The UK Biobank Study: Systematic Review of the Reporting Quality in UK Biobank Studies

Emma Anderson, George Davey Smith, Katie Drax, Mark Gibson, Marcus Munafò, Robert Thibault, Benjamin Woolf, and Rebecca Richmond

04 November 2021

# 1 Abstract

[insert]

# 2 Background

There is a general move towards making data freely accessible, or ‘open’, by academic and governmental institutions alike (Vasilevsky, Minnier, Haendel, & Champieux, 2017; World Wide Web Foundation, 2018). Completely open data allows any individual, regardless of their motivation, discipline or skillset, to access and analyse a dataset for free. This accessibility offers researchers new opportunities but there is also uncertainty about how the nature of open data may impact the reliability of research findings. One method by which the reliability of results from open data may be investigated is through replication. Direct replication of secondary data analyses is aided by the publication of an original analysis plan, in which the authors describe their statistical analysis in full. However, recent evidence suggests that it is unclear whether this is sufficient to ensure direct replicability of findings (Hardwicke et al., 2018; Naudet et al., 2018).

To investigate the reliability of findings from open data, we will attempt to replicate all eligible research articles that use UK Biobank (UKB) data. The UKB is a prospective cohort study which collected a wide range of health data from 500,000 people aged 40-69 in the UK during 2006-2010 (UK Biobank Coordinating Centre, 2007). Importantly, UKB data is partially open data because it is only available to bona fide researchers conducting research in the public interest on a cost-recovery basis. The UKB reserves the right to reject access applications for projects it deems inappropriate (UK Biobank Coordinating Centre, 2011). We first conducted a systematic review to (1) assess the completeness of UKB study reporting, (2) identify articles that are suitable for replication, and (3) inform reporting guidelines for other articles using the UKB and similar resources. We will evaluate reproducibility of studies conducted in the UKB given the abundance of published articles using data from this cohort and their contemporary nature, increasing the likelihood that authors will be contactable. Furthermore, they all use the same, professionally collected, maintained and diverse open resource. This guarantees data accessibility and usability and means that differences between the original and replication results cannot be explained by poor data curation or heterogeneity between studies.

Many previous systematic reviews of reporting standards across a range of disciplines used guidelines to assess the completeness of reporting [e.g. Agha et al. (2014); Cook, Levinson, & Garside (2011); Plint et al. (2006)). We used the STROBE (STrengthening the Reporting of Observational Studies in Epidemiology) Statement, which provides reporting guidelines for observational studies, to evaluate articles arising from UKB and to make comparisons between the articles. The STROBE Statement has been endorsed by over 100 journals and, as no major changes have been made to the original version published in 2007 (Cevallos & Egger, 2014), these guidelines are relevant to all studies conducted using the UKB. Given that studies conducted using UKB may be cohort, cross-sectional or ‘nested’ case-control, the review used the combined STROBE Statement for cohort, cross-sectional and case-control designs, and its extensions, to assess the reporting standards in UKB articles.

We aimed to conduct a systematic review of articles reporting studies using data from the UKB with the aim of answering the following questions:

1. What is the quality of reporting of findings from the UKB (in terms of the detail, clarity and completeness with which authors report their design, analysis and results)?
2. Does reporting quality vary across the type of studies conducted?

# 3 Methods

## 3.1 Protocol registration

This study was preregistered before we began to screen the search results (Drax, Richmond, Woolf, Smith, & Munafo, 2019).

## 3.2 Eligibility criteria

Eligible articles are full research articles that report analyses of UKB data. Each article that is excluded at the full-text level of screening will be listed with the reason for exclusion.

Inclusion criteria:

* Publication date: After 30 March 2012. This is the date the UKB was launched and UKB data was first available beforehand.
* Publication type: Full research articles (including simulation and modelling articles).
* Language: English. Relevant studies written in another language may be eligible but, to ensure accurate and precise evaluation, the article would require a professional translation which is not feasible given the review’s resources.
* Data source: Article’s findings were produced from analysis(es) of data from the UKB. This includes studies which reported findings from other studies alongside findings from UKB data.

Exclusion criteria:

* Article findings were produced from analysis(es) of pooled data from multiple studies. These are studies which have used data from participants from the UKB and other datasets in the same analysis.
* Meta-analyses; narrative and systematic reviews; preprints; post-prints; replies; letters to the editor; corrections; and any other publication types that are not full research articles.
* Retracted articles

## 3.3 Information sources and search strategy

Four databases were electronically searched on 15/01/2019 for UKB articles: PubMed, EMBASE, Web of Science Core Collection (WoSCC) and PsycINFO. The search strategy was limited to articles published from 2012 onwards and there was no restriction on language or publication type.

The full search strategy for each database is included in Appendix A of the protocol [osf ref]. To summarise, the terms “UK Biobank”, “UKB”, “UKBiobank”, and “UKB Resource” were included in the “Title”, “Abstract”, “Keywords”, and “All Fields” fields or equivalent of each database.

## 3.4 study selection

We conducted four stages of screening for eligibility. First, using Endnote, KD removed duplicate results, checked the results for completeness, updated the metadata if necessary, and retrieved the full-texts of all results. KD checked completeness using a list of published full research articles that have used the UKB Resource, provided by the UKB on 27 November 2018. Using the full-texts of the results, KD assessed if results were full research articles, and excluded reviews, corrections, conference abstracts, etc.

Second, the results were imported into Rayyan (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016) and screened by BW and KD to determine whether they met the inclusion criteria. If insufficient information was available to determine a paper’s eligibility, the authors were contacted to request the required information. If this information was unobtainable, the article was excluded. Time constraints meant we excluded some articles with insufficient information without contacting the authors.

Third, We classified studies as “traditional epidemiology”, “Mendelian randomisation” and “Other”, according to the type of data they analysed. After being piloted on 20 studies, the classification criteria were:

* Traditional epidemiology – studies investigating associations between exposures and health outcomes using the UK Biobank data alone.
* Mendelian randomisation – studies that conduct Mendelian randomisation analyses on UK Biobank data alone
* Other – articles that do not meet the ‘Traditional epidemiology’ or ‘MR’ classification criteria described above or include analyses of genetic or imaging data.

Articles abstracts were classified independently by MG and KD. Full texts were examined where classification is unclear from the abstract. Any conflicts were discussed by KD and BW and resolved by mutual consent if possible. If not resolved after discussion RR assign the category. Articles could be allocated into more than one classification group. We then excluded all but the traditional epidemiology studies. The Mendialian randomisation studies were not relevant to this paper but we included it as a classification because the data may be used for one of our author’s future projects.

Only including observational epidemiological studies is a major departure from the protocol in which we stated that studies would be “classified according to their data type” during data extraction but that all classifications would be included. This was so we could analyse articles according to their appropriate STROBE extension, such as the strengthening the reporting of genetic association studies (STREGA) statement (Little et al., 2009). Given that it would be too time-consuming to extract data from all studies in all included articles and to design multiple data extraction forms, we decided to extract data and assess the reporting quality of studies that could be assessed using the original STROBE guidelines only, i.e. traditional epidemiology studies.

Fourth, we catergorised the 178 studies classified as observational epidemiological studies into those with cohort, case-control and cross-sectional study designs using the Study Design Form described below. During this process KD also examined the notices attached to the online version of the observational epidemiology articles to identify any retracted articles. None were identified. However, we did identify studies that should be excluded, these studies were:

* Studies incorrectly classified as observational epidemiological studies.
* Studies with inaccessible supplementary material
* Studies we did consider to be observational epidemiological studies but could not consider them to have cohort, case-control or cross-sectional designs - for example those using prediction models

## 3.5 Data collection process

### 3.5.1 Study Design Form

Different STROBE items are relevant for cohort, case-control and cross-sectional study designs. To ensure coders completed the same STROBE items for each article KD and MG independently identified the study designs used by each of the 178 “observational epidemiology” articles by completing the Study Design Form for each article. In the form, extractors also indicated if all supplementary material can be accessed via University of Bristol subscriptions. Any conflicts were discussed by the two extractors and resolved by mutual consent if possible. If not resolved after discussion, RR determined the study designs. The Study Design Form was hosted on Qualtrics but a PDF version can be found at <https://osf.io/jfk24/>.

The articles were randomly assigned an ID number from 1-178. The articles with an ID number between 1 and 80 were selected for the first round of data extraction. KD was assigned as the first extractor to all 80 articles. The second extractor was randomly assigned from MG, RR or BW. We initially randomly selected 80 articles with the hope that we would randomly select more studies once data extraction for the first 80 were complete. Unfortunately we did not have time to assess more studies.

### 3.5.2 Data Extraction Form

The Data Extraction Form was hosted on Qualtrics and piloted twice. This form combined the data items described as part of the “Data Extraction Form” and “Reporting Quality Assessment” in the protocol, but not all were kept. In the first, 5 observational epidemiology articles were piloted by MG, KD and/or RR. Major changes were made so the Data Extraction Form will be completed again for the articles in the first pilot. In the second, KD piloted the changed form on 3 articles. No changes were made so the responses for articles in the second pilot were retained.

The extractors independently completed the Data Extraction Form for each article assigned to them. Initially, any conflicts were discussed by the two extractors and resolved by mutual consent if possible. If not resolved after discussion, RR resolved the conflict. However, given the number of conflicts this could not be sustained. To reduce the time-burden on other authors KD and RR devised a series of “rules” and general guidelines KD could use to resolve conflicts by herself. This means most conflicts were resolved by these “Rules” or KD’s own judgement. An explanation of the rules and guidelines can be found at <https://osf.io/hngcj/>.

### 3.5.3 Other data collection

The items in Table 1 were extracted manually by KD or automatically by Endnote which contained the article’s metadata exported from the database the article was located in.

### 3.5.4 Unclear and Missing Information.

In our protocol we planned to locate and examine publications linked to the article, e.g. related full research articles, comments, corrections, etc, so that we could retrieve any missing items in the Data Extraction Form from them. If the missing information is not contained in a linked publication we planned to contact the article’s authors. However, we realised this would give an inaccurate assessment of the verision of record. Also few articles had linked publications and that there was a lot of missing or unclear information. Therefore, we did not attempt to retrieve any missing items in the Data Extraction Form, from either the authors or linked articles.

# 4 Results

## 4.1 Study selection

[insert]

## 4.2 Study characteristics

Of the 70 included studies 4 contained case-control designs, 33 cohort, and 38 cross-sectional. These numbers add up to more than 70 because some articles contained cross-sectional and cohort studies. UK based researchers led the vast majority with 53 having authors affiliated with institutions in the UK. 3 studies were published in 2014, 6 in 2015, 10 in 2016, 17 in 2017, 30 in 2018, and 4 in 2019.

## 4.3 STROBE Reporting quality

We originally assessed reporting quality separately for all subdivisions of the STROBE items and for the “starred” items associated with any STROBE items that had an asterisk (\*) next to it. For the final analysis we excluded all starred subdivisions because inter-rater reliability was so poor and the applicability of the items was often uncertain. We then summed the subdivisions to indicate the reporting quality of each STROBE item, these results are presented in Figure 1. For completeness, the results for the subdivisions are available in Supplementary Table 2.

Chart

Description automatically generated

We calculated kappa statistics to estimate the inter-rater reliability for each of the STROBE items, these are presented in Table 1.

Table 1. Inter-rater reliability for the STROBE items (continued below)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| STROBE  Item | Kappa | Z statistic | p-value | Matches | Conflicts | Not Applicable/  Unsure |
| 1a | 0.56 | 5.66 | 0 | 52 | 17 | 1 |
| 1b | 0.3 | 2.67 | 0.01 | 49 | 21 | 0 |
| 2 | 0.47 | 4.09 | 0 | 66 | 4 | 0 |
| 3 | 0.32 | 4.99 | 0 | 66 | 4 | 0 |
| 4 | -0.03 | -0.75 | 0.46 | 43 | 27 | 0 |
| 5 | 0 | 0 | 1 | 68 | 2 | 0 |
| 6a | NA | NA | NA | 70 | 0 | 0 |
| 6b | 0.38 | 3.24 | 0 | 67 | 3 | 0 |
| 7 | NA | NA | NA | 70 | 0 | 0 |
| 8 | NA | NA | NA | 70 | 0 | 0 |
| 9 | 0.24 | 2.27 | 0.02 | 41 | 28 | 1 |
| 10 | NA | NA | NA | NA | NA | NA |
| 11 | 0.06 | 1.19 | 0.24 | 51 | 19 | 0 |
| 12a | 0.19 | 1.88 | 0.06 | 57 | 13 | 0 |
| 12b | 0.5 | 6.03 | 0 | 48 | 22 | 0 |
| 12c | 0.37 | 4.54 | 0 | 38 | 29 | 3 |
| 12d | 0.52 | 4.38 | 0 | 64 | 6 | 0 |
| 12e | 0.16 | 1.36 | 0.17 | 25 | 14 | 31 |
| 13a | 0.3 | 3.44 | 0 | 44 | 26 | 0 |
| 13b | 0.29 | 3.35 | 0 | 47 | 23 | 0 |
| 13c | 0.87 | 7.55 | 0 | 67 | 3 | 0 |
| 14a | 0.26 | 2.43 | 0.02 | 55 | 15 | 0 |
| 14b | 0.39 | 3.8 | 0 | 39 | 14 | 17 |
| 14c | 0.37 | 2.72 | 0.01 | 25 | 7 | 38 |
| 15 | 0.3 | 3.51 | 0 | 62 | 8 | 0 |
| 16a | 0.21 | 2.17 | 0.03 | 64 | 6 | 0 |
| 16b | 0.06 | 0.6 | 0.55 | 23 | 19 | 28 |
| 16c | 0.55 | 5.06 | 0 | 38 | 3 | 29 |
| 17 | 0.41 | 4.31 | 0 | 59 | 11 | 0 |
| 18 | 0 | 0 | 1 | 68 | 2 | 0 |
| 19 | 0.18 | 1.53 | 0.13 | 43 | 27 | 0 |
| 20 | 0.09 | 1.11 | 0.26 | 51 | 18 | 1 |
| 21 | 0.51 | 5.16 | 0 | 52 | 18 | 0 |
| 22 | 0.56 | 6.18 | 0 | 61 | 9 | 0 |
| *Note. NA indicates kappa could not be calculated. 10 is all NA because all we judged the item to be not applicable for secondary analyses.* | | | | | | |

## 4.4 Other reporting quality

We collected reporting quality data in addition to the STROBE items. We intended to conduct exploratory analyses using these variables, for example comparing articles that reported using some form of reporting guidelines to those who did not. We did not conduct an exploratory analyses into the relationships between these variabels because subgroup analyses for all of these variables would have very small subgroup samples. Instead we report the descriptive statistics for each variable in Table 2 below.

*Table 2. Descriptive statistics for additional reporting quality variables*

|  |  |
| --- | --- |
| **Variable** | **N (%)** |
| Report shared analysis code = Yes | 2 (2.9) |
| Exposures reported as exposures = Yes | 40 (57.1) |
| Outcomes reported as outcomes = Yes | 51 (72.9) |
| Report analyses registered = Yes but time unknown | 2 (2.9) |
| Report using STROBE guidelines = Yes | 3 (4.3) |
| Cite STROBE if used = Yes | 1 (33.3) |
| Report using other guidelines if not STROBE = Yes | 0 (0) |
| Report UK Biobank credit statement correctly | |
| No | 11 (15.7) |
| Yes -- almost exact | 22 (31.4) |
| Yes -- exact | 37 (52.9) |
| Report UK Biobank variable IDs (%) | |
| Always | 1 (1.4) |
| Most | 2 (2.9) |
| Never | 60 (85.7) |
| Sometimes | 7 (10.0) |
| Has supplementary material = Yes | 49 (70.0) |
| Published as open access | |
| No | 7 (10.0) |
| No but available elsewhere | 11 (15.7) |
| Yes | 52 (74.3) |
| Has a correction = Yes | 3 (4.3) |
| Report UK Biobank application number = Yes | 41 (58.6) |
| Has email address for corresponding author = Yes | 69 (98.6) |
| Has keywords = Yes | 40 (57.1) |
| Report conflict of interest statement = Yes | 62 (88.6) |

# 5 Discussion

## 5.1 Summary

Our results indicate that the STROBE items best reported by UK Biobank studies are those regarding background (item 2), objectives (item 3), explaining design early (item 4), main statistical methods (item 12a), and summarising results (item 18) with nearly all articles reporting each item fully. The good reporting of the main statistical methods is encouraging for the reproducibility of the analyses. Most articles also fully reported information on generalisability (item 21).

Partial reporting dominated items regarding setting, recruitment, location, and dates (item 5); assessment methods (item 8); variable analysis and grouping (item 11); addressing missing data (item 12c); main results and definitions of confounders (item 16a); limitations and bias (item 19); and funder details (item 22). For some items partial reporting is not a major concern because the information is available elsewhere. Although it does not meet the STROBE guidelines, all information for item 5 and 8 is fully reported in the publicly available UK Biobank documentation. For others, partial reporting casts doubt on the reproducibility of the included studies. Authors poorly reported how they handled their variables (item 11) and missing data (item 12c). Also, most reporting of the methods (items 12b and 12e) and results (item 17) from sensitivity, subgroup, interaction, and other analyses was incomplete. Both issues could limit the reproducibility of UK Biobank analyses.

Other items that were poorly reported pose less of a threat to reproducibility but could still be cause for concern. For one, most authors partially reported item 16a. Supplementary Table 2 shows that while nearly all authors fully report their confounder-adjusted estimates most are failing report their unadjusted elements and to justify why they included the confounders they did. This information may be less vital to reproducing analyses than other items, but it does suggest a widespread norm of researchers including confounders without a stated justification. Similarly, the significant lack of reporting for item 9, 19 and 22 suggests it is acceptable for researchers not to provide detailed analysis of potential bias in their studies or report the role of funders in their studies. Lastly, many authors failing to mention the design of their study in their title or abstract (item 1a) could create problems in retrieving studies during searches.

The worst reported items were 12d, 13c, absolute risk estimates (item 14b) and amounts of missing data (item 16c), with most authors reporting no information about them. Again the implications are of varying concern for different items. Figure 1 shows few authors reported addressing loss to follow up (item 12d). This could be because many studies used NHS data linked to the UK Biobank as follow up data so the studies are likely to have low loss to follow up. Although confirming this will require further investigation into the individual studies. The impact of authors’ rarely providing flow diagrams (item 13c) is also tempered by the good results for 13a and 13b, as it suggests authors do report the information just not in a flow diagram.

We designed variables to measure reporting quality of information not covered by the STROBE guidelines. 62 reported a conflict-of-interest statement, 41 provided their studies' UK Biobank application number, and 59 reported the UK Biobank credit statement correctly. However, the reporting was poor for the variables we included to assess the reproducibility of the analyses. Only 2 papers shared any part of their analysis code, 2 reported having registered their analyses though whether these registrations were before or after seeing the data was unclear, and 10 reported at least one UK Biobank variable ID.

## 5.2 Limitations

[plan – see bullets]

* appropriateness of strobe for secondary data analysis - e.g. item 10 excluded completely
* kappa statistics -- limitations of double coding for something so subjective
* deviations from systematic review method & protocol
* reverse engineering of reporting guidelines = inappropriate to use as reporting assessment?
  + Releasing ‘adherence checklists’ with reporting guidelines would help reporting assessments

## 5.3 Conclusions

[plan see bullets]

* Good reporting of statistical methods will aid replications of the main results. Comparing replication results to the original may be difficult given the poor reporting of unadjusted estimates. The lack of shared code will also make replications more difficult.

# 6 Funding Source

This review is being funded by the John Climax Benevolent Fund. The funder will support the conduct of the review by paying KD’s stipend. The funder had no input on any aspect of the project, such the protocol design, data collection, data analysis nor interpretation or publication of results.

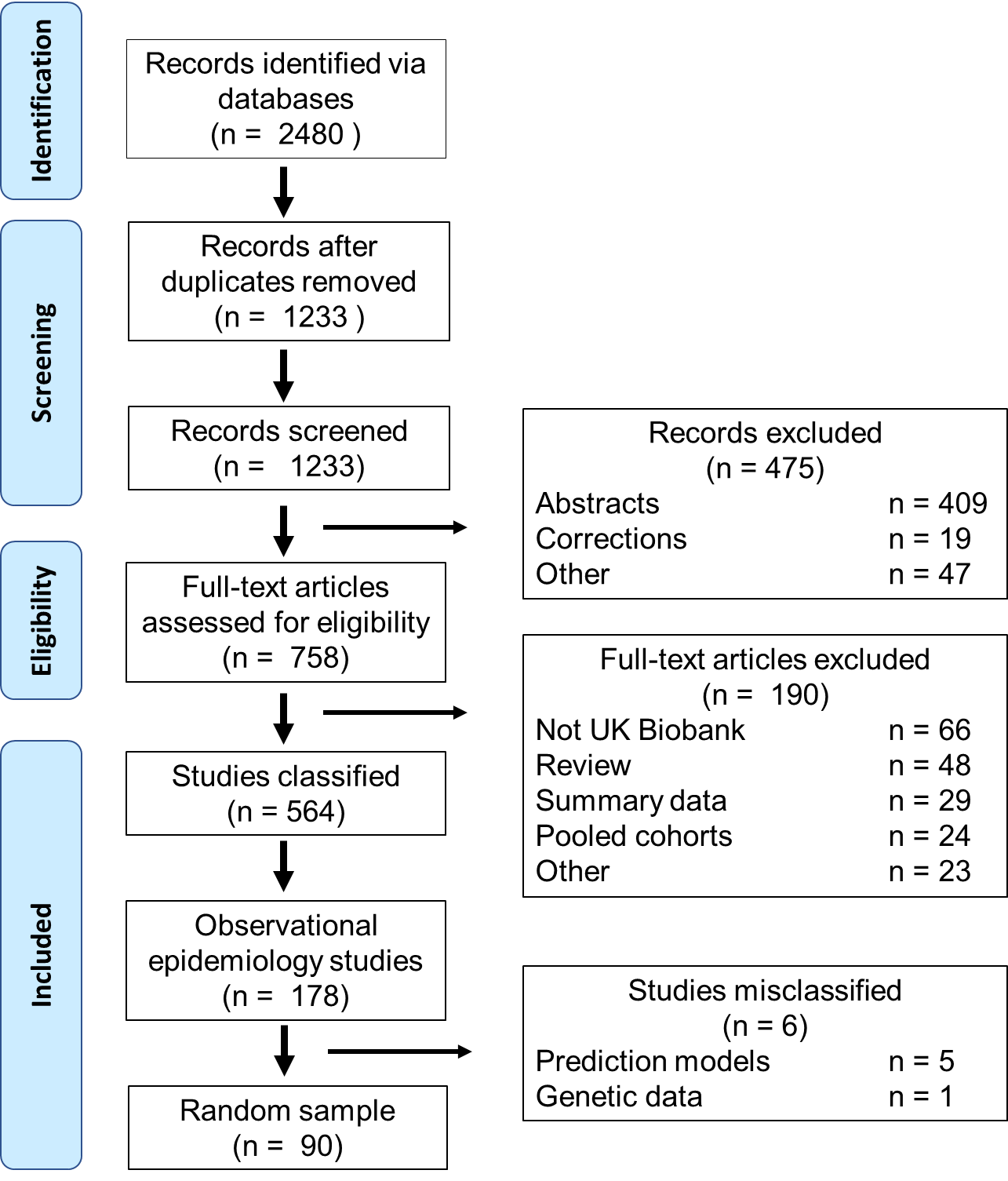
# 7 Conflicts of Interest

There are no conflicts of interest to report.

# 8 Supplementary Material

## 8.1 Supplementary Figure 1

*Supplementary Figure 1. Flow diagram of included articles*



## 8.2 Supplementary Table 2

Supplementary Table 2. Descriptive statistics for reporting quality of subdivisions of all STROBE items

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **STROBE Item** | **No** | **Partially** | **Partially-External** | **Yes** |  | **Not Applicable** |
| **1a** | 31 | 5 | 0 | 34 |  | 0 |
| **1b\_i** | 0 | 13 | 0 | 57 |  | 0 |
| **1b\_ii** | 0 | 12 | 0 | 58 |  | 0 |
| **2\_i** | 0 | 0 | 0 | 70 |  | 0 |
| **2\_ii** | 0 | 1 | 0 | 69 |  | 0 |
| **3** | 1 | 0 | 0 | 69 |  | 0 |
| **4** | 0 | 2 | 0 | 68 |  | 0 |
| **5\_i** | 0 | 32 | 2 | 36 |  | 0 |
| **5\_ii** | 4 | 1 | 0 | 65 |  | 0 |
| **5\_iii** | 22 | 1 | 0 | 47 |  | 0 |
| **5\_iv** | 17 | 21 | 0 | 31 |  | 1 |
| **5\_v** | 8 | 5 | 0 | 25 |  | 32 |
| **5\_vi** | 7 | 23 | 0 | 10 |  | 30 |
| **6a\_i\_cc** | 0 | 0 | 0 | 4 |  | 66 |
| **6a\_i\_coh\_cs** | 5 | 6 | 0 | 55 |  | 4 |
| **6a\_ii\_cc** | 0 | 0 | 0 | 4 |  | 66 |
| **6a\_ii\_coh\_cs** | 0 | 1 | 1 | 64 |  | 4 |
| **6a\_iii\_cc** | 0 | 1 | 0 | 3 |  | 66 |
| **6a\_iii\_coh\_cs** | 1 | 2 | 0 | 7 |  | 60 |
| **6a\_iv\_cc** | 3 | 0 | 0 | 1 |  | 66 |
| **6a\_iv\_coh** | 0 | 7 | 1 | 24 |  | 38 |
| **6b\_i\_cc** | 0 | 0 | 0 | 1 |  | 69 |
| **6b\_i\_coh** | 0 | 0 | 0 | 0 |  | 70 |
| **6b\_ii\_cc** | 0 | 0 | 0 | 1 |  | 69 |
| **6b\_ii\_coh** | 0 | 0 | 0 | 0 |  | 70 |
| **7\_i** | 0 | 4 | 0 | 66 |  | 0 |
| **7\_ii** | 1 | 1 | 0 | 68 |  | 0 |
| **7\_iii** | 0 | 0 | 0 | 0 |  | 70 |
| **7\_iv** | 1 | 15 | 3 | 51 |  | 0 |
| **7\_v** | 4 | 1 | 1 | 44 |  | 20 |
| **7\_vi** | 2 | 1 | 26 | 11 |  | 30 |
| **8\_i** | 0 | 16 | 6 | 48 |  | 0 |
| **8\_ii** | 0 | 45 | 12 | 13 |  | 0 |
| **8\_iii** | 0 | 0 | 0 | 0 |  | 70 |
| **9** | 28 | 1 | 0 | 41 |  | 0 |
| **10** | 0 | 0 | 0 | 0 |  | 70 |
| **11\_i** | 3 | 26 | 3 | 38 |  | 0 |
| **11\_ii** | 1 | 8 | 1 | 52 |  | 8 |
| **11\_iii** | 29 | 28 | 1 | 4 |  | 8 |
| **12a** | 0 | 8 | 0 | 62 |  | 0 |
| **12b\_i** | 5 | 10 | 0 | 23 |  | 32 |
| **12b\_ii** | 5 | 19 | 0 | 17 |  | 29 |
| **12c** | 18 | 30 | 0 | 22 |  | 0 |
| **12d\_cc** | 0 | 0 | 0 | 1 |  | 69 |
| **12d\_coh** | 28 | 0 | 0 | 3 |  | 39 |
| **12d\_cs** | 0 | 0 | 0 | 0 |  | 70 |
| **12e** | 4 | 12 | 0 | 32 |  | 22 |
| **13a** | 1 | 23 | 0 | 46 |  | 0 |
| **13b** | 7 | 9 | 0 | 54 |  | 0 |
| **13c** | 55 | 0 | 0 | 15 |  | 0 |
| **14a\_i** | 0 | 4 | 0 | 66 |  | 0 |
| **14a\_ii** | 0 | 14 | 0 | 56 |  | 0 |
| **14b** | 34 | 12 | 0 | 9 |  | 15 |
| **14c\_coh** | 2 | 2 | 0 | 28 |  | 38 |
| **15\_cc** | 0 | 0 | 0 | 4 |  | 66 |
| **15\_coh** | 1 | 3 | 0 | 29 |  | 37 |
| **15\_cs** | 2 | 3 | 0 | 33 |  | 32 |
| **16a\_i** | 38 | 6 | 0 | 24 |  | 2 |
| **16a\_ii** | 1 | 5 | 0 | 62 |  | 2 |
| **16a\_iii** | 4 | 6 | 0 | 59 |  | 1 |
| **16a\_iv** | 1 | 3 | 0 | 65 |  | 1 |
| **16a\_v** | 48 | 18 | 0 | 3 |  | 1 |
| **16b** | 6 | 16 | 0 | 32 |  | 16 |
| **16c** | 49 | 1 | 0 | 2 |  | 18 |
| **17\_i** | 2 | 3 | 0 | 33 |  | 32 |
| **17\_ii** | 3 | 13 | 0 | 26 |  | 28 |
| **17\_iii** | 2 | 15 | 0 | 31 |  | 22 |
| **17\_iv** | 0 | 3 | 0 | 9 |  | 58 |
| **18** | 0 | 0 | 0 | 70 |  | 0 |
| **19\_i** | 0 | 1 | 0 | 69 |  | 0 |
| **19\_ii** | 30 | 28 | 0 | 12 |  | 0 |
| **20** | 0 | 15 | 0 | 55 |  | 0 |
| **21** | 17 | 3 | 0 | 50 |  | 0 |
| **22\_i** | 13 | 0 | 0 | 57 |  | 0 |
| **22\_ii** | 44 | 0 | 0 | 22 |  | 4 |
| **22\_iii** | 54 | 0 | 0 | 16 |  | 0 |
| **22\_iv** | 58 | 1 | 0 | 11 |  | 0 |

# 

# 9 References

Agha, R. A., Camm, C. F., Doganay, E., Edison, E., Siddiqui, M. R. S., & Orgill, D. P. (2014). Randomised controlled trials in plastic surgery: A systematic review of reporting quality. *European Journal of Plastic Surgery*, *37*(2), 55–62. <https://doi.org/10.1007/s00238-013-0893-5>

Cevallos, M., & Egger, M. (2014). STROBE (STrengthening the Reporting of Observational studies in Epidemiology). In D. Moher, D. G. Altman, K. F. Schulz, I. Simera, & E. Wager (Eds.), *Guidelines for Reporting Health Research: A User’s Manual* (pp. 169–179). <https://doi.org/10.1002/9781118715598.ch17>

Cook, D. A., Levinson, A. J., & Garside, S. (2011). Method and reporting quality in health professions education research: A systematic review. *Medical Education*, *45*(3), 227–238. <https://doi.org/10.1111/j.1365-2923.2010.03890.x>

Drax, K., Richmond, R., Woolf, B., Smith, G., & Munafo, M. (2019). *UK Biobank Study - Systematic Review*. <https://doi.org/10.17605/OSF.IO/SF5VJ>

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., … Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. *Royal Society Open Science*, *5*(8), 180448. <https://doi.org/10.1098/rsos.180448>

Little, J., Higgins, J. P. T., Ioannidis, J. P. A., Moher, D., Gagnon, F., Von Elm, E., … Birkett, N. (2009). STrengthening the REporting of genetic association studies (STREGA)- An extension of the STROBE statement. *Genetic Epidemiology*, *33*(7), 581–598. <https://doi.org/10.1002/gepi.20410>

Naudet, F., Sakarovitch, C., Janiaud, P., Cristea, I., Fanelli, D., Moher, D., & Ioannidis, J. P. A. (2018). Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: Survey of studies published in *The BMJ* and *PLOS Medicine*. *BMJ*, k400. <https://doi.org/10.1136/bmj.k400>

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>

Plint, A., Moher, D., Morrison, A., Schulz, K., Altman, D., Hill, C., & Gaboury, I. (2006). Does the CONSORT checklist improve the quality of reports of ... *Medical Journal of Australia*, *185*(5), 263. <https://doi.org/10.5694/J.1326-5377.2006.TB00557.X>

UK Biobank Coordinating Centre. (2007). *UK Biobank: Protocol for a large-scale prospective epidemiological resource UK Biobank Coordinating Centre Stockport* (No. March; pp. 1–112).

UK Biobank Coordinating Centre. (2011). *Access Procedures: Application and review procedures for access to the UK Biobank Resource* (pp. 1–36).

Vasilevsky, N. A., Minnier, J., Haendel, M. A., & Champieux, R. E. (2017). Reproducible and reusable research: Are journal data sharing policies meeting the mark? *PeerJ*, *5*, e3208. <https://doi.org/10.7717/peerj.3208>

World Wide Web Foundation. (2018). *Open Data Barometer - Leaders Edition*. Washington DC: World Wide Web Foundation.