | 109 (14)[a] |
| Sensitivity (MI δ = 1.5) | 105 (14)[a] | 110 (14)[a] |
| Sensitivity analyses | | |
| No restrictions ($n$ = 154) | 107 (10) | 108 (11) |
| Any restrictions ($n$ = 33) | 106 (8) | 103 (8) |
| Hypothetical scenario with data imputed (100 imputations) ($n$ = 249) randomized groups imputed separately | 106 (10)[a] | 107 (11)[a] |
| Sensitivity (MI δ = 0.5) | 106 (10)[a] | 108 (11)[a] |
| Sensitivity (MI δ = 1.0) | 106 (10)[a] | 108 (11)[a] |
| Sensitivity (MI δ = 1.5) | 106 (10)[a] | 109 (11)[a] |

SFS, social functioning scale; MI, multiple imputation.

Using data from participants who provided sufficient data to provide a score (complete case, treatment policy estimand) the adjusted treatment effect estimate was a nonsignificant higher score in the reduction group in comparison to the maintenance group (0.51; 95%CI -1.33, 2.35). This was attenuated slightly when including a categorical variable to indicate level of COVID-19 restrictions.

[a] SD the mean of the SD from each imputed dataset.

score between the randomized groups, exacerbating the trend shown in the multiply imputed result before the delta method was carried out. This sensitivity analysis considered that those in the maintenance group who had missing SFS data (mostly due to COVID-19 restrictions) had even better social functioning than that shown by multiple imputation alone.

The treatment effect estimates for a pandemic restriction-free world was lower than the prespecified clinically important difference in SFS score of four points used to power the trial [17]. However, the 95% confidence interval encompassed a decrease of more than four points in the primary reanalysis in this paper.

### *4.1. Strengths and limitations*

We examined a number of scenarios, which could be considered the best and worst cases based on the timing of when the SFS was completed in relation to COVID-19 restrictions and by considering the hypothetical intercurrent event. This gave a range of possible differences between randomized groups. It is likely that the difference in a pandemic restriction-free world would have been greater than that shown in the trial paper [6] but may not be as large as our maximum difference with imputed SFS under the delta method, and therefore, the estimate may not have reached the clinically important difference of 4, though this is likely to have been included in the 95% CI.

In our approach, we imputed values for the total SFS score. It would have been preferable to impute values for each individual SFS item that had missing data, including those that were set to missing to take account of the COVID-19 restrictions. However, this was not possible, because this made the imputation models too complex, and it would not converge. Using simpler imputation models may mean that we lost some accuracy in our estimates for the SFS in a pandemic-free world.

Multiple imputations make a strong assumption that the missing data are missing at random. While this seems to be a reasonable assumption in this trial for participants whose SFS data were set to missing, because of being collected during COVID-19 restrictions, this may not hold for those who did not provide data at other times during the trial. However, this is a common assumption used in many trials and will have been partially mitigated by including predictors of missingness. For participants who should have been followed up during COVID-19 restrictions but were not, it is not clear whether this was because of COVID-19 — directly or indirectly — or another reason unrelated to COVID-19, and therefore whether their data were likely to have been missing at random or due to another missing data mechanism.

**Table 3.** Substantive model in terms of the reduction group under different data assumptions

| Model | Treatment coefficient[a] | 95% CI |
|---|---|---|
| Treatment policy estimand[b] | | |
| Data and methods in the main paper ($N = 180$) [6] | 0.51 | (−1.33, 2.35) |
| Including a categorical variable to indicate the level of COVID-19 restrictions (3 categories) ($N = 180$) | 0.50 | (−1.36, 2.37) |
| Including a categorical variable to indicate the level of COVID-19 restrictions (2 categories) ($N = 180$) | 0.49 | (−1.37, 2.35) |
| Sensitivity analysis similar to the main paper, including a categorical variable to indicate the level of COVID-19 restrictions lagged (2 categories) ($N = 180$) | 0.34 | (−1.53, 2.21) |
| Hypothetical estimand[b] | | |
| Data collected when there were no COVID-19 restrictions (complete case) ($N = 101$) | −0.81 | (−3.45, 1.83) |
| Data collected when there were no COVID-19 restrictions, including predictors of missingness (complete case) ($N = 100$) | −0.82 | (−3.42, 1.79) |
| After multiple imputation, using hypothetical intercurrent events 100 imputations group imputed separately | −3.01 | (−7.22, 1.20) |
| Sensitivity (MI $\delta = 0.5$) ($N = 248$) | −3.43 | (−7.64, 0.79) |
| Sensitivity (MI $\delta = 1.0$) ($N = 248$) | −3.84 | (−8.06, 0.37) |
| Sensitivity (MI $\delta = 1.5$) ($N = 248$) | −4.26 | (−8.48, −0.04) |
| Sensitivity analysis only including data not collected during lagged COVID-19 restrictions (complete case) ($N = 143$) | −0.05 | (−2.19, 2.09) |
| Sensitivity analysis only including data not collected during lagged COVID-19 restrictions, including predictors of missingness (complete case) ($N = 140$) | −0.50 | (−2.67, 1.67) |
| After multiple imputation, using hypothetical intercurrent events with changed dates of restriction (100 imputations) groups imputed separately ($N = 248$) | −1.22 | (−3.76, 1.33) |
| Sensitivity (MI $\delta = 0.5$) ($N = 248$) | −1.67 | (−4.22, 0.87) |
| Sensitivity (MI $\delta = 1.0$) ($N = 248$) | −2.13 | (−4.68, 0.41) |
| Sensitivity (MI $\delta = 1.5$) ($N = 248$) | −2.59 | (−5.13, −0.04) |

All imputation analyses used 100 imputations, and imputation for each treatment group was conducted separately; MI, multiple imputation.
[a] In terms of the reduction group.
[b] These strategies for dealing with intercurrent events are answering different research questions.

We were most interested in an estimand in the absence of COVID-19 restrictions, not COVID-19 in its entirety; for example, there was no intercurrent event related to a participant contracting COVID-19 infection and this is beyond the scope of this study. With the delta-based method of multiple imputation, the same adjustment was used for all data imputed in the maintenance arm, which may not have been appropriate for those in the maintenance groups whose SFS score was imputed.

We used a principal stratum strategy to handle death as an intercurrent event. This was the approach used in the main trial. While this may not be optimal, we employed this approach to align with the main paper. As there were only four deaths, it is unlikely that another intercurrent event strategy would have changed the result.

Some trials in the pandemic replaced the participants who were affected by COVID-19. [18]. This was not feasible for RADAR because funding was limited and many participants in RADAR were affected for at least one follow-up over the 2-year follow-up. Mitigating through statistical methods was deemed to be a more feasible solution in these circumstances.

This unexpected intercurrent event, which could not have been fully mitigated against in RADAR because of the nature of the outcome, must be reported correctly; to this end, the CONSERVE statement [19] was constructed in 2020 as an extension to CONSORT and STROBE to ensure that all important details related to a deviation from the protocol because of pandemics or other extenuating circumstances are fully explained.

## 5. Conclusion

This study demonstrates the potential impact of the pandemic on the results of the RADAR trial. We demonstrated how the intervention effect can change considerably when estimating the intervention effect in a pandemic world vs a pandemic restriction-free world (hypothetical

estimand) and that estimates are sensitive to imputation and input assumptions. Trialists should be aware of potential intercurrent events and plan the analysis accordingly to take them into account.

## CRediT authorship contribution statement

**Louise Marston:** Writing − review & editing, Writing − original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Joanna Moncrieff:** Writing − review & editing, Project administration, Investigation, Conceptualization, Funding acquisition. **Stefan Priebe:** Writing − review & editing, Funding acquisition. **Suzie Cro:** Writing − review & editing, Supervision, Methodology. **Victoria R. Cornelius:** Writing − review & editing, Methodology, Conceptualization.

## Ethical approval

Written informed consent was obtained prior to participation in the trial, following a full explanation of the aims, methods, anticipated benefits, and hazards of the trial. An assessment of each participant's capacity to provide consent was completed first. Following that, researchers asked for verbal consent at each follow-up assessment. All trial data were handled according to the UK Data Protection Act 1998 and UK General Data Protection Regulation (UK GDPR). The trial was approved by the Health Research Authority London - Brent Research Ethics Committee (reference 16/LO/1507) on 27/10/2016.

The trial was registered with the International Standard Randomised Controlled Trials register on 07/02/2017 (ISRCTN90298520), with ClinicalTrials.gov on 15/06/2018 (NCT03559426) and EudraCT number 2016-000709-36.

## Declaration of competing interest

L.M. reports financial support was provided by National Institute for Health and Care Research. S.C. reports a relationship with National Institute for Health and Care Research that includes funding grants. There are no competing interests for any other authors.

## Acknowledgments

We would like to thank all participants for taking part in the study.

## Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2025.111753.

## Data availability

Data will be made available on request.

## References

[1] Degtyarev E, Rufibach K, Shentu Y, Yung G, Casey M, Englert S, et al. Assessing the impact of COVID-19 on the clinical trial objective and analysis of oncology clinical trials—application of the estimand framework. Stat Biopharm Res 2020;12:427−37. https://doi.org/10.1080/19466315.2020.1785543.

[2] Meyer RD, Ratitch B, Wolbers M, Marchenko O, Quan H, Li D, et al. Statistical issues and recommendations for clinical trials conducted during the COVID-19 pandemic. Stat Biopharm Res 2020;12:399−411. https://doi.org/10.1080/19466315.2020.1779122.

[3] Medicines and Healthcare Products Regulatory Agency. Guidance on minimising disruptions to the conduct and integrity of clinical trials of medicines during COVID-19. MHRA; 2020:1−8. Available at: https://www.gov.uk/guidance/guidance-on-minimising-disruptions-to-the-conduct-and-integrity-of-clinical-trials-of-medicines-during-covid-19. Accessed July 10, 2024.

[4] Administration USF and D. FDA guidance on conduct of clinical trials of medical Products during COVID-19 pandemic: guidance for industry, investigators, and institutional review boards. Rockville, MD: United States Food and Drug; 2020.

[5] Agency EM. Implications of coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials | European Medicines Agency. 2020. Available at: https://www.ema.europa.eu/en/implications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical-trials. Accessed January 6, 2022.

[6] Moncrieff J, Crellin N, Stansfeld J, Cooper R, Marston L, Freemantle N, et al. Antipsychotic dose reduction and discontinuation versus maintenance treatment in people with schizophrenia and other recurrent psychotic disorders in England (the RADAR trial): an open, parallel-group, randomised controlled trial. Lancet Psychiatry 2023;10:848−59. https://doi.org/10.1016/S2215-0366(23)00258-4.

[7] Moncrieff J, Lewis G, Freemantle N, Johnson S, Barnes TRE, Morant N, et al. Randomised controlled trial of gradual antipsychotic reduction and discontinuation in people with schizophrenia and related disorders: the RADAR trial (Research into Antipsychotic Discontinuation and Reduction). BMJ Open 2019;9:1−8. https://doi.org/10.1136/bmjopen-2019-030912.

[8] Birchwood M, Smith J, Cochrane R, Wetton S, Copestake S. The social functioning Scale the development and validation of a new Scale of social adjustment for use in family intervention programmes with schizophrenic patients. Br J Psychiatry 1990;157:853−9. https://doi.org/10.1192/BJP.157.6.853.

[9] European Medicines Agency. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-and-sensitivity-analysis-clinical-trials-guideline-statistical-principles-clinical-trials-step-5_en.pdf. Accessed July 10, 2024.

[10] Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum. BMJ 2024;384:E076316. https://doi.org/10.1136/BMJ-2023-076316.

[11] Kahan BC, White IR, Edwards M, Harhay MO. Using modified intention-to-treat as a principal stratum estimator for failure to initiate treatment. Clin Trials 2023;20:269−75. https://doi.org/10.1177/17407745231160074.

[12] Cro S, Morris TP, Kahan BC, Cornelius VR, Carpenter JR. A four-step strategy for handling missing outcome data in randomised trials affected by a pandemic. BMC Med Res Methodol 2020;20:1−12. https://doi.org/10.1186/S12874-020-01089-6/FIGURES/2.

[13] Van Lancker K, Tarima S, Bartlett J, Bauer M, Bharani-Dharan B, Bretz F, et al. Estimands and their estimators for clinical trials impacted by the COVID-19 pandemic: a report from the NISS ingram olkin forum series on unplanned clinical trial disruptions. Stat

Biopharm Res 2022;15:94−111. https://doi.org/10.1080/19466315.2022.2094459.

[14] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30:377−99. https://doi.org/10.1002/SIM.4067.

[15] White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. Comput Stat Data Anal 2010;54:2267−75. https://doi.org/10.1016/J.CSDA.2010.04.005.

[16] Rubin DB. Multiple imputation for nonresponse in surveys. New York, NY: Wiley; 1987.

[17] Jaracz K, Górna K, Kiejda J, Grabowska-Fudala B, Jaracz J, Suwalska A, et al. Psychosocial functioning in relation to symptomatic remission: a longitudinal study of first episode schizophrenia. Eur Psychiatry 2015;30:907−13. https://doi.org/10.1016/J.EURPSY.2015.08.001.

[18] Marston L, Le Novere M, Ricciardi F, Nazareth I, Carson A, Edwards M, et al. COVID-19 and the Physio4FMD trial: impact, mitigating strategies and analysis plans. Contemp Clin Trials Commun 2023;33:101124. https://doi.org/10.1016/J.CONCTC.2023.101124.

[19] Orkin AM, Gill PJ, Ghersi D, Campbell L, Sugarman J, Emsley R, et al. Guidelines for reporting trial protocols and completed trials modified due to the COVID-19 pandemic and other extenuating circumstances: the CONSERVE 2021 statement. JAMA 2021;326:257−65. https://doi.org/10.1001/JAMA.2021.9941.