Statistical analysis protocol for an evaluation of COVID Oximetry @ home using a Regression Discontinuity Design

Therese Lloyd, Will Parry

March 2021

The Health Foundation

Background

Overview

Coronavirus disease 2019 (COVID-19) has led to many individuals in England suffering from severe health degradation, complications, and deaths. One issue lies with people presenting to hospital with low oxygen saturation levels, often without accompanying breathlessness (known as silent hypoxia). Delays in escalating and admitting these individuals to hospital can lead to invasive treatment, prolonged hospital stay, and an increased risk of death.

Remote home monitoring models, which aim to remotely monitor COVID-19 patients at risk, have been implemented in several countries, including England, in response to COVID-19. The aim of these models is twofold:

- avoid unnecessary hospital admissions ('appropriate care in the appropriate place')
- escalate cases of health deterioration earlier to avoid invasive ventilation and ICU admission.

The National Health Service (NHS) pilot of remote home monitoring models, which included the use of pulse oximeters that measure a person's blood oxygen saturation levels, was launched in eight settings in England during the first wave of the pandemic. The pilots encompassed various models, with differing settings (eg primary care providing pre-hospital support or secondary care providing post-discharge support) and different mechanisms for triage and monitoring.[1]

National roll-out of CO@h

The national roll-out in England of the COVID Oximetry @home (CO@h) programme, consisting of predominately pre-hospital remote monitoring of oxygen saturation levels of people diagnosed with COVID-19 and at risk of health deterioration due to silent hypoxia, was launched in November 2020. The plan was for all Clinical Commissioning Groups (CCGs) to have implemented the programme by the end of December 2020. However, in practice the introduction of the service was staggered over time and geography, with some areas having implemented the model earlier but other CCGs reporting a delay in providing the programme across the population under their remit. As CO@h is not a mandatory intervention it may also not have been rolled out across a whole CCG.

According to the CO@h standard operating procedures (SOP),[2] individuals eligible for onboarding (ie enrolment in the programme) are those diagnosed with COVID-19, either clinically or via a positive test result, who are symptomatic and (i) either 65 years of age or over, or (ii) clinically extremely vulnerable (CEV) to COVID-19. The model of CO@h delivery is primarily implemented in primary care; referrals may additionally be received from the NHS111 COVID Clinical Assessment Service, NHS Test and Trace, and Accident & Emergency (A&E) hospital departments. Patients are encouraged to record three oximeter readings a day; patients should receive text or email prompts or check-in calls to confirm that they are using

the oximeter correctly and that the readings are 95% peripheral oxygen saturation (SpO2) or above. If patients do not show signs of health deterioration within 14 days of onset of symptoms, they are discharged from the programme.

It is estimated that 4% of the population in England are CEV,[3] and that approximately 2.4% of the population are CEV and 65 years of age or under. Patients are normally added to the CEV COVID-19 list through an automated process, based on information from patient records. However, GPs are required to periodically review the list to identify registered patients to be added to or removed from it and hospital specialists may also identify additional patients at high risk of health complications from COVID-19. Therefore, clinical judgement may also influence the process, bringing into consideration multiple additional COVID-19 risk factors.

CO@h is also available to care home residents. Training and support for using pulse oximetry is available and there is a CO@h diary (for recording oxygen saturation levels) that has been tailored for care home usage.

At the beginning of February 2021, the CO@h Programme Board recommended to the National Incidence and Response Board (NIRB) to extend the age criteria for CO@h to people aged 50 and over and place more emphasis on clinical judgement. Anecdotally, some areas were already doing this.

Other related interventions

Vaccinations

The roll-out of COVID-19 vaccinations in England started on 8 December 2020. The order of priority for vaccination is, as of 11 February 2021 (with the four highest priority groups being offered a first vaccination by mid-February 2021):[4,5]

1. Residents in a care home for older adults and their carers
2. Individuals 80 years of age or over and front-line health and social care workers
3. Individuals 75 years of age or over
4. Individuals 70 years of age or over and CEV individuals
5. Individuals 65 years of age or over
6. Individuals 16 years to 64 years of age with underlying health conditions which put them at higher risk of serious complications and mortality
7. Individuals 60 years of age or over
8. Individuals 55 years of age or over
9. Individuals 50 years of age or over

COVID virtual wards

COVID virtual wards (CVWs) is a national programme that aims for early supported discharge from hospital by providing additional support at home (including, but not limited to, pulse oximeters) to patients hospitalized due to COVID-19.[6] CVWs is a complementary but separate programme to CO@h, predominately led from a hospital setting and targeting patients with higher acuity or complex conditions than the CO@h programme.

The CVWs programme was introduced at the end of December 2020 and is subsequently being rolled out across the country. It does not fall within the scope of the present evaluation.

PRINCIPLE trial

The PRINCIPLE trial is a national clinical trial to find effective COVID-19 drug treatments that can be taken at home; for example, one treatment included is a commonly used inhaled corticosteroid.[7] The eligibility criteria are similar to CO@h: patients are enrolled if they are either symptomatic of or testing positive for COVID-19 within the previous 14 days, are aged over 65, or aged over 50 and have certain underlying health conditions (which overlap but differ from the CEV list). Enrolment to PRINCIPLE started on 17 April 2020; as of 11 February 2021, 4,053 patients had been recruited to the trial.

Aim of the evaluation

The IAU will carry out two separate evaluations to quantify the impact of the roll-out of CO@h in England during the second wave of the COVID-19 pandemic on hospital activity and mortality. One evaluation will explore the impact of CO@h at the CCG level; the other is a patient-level analysis that will take advantage of the 65 years of age criterion for eligibility to estimate treatment effects. The latter evaluation is the subject of the remainder of this document.

Methods

Study design

This study evaluates the COVID oximetry @home (CO@h) programme, which is predominantly a community-based initiative that aims to support patients who are COVID-19 positive pre-hospital (though some deployments of CO@h are being run by acute trusts). The current study will focus primarily on the pre-hospital use of oximeters by limiting the evaluation to sites that have implemented the 'standard' CO@h model, as detailed in the SOP.[2] The sites will be identified either at CCG level or, if we have access to data at a more granular level, areas within CCGs implementing the standard CO@h model.

This evaluation will take advantage of the 65-year age threshold for eligibility for CO@h for patients with a confirmed COVID-19 diagnosis, by using a Regression Discontinuity Design (RDD).[8,9] RDDs are quasi-experimental approaches to causal inference and are appropriate in situations where eligibility for an intervention changes sharply at a predefined threshold of a 'running' variable (in this case, age). This design largely avoids problems of observed

or unobserved confounding, as the expected outcomes of patients just below and above the threshold are assumed to be equivalent in the absence of intervention. In order to defend this assumption, baseline characteristics can be compared to see whether they are comparable either side of the threshold, suggesting that the outcomes should also be comparable were there to be no effect from the intervention. Under these circumstances we can therefore attribute any difference in outcomes at the threshold to receipt of the intervention.

As clinically extremely vulnerable (CEV) patients are eligible for enrolling in CO@h (the 'intervention'), whether under or over 65 years of age, these patients will be excluded from the analysis.

For patients who have tested positive for COVID-19 and are not CEV, we will evaluate the effect of CO@h on outcomes relating to hospital use and mortality in the 28 days following a COVID-19 positive test by comparing patients aged 65 or just over with those who are not quite 65 years old. The risk of severe complications from COVID-19 is highest in the first 28 days and this time period is consistent with Public Health England's definition of COVID-19 death.[10,11]

There are likely to be patients who have tested positive for COVID-19 who are 'non-compliers', ie they did not receive the intervention even though they were aged 65 or over, or they did receive it even though they were under 65 years of age. We will (if it is possible to do so with the data available) ascertain whether this is the case, and if so, will employ a 'fuzzy' RDD. This method allows for treatment receipt to differ from treatment assignment in the 'window' surrounding the age threshold, as long as the probability of receiving treatment still changes abruptly at the age 65 threshold. The analysis will therefore estimate a treatment effect for compliers who are not CEV in a range or 'window' of ages around the 65-year threshold.

The evaluation study period – ie the calendar period over which we will be evaluating CO@h – can differ between CCGs, with each CCG only contributing data to this analysis once (a) the intervention is fully operational within the CCG, and (b) data on which patients have been enrolled in CO@h (onboarding data) are available. Due to the COVID-19 crisis, the resulting pressures within health care services, and the ongoing work to improve support to both patients and health services, the environment within which we are planning this study is continuously changing. In particular, there are potential changes to the eligibility criteria (the lowering of the age criteria to patients over 50 years old, and the placing of more emphasis on clinical judgement) and other interventions with similar target groups (such as the PRINCIPLE trial) occurring in tandem. Furthermore, there are some potential issues with accessing complete and reliable onboarding data. If we cannot access complete and reliable onboarding data, we may not be able to reliably identify patients who were not onboarded (which is necessary for the proposed analysis). To mitigate this, we will if necessary – and if possible – make some pragmatic decisions, for example relating to the study period and/or which areas to include in the study, once we have received the data.

At the time of writing, it is unclear what volume of patients have been referred to the CO@h pathway, and what onboarding data will be available due to pressure on the system to cope with the pandemic. As a result, the study may be underpowered. This risk is further exacerbated by the fact that the primary outcomes relate to use of critical care beds and mortality. The mixed-methods study of pilots of remote home monitoring models in England

found that for the pre-hospital model of CO@h, although 10% (174/1,737) of patients that were monitored deteriorated and were escalated, 0.2% (3/1,737) of patients were admitted to ICU and 1% (20/1,737) died.[1] See the Power calculation section for further details

All decisions taken after this statistical analysis protocol (SAP) has been finalised will be documented and justified in an addendum to it. We may run interim analyses on incomplete data to provide the CO@h programme team with some preliminary indications of results. The results of the final analysis will be published in line with the Health Foundation's Improvement Analytic Unit's (IAU's) aim of openness and transparency.

Study outcomes

We will analyse outcomes in the 28-day period following a COVID-19 positive test (ie equal to or less than 28 days from a positive test, with the date of the first COVID-positive test being equal to day zero). We are limiting most of the outcomes to binary variables (ie capturing whether there was an A&E attendance rather than the number of attendances); this is because of the short follow-up period and likely sparsity in hospital event counts other than 0 and 1 (which we will confirm on receipt of the data). The following outcomes will be included:

- Type 1 A&E attendance (binary outcome yes/no)
- Emergency admission (binary outcome yes/no)
- Critical care bed use (binary outcome yes/no)
  - This will capture a stay in ICU at any point during an admission, which could be as a result of either an emergency or an elective admission.
  - This variable includes level 2 and level 3 beds, as well as beds under code 5 (flexible critical care beds where there is a mix of level 2 and level 3 beds) and code 90 (temporary use of non-critical care beds).[12]
- Total emergency hospital bed days
  - This includes all hospital bed days during emergency admissions in the 28 days from first COVID-19 positive test. If the patient is still in hospital on day 28 then bed days will be truncated on day 28. Except for same-day admissions, bed days are counted as number of nights in hospital. Thus, an admission on day one and discharge on day three would equate to two bed days. Same day admission-discharges will be treated as a half-day.
  - We will evaluate total hospital bed days as opposed to average length of stay (LOS) in order to better capture the hospital resources required to care for a patient. For example, a patient's single stay of 3 days or two stays of 3 days would both result in an average LOS of 3 days, whereas the total number of hospital bed days would be 3 and 6 days, respectively.

- Total hospital bed days in critical care
  - This includes all bed days in critical care (see definition above) in the 28 days from first COVID-positive test. If the patient is still in a critical care bed on day 28 then bed days will be truncated on day 28. Same-day admission-discharges will be treated as a half-day.
- Death (binary yes/no)
  - This will include deaths recorded within the 28-day follow-up period. This is consistent with the definition of COVID-19 deaths used for reporting by Public Health England.

Hypothesised outcomes

The primary outcomes for this study are whether a patient was allocated a critical care bed, total hospital bed days in critical care, and death. The CO@h programme's logic model hypothesises that early escalation of deteriorating patients will result in fewer patients needing critical care, less time spent in critical care, and fewer deaths. It may be, however, that a decrease in mortality results in longer stays in hospital, on average, due to more patients surviving while still becoming severely ill.

It is hypothesised that A&E attendances may go down if patients who do not need urgent care are reassured by their pulse oximetry readings and the remote support received from CO@h clinical teams. They may also go down if CO@h clinical teams are admitting patients straight to wards, bypassing A&E (although this is not expected to be a common occurrence). A&E attendances may also increase if pulse oximeters are giving inaccurate readings,[13,14] leading to unnecessary attendances. We do not anticipate a marked difference in emergency admissions – patients on CO@h may be admitted earlier, but we would expect all patients requiring emergency admission to be admitted within the 28-day follow-up period from first positive test. If this is not the case, we may find that treated patients are more likely to experience an emergency admission. If some patients are admitted directly as a result of an inaccurate reading, this could also lead to an increase in emergency admissions. Total hospital bed days may decrease thanks to earlier escalation but could also increase if more patients survive but remain critically ill. CVWs, which is being rolled out nationally in January 2021, could also affect hospital bed days.

Sources of data

NHS Digital will provide pseudonymised, patient-level linked data from several sources. Pseudonymisation is where data sets are stripped of the personal information used for identification, such as name, full date of birth and address. A unique person identifier (such as an NHS number) is used but replaced with a consistent, randomised identifier. This identifier is used to link data sources and hospital records for the same person over time.

To further limit the patient information sourced for this study, our data set will be restricted to patients between 50–79 years of age (up to the day before their 80th birthday) at the time of their first COVID-positive test.*

The data sources used will include:

- **General Practice Extraction Service (GPES) Data for Pandemic Planning and Research (GDPPR)** – this data set comprises general practice data on patients registered in 97.5% of all practices in England.[15] This will be the source of data on baseline characteristics for the patients included in our analysis. The age of patients (in whole weeks relative to a reference date) will be calculated by NHS Digital using this data set. From this, we will derive the age at first COVID-19 positive test (eg 65 years and 3 weeks). We will request data for patients born between 1 October 1940 (ie turning 80 on 1 October 2020, the earliest potential study start date) and 1 May 1971 (ie turning 50 on 1 May 2021).
- **Second Generation Surveillance System (SGSS)** – this system holds data on COVID-positive tests, including the full date of first COVID-positive test. SGSS is the national laboratory reporting system used in England to capture routine laboratory data (mainly on infectious diseases and antimicrobial resistance). Only the first COVID-positive test is captured in this database. We will also be provided testing data from the COVID-19 UK Non-hospital Testing Results ('pillar 2'). These sources will be used to identify our study cohort.
- **Shielded Patient List** – this data set identifies patients regarded as clinically extremely vulnerable (CEV) and at high risk of serious illness from COVID-19.
- **CO@h onboarding data** – this data set lists the date of patient onboarding (ie enrolment onto the CO@h intervention). These data will allow us to check the assumptions of a discontinuity in onboarding at age 65 and allow us to make use of a fuzzy RDD. In addition, by triangulating with hospital data, we will be able to check other assumptions – eg that onboarding is primarily occurring pre-hospital. Onboarding data were due to be collected across all CCGs from 1 December 2020, with the option of retrospectively collected data from as early as 1 October 2020 from sites that started rolling-out CO@h earlier. However, due to COVID-19 pressures, as of mid-January 2021 these patient-level data are not currently being reliably recorded nationally. Thus, the quality of this data set is a key risk to the execution and validity of the study and its ability to identify a treatment effect.
- **CO@h offboarding data** – these data will be used to check that our proposed 28-day follow-up period adequately captures the duration of CO@h use (which is intended to usually last up to 14 days from first symptoms, according to the SOP). The various quality issues affecting the onboarding data also affect this data set.
- **Hospital Episode Statistics (HES**) – this data set will include information on hospital admissions and critical care spells.

* Although we will estimate the intervention effect at age 65, we will require data on either side of the threshold to do so. The range or window of ages required will be determined using data-driven methods, but it is expected that 50–79 will be a large enough range not to restrict this choice prematurely.

- **Emergency Care Data Set (ECDS)** – this will be used to provide information on A&E attendances.
- **Office for National Statistics (ONS) mortality data** – this data set will include the registered date of death of any patients that died during the study period.
- **Data used to flag care home residents** – based on patient registration information derived from pseudonymised National Health Application and Infrastructure Services (NHAIS) data.

We will also receive information from Kent, Surrey & Sussex Academic Health Science Network (AHSN) on when each CCG began to implement CO@h and when they were considered 'fully operational' (see the later section on Study period, index dates and follow-up period). We will also access publicly available data that maps GP practices to CCGs at the earliest date of study start.

If onboarding data are delayed or unavailable, we may receive aggregate data on number of onboarded patients by CCG. This would allow us to determine whether areas are implementing CO@h at a scale sufficient to potentially identify an effect based on a 'sharp' RDD, whereby a treatment assignment effect is estimated (see later section on Statistical methods for more detail).

Study cohort

The study cohort consists of patients in England with a first COVID-19 positive test that were aged 50–79 at the time of test, with the following patient-level and area-level exclusions.

Patient-level exclusions:

- Patients with unlinked records
- Patients who are not captured in the GDPPR
- Patients with missing age or sex information
- Patients onboarded onto CO@h who do not have a first COVID-positive test
- Patients who are CEV
- Potential additional exclusion: patients living in care homes

Although CEV patients are an important target group for CO@h and may particularly benefit from it, the current study relies on a discontinuity in treatment status by age. As the intention of CO@h is to enrol all CEV patients, it is likely there will be no discontinuity to exploit for this group. Although the fuzzy RDD allows for there to be a proportion of 'non-compliers' (ie where patients' receipt of the intervention contradicts their eligibility due to age), it relies on the assumption that 'treatment eligibility' jumps from 0 to 1 at the threshold. For CEV patients, eligibility for CO@h is 1 across all ages. Furthermore, including CEV patients under 65 years of age as non-compliers would bias the resulting estimation of the treatment effect, as they are a cohort with specific baseline characteristics which are predictive of outcomes.

It is estimated that 4% of the general population are CEV.[3] We will check the proportion of CEV patients in our data set by age and onboarding status before we exclude them. A discontinuity in CEV patients about the threshold would imply that eligibility criteria are not being followed closely by the NHS clinicians responsible for onboarding patients.

The rationale for the exclusion of CEV patients also applies to patients living in care homes, where a large proportion of care home residents under 65 years of age may also be CEV. Care home residents will therefore be checked by age and onboarding status and, if no age-related discontinuity is identified, will also be excluded from the analysis.

As our study cohort is based on patients being followed for 28 days following a first COVID-19 positive test result, any patients who were onboarded onto CO@h without a COVID-19 positive test will automatically be excluded.

Patients are assigned to a CCG based on their registered GP practice. Area-level exclusions based on CCGs or areas within CCGs include:

- Patients in CCGs that did not implement CO@h or were not fully operational (ie available to all primary care networks (PCNs) in the area) within the study period.
- Patients within particular areas of CCGs that were not offering CO@h during the study period. As CO@h is not a mandatory programme, it may not be rolled out across the whole CCG. If and where we have relevant and reliable data from the AHSN on which areas within a CCG have implemented CO@h, we will limit the analysis to these areas.
- Patients in CCGs that are considered to be implementing pulse oximetry in a substantially different way to that set out in the CO@h SOP.[2] These CCGs will be identified through a combination of sources from the CO@h programme team and checks using the onboarding data. The exact exclusions will be agreed with the CO@h programme team, but are likely to include:
  - CCGs that do not use the 65+ years of age criterion, as set out in the SOP, eg if they only enrol CEV patients. This is because an RDD approach would not be applicable here. If the CO@h programme team or the qualitative evaluation suggests that this may be the case, we can identify such CCGs by checking for a discontinuity in onboarding at age 65 for each CCG independently (data permitting).
  - CCGs that only or mostly implemented a step-down model of CO@h (ie were only providing pulse oximeters at discharge from hospital), as opposed to the pre-hospitalization intervention specified in the SOP. (Note: the step-down model of CO@h differs from the CVWs programme, which is being implemented nationally from January 2021.[6])

Study period, index dates and follow-up period

The study period start date will differ between CCGs, depending on the following:

- the date CO@h was 'fully operational'
- when patient-level onboarding data are available.

The date of a CO@h service being fully operational is determined by the sites themselves and collected by the Kent, Surrey & Sussex AHSN. It is the date the service was available to all PCNs in the geographical area covered by CO@h (eg the CCG) and does not necessarily mean that that the intervention was being provided to all eligible patients by that date.

We expect that the earliest study period start dates may be at the beginning of December 2020 for some CCGs, as preliminary information suggests CO@h was fully operational in some CCGs at this point and plans were in place to collect patient onboarding data from 1 December 2020. However, due to intense service pressures in January 2021, roll-out may have been delayed in certain areas and there may also be substantial delays to their submissions of onboarding data.

The study period end date can differ between CCGs and may depend on various factors, including some that could affect the validity of the study:

- Changes to the CO@h eligibility criteria. For example, if there is a change to the age threshold, the study period will end prior to the date of this change.
- Roll-out of vaccinations and other interventions (such as the PRINCIPLE trial, see Background section).
- The number of (documented) onboarded patients available for analysis.
- The requirements of the CO@h programme team to provide timely results across the evaluation programme.

We will monitor the vaccination roll-out to determine at what point we expect it to affect our study. The vaccinations are being rolled out by decreasing priority group and the government is aiming to vaccinate all patients who are aged 70 or over or CEV (priority group 4) by 15 February 2021. Although the window of ages around the threshold that we will use for this analysis will be determined through a data-driven approach, we anticipate that 5 years either side will likely be sufficient to create the window, so are expecting that we can use data collected until at least 15 February 2021 (subject to checks).

The PRINCIPLE trial is particularly problematic for this study, as the eligibility criteria are similar to CO@h: patients are enrolled if they are aged over 65, or aged over 50 and have certain underlying health conditions (which overlap but differ from the CEV list). These criteria imply there will be the same age-related discontinuity in patients enrolled on the PRINCIPLE trial as there is with CO@h. This will cause any findings from the current study to arguably comprise a mixture of effects of both PRINCIPLE and CO@h. To identify the potential for confounding, we will attempt to examine the likely overlap of patients between CO@h and PRINCIPLE.

As of 14 February 2021, there were 4,182 patients with COVID-19 symptoms within the last 14 days enrolled in the PRINCIPLE trial. We will attempt to source aggregate numbers of trial participants by age band and enrolment period, which should enable us to gauge the extent to which the two interventions are intertwined, but we would not be able to exclude trial participants, as this would introduce bias within the RDD. Therefore, we will examine the numbers of patients on PRINCIPLE and CO@h within the window of ages used in the analysis and aim to end the study period before an estimated 5% of patients in this range are likely to have been enrolled onto PRINCIPLE. This should limit the risk that these patients have a substantive effect on the outcomes of our evaluation. We may also vary the study period end date by area, depending on the roll-out of the PRINCIPLE trial, if data are available to assess this geographically. We will confirm the area-level study start and end dates and document them in an addendum to this SAP following receipt of relevant information and data.

Each patient is assigned an index date – the date of their first COVID-19 positive test – which determines the patient's baseline and the start of their follow-up period. We will use the date of the first COVID-19 positive test rather than the date of onboarding because this allows consistent start dates for both patients who received the intervention and those who did not. This evaluation examines the effectiveness of CO@h during a 28-day period from this date, which is likely to include a period of time before patients had access to a pulse oximeter. This will allow us to evaluate the effectiveness of CO@h as it is used in practice, which may be dependent on whether pulse oximeters are distributed in time to support patients' needs and care. We considered using the date of first onset of symptoms as the index date, as it is used by implementation teams to determine for how long oximeters should be used (ie 14 days from start of onset, unless there are signs of deterioration). However, this information is not routinely collected.

A patient's follow-up period will end 28 days from the index date (where the date of the test is day 0). The risk of severe complications from COVID-19 is highest in the first 28 days and, nationally, 88% of deaths with COVID-19 recorded on the death certificate happen within 28 days of testing positive.[11] This follow-up period matches the main definition of COVID death used for reporting by Public Health England and should comfortably cover the period of up to 14 days during which the pulse oximeters are expected to be used (standard practice is to provide an oximeter as soon as possible after the positive test but this is not something we can check empirically).

Statistical methods

Local randomisation approach

There are two main conceptual approaches to RDD: (1) the continuity-based approach, whereby smooth polynomial functions are estimated either side of the threshold in order to identify the average treatment effect through extrapolation toward the threshold value, and (2) the local randomisation-based approach, where patients are assumed to be as-if randomised to the treatment or control group within a small window around the threshold.[16,17]

Where the running variable is continuous and there are very many unique data points, the continuity-based approach is the logical choice. When the running variable is discrete or the sample size is relatively small, the local randomisation-based approach is more appropriate.[17]

In the current study, the running variable is age, which we will be able to determine to a weekly accuracy. Therefore, our age variable is discrete and will contain relatively few 'mass points' (ie discrete values of the running variable, shared by many subjects/patients: up to 52 mass points per year of age). In the continuity-based approach, discrete mass points represent the sampling units available for estimation (rather than the number of individual data points). Therefore, the presence of mass points greatly reduces the number of unique values available for estimation of the polynomial function. Due to the paucity of unique values of weekly age, the local randomisation approach is the logical choice in the current study. Further, there will be a week of age across which it will not be possible to determine whether the patient was (a) aged 65, or (b) just under 65 at the date of the first COVID-19 positive test; these patients will be excluded. We would expect these patients to be just as likely eligible as ineligible for CO@h, and so would not expect their exclusion to affect one group more than the other.

The local randomisation approach rests on the assumption that there is a window around the threshold where treated and control patients have the same expected outcomes and can be considered as-if randomised to treatment. Within this window, the age of the patient is assumed to be unrelated to the outcomes of interest. Although the conceptualisation of RDDs as local randomised experiments dates back many years,[8,18] the framework for the local randomisation approach used in the current study is a relatively recent development. Our approach is based primarily on the work of Cattaneo et al.[16,17,19,20]

Sharp and fuzzy designs

There are two RDD approaches for dealing with treatment assignment: the 'sharp' and the 'fuzzy' design. The sharp RDD assumes that the probability of treatment allocation goes from 0 to 1 at the threshold (ie that everyone below the threshold is untreated and everyone above is treated) and estimates the effect of treatment eligibility on the outcomes. This means that a sharp design can be viewed as similar to an intention-to-treat analysis in a randomised trial. In contrast, the fuzzy RDD allows for treatment compliance to be imperfect (ie for treatment receipt to differ from treatment eligibility) as long as the probability of receiving treatment still jumps abruptly at the threshold.

In the current study, the sharp design estimates the average effect of CO@h on all patients who are eligible compared to all patients who are not (regardless of oximeter receipt and use). The fuzzy design will allow us to factor in treatment receipt: a proportion of patients aged 65 and over will not receive an oximeter and some patients aged under 65 will. In other words, compliance in the current study means that patients' receipt of oximeters aligns with their eligibility.

The fuzzy design can be used to estimate the average treatment effect for these 'complier' patients (within the window of ages used in the analysis) under certain assumptions: either (a) local independence between the treatment effect and compliance behaviour, or (b) a monotonicity condition. The monotonicity condition assumes that the 'compliance behaviour' for a patient (ie of being enrolled if they are aged 65 or over or not being enrolled if they are under 65) would be the same whether that patient was eligible or not, ie it is monotonic with respect to the running variable. This means that there are no 'local defiers', ie patients within the window around the threshold who are more likely to be enrolled in CO@h because they are below the age threshold (or vice versa).[21]

Assuming reliable patient-level onboarding data are available, we will aim to check how many patients are non-compliers due to incongruent oximeter receipt and make use of a fuzzy RDD in our main analysis. We will not have information on oximeter use by patients, and so this is not something we can analyse or incorporate in the RDD. The resulting average treatment effect estimates will relate only to compliers; the method does not allow any estimates to be made for the potential effect of the treatment on patients that would always either (a) receive the treatment, or (b) not receive the treatment, regardless of eligibility. (Note: CEV patients, who are expected to receive the intervention regardless of age, have been excluded from the analysis.)

In the local randomisation approach to RDD, the fuzzy design is operationalised using the Anderson-Rubin statistic, which is the difference in the adjusted, transformed responses between units assigned to treatment and control.[19] In other words, the treatment effect on the treated is identified by estimating the difference between two regression equations relating the outcome to the running variable either side of the threshold. The fuzzy approach to RDD will rely on access to good quality, patient-level onboarding data.

Basic descriptive analysis and diagnostic checks for RDD validity

Before the main analyses are carried out, diagnostic checks will be undertaken to confirm that the various assumptions of the RDD are supported.

1. **Identify the discontinuity in CO@h enrolment at age 65 and whether a sharp or fuzzy RDD is appropriate.** We will check that there is a discontinuity in the proportion of patients enrolled (ie onboarded) in CO@h at 65 years of age. We will also identify the size of this discontinuity (ie the range of discontinuity) at the threshold and whether a fuzzy design is appropriate. We will graphically show the proportion of individuals who received CO@h by weekly age. We will overlay plots of a local polynomial function either side of the threshold. These plots will be built using the 'rdplot' function from the 'rdrobust' package in R.

2. **Descriptive analysis of baseline variables, outcomes, treatment assignment and enrolment.** We will present means and standard deviations or medians, 25th and 75th quantile statistics (as appropriate). We will plot the means of the baseline variables and outcomes by week of age. We will overlay plots of a local polynomial function either side of the threshold. These plots will be built using the 'rdplot' function from the 'rdrobust' package in R, as above.

3. **Check that there is no discontinuity in the number of people by age around the threshold.** This unlikely scenario would indicate that patients could manipulate their age, potentially to gain access to the CO@h pathway. We will examine this graphically in a histogram by plotting the density of patients by weekly age in the window 1 year either side of the threshold. Following window selection (see below), we will also test whether the proportions of patients within the window are statistically comparable using a binomial test.[22]

4. **Check that CEV and care home patients should be excluded.** We will also check for an age-related discontinuity in onboarding of CEV and care home residents and exclude these patients from the analysis if there is no discontinuity to exploit at age 65:

- CEV (yes/no)
- Care home residents (yes/no)

Window selection

Although we expect there to be a trend in the proportion of patients with certain characteristics by age (eg multimorbidity, diabetes), the local randomisation approach assumes there is a window around the threshold where there are no statistically significant differences between treated and control patients. The symmetric window around the threshold that we will use in our analysis will be identified through a data-driven approach formulated by Cattaneo et al.[17] The baseline variables will be exploited to identify the largest window around the age 65 threshold within which there is no evidence that the patients vary significantly between groups in terms of those baseline variables. By identifying this window, an age range is identified across which the groups of patients can be said to be comparable at baseline. These patients are then used in the main analysis to estimate treatment effects. This method assumes that the baseline variables adequately account for differences in pre-intervention characteristics within this window that might affect outcomes and vary by age. This assumption implies that the outcomes are conditionally independent of the baseline variables.

The window will be identified by first conducting tests of mean differences in the baseline variables at the narrowest window – ie 1 week either side of the age 65 threshold, and with a minimum of 10 observations (patients) per group. Each baseline variable will be tested and the smallest resulting p-value reported. The window will then be increased by 1 week either side of the threshold and the process repeated. The maximum window used in the main analysis will be the smaller of (a) the one immediately prior to the window where a statistically significant

difference between groups is first identified, or (b) the one where the sample size achieved should comfortably identify a statistically significant difference in the outcome based on power calculations*.

Although we believe that the window will be determined by (a), the criterion in (b) will ensure that any potential for unobserved confounding caused by the use of a very wide window is limited.

For the procedure in (a), Cattaneo et al recommend that the p-value chosen for these tests is at least 0.10 and preferably 0.15.[17] As we have a large number of baseline characteristics, a p-value of 0.10 should provide a reasonably conservative threshold, given that the chances of identifying a significant difference in at least one variable increase with the number of variables tested.[23] The results of these tests will be tabulated and plotted, showing how the minimum p-value changed against each window length from a window length of 1 week either side, up until the resulting minimum p-value remains consistently below 0.05.

Although the window selection can be done directly through the 'rdrandinf' command (within the 'rdlocrand' R package), we will calculate it using the separate window selection function 'rdwinselect', so that the window selection process is completed independently of the main analysis.[24]

Baseline variables

The baseline variables that we will check and use to identify the age window for analysis are:

- Gender
- Ethnicity
- IMD quintile
- IMD health deprivation and disability quintile
- Smoker (yes/no)
- Obesity (yes/no)
- Diabetes (yes/no)
- Multimorbidity (yes/no), ie whether a patient had two or more long-term conditions. We will use the list of conditions in the Cambridge Multimorbidity Score,[25] based on a list of conditions recorded in GP records and available in GDPPR and also used in Stafford et al[26].
- Region – as a proxy for other unobservable characteristics, as regions may have differing rates of COVID-19 and therefore also hospital pressures, which may affect patients' outcomes.

* Using power = 80%, alpha = 5%, and two-sided test.

Main analysis

In the main analysis, we will estimate a treatment effect based on the difference in means between the treated and untreated. Using randomisation inference, 10,000 simulated draws based on Fisher's exact approach to null-hypothesis testing will be used to estimate p-values of zero effect. We do not expect to have to transform the outcomes – as we do not expect the outcomes to vary with age within the chosen window – therefore we will use a polynomial of order zero (ie a constant within-group average outcome, which is the standard in the local randomisation approach. The plots created in the prior descriptive analysis will be used to confirm this).

The null hypothesis for the Fisher-based simulation will assume complete randomisation of patients to treatment within the window. A confidence interval will be estimated for the treatment effect by inverting the calculation for the Fisher-based p-value and identifying values of treatment effect which are not significantly different to the estimated effect. As a comparison to the Fisher-based approach, we will also estimate the p-value using the large-sample Neyman approach based on normal approximation, and also report the power of the Neyman test to inform interpretation.[17] If the results of the Fisher and Neyman approaches differ, we will consider the number of observations within the window to determine which approach is most valid, with the Fisher framework being preferred where sample sizes within the window are small.

Many of our outcomes are binary. In these cases, the estimated treatment effects will comprise differences in mean outcomes on the per cent scale. For our other outcomes, the treatment effects will be differences in mean bed days. As well as the treatment effects, we will also present the point estimates for the treated and control groups in each case. We will not control for the baseline variables in the model, as the window selection procedure will already have been used to identify groups which are comparable in terms of these variables.

Calculation of intervention estimates where outcomes happen before enrolment

This evaluation aims to estimate the effect of pre-hospitalisation CO@h, as detailed in the SOP.[2] Although we will exclude any CCGs which are identified as only offering enrolment in CO@h at hospital discharge, it is conceivable that other CCGs may offer a combination of pre-hospital and post-discharge as part of the CO@h model, especially in those areas that had implemented a post-discharge model early on in the pandemic. There may therefore be some patients enrolled in CO@h at hospital discharge that we can identify by triangulating date of onboarding with date of hospital discharge. These patients will be considered as not having received the CO@h intervention (ie treatment receipt = 0), similar to those patients enrolled in CVWs (for whom we do not have data).

Patients who have a COVID-19 positive test and are subsequently enrolled without any intermediate hospital activity will be considered 'treated' for the purpose of all hospital outcome analyses which occur within 28 days of the date of their first COVID-19 positive test. If a patient attends A&E after enrolment but before they received a pulse oximeter, they will still be considered as having received the intervention. This is because the evaluation seeks to understand the usefulness of CO@h as it is used in practice.

However, A&E departments are also one of the services that can refer patients to CO@h. In this case, the A&E attendance by definition pre-dates onboarding (as the pulse oximeter was dispensed following A&E attendance). Here, it is not appropriate to count the patient as having had the intervention when analysing A&E attendances. However, it is also not appropriate to exclude these patients from the analysis for outcomes that occur before onboarding, as this would introduce bias compared with patients not enrolled in CO@h. We can identify the cases described above by triangulating the date of onboarding with any dates of A&E, thereby determining whether patients were (most likely) enrolled through A&E. Using these data, we can set the intervention receipt (which is used in the fuzzy part of the design) to 1 (enrolled) or 0 (not enrolled) depending on the outcome – ie for a patient enrolled following an A&E attendance, this attendance would still count towards the outcome in the model for A&E attendances, but treatment receipt (enrolled) would be set to 0. However, for the emergency admission outcome, the same patient's subsequent emergency admission (if one occurred) would have enrolment set to 1.

Sensitivity analyses

We will undertake a range of sensitivity analyses to identify whether the resulting treatment effect estimates are reliable. These will be carried out on each of the primary outcomes (critical care bed use, total bed days in critical care, and death) but only on the secondary outcomes if these show statistically significant treatment effects. Unless otherwise stated, the design choices used in the main analysis (eg fuzzy design, window, order of polynomial, etc) will be the same as in the main analysis. The sensitivity analyses will include:

1. Intention-to-treat analysis – we will use a sharp design (without taking account of compliance/oximeter receipt) to estimate the average effect of intervention eligibility at age 65, ie an intention-to-treat analysis.

2. Placebo outcomes – we will undertake statistical tests for treatment effects for each of the baseline variables and plot the means for each group, weekly averages and fitted constants to identify any evidence of infringement of the assumption of comparability of the groups.

3. Test of treatment assignment – we will use a binomial test to check that, within the specified window, treatment can be considered randomly assigned.

4. Influence of window choice – we will test a selection of windows smaller than the one used in the main analysis (minimum 10 patients either side of the threshold). We will not use larger windows as this may lead to biased estimates due to imbalances in baseline variables.[17] By using smaller windows, this problem is avoided while also identifying whether restriction of the cohort to those increasingly closer to the threshold affects the treatment effect estimates. We will plot the resulting point estimates and confidence intervals.

5. Placebo thresholds – we will estimate treatment effects at both a lower and a higher threshold than age 65 to check that these spurious, placebo thresholds result in a treatment effect of zero. For these checks, we will only use control and treated patients,

respectively.[17] We will test at thresholds centred on the first mass point past the end boundaries of the window used in the main analysis. We will use the same window size (ie number of mass points either side) as in the main analysis.[17]

6. Placebo study period – we will use data for a period before CO@h was implemented (ie in CCGs before they implemented CO@h) to check that there is no discontinuity in outcomes in a window around age 65 when no treatment was available. This will be a sharp RDD (as there will, by definition, be no intervention available). This sensitivity analysis will only be conducted if the intention-to-treat analysis using a sharp RDD (sensitivity analysis #1) is statistically significant.

7. Sensitivity to unmeasured confounding – we will calculate Rosenbaum's sensitivity bounds, within the context of a local randomisation RDD, to estimate how large an unobserved confounder would need to be to affect the conclusions of the study.[19,27]

Where relevant, we will make use of the Benjamini-Hochberg procedure[28] to adjust for the increase in the false discovery rate (type 1 errors) due to multiple testing.[23]

Descriptive analyses of patient flow

For context, we will create a flow diagram based on the CONSORT standard[29] of how patients move between receiving a COVID-19 positive test, are onboarded and use hospital services. This will help visualise some of the implicit decisions that we make, for example by showing the number (and percentage) of patients that were first enrolled following A&E. We will also check and document the number (and percentage) of patients who were onboarded without a COVID-19 positive test, and therefore excluded from our analysis.

We will also use the onboarding data to understand the enrolled population: the number of patients who are enrolled due to age, CEV status, both or neither. Likewise, we will consider these categories in those patients who were not enrolled. As long as these patients are flagged as CEV in the GDPPR data, we will be able to identify these patients.

Subgroup analyses

As care home residents will have access to support to use the pulse oximeters and record measurements correctly, care home residency may act as an effect modifier. However, as it is possible that many care home residents below 65 years of age will also be eligible for CO@h (eg because of learning disabilities, according to the SOP) it will probably not be possible to do a subgroup analysis on care home residents using the RDD method and we may exclude these patients from the study cohort. If they are not excluded, we may still run an RDD in the subgroup of patients not in care homes, which would potentially allow for an analysis with a sharper discontinuity at age 65.

Given the inequalities that COVID-19 has exposed, it is important to understand whether technologies such as the use of remote pulse oximeters are helping the communities worst affected by the pandemic. If the study is sufficiently powered, we will estimate the intervention effect in patients who live in the most socio-economically deprived areas. We could look at either the most deprived IMD quintile or the aggregate of the two most deprived IMD quintiles. We may also estimate the intervention effect in patients of different ethnic groups

(White, Black, Asian, Mixed and Other). Depending on the achieved power, this may be limited to a broader group of ethnicities (eg 'ethnicities other than White'). Once we know more about the quality of onboarding data and numbers of patients enrolled, we will make an assessment as to what is possible in terms of subgroup analyses.

Reporting

Our reporting of the study will aim to meet the requirements of the STROBE checklist for the reporting of cohort studies.[30]

Power calculation

A set of power calculations for binary outcomes (eg ICU: yes/no) are shown in Table 1, below.

Table 1: Total sample sizes resulting from power calculations

Total sample sizes (N)		Relative risk in treated group		
		50%	75%	90%
Baseline risk	10.0%	870	4,010	26,990
	5.0%	1,812	8,404	56,816
	4.0%	2,282	10,602	71,730
	3.0%	3,068	14,266	96,584
	2.0%	4,638	21,590	146,294
	1.0%	9,346	43,568	295,422
	0.5%	18,766	87,522	593,680
Note: equal sized treatment and control groups of N/2; full compliance assumed, two-sided tests; power 80%; alpha 5%.				

The power calculations are based on standard large sample approximations. Power calculations using Fisher-based randomisation inference produce similar results to those based on large sample approximations. For example, assuming two equally sized groups, a two-sided test at the 5% significance level, and an outcome of 2% in the control group and 1.5% in the treated group, a sample size of around 22,000 (11,000 in each group) would be required to achieve a power of 80% (based on a simulation with 10,000 draws).

These power calculations demonstrate that relatively large samples will be required to identify treatment effects of CO@h on the primary outcomes, particularly for infrequent outcomes such as ICU admittance. For reference, the mixed-methods study of pilots of remote home monitoring models in England found that for the pre-hospital model of CO@h, approximately 8% of onboarded patients attended A&E, 5% were admitted to hospital in an emergency, 0.2% were admitted to ICU and 1% died.[1]

Strengths and limitations

RDDs avoid the assumption that all relevant confounders are observed and accounted for. Instead, they take advantage of the fact that patients just above and below the threshold should, in general, have similar expected outcomes and baseline characteristics. Therefore, any discontinuity in the outcome variable at the threshold can be attributed to differences in the treatment received.

The CO@h model is being rolled out nationally, which makes finding a comparison group problematic using area-level quasi-experimental approaches. However, the RDD approach creates a counterfactual from within the same health care locations and providers, allowing for a robust counterfactual analysis even if all CCGs implement the CO@h intervention at the same time.

For example, hospital admission thresholds may differ depending on bed pressures, and these may differ between hospitals. This could introduce area-level differences. However, in the current study both patients aged above and below 65 have access to the same services in the same hospitals during the same period, thereby limiting the risk of bias.

The local randomisation RDD estimates a treatment effect in a window around the 65-year threshold (given that we will restrict the study to the period in which this was the criterion for onboarding on CO@h). The results are not necessarily generalisable to other ages, non-compliers or excluded patient groups (ie CEV or care home residents). It is always the case with RDD that results may not be generalisable to other ages, but this may particularly be the case here where adherence to the CO@h model often relies on patient input – for example, older patients may find it more difficult to remember to take measurements and read and record them. Other evaluation partners may be able to investigate the adherence and efficacy across differing ages through the qualitative work they are undertaking, which would provide useful context when interpreting the results of the RDD.

A major limitation of this analysis is that it will not include patients who were CEV. Although CEV patients account for only about 4% of the population, this patient group is at increased risk of severe complications from COVID-19 and could therefore particularly benefit from CO@h. However, we are unlikely to be able to estimate a treatment effect for CEV patients due to there being no discontinuity in treatment status for this group. By excluding CEV patients, we will be able to provide a more accurate estimate of the effect of receiving the CO@h intervention at age 65 for non-CEV patients.

As care home residents will have access to support to use the pulse oximeters and record measurements correctly, care home residency may act as an effect modifier. It is not possible to do an RDD analysis on the subgroup of patients living in care homes (as there may not be a large enough discontinuity at age 65 among care home residents, as residents younger than 65 may also be eligible for CO@h). Indeed, if to a large extent all care home residents are eligible, this group may be excluded from the main analysis. However, evaluation partners are investigating this patient group.

As we are using routinely recorded data, we will not be able to identify patients who were clinically diagnosed as COVID-19 positive without a test. Any such patients will therefore be excluded from our analysis. We will check and present within the flow diagram the number of patients who were onboarded without a COVID-19 positive test, to get an indication of the proportion of patients that we are missing from our analysis.

This study does not address whether individuals use their pulse oximeter as indicated (eg taking three readings a day). The RDD will instead evaluate the effectiveness of pulse oximeters as they are used 'in practice', ie independent of both any usage compliance issues and any potential delays in providing pulse oximeters. If the pulse oximeters are not used as indicated or there are substantial delays in receiving them, then it will be less likely that we will find a significant treatment effect, despite this being an accurate reflection of their usefulness. The qualitative evaluation may be able to explore such issues, which would provide context for the interpretation of the results.

We are only including data from patients from CCGs from the period at which they have reported that CO@h is fully operational. By excluding data from areas that have not, or have not yet, implemented CO@h, or where they are just ramping up the service, we want to evaluate the intervention once implemented, while also ensuring the best chance of finding a discontinuity in treatment by age in order to be able to use an RDD. It may be that these 'adopters' are different from 'non-adopters' or 'late adopters', and that the effect may be different when the intervention has had time to embed (new pathways established, new ways of working etc), or after the eligibility criteria have changed, which may affect the generalisability of our results. However, as the comparison group is from within the same areas, this does not affect the internal validity or invalidate the RDD.

Due to severe pressures within the NHS, collection of onboarding data in many areas has been delayed or diminished. Unless these data are retrospectively and – at least in some areas comprehensively – collected within the time frame of this evaluation, this study could be seriously undermined. Although a sharp design is still possible, in that we can evaluate the effect of being 'eligible' for enrolment (ie in effect treating all patients aged 65 or over as having the intervention and all under 65 years of age as not) in those CCGs that are expected to have had higher levels on onboarding, this would understate the treatment effect, as treatment group non-compliers (patients not receiving oximeters when they should have) would not be accounted for using the 'fuzzy' approach.

By limiting the analysis to patients with a COVID-19 positive test in those areas that were fully operational, that abided by the original CO@h criteria and provided onboarding data before the change in age criteria was introduced, we may find that our sample size and in particular the number of patients with primary outcomes may be very small. As a result, the study may be underpowered.

We do not expect the inclusion/exclusion criteria to introduce any bias in any observed or unobserved confounders between patients above and below the threshold. We will check for bias in observed characteristics that may act as confounders. However, there may be some selection bias compared to the overall population, for example due to our patient cohort consisting of patients who have taken a COVID-19 test or not including patients with data linkage issues.

The length of time from date of first onset of symptoms to date of test may vary, especially over time and between areas, depending on supply of and demand for tests. This can introduce bias if the intervention and control groups differ. However, the RDD, by including patients for both groups from the same areas in the same period, limits this risk. Although we will not have data on first onset of symptoms, we do not expect that the length of time from first onset to date of test will vary substantively across ages, and therefore do not anticipate a discontinuity at age 65.

There is some evidence of false positive swab tests of between 0.8% and 4%.[31] Although we would expect the rate of false positives to be smooth across age, this is not something that we can verify.

The logic model for CO@h stipulates that regular pulse oximetry readings will identify patients who are deteriorating by their decreasing oxygen levels and that the CO@h clinical team can quickly escalate their care, thereby avoiding more serious complications, invasive treatment, and death. This study is based on this logic model by evaluating the effect of being enrolled on CO@h on outcomes relating to serious disease. However, we do not have information on the intermediate steps (ie how pulse oximeters are used, whether oxygen saturation readings informed the decision to escalate care or indeed how or by who decisions about patients' care were made). Exploring the correlation between pulse oximetry readings and hospital admissions would be useful in order to understand how oximeter readings impact on the decision to admit a patient to hospital. This is outside the scope of this study but may be something other evaluation teams can explore.

COVID-19 vaccinations commenced in December 2020, starting with older care home residents and care home staff, then moving down the priority list. By 15 February 2021, all patients aged 70 and over or CEV should have been offered a vaccination. If vaccinations are effective in stopping people getting COVID-19, this will affect the size of our cohorts, rather than confound our results. However, if vaccinated patients can still get COVID-19 but with milder symptoms (and therefore less likely to require hospital care), this could confound our analysis. We will be provided with information on the roll-out of the vaccinations and will aim to end the study period before or at the point vaccinations start being routinely provided to patients aged 65–70. This will ensure that there is limited confounding close to the threshold. The study period can be, if appropriate, tailored to different areas depending on local vaccination rates.

Some of the outcomes of this study may be difficult to interpret, either because of opposing mechanisms at play or because of other interventions, such as CVWs. However, our primary outcomes, relating to critical care and mortality, are expected to be easier to interpret. The qualitative analysis may also be able to provide contextual information to help interpret our results.

The aim of CVWs is to discharge patients earlier to alleviate the current pressures in hospitals. This could therefore impact on one of the outcomes of the CO@h evaluation: hospital bed days. As we will not have access to data on virtual ward enrolment, we are not able to assess whether CVWs affect patients' total hospital bed days in the CO@h intervention or control groups differently, and therefore whether any difference – or lack of difference – in outcomes is due to CO@h or CVWs. Although patients who are discharged early with CVWs may in general have different characteristics, eg in age, from those that are not, we do not expect a

discontinuity at age 65. Likewise, while CVWs may be used more in areas where there is the most pressure on hospital beds, by virtue of 'treated' and 'control' patients being found within the same areas and hospitals, we do not expect this to cause any bias in our analyses.

As the PRINCIPLE trial also used an age threshold of age 65 for trial inclusion, there is a risk that this will introduce bias in our analysis. We will limit the risk of the PRINCIPLE trial biasing our results by checking the number of patients enrolled in the trial by area and age band and end the study period of our evaluation before a maximum of 5% of patients around the threshold have been enrolled in the PRINCIPLE trial.

This analysis builds on the assumption that there are no other differences in care of patients above and below 65 years of age that would affect the hospital outcomes of COVID-19 positive patients. We do not know of any such interventions but, if they exist, they may bias our results. We may be able to indirectly examine this to a certain extent by checking that there is no discontinuity in outcomes at age 65 in a period before the CCGs implemented CO@h (ie the 'placebo outcomes' sensitivity analysis).

There is a substantial risk to the execution and validity of the study if we cannot access onboarding data. Currently onboarding data is potentially going to be delayed, especially from CCGs that have been badly hit by the third pandemic wave, and/or may be incomplete. These data are necessary to allow us to estimate the effect of being enrolled in CO@h. In its absence, we can run an intention-to-treat analysis, estimating the effect of treatment eligibility, ie treat all non-CEV patients aged 65 or above as receiving CO@h and all non-CEV patients below as not receiving the intervention (a 'sharp' design). However, it is possible that compliance will not be high; therefore, this analysis is less likely to find an effect even if onboarding did affect outcomes, due to the added noise of non-compliers. It may however still provide a useful indication in the absence of onboarding data: a significant intervention effect would indicate that CO@h affected patients' outcomes (subject to sensitivity analyses), while a lack of a statistically significant effect would be non-informative as it would not be possible to unpick whether this is due to low uptake of the CO@h programme, the 65+ threshold not being complied with, or the CO@h programme not having the anticipated effect on outcomes.

If interim analyses are performed before receipt of final data, only the final analyses may be published.

Acknowledgements

We are grateful to Sara Genelitti and Jake Anders for their peer review comments on earlier versions of this SAP and to Matias Cattaneo for answering our methodological questions.

References

1. Vindrola-Padros C, Sidhu M, Georghiou T, Sherlaw-Johnson C, Singh K, Tomini S, et al. The implementation of remote home monitoring models during the COVID-19 pandemic in England. *medRxiv.* 2020; doi: 10.1101/2020.11.12.20230318.
2. NHS England and NHS Improvement. *Novel Coronavirus (COVID-19) Standard Operating Procedure: COVID Oximetry @home.* NHS England and NHS Improvement; 2020 (www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/11/C0817-sop-covid-oximetry-@home-november-2020.pdf).
3. Dreyer K, Peytrignet S. Understanding the needs of those most vulnerable to COVID-19. Health Foundation; 2020 (www.health.org.uk/news-and-comment/charts-and-infographics/understanding-the-needs-and-experiences-of-those-most-clinic).
4. Department of Health and Social Care. Joint Committee on Vaccination and Immunisation: Advice on Priority Groups for COVID-19 Vaccination, 30 December 2020 [webpage]. *DHSC*; 6 January 2021 (www.gov.uk/government/publications/priority-groups-for-coronavirus-covid-19-vaccination-advice-from-the-jcvi-30-december-2020/joint-committee-on-vaccination-and-immunisation-advice-on-priority-groups-for-covid-19-vaccination-30-december-2020).
5. Roxby P. Covid: When will I get the vaccine? [webpage]. *BBC*; 15 February 2021 (www.bbc.co.uk/news/health-55045639).
6. NHS England and NHS Improvement. *Novel Coronavirus (COVID-19) Standard Operating Procedure: COVID Virtual Ward.* NHS England and NHS Improvement; 2021 (www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2021/01/C1042-sop-discharge-covid-virtual-ward-13-jan-21.pdf).
7. University of Oxford. PRINCIPLE - Platform Randomised trial of INterventions against COVID-19 In older peoPLE [webpage] (www.principletrial.org).
8. Thistlewaite DL, Campbell DT. Regression-discontinuity analysis: An alternative to the ex-post facto experiment (reprint). *Obs Stud.* 2016; 2: 124–128 (https://obsstudies.org/reprint-regression-discontinuity-analysis-an-alternative-to-the-ex-post-facto-experiment/).
9. Imbens GW, Lemieux T. Regression discontinuity designs: A guide to practice. *J Econom.* 2008; 142(2): 615–635. doi: 10.1016/j.jeconom.2007.05.001.
10. Public Health England. *Public Health England Data Series on Deaths in People with COVID-19: Technical Summary.* PHE; 2020 (www.gov.uk/government/publications/phe-data-series-on-deaths-in-people-with-covid-19-technical-summary).
11. Department of Health and Social Care. New UK-wide methodology agreed to record COVID-19 deaths. *DHSC*; 12 August 2020 (www.gov.uk/government/news/new-uk-wide-methodology-agreed-to-record-covid-19-deaths).
12. NHS Digital. HES *Data Dictionary - Adult Critical Care.* (https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary).
13. Schnapp LM, Cohen NH. Pulse oximetry: Uses and abuses. *Chest J.* 1990; 98(5): 1244–1250. doi: 10.1378/chest.98.5.1244.
14. Hidalgo DC, Olusanya O, Harlan E. Critical care trainees call for pulse oximetry reform. *Lancet Respir.* Published online 2021. doi: 10.1016/S2213-2600(21)00102-8.

15. NHS Digital. General Practice Extraction Service (GPES) Data for pandemic planning and research: A guide for analysts and users of data. Data coverage - management information [webpage]. *NHS Digital*; 2020 (https://digital.nhs.uk/coronavirus/gpes-data-for-pandemic-planning-and-research/guide-for-analysts-and-users-of-the-data#data-coverage-management-information).
16. Cattaneo MD, Idrobo N, Titiunik R. *A Practical Introduction to Regression Discontinuity Designs: Foundations (Elements in Quantitative and Computational Methods for the Social Sciences).* Cambridge University Press; 2020. doi: 10.1017/9781108684606.
17. Cattaneo MD, Idrobo N, Titiunik R. *A Practical Introduction to Regression Discontinuity Designs: Volume II.* 2018. doi: 10.1017/9781108684606.
18. Lee DS, Lemieux T. Regression discontinuity designs in economics. *J Econ Lit.* 2010; 48: 281–355. doi: 10.1257/jel.48.2.281.
19. Cattaneo MD, Titiunik R, Vazquez-Bare G. Inference in regression discontinuity designs under local randomization. *Stata J.* 2016; 16(2): 331–367. doi: 10.1177/1536867X1601600205.
20. Cattaneo MD, Frandsen BR, Titiunik R. Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. Senate. *J Causal Inference.* 2015; 3(1): 1–24. doi: 10.1515/jci-2013-0010.
21. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* 1996; 91(434): 444–455.
22. Frandsen BR. Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. *Adv Econom.* 2017; 281–315. doi: 10.1108/S0731-905320170000038012.
23. Herzog MH, Francis G, Clarke A. The multiple testing problem. In: *Understanding Statistics and Experimental Design.* Springer; 2019: 63–66. doi: 10.1007/978-3-030-03499-3_5.
24. Cattaneo AMD, Titiunik R, Vazquez-Bare G. *Package 'Rdlocrand'*; 2020 (https://cran.r-project.org/web/packages/rdlocrand/rdlocrand.pdf).
25. Payne RA, Mendonca SC, Elliott MN, et al. Development and validation of the Cambridge Multimorbidity Score. *CMAG*. 2020; 192(5): E107–E114. doi: 10.1503/cmaj.190757.
26. Stafford M, Steventon A, Thorlby R, Fisher R, Turton C, Deeny S. *Understanding the health care needs of people with multiple health conditions*. Health Foundation; 2018 (www.health.org.uk/sites/default/files/upload/publications/2018/Understanding the health care needs of people with multiple health conditions.pdf).
27. Rosenbaum PR. *Observational Studies.* Springer; 2002.
28. Benjamini Y, Hochberg Y. Controlling the false discovery rate : A practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 1995; 57(1): 289–300.
29. Consort: Transparent reporting of trials. (http://www.consort-statement.org/).
30. STROBE Statement - checklist of items that should be included in reports of cohort studies. 2007 (www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_v4_cohort.pdf).
31. Surkova E, Nikolayevskyy V, Drobniewski F. False-positive COVID-19 results: hidden problems and costs. *Lancet Respir Med.* 2020; 8(12): 1167–1168. doi: 10.1016/S2213-2600(20)30453-7.