





# Characteristics of non-randomised studies of drug treatments: cross sectional study

Sally Yaacoub <sup>1,2,3</sup> Raphael Porcher,<sup>1,2</sup> Anna Pellat,<sup>1,4</sup> Hillary Bonnet,<sup>1,2,3</sup> Viet-Thi Tran <sup>1,2</sup>  
Philippe Ravaud,<sup>1,2,3</sup> Isabelle Boutron<sup>1,2,3</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjmed-2024-000932>).

For numbered affiliations see end of article.

Correspondence to: Dr Sally Yaacoub, Inserm, INRAe, Centre for Research in Epidemiology and Statistics (CRESS), Université Paris Cité, Hôpital Hôtel Dieu, 1 place du Parvis Notre-Dame, 75004 Paris, France; [sally.yaacoub@gmail.com](mailto:sally.yaacoub@gmail.com)

Cite this as: *BMJMED* 2024;3:e000932. doi:10.1136/bmjmed-2024-000932

Received: 26 March 2024

Accepted: 20 September 2024

## ABSTRACT

**OBJECTIVE** To examine the characteristics of comparative non-randomised studies that assess the effectiveness or safety, or both, of drug treatments.

**DESIGN** Cross sectional study.

**DATA SOURCES** Medline (Ovid), for reports published from 1 June 2022 to 31 August 2022.

### ELIGIBILITY CRITERIA FOR SELECTING STUDIES

Reports of comparative non-randomised studies that assessed the effectiveness or safety, or both, of drug treatments were included. A randomly ordered sample was screened until 200 eligible reports were found. Data on general characteristics, reporting characteristics, and time point alignment were extracted, and possible related biases, with a piloted form inspired by reporting guidelines and the target trial emulation framework.

**RESULTS** Of 462 reports of non-randomised studies identified, 262 studies were excluded (32% had no comparator and 25% did not account for confounding factors). To assess time point alignment and possible related biases, three study time points were considered: eligibility,

treatment assignment, and start of follow-up. Of the 200 included reports, 70% had one possible bias, related to: inclusion of prevalent users in 24%, post-treatment eligibility criteria in 32%, immortal time periods in 42%, and classification of treatment in 23%. Reporting was incomplete, and only 2% reported all six of the key elements considered: eligibility criteria (87%), description of treatment (46%), deviations in treatment (27%), causal contrast (11%), primary outcomes (90%), and confounding factors (88%). Most studies used routinely collected data (67%), but only 7% reported using validation studies of the codes or algorithms applied to select the population. Only 7% of reports mentioned registration on a trial registry and 3% had an available protocol.

**CONCLUSIONS** The findings of the study suggest that although access to real world evidence could be valuable, the robustness and transparency of non-randomised studies need to be improved.

## Introduction

Randomised controlled trials have long been considered the gold standard for assessing the effects of drug treatments. Whether these trials comprehensively describe the scope of real world clinical practice, however, has been questioned.<sup>1</sup> Also, randomised controlled trials might not be feasible for dealing with specific clinical questions, including a particular population, or providing timely evidence.<sup>2,3</sup>

The prominence of non-randomised studies has risen in recent years, specifically with the increase in real world data.<sup>4,5</sup> Subsequently, non-randomised studies can provide evidence on broader patient populations, various treatment regimens, long term outcomes, rare events, and harms.<sup>6,7</sup> These studies can have a role in generating timely and cost effective evidence for comparative effectiveness research, providing insight for decision making on drug treatments in the real world setting.<sup>8-10</sup> A study summarising the levels of evidence supporting clinical practice guidelines in cardiology found that 40% of 6329 recommendations were supported by level of evidence B (ie, supported by data from observational studies or one randomised controlled trial), with only few recommendations supported by evidence from randomised trials.<sup>11</sup>

Non-randomised studies are susceptible to numerous limitations related to their design and analysis choices, which could result in effect estimates

## WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ The literature has highlighted particular concerns in non-randomised studies, such as reporting and bias, and occasionally has focused on specific conditions
- ⇒ Heterogeneity and limitations in the conduct, analysis, and reporting of non-randomised studies are not as well studied

## WHAT THIS STUDY ADDS

- ⇒ Only 11% of non-randomised studies that assessed the effectiveness or safety, or both, of drug treatments had a comparator, accounted for confounding factors, and had no biases related to time point misalignment (or all biases were dealt with)
- ⇒ In a representative sample of 200 reports of non-randomised studies indexed in Medline, most studies had one possible bias related to time point misalignment (70%) and only 2% reported all six of the key elements considered
- ⇒ Most studies used routinely collected data (67%), but few reported using validation studies of the codes or algorithms applied to select the population (7%), mentioned registration on a trial registry (7%), or had an available protocol (3%)

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

- ⇒ The robustness and transparency of non-randomised studies should be improved by providing tools for researchers to take advantage of the availability of routinely collected data, comprehensively report study elements, and facilitate the adoption of the target trial emulation framework

### BOX 1 | DEFINITIONS OF SPECIFIC ITEMS

#### Validation studies of codes or algorithms

We considered reporting of using validation studies assessing the sensitivity and specificity of the codes or algorithms used to select the population in studies that used routinely collected data (when studies explicitly stated it or cited validation studies). Appropriate sensitivity and specificity are important to avoid misclassification and maximise detection of the population of interest.

#### Eligibility for any treatment arm

We considered that studies reported eligibility for any treatment arm when the authors explicitly stated that participants should not have any contraindications to any treatment arm or that participants could receive any treatment arm, to ensure clinical equipoise (ie, having an equal probability of individuals being allocated to any of the treatment groups).

#### Negative controls

We determined whether negative controls were explicitly stated. Negative controls are proxies for an unmeasured confounder. A negative control outcome is a variable known not to be causally affected by the treatment of interest. A negative control exposure is a variable known not to causally affect the outcome of interest.

#### E value

We determined whether E values were computed. The E value is a sensitivity analysis method that represents the extent to which an unmeasured confounder would have to be associated with both treatment and outcome to nullify the observed treatment-outcome association. The E value is a useful concept for assessing the robustness of non-randomised studies.

that are biased.<sup>12</sup> Several guidelines have been developed for reporting non-randomised studies,<sup>13–15</sup> and the target trial emulation framework has been developed to overcome the avoidable methodological pitfalls of traditional causal analysis of observational data, thus reducing the risk of bias.<sup>16</sup> The framework suggests that a non-randomised study should be conceptualised as an attempt to emulate a hypothetical randomised controlled trial addressing a research question of interest, to make causal inference with observational data.<sup>16</sup> This framework requires specifying key components of the target trial, such as time points of eligibility, treatment assignment, and start of follow-up. Failure to align these time points would impose a risk of bias in effect estimates.<sup>16</sup>

The literature highlights particular concerns in non-randomised studies, such as inadequate reporting, and occasionally focuses on specific

conditions.<sup>17–20</sup> But the heterogeneity and limitations in the conduct, analysis, and reporting of non-randomised studies, in a representative sample of reports, has not been well studied. In this study, our aim was to examine the characteristics of comparative non-randomised studies that assessed the effectiveness or safety, or both, of drug treatments. We focused on general characteristics, reporting characteristics, and time point alignment, and possible related biases.

## Methods

### Design

This cross sectional study analysed a representative sample of reports, indexed by Medline, of comparative non-randomised studies accounting for confounding, that assessed the effectiveness or safety, or both, of drug treatments, and that were published in June–August 2022. The protocol is registered in Open Science Framework (<https://osf.io/sjauh>).

### Eligibility criteria and search strategy

We included reports of non-randomised studies: conducted in humans; aimed at assessing the effectiveness or safety, or both, of drug treatments; with a comparator arm (eg, active drug treatment comparator, standard of care, or no treatment); and reporting methods to account for at least one confounding factor (eg, multivariable regression, matching, or weighting). We excluded reports of specific publication types (ie, editorials, letters, and opinions piece); reports of studies only assessing non-drug treatments (eg, alternative treatment, surgery, or vaccines); reports of specific study types (trials, case series, case reports, interrupted time series, and guidelines); and reports not written in English. We searched Medline (Ovid) on 29 September 2022. We developed a search strategy with the help of a medical librarian that included both medical subject headings (MeSH) and keywords (online supplemental appendix 1). We combined terms for non-randomised studies (eg, cohort and real world evidence) and for pharmacological treatments (eg, drug treatments).

### Study selection

The records identified from the search were exported to Microsoft Excel (Microsoft Corporation, Redmond, WA) and ranked in random order with a random number generator. One reviewer assessed the eligibility of the abstracts and full texts, with 20% done in duplicate and independently by a second reviewer. Any disagreements were resolved by discussion between the two reviewers or with a third reviewer. We sequentially screened records in batches of 500 until we identified 200 reports of eligible studies.

### Data extraction

We developed and piloted a standardised data extraction form (online supplemental appendix 2) inspired by RECORD (Reporting of studies Conducted using Observational Routinely collected health Data),<sup>13</sup> STROBE (Strengthening the Reporting of Observational Studies in Epidemiology),<sup>14</sup> CONSORT (Consolidated Standards of Reporting Trials),<sup>21</sup> target trial emulation framework,<sup>16</sup> STaRT-RWE (structured template and reporting tool for real world evidence),<sup>22</sup> ROBINS-I (Risk Of Bias in Non-randomised Studies - of Interventions),<sup>23</sup> and previous studies.<sup>24</sup> Data were extracted in duplicate (independently for 20% and 80% as data verification) by two trained researchers. Any disagreements were resolved by discussion or with a third reviewer.

### General characteristics

We extracted study characteristics (eg, medical area, region, and contribution of a statistician or methodologist, determined from author affiliations or explicit reporting). We also extracted data on the research question (eg, explicit statement) and its elements: population (eg, patients with chronic diseases), intervention (eg, start of treatment), comparator (eg, active comparator), and outcomes (eg, effectiveness). We extracted data on study design (eg, cohort), type of data used (eg, routinely collected data), and sources (eg, electronic health records). For the study design, we did not use the descriptions reported by the authors but instead we judged the study design based on prespecified criteria (online supplemental appendix 3). We extracted data on the funding source, conflict of interest statement, setting (eg, primary), centre (eg, number of centres), participants (eg, number analysed), and follow-up time. We extracted data on any reference to registration, access to the protocol, content of reported changes to the protocol, data sharing statement (when available), access to codes and algorithms, and reference to ethical review.

### Reporting characteristics

We considered six key study elements for reporting: (1) eligibility criteria (ie, explicit reporting of inclusion and exclusion criteria); (2) description of treatment (ie, explicit reporting of dose, frequency, and length of treatment); (3) deviations in treatment (ie, explicit reporting of differences between the definition of treatment and the actual or received treatment); (4) causal contrast or estimand (ie, explicit reporting); (5) primary outcome (ie, explicit reporting of an outcome to be the main or primary outcome in the methods section of the paper); and (6) confounding factors (ie, explicit reporting of the confounding factors accounted for, as part of the methods or results sections in the main text or in the online supplemental material). We also noted whether the reports explicitly stated the use of a

**Table 1 | General characteristics of included reports of non-randomised studies (n=200)**

Characteristics	No (%)
<b>Research question</b>	
Population:	
Patients with chronic disease	118 (59.0)
Healthy individuals with acute disease	55 (27.5)
Patients with acute disease and chronic conditions	27 (13.5)
Drug approval or regulation:	
Drugs with long standing approval	181 (90.5)
Recently approved drugs	19 (9.5)
Interventions:	
Start of treatment	88 (44.0)
Static strategy	43 (21.5)
Dynamic strategy	32 (16.0)
Length of treatment	11 (5.5)
Delay to start of treatment	7 (3.5)
Dose reduction	3 (1.5)
Stopped treatment	2 (1.0)
>1 intervention	12 (6.0)
Other	2 (1.0)
Comparators:	
Usual care or no treatment	72 (36.0)
Other active treatment	78 (39.0)
Different treatment regimen	23 (11.5)
Combination	27 (13.5)
Outcomes:	
Effectiveness and safety	99 (49.5)
Effectiveness	64 (32.0)
Safety	37 (18.5)
<b>Study design and data</b>	
Study design:	
Cohort*	189 (94.5)
Case-control	11 (5.5)
Data used (n=188):	
Routinely collected data	126 (67.0)
Standardised data collection from an existing cohort	17 (9.1)
Standardised data collection for the purpose of the study	14 (7.4)
>1 type of data	11 (5.9)
Other	20 (10.6)
Sources of routinely collected data (n=137):	
Electronic health records	49 (35.8)
Registry	32 (23.3)
Administrative data	30 (22.0)
>1 source	15 (10.9)
Not reported	11 (8.0)

\*Two reports were of target trial emulation studies, as reported by the authors.

reporting guideline (eg, RECORD). We extracted additional data on specific items related to the participants, treatment, outcomes, and confounding (box 1). Online supplemental appendix 2 shows the data extraction instruction sheet.

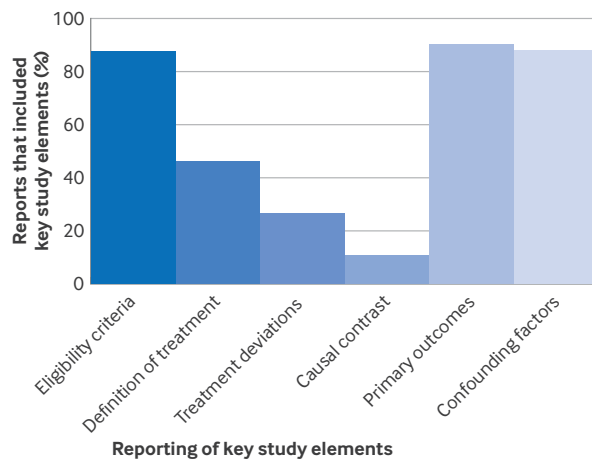
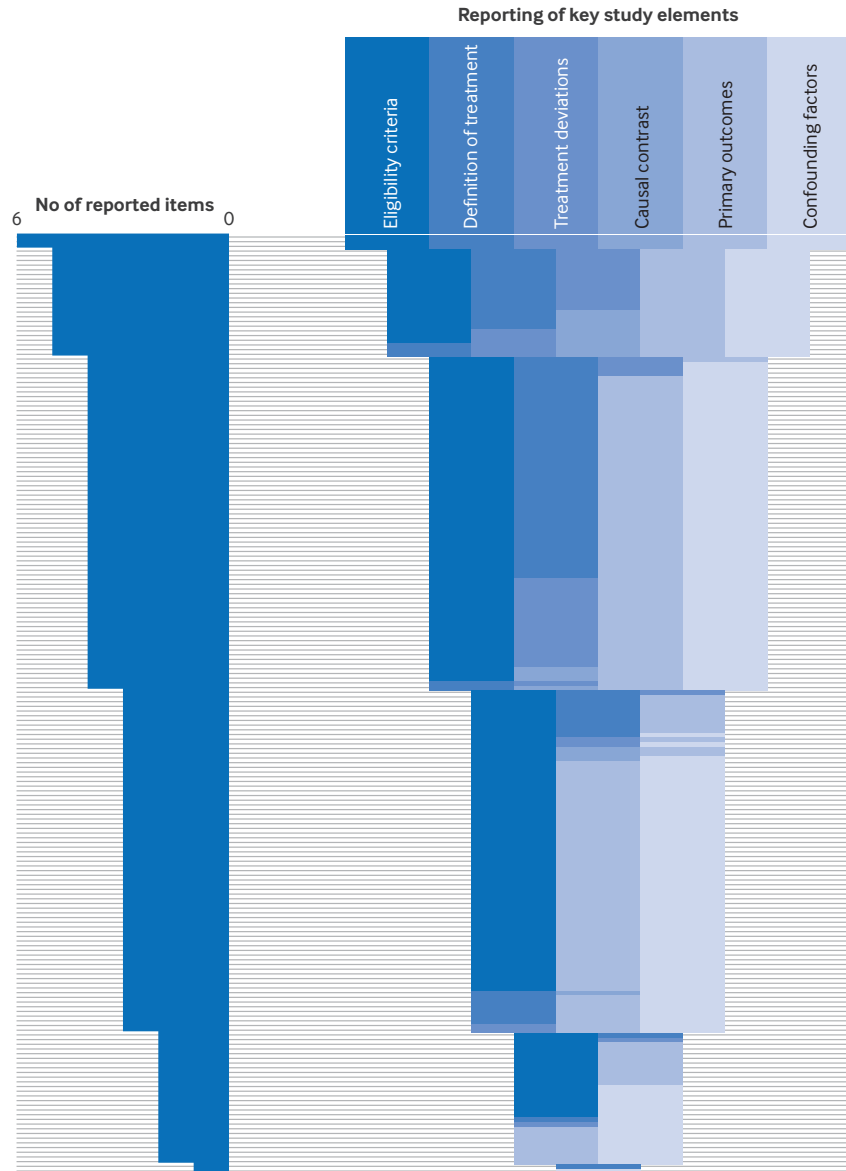


Figure 1 | Reporting of key study elements for each report (n=200). Each horizontal line corresponds to one included report. Top right panel: a specific colour was attributed to each of the six key study elements. The colour band shows which of these items were reported for each included report. The 200 included reports of non-randomised studies were sorted according to the total number of reported items, in decreasing order. Top left panel=distribution of total number of reported items for the 200 included reports. Bottom panel=proportion of reports that reported each element

**Table 2 | Reporting characteristics of the included reports (n=200)**

Reporting characteristics	No (%)
<b>Participants</b>	
Eligibility criteria	174 (87.0)
Sources for selection of participants	189 (94.5)
Use of validation studies of codes and algorithms (n=137)	10 (7.3)
Having no contraindications to any of the treatment arms	25 (12.5)
Sample size calculation	20 (10.0)
Structured sampling method*	41 (20.5)
<b>Treatment</b>	
Description of treatment (explicit reporting of dose, frequency, and length of treatment)	92 (46.0)
Definition for treatment deviations	53 (26.5)
Occurrence of treatment deviations	52 (26.0)
Dealing with treatment deviations (n=64):	
Excluded	22 (34.4)
Censored	20 (31.2)
Included in the allocated (original) arm	14 (21.9)
Included in the received treatment arm	3 (4.7)
Other	5 (7.8)
<b>Causal contrast, outcomes, and confounding factors</b>	
Causal contrast or estimand (n=21, 10.5%):	
Intention-to-treat analysis	12 (57.1)
Per protocol analysis	2 (9.5)
Intention-to-treat and per protocol analysis	7 (33.4)
Primary outcome or outcomes:	
Identified and defined	157 (78.5)
Identified only, not defined	22 (11.0)
Not identified	21 (10.5)
Confounders and covariates	
Identifying confounders and covariates (n=186, 93.0%):	
Listed with no justification	115 (61.8)
Statistical methods	43 (23.1)
Literature	15 (8.1)
>1 method	10 (5.4)
Other	3 (1.6)
Statistical method of identifying confounding factors (n=52, 26.0%):	
Variables with P value <X in univariate analysis†	20 (38.5)
Variables with P value <0.05 in univariate analysis	17 (32.7)
Backward approach in the model	6 (11.5)
Forward approach in the model	2 (3.8)
Other	7 (13.5)
Method to deal with confounding factors:‡	
Matching	73 (36.5)
Matching with propensity score	62 (84.9)
Stratification or regression	178 (89.0)
Stratification or regression with propensity score	3 (1.7)
Inverse probability weighting	29 (14.5)
Inverse probability weighting with propensity score	23 (79.3)
E value	3 (1.5)
Negative control outcomes	2 (1.0)

\*Structured sampling method includes random sampling of participants or centres, or inclusion of a representative sample of the population (eg, based on data from a registry that includes 90% of the population).  
†X is a P value different from 0.05.  
‡Percentages add up to >100 because reports have more than one method.

**Time point alignment and possible related biases**  
We relied on the target trial emulation framework.<sup>16</sup> We determined whether the three time points were identifiable: eligibility (when patients fulfil the eligibility criteria), treatment assignment (when patients are assigned to one of the treatment strategies), and (3) start of follow-up (when outcomes in participants are assessed). We then considered if these time points were aligned and, if not, the biases that could occur, such as bias related to: inclusion of prevalent users (selection bias), post-treatment eligibility (selection bias), immortal time periods, and classification of treatment arms. We also recorded the methods reported to deal with these possible biases. We recorded whether the reports explicitly mentioned these biases. Online supplemental appendix 4 provides a detailed description of biases related to time point misalignment and online supplemental appendix 5 describes the methods to deal with bias related to immortal time periods or classification of treatment arms.

### Data synthesis

The data were synthesised narratively and in tabular formats. We report descriptive statistics with frequencies and percentages for categorical outcomes and medians with interquartile ranges (25th-75th percentiles) for continuous outcomes. We used IBM SPSS Statistics version 21 for the data analysis.

### Patient and public involvement

The study was about the methods of studies and needed methodological and statistical expertise, therefore the involvement of patients and the public was not possible. The results of this study will be disseminated on the research institute's website and publicised on social media.

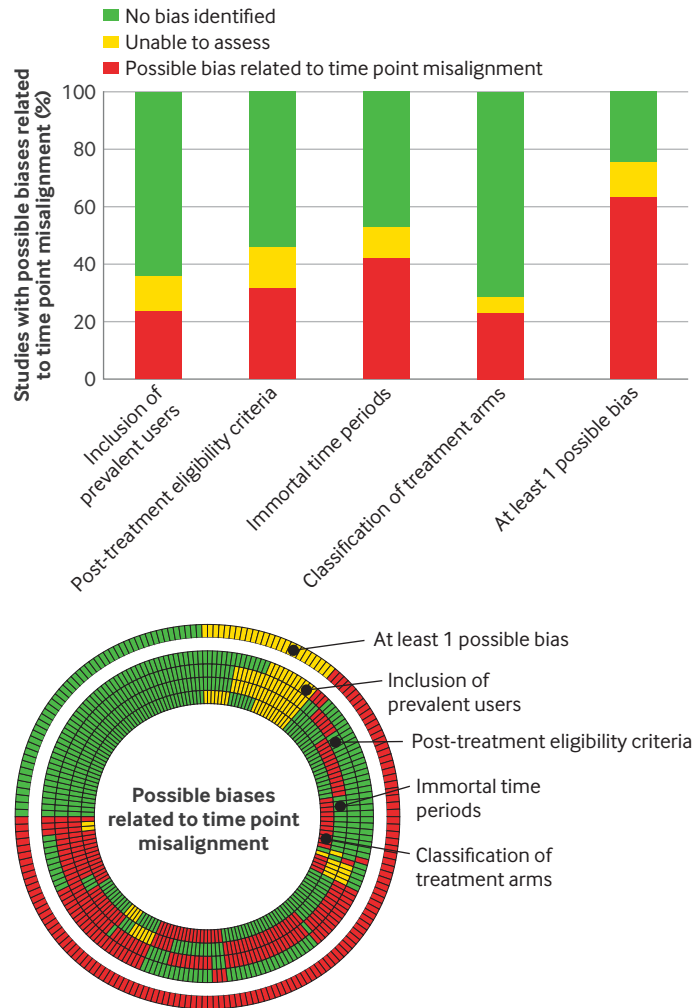
### Results

Of the 26 123 reports retrieved from the search and randomly ordered, we screened 6800 records and identified 462 reports of non-randomised studies that assessed the effectiveness or safety, or both, of drug treatments. Of these, 57% were excluded: 148 (32%) reports had no comparator and 114 (25%) reports did not account for confounding factors. Overall, we extracted 200 reports of non-randomised studies that had a comparator and accounted for confounding factors. Online supplemental appendix 6 shows a flowchart of the selection of the reports included in our study and online supplemental appendix 7 lists the reports of the 200 non-randomised studies.

### General characteristics

#### Study characteristics

The reports mainly assessed treatments in the specialties of oncology (n=54, 27%), infectious diseases (n=41, 21%), and cardiology (n=24, 12%).



**Figure 2 | Presence of possible biases related to time point misalignment (n=200).** Possible biases for each report of non-randomised studies are summarised for bias related to: inclusion of prevalent users, post-treatment eligibility criteria, immortal time periods, and classification of treatment arms. Each spoke represents one report. The bricks are a visual representation of the possible bias related to time point misalignment: possible bias, could not assess, or no bias. Every concentric circle represents one of the biases, with bias related to inclusion of prevalent users being the furthest circle from the centre, and bias in classification of treatment arms is the central circle. The most external circle represents an overview of the possible biases for each report (at least one possible bias exists, could not assess, and when no bias exists). The histogram summarises the possible biases for the 200 reports

The studies were conducted in Central and East Asia, and the Pacific region (n=82, 41%), Europe (n=59, 30%), and North America (n=50, 25%). Most trials were published in specialised medical journals (n=15, 76%) and only 47% (n=94) included a statistician or methodologist among the authors or in the acknowledgements. Online supplemental appendix 8 shows a summary table of the general characteristics of the included reports of non-randomised studies.

*Research question*

The research question was explicitly stated in 83% of reports (n=166). The population in most reports were patients with chronic diseases (n=118, 59%). Most studies assessed drugs with long standing approval (n=181, 91%). Diverse types of treatment strategies were assessed, such as the start of treatment (n=88,

44%), static treatment strategies (n=43, 22%) (eg, antibiotic treatment until discharge), dynamic treatment strategies (n=32, 16%) (eg, treat-to-target strategies), and different lengths of treatment (n=11, 6%). The comparator was mainly an active comparator (n=78, 39%) and usual care or no treatment (n=72, 36%). Also, half of the reports focused on both effectiveness and safety (n=99, 50%). Table 1 provides a summary of the characteristics of the reports of non-randomised studies.

*Study design and data*

The reports were mostly of cohort studies (n=189, 95%). Most reports used routinely collected data (n=126/188, 67%) from various sources, such as electronic health records (n=49/137, 36%), registries (n=32/137, 23%), and administrative data (n=30/137, 22%) (table 1). More than half of the

studies were conducted in a tertiary setting (n=102, 54%) (online supplemental appendix 8). Of those that reported the number of centres, half were conducted in one centre (n=72/135, 53%). The median number of participants included was 949 (interquartile range 288-9881) and the median number of participants analysed was 633 (216-7708). Median follow-up time was 17.6 months (3-43 months). The study was funded by governmental sources in 31% of reports (n=52/168) and received no funding in 24% (n=41/168). In 72% of reports (n=141/196), the authors declared no conflict of interest (online supplemental appendix 8).

Only 7% (n=14) of reports mentioned registration in a trial registry and 3% (n=5) had an available protocol. More than half had a data sharing statement (n=123, 62%), with the most common being that data are available on reasonable request (n=61/123, 50%) and that data might be obtained from a third party (n=27/123, 22%). Only 5% (n=9) provided access to the codes or algorithms used to classify interventions and outcomes. Most reports mentioned obtaining ethical approval (n=167/190, 88%). In the abstract, a third of the reports used causal language (n=69, 35%). Online supplemental appendix 8 provides a summary of the characteristics.

#### Reporting characteristics

**Figure 1** shows the reporting of key study elements. Only 2% of studies reported all of the key study elements (n=3). Only 11% (n=21) of reports mentioned adherence to reporting guidelines.

#### Participants

Eligibility criteria and sources for selection of participants were reported in most reports (n=174 (87%) and n=189 (95%), respectively) (**table 2**). Only 7% (n=10/137) of reports based on routinely collected data reported using validation studies of the codes or algorithms applied to select the population. Only 13% (n=25) reported that participants did not have any contraindications to any of the treatment arms (ie, participants should be eligible for all treatment arms). Some reports mentioned sample size calculation (n=20, 10%) and 21% (n=41) reported a structured sampling method of the population.

#### Treatment

Less than half of the reports described (ie, an explicit report of dose, frequency, and length of treatment) the treatment (n=92, 46%) (**table 2**). Deviations in treatment were defined in 27% (n=53) and reported in 26% (n=52) of reports, of which 34% (n=22/64) excluded participants from the analysis.

#### Causal contrast, outcomes, and confounding factors

The causal contrast or estimand was reported in 11% (n=21) of reports (**table 2**). Primary outcomes were identified in 90% of reports (n=179), and were

identified and defined (ie, included details on the method of assessment or on prespecified time points of assessment) in 79% (n=157). Confounding factors were clearly reported in 88% (n=175) of reports, but were mostly listed without justification (n=115/186, 62%). Of the 52 reports that used statistical methods to identify confounding factors, the most common method was choosing variables from the univariate analysis, with a P value <0.05 (n=17/52, 33%) or a different cut-off value (n=20/52, 39%). Several methods were used to account for confounding factors, and in 39% (n=78) of reports more than one method was used. These methods included: matching (n=73, 37%; of which 85% (n=62) used propensity scores); stratification or regression (n=178, 89%); and inverse probability weighting (n=29, 15%; of which 79% (n=23) used propensity scores). Only three reports (2%) mentioned the E value and two (1%) mentioned using negative control outcomes.

#### Time point alignment and possible related biases

In most reports (n=189, 95%), the time points for eligibility, treatment assignment, and start of follow-up were identifiable (online supplemental appendix 9). Figures displaying the three time points were presented as study design diagrams (n=12, 6%) or participant flowcharts (n=51, 26%). The time points were not aligned in 72% (n=143) of reports. Methods were applied in 11% (n=15/143) to deal with possible biases imposed from the misalignment, but biases were only completely dealt with in three reports. Online supplemental appendix 5 lists the methods used to deal with bias.

Overall, 70% (n=140) of reports had at least one possible bias, 6% (n=11) could not be assessed because of inadequate reporting, and only 25% (n=49) had no bias (**figure 2**). We identified bias related to inclusion of prevalent users in 24% (n=47) of reports, post-treatment eligibility in 32% (n=63), immortal time periods in 42% (n=84), and classification of treatment in 23% (n=46).

Post-treatment eligibility included having a specific length of follow-up time (n=28, 14%) or an event during follow-up (n=39, 20%), or both (online supplemental appendix 10). Immortal time periods were related to sequential eligibility criteria (n=27, 14%), a requirement to use the treatment during follow-up (n=43, 22%), and presence of grace periods (n=48, 24%). The median grace period was 1 month (interquartile range 0.1-5.7), and 44% (n=21/48) of reports had a difference in grace periods between treatment arms. The authors explicitly stated the presence of at least one possible bias in only 17% (n=24/140) of reports and all possible biases in 4% (n=5/140) of reports.

## Discussion

### Principal findings

Our study provides a detailed description of the general characteristics, reporting characteristics, and time point alignment and possible related biases, in a representative sample of non-randomised comparative studies assessing drug treatments, indexed in Medline. Most of the reports were of cohort studies conducted in patients with chronic diseases. The reports commonly compared start of treatment with usual care or no treatment, or other active treatments, assessing both effectiveness and safety outcomes. Most of the reports used routinely collected data, but half were conducted in one center or in a tertiary setting. Also, reporting of key study elements, such as a description of the treatment, was often missing. Most of the reports had at least one possible bias. In summary, if we consider the reports of studies that assessed the effectiveness or safety, or both, of drug treatments, only 11% had a comparator, accounted for confounding factors, and had no possible bias related to time point misalignment (or all biases were dealt with).

### Comparison with other studies

Our findings are in agreement with the literature. One study evaluating the reporting quality of cohort studies based on real world data reported limited transparency, and only 24% of studies had an available study protocol and 20% had available raw data.<sup>25</sup> Also, a study found poor reporting of eligibility criteria in cohort studies (14%).<sup>26</sup> Other key areas with suboptimal reporting were variables and their assessment, description of outcomes, statistical methods, biases, and confounding.<sup>25–27</sup> Similarly, in target trial emulation studies, a recent systematic review found that the reporting of how the target trial was emulated was inconsistent across the studies identified.<sup>20</sup> The literature also highlights the presence of biases in non-randomised studies. One study found that 25% of studies were at high risk of selection bias or immortal time bias, or both, and only five of these studies described solutions to mitigate these biases.<sup>24</sup> Also, a scoping review of pharmacoepidemiological studies analysing healthcare data found that 25% of 117 studies mentioned the presence of immortal time bias,<sup>28</sup> which was not the case in our study, because only some reports stated the presence of risk of bias.

### Interpretation of the findings of this study

The use of routinely collected data has been encouraged in the past, with claims of increasing generalisability to the real world and better adaptation in assessing long term outcomes.<sup>3 29</sup> Our findings showed that most non-randomised studies had important limitations related to the accessibility or quality of the available routinely collected data, poor methodological conduct of the studies, or poor

reporting of the studies, or a combination of these factors. In addition, elements that increase confidence in the results (eg, use of validation studies and eligibility to any treatment arm) were rarely reported.

Reports of non-randomised studies also lacked transparency as key study elements were not reported adequately. The reasons behind poor reporting should be questioned because these key elements are the basis of every study and, if completed, would improve the quality, reproducibility, and applicability of research.<sup>30 31</sup>

The possible biases identified can be avoided by thorough planning and explicit reporting to align the three time points. Although aligning the time points of eligibility, treatment assignment, and start of follow-up might sometimes be challenging, many approaches have been proposed in the target trial emulation framework. Moreover, although we only included reports of studies that accounted for confounding factors, we have highlighted the inadequacy of the methods to select these factors (eg, including significant variables from univariate analysis), raising concerns on the presence of bias related to confounding factors. Also, with limited access to study protocols, we could not compare what was planned to what was conducted, raising the possibility of selective outcome reporting.

### Strengths and limitations of this study

Our study had several strengths. We described a representative sample of non-randomised studies indexed in Medline, covering all medical disciplines, without focusing on specific conditions. Also, we included a wide scope of information from the reports, while using rigorous quality control measures for data extraction.

Our study had some limitations. We did not assess all of the biases that might have been present, particularly those that were more relevant to case-control studies, such as inappropriate adjustment for covariates, because we only focused on possible biases related to time point misalignment. Our data extraction and assessments were based on the reporting of studies, which might not always reflect how the study was truly conducted. Also, data extraction was done in duplicate and independently for only 20% of reports and 80% as data verification. We included only those studies indexed in Medline in a specific period of time (three months in the year 2022).

### Study implications for practice

Researchers should take advantage of the availability of real world data to show the true value of real world evidence. One approach is to provide tools for researchers specific to the use of routinely collected data, covering elements of the whole study process (eg, conception, design, and conduct). Also, although reporting guidelines for non-randomised studies (eg, RECORD and ESMO GROW (European Society for



Medical Oncology-Guidance for Reporting Oncology real World evidence)<sup>32</sup>) and protocol harmonisation (eg, HARPER (HARmonised Protocol Template to Enhance Reproducibility)<sup>33</sup>) have been emphasised, we advocate for the development of more comprehensive guidelines that include elements specific to non-randomised studies in comparative effectiveness research because they have distinct methodological problems that require other considerations. Moreover, when applicable, researchers should adopt and properly apply the target trial emulation framework, which highlights the need to clearly define the research question, have a well designed protocol with explicitly stated components, and implement appropriate statistical analysis methods. Tools should be developed to facilitate and guide the planning of studies, urging researchers to explicitly define the research question and comprehensively state the components of the study, along with determining an appropriate statistical analysis plan.

## Conclusions

Non-randomised studies assessing the effectiveness, safety, or both, of drug treatments are becoming increasingly important as a source of evidence. As the literature shifts more towards non-randomised studies, however, specifically with the increase in access to routinely collected data, reassessing their conduct and reporting is important. While recognising the value of real world evidence, the robustness, quality, and transparency of non-randomised studies need to be improved.

## AUTHOR AFFILIATIONS

<sup>1</sup>Université Paris Cité et Université Sorbonne Paris Nord, Inserm, INRAe, Centre for Research in Epidemiology and Statistics (CRESS), Paris, France

<sup>2</sup>Hôpital Hôtel-Dieu Centre d'Épidémiologie Clinique, Paris, France

<sup>3</sup>Cochrane Centre France, Paris, France

<sup>4</sup>Gastroentérologie et Oncologie Digestive, Hôpital Cochin, Paris, France

**Acknowledgements** We thank Colin Sidre, a medical librarian at Université Paris Descartes, for his support in developing the search strategy. We thank Elise Diard from CRESS for developing figures 1 and 2. We thank Carolina Riveros, Malamati Voulgaridou, Elodie Perrodeau, and Gabriel Baron for their contribution to quality control.

**Contributors** SY, PR, and IB contributed to the conception and design of the study. SY, AP, and HB selected studies and extracted data. SY analysed the data. SY, RP, V-TT, and IB interpreted the data. SY and IB wrote the manuscript. All authors reviewed and approved the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. IB is the guarantor. Transparency: The lead author (the guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial, or not-for-profit sectors.

**Competing interests** All authors have completed the ICMJE uniform disclose form at [www.icmje.org/disclosure-of-interest/](http://www.icmje.org/disclosure-of-interest/) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the

submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

**Ethics approval** Ethical approval was not required for this study.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. All datasets are available online on <https://osf.io/sjauh>.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Sally Yaacoub <http://orcid.org/0000-0003-0819-1561>

Viet-Thi Tran <http://orcid.org/0000-0003-1863-6739>

## REFERENCES

- Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet* 2005;365:82–93. 10.1016/S0140-6736(04)17670-8
- Monti S, Grosso V, Todoerti M, *et al*. Randomized controlled trials and real-world data: differences and similarities to untangle literature data. *Rheumatology (Sunnyvale)* 2018;57:vii54–8. 10.1093/rheumatology/key109
- Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* 2016;188:E158–64. 10.1503/cmaj.150653
- Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther* 2017;102:924–33. 10.1002/cpt.857
- Dang A. Real-World Evidence: A Primer. *Pharmaceut Med* 2023;37:25–36. 10.1007/s40290-022-00456-6
- Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215–8. 10.1136/bmj.312.7040.1215
- Feinstein AR. An Additional Basic Science for Clinical Medicine: II. The Limitations of Randomized Trials. *Ann Intern Med* 1983;99:544. 10.7326/0003-4819-99-4-544
- Bain MR, Chalmers JW, Brewster DH. Routinely collected data in national and regional databases--an under-used resource. *J Public Health Med* 1997;19:413–8. 10.1093/oxfordjournals.pubmed.a024670
- Tugwell P, Knottnerus JA, Idzerda L. Has the time arrived for clinical epidemiologists to routinely use "routinely collected data"? *J Clin Epidemiol* 2013;66:699–701. 10.1016/j.jclinepi.2013.04.004
- Deeny SR, Steventon A. Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data. *BMJ Qual Saf* 2015;24:505–15. 10.1136/bmjqs-2015-004278
- Fanaroff AC, Califf RM, Windecker S, *et al*. Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008–2018. *JAMA* 2019;321:1069–80. 10.1001/jama.2019.1122
- Usman MS, Pitt B, Butler J. Target trial emulations: bridging the gap between clinical trial and real-world data. *Eur J Heart Fail* 2021;23:1708–11. 10.1002/ehfj.2331
- Benchimol EI, Smeeth L, Guttman A, *et al*. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12:e1001885. 10.1371/journal.pmed.1001885
- von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9. 10.1016/j.jclinepi.2007.11.008

- 15 Langan SM, Schmidt SA, Wing K, *et al.* The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ* 2018;363:k3532. 10.1136/bmj.k3532
  - 16 Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA* 2022;328:2446–7. 10.1001/jama.2022.21383
  - 17 Bykov K, He M, Franklin JM, *et al.* Glucose-lowering medications and the risk of cancer: A methodological review of studies based on real-world data. *Diabetes Obes Metab* 2019;21:2029–38. 10.1111/dom.13766
  - 18 Bykov K, Patorno E, D'Andrea E, *et al.* Prevalence of Avoidable and Bias-Inflicting Methodological Pitfalls in Real-World Studies of Medication Safety and Effectiveness. *Clin Pharmacol Ther* 2022;111:209–17. 10.1002/cpt.2364
  - 19 Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care* 2012;35:2665–73. 10.2337/dc12-0788
  - 20 Hansford HJ, Cashin AG, Jones MD, *et al.* Reporting of Observational Studies Explicitly Aiming to Emulate Randomized Trials: A Systematic Review. *JAMA Netw Open* 2023;6:e2336023. 10.1001/jamanetworkopen.2023.36023
  - 21 Schulz KF, Altman DG, Moher D, *et al.* CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;152:726–32. 10.7326/0003-4819-152-11-201006010-00232
  - 22 Wang SV, Pinheiro S, Hua W, *et al.* STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ* 2021;372:m4856. 10.1136/bmj.m4856
  - 23 Sterne JA, Hernán MA, Reeves BC, *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919. 10.1136/bmj.i4919
  - 24 Nguyen VT, Engleton M, Davison M, *et al.* Risk of bias in observational studies using routinely collected data of comparative effectiveness research: a meta-research study. *BMC Med* 2021;19:279. 10.1186/s12916-021-02151-w
  - 25 Zhao R, Zhang W, Zhang Z, *et al.* Evaluation of reporting quality of cohort studies using real-world data based on RECORD: systematic review. *BMC Med Res Methodol* 2023;23:152. 10.1186/s12874-023-01960-2
  - 26 Wiehn E, Ricci C, Alvarez-Perea A, *et al.* Adherence to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist in articles published in EAACI Journals: A bibliographic study. *Allergy* 2021;76:3581–8. 10.1111/all.14951
  - 27 Bruggesser S, Stöckli S, Seehra J, *et al.* The reporting adherence of observational studies published in orthodontic journals in relation to STROBE guidelines: a meta-epidemiological assessment. *Eur J Orthod* 2023;45:39–44. 10.1093/ejo/cjaco045
  - 28 Prada-Ramallal G, Takkouche B, Figueiras A. Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. *BMC Med Res Methodol* 2019;19:53. 10.1186/s12874-019-0695-y
  - 29 Welk B. Routinely collected data for population-based outcomes research. *Can Urol Assoc J* 2020;14:70–2. 10.5489/cuaj.6158
  - 30 Cobey KD, Haustein S, Brehaut J, *et al.* Community consensus on core open science practices to monitor in biomedicine. *PLoS Biol* 2023;21:e3001949. 10.1371/journal.pbio.3001949
  - 31 Wang SV, Sreedhara SK, Schneeweiss S, *et al.* Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. *Nat Commun* 2022;13:5126. 10.1038/s41467-022-32310-3
  - 32 Castelo-Branco L, Pellat A, Martins-Branco D, *et al.* ESMO Guidance for Reporting Oncology real-World evidence (GROW). *Ann Oncol* 2023;34:1097–112. 10.1016/j.annonc.2023.10.001
  - 33 Wang SV, Pottegård A, Crown W, *et al.* HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force. *Pharmacoepidemiol Drug Saf* 2023;32:44–55. 10.1002/pds.5507
  - 34 Open Science Framework. Datasets (Epidemiology and reporting characteristics of non-randomized studies of pharmacologic treatment: a cross-sectional study), Available: <https://osf.io/n8mbd>
- Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjmed-2024-000932>).