



BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Accurate diagnosis of iron deficiency in children: an observational study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-046865
Article Type:	Original research
Date Submitted by the Author:	12-Nov-2020
Complete List of Authors:	Sezgin, Gorkem; Macquarie University Faculty of Medicine and Health Sciences, Centre for Health Systems and Safety Research, Australian Institute of Health Innovation Li, Ling; Macquarie University, Westbrook, Johanna; Macquarie University Faculty of Medicine and Health Sciences, Australian Institute of Health Innovation Wearne, Elisabeth; Gippsland Primary Health Network Azar, Denise; Gippsland Primary Health Network McLeod, Adam; Outcome Health Pearce, Christopher; Melbourne East General Practice Network, General Practice Ignjatovic, Vera; The University of Melbourne, Department of Paediatrics; Murdoch Childrens Research Institute Monagle, Paul; University of Melbourne, Paediatrics; Murdoch Childrens Research Institute Georgiou, A; Macquarie University, Australian Institute of Health Innovation
Keywords:	CHEMICAL PATHOLOGY, Anaemia < HAEMATOLOGY, Paediatric pathology < PAEDIATRICS, Clinical chemistry < PATHOLOGY, Paediatric pathology < PATHOLOGY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

## Accurate diagnosis of iron deficiency in children: an observational study

Gorkem Sezgin<sup>a</sup>, Ling Li<sup>a</sup>, Johanna Westbrook<sup>a</sup>, Elisabeth Wearne<sup>b</sup>, Denise Azar<sup>b</sup>,  
Adam McLeod<sup>c</sup>, Chris Pearce<sup>c</sup>, Vera Ignjatovic<sup>d,e</sup>, Paul Monagle<sup>d,e,f</sup>, Andrew  
Georgiou<sup>a</sup>

**Affiliations:** <sup>a</sup> Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia ; <sup>b</sup> Gippsland Primary Health Network, Traralgon, Victoria, Australia; <sup>c</sup> Outcome Health, East Burwood, Victoria, Australia ; <sup>d</sup> Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia; <sup>e</sup> Murdoch Children's Research Institute, Parkville, Victoria, Australia ; and <sup>f</sup> Royal Children's Hospital Melbourne, Parkville, Victoria, Australia

**Address correspondence to:** Gorkem Sezgin, Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia, gorkem.sezgin@mq.edu.au, +61 2 9850 2940

**Word Count:** 2688

**ABSTRACT**

Background and objective: Serum iron results are not indicative of iron deficiency yet may be incorrectly used to diagnose iron deficiency instead of serum ferritin results. Our objective was to determine the association between serum iron test results and iron deficiency diagnosis in children by general practitioners.

Design, setting, patients, main outcome measures: A retrospective observational study of 13,773 children aged 2 to 18 years with serum ferritin and serum iron test results from 137 general practices in Victoria, Australia, between 2008 and 2018. Generalised estimating equation models calculating odds ratios were used to determine the association between serum iron test results (main exposure measure) and iron deficiency diagnosis (outcome measure) in two population groups: 1) iron deplete population, defined as having a serum ferritin <12 µg/L; and 2) iron replete population, defined as having a serum ferritin >30 µg/L.

Results: 2,252 tests were iron deplete and 15,047 were iron replete. Iron deplete children were less likely to be diagnosed with iron deficiency if they had normal serum iron levels (AOR: 0.70; 95% CI 0.56-0.87). Iron replete children had greater odds of an iron deficiency diagnosis if they had low serum iron results (AOR: 3.96; 95%CI 3.26-4.82). Other contributors to an iron deficiency diagnosis were female sex and having anaemia.

Conclusion: Serum ferritin alone remains the best means of diagnosing iron deficiency. Reliance on serum iron test results by general practitioners is leading to significant over- and under-diagnosis of iron deficiency in children.

**STRENGTHS AND LIMITATIONS OF THIS STUDY**

Our study utilises general practice electronic health record data which contains a large sample size to draw our analysis from

Our study contains data from 137 general practices across a large region

Through general practice data, our study is able to correlate diagnostic procedures associated with pathology test orders, and to an extent, treatment thereafter.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

General practice data is limited to information recorded in the electronic health software, which is not always complete for all entry fields

For peer review only

INTRODUCTION

Iron deficiency is the most common nutritional deficiency worldwide and the highest contributing factor to anaemia. The global prevalence of anaemia was estimated as 32.9% in 2010, with the highest burden among young children.[1] Missed diagnosis of iron deficiency can be debilitating for young children, with studies suggesting a link between iron deficiency anaemia and cognitive disorders.[2] Iron deficiency occurs some time before patients become anaemic, marked by a process in which red cells become microcytic and hypochromic.[3] Therefore, anaemia can be prevented with early diagnosis and treatment of iron deficiency.

Diagnosing iron deficiency can be challenging: patients may present with non-specific clinical symptoms, and interpretation of laboratory tests in diagnosing iron deficiency can be difficult.[4] Low serum ferritin levels (a measure of iron stores in the body) are clinically indicative of iron deficiency.[5] In general practice serum ferritin is often ordered as part of an iron panel test, which also contains serum iron, transferrin, and total iron binding capacity. The provision of these accompanying tests can complicate the interpretation of results for diagnosing iron deficiency, particularly due to the presence of serum iron (a measure of circulating iron), which is not clinically indicative of iron deficiency. Presence of inflammation can further complicate the diagnosis of iron deficiency, as serum ferritin levels rise under this condition [6], and can mask an underlying iron deficiency. Where inflammation is suspected, clinicians may either delay testing for iron deficiency until inflammation is resolved, or order a C- reactive protein (CRP) test, an inflammation marker, and interpret ferritin results in context of the CRP result.[7]

Research focusing on diagnostic errors among paediatric patients is limited, particularly in general practice. Failure to order appropriate laboratory tests and incorrect interpretation of test results commonly go unrecognized,[8] and are a considerable cause of diagnostic errors and adverse outcomes to patients.[9] Our aim was to determine whether serum iron tests results (which are often ordered alongside serum ferritin) were mistakenly used in the diagnosis of iron deficiency for paediatric patients in general practice. We therefore assessed the association between serum iron test results and the diagnosis of iron deficiency.

## METHODS

We conducted a retrospective observational study of the association between serum iron test results for children, and their subsequent diagnosis of iron deficiency. We obtained our data from general practice electronic health records from 137 practices in Victoria, Australia.[10]. This study was conducted and reported according to Reporting of Studies Conducted Using Observational Routinely Collected Health Data Statement for Pharmacoepidemiology (RECORD-PE) guidelines.[11]

### Study population and variables

Our sample population consisted of paediatric patients aged between 2 and 18 years with iron panel test results. We included laboratory tests containing both serum iron and serum ferritin test results between 2008 and 2018. Newborns and toddlers were excluded due to uncertain diagnostic criteria for iron deficiency among this population group.



Serum ferritin, serum iron, CRP tests, and red blood cell indices were identified from laboratory test records by searching for related terms in Logical Observation Identifiers Names and Codes (LOINC) [12] and test name fields. We determined ages by subtracting the years between test request date and patient year of birth. Patients under the age of 2 and over 18, and laboratory tests conducted outside of 2008 and 2018 were excluded.

Diagnostic records of iron deficiency were identified by searching SNOMED (Systematized Nomenclature of Medicine) [13] terms for “iron” related to iron deficiency, and additionally, text fields for “iron”, “ID”, and “IDA” related to iron deficiency. Specialist referrals were identified by searching for the terms “referral” and “specialist”. Medications used in the treatment of iron deficiency were identified by searching for commercial and generic names of iron supplements. Due to limitations related to electronic health records, the association between test results and related diagnoses, referrals, and medication are not recorded. The closest iron deficiency diagnosis, referral, and medication within a cut-off of 14 days after the test request was considered as being associated with the test result. Treatment for iron deficiency is recommended to be followed up with a repeat test within a three month period. We therefore identified test requests with a follow up serum ferritin test within a 90 day period, which may indicate that the patient is being treated.

**Inclusion/Exclusion criteria**

To ensure our study population contained only new cases, we excluded cases where a serum ferritin or serum iron test was ordered within the last 12 months. Prior serum ferritin results may influence the diagnoses of iron deficiency due to a variety of reasons, including: continuation of care to assess the effect of any prior treatments, or

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).

re-testing to confirm results. Although re-test interval is recommended as three months, our criteria was purposefully wide to ensure no potential known-cases were included. Likewise, we excluded patients who had a past iron medication or a past diagnosis for iron deficiency or anaemia at any time. We also excluded test results accompanied by high CRP results, as indicated by the pathology provider's report or having a result of  $>10$  mg/ml to account for possible inflammatory diseases.

Thalassemia can often be mistaken for iron deficiency. We therefore excluded patients who had a known thalassemia diagnosis. Only test results containing both serum iron and serum ferritin results were included.

### Statistical analyses

We set out to determine association between serum iron test results and iron deficiency diagnoses in two population groups: (i) in an iron deplete population group; and (ii) in an iron replete population group. Iron deplete was defined as having low serum ferritin levels: denoted as either the pathology provider reporting the serum ferritin results as low (pathological), or the serum ferritin result value being less than  $12\text{ }\mu\text{g/L}$ .<sup>[14]</sup> A patient was considered as iron replete if their serum ferritin result was either denoted as normal or high by the pathology provider, or had a ferritin result value above  $30\text{ }\mu\text{g/L}$ .<sup>[14]</sup> Results for other thresholds ( $<15\text{ }\mu\text{g/L}$  and  $<20\text{ }\mu\text{g/L}$  for iron deplete;  $>50\text{ }\mu\text{g/L}$  and  $>100\text{ }\mu\text{g/L}$  serum ferritin for iron replete) can be found in the supplementary document.

In our iron deplete population group, we tested if having normal serum iron results decreased the odds of a diagnosis; which indicates under-diagnosis. In our iron replete population group, we tested if having low serum iron results increased the odds of an iron deficiency diagnosis; which indicates over-diagnosis (Figure 1). For the purposes

of our study, the term ‘diagnosis’ comprises a diagnostic record, referral, medication, or treatment related to iron deficiency; identified as described under the study population and variables heading within the methods section.

Generalised estimating equations were formed to model the probability of an iron deficiency diagnosis associated with serum iron test results, fitted as a binomial distribution with a logit link function and robust standard error estimations. Clusters by general practice and patient were formed, with exchangeable within group correlation. The model was used to report odds ratios, p-values, and 95% confidence intervals: with iron deficiency diagnosis as the binary outcome variable (diagnosed; not diagnosed). Serum iron test result was our main exposure variable of interest, as a binary variable (pathology provider noted low serum iron levels; pathology provider noted normal or high serum iron levels). Covariates included in our model were age (continuous), sex (binary), serum ferritin result value (continuous), and mean corpuscular volume, mean corpuscular haemoglobin, and mean corpuscular haemoglobin concentration values reported by pathology provider (binary variable as: low result value; normal or high result value), as well as red cell distribution width (binary variable as: high result value; normal or low result value). We also included anaemia status as a covariate, identified as having a haemoglobin concentration below 12 g/dL (120 g/L) for females, and 13 g/dL (130 g/L) for males. All statistical analyses were conducted using Stata MP 16.0 (StataCorp, TX, USA).

**Patient and public involvement**

Patients were not involved in this study.

**Ethics approval**

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

Approval to use the de-identified general practice electronic health record data in research was obtained by the data custodians, Outcome Health, from the RACGP National Research and Evaluation Ethics Committee (NREEC 17–008). Ethics approval for the use of the de-identified data was obtained from Macquarie University Human Research Ethics Committee (5201700872).

## RESULTS

A total of 17,299 test results from 13,773 patients were included in our analyses. Females accounted for 59.6% of the test results, and the median age was 14 years (10–17 years).

A total of 2,252 test results with low serum ferritin results (i.e. iron deplete) were identified (Figure 1). There were 1,480 test results with normal serum iron, of which 396 were diagnosed with iron deficiency. Our analysis revealed a decrease in the odds of an iron deficiency diagnosis by 0.70 (95% CI 0.56–0.87; p-value 0.001) when serum iron results were normal (Table 1). Having a low MCHC result (AOR 0.71; 95%CI 0.55–0.90; p-value 0.006) and increasing serum ferritin result value (AOR 0.93; 95%CI 0.89–0.97; p-value 0.001) also decreased the odds of a diagnosis. Having anaemia also increased the odds of diagnosis, although the association is not strong (AOR 1.27; 95%CI 1.00–1.60; p-value 0.048).

Table 1. Associations between normal iron test results and iron deficiency diagnosis for the iron deplete population. Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between serum iron result and diagnosis, with all co-variates included in the model.

Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.51 (0.44–0.6; <0.001)

Adjusted Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.7 (0.56-0.87; 0.001)
Female sex	1.25 (0.94-1.66; 0.127)
Increasing age	1.02 (1-1.04; 0.101)
Anaemia	1.27 (1-1.6; 0.048)
Low MCV	1.21 (0.89-1.65; 0.217)
Low MCH	1.21 (0.92-1.59; 0.183)
Low MCHC	0.71 (0.55-0.9; 0.006)
High RDW	1.14 (0.85-1.54; 0.38)
Increasing serum ferritin value	0.93 (0.89-0.97; 0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width

A total of 15,047 test results had normal or high serum ferritin levels (i.e. iron replete). Low serum iron results accompanied 1,365 of these tests, with a further 177 iron deficiency diagnoses recorded among these tests (Figure 1). We observed a statistically significant increase in the odds of iron deficiency diagnosis with low serum iron results by 3.96 (95% CI 3.26-4.82; p-value <0.001) (Table 2). Female sex (AOR 1.49; 95%CI 1.25-1.78; p-value<0.001), having anaemia (AOR 1.59; 95%CI 1.23-2.05; p-value<0.001), and having high red cell distribution width (AOR 2.24; 95%CI 1.16-4.35; p-value 0.017) were also associated with increased odds a diagnosis.

Table 2. Associations between low iron test results and iron deficiency diagnosis for the iron replete population. Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between serum iron result and diagnosis, with all co-variables included in the model.

Odds Ratio (95% CI ; p-value)	
Low serum iron result	4.35 (3.77-5.02; <0.001)
Adjusted Odds Ratio (95% CI ; p-value)	
Low serum iron result	3.96 (3.26-4.82; <0.001)
Female sex	1.49 (1.25-1.78; <0.001)
Increasing age	1.02 (1-1.04; 0.033)
Anaemia	1.59 (1.23-2.05; <0.001)
Low MCV	0.79 (0.38-1.66; 0.533)

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

Low MCH	0.84 (0.52-1.35; 0.474)
Low MCHC	1.05 (0.73-1.53; 0.777)
High RDW	2.24 (1.16-4.35; 0.017)
Increasing serum ferritin value	1.004 (1.001-1.01; 0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width

## DISCUSSION

Our study identified both under and over diagnosis of iron deficiency among children treated in general practice related to the interpretation of serum iron results. Children without iron deficiency were nearly 4 times more likely to be erroneously diagnosed (over-diagnosed) when they had low serum iron results. Children with iron deficiency were 0.7 times less likely to be diagnosed (under-diagnosed) when they had normal serum iron levels.

Our study demonstrates serum iron results being erroneously used in the diagnosis of iron deficiency, consequently leading to diagnostic errors. Our results suggest a misconception of the changed role of iron levels in diagnosing iron deficiency; serum iron being confused as the biomarker of iron deficiency where in fact, serum ferritin is. Serum iron tests performed as part of a panel act as a distractor in interpreting the test.

We also observed an increase in the likelihood of an iron deficiency diagnosis with low haemoglobin results. Low haemoglobin levels are indicative of anaemia, and the most common cause of anaemia is iron deficiency.[15] Based on Croskerry's [16] study on biases in decision making, our observation may be a result of 'availability and non-availability' bias; defined as a tendency to make a judgement based on a thought coming more readily to mind due to its frequency. Alternatively, it may a

result of ‘information distortion bias’, defined by Boyle and Purdon [17] as using information to support of a course of action more so than is warranted by the available information. Anaemia is frequently caused by iron deficiency and may have consequently led to the misdiagnosis. We also observed females being more likely to be diagnosed with iron deficiency. Females are known to have a high incidence of iron deficiency in the general population, predominantly thought to be a result of menses.[18] Similar to haemoglobin, female gender could also be resulting in bias in interpretation of the iron panel tests due to its stronger association with the disease.

Laboratory testing related errors are one of the highest contributing factors to diagnostic errors, as demonstrated by Schiff *et al.*’s [19] study. Schiff *et al.* asked clinicians across the US to report three cases of diagnostic errors, and describe their cause, significance, and frequency. They found that laboratory testing related errors, including erroneous interpretation of test results, was the highest contributor to diagnostic errors: with 44% of the errors accounted to it. In Hickner *et al.*’s survey [20], doctors reported uncertainty in test result interpretation 8.3% of the time. Some of the reported challenges related to nomenclature of the tests, such as different tests with the same name, or different names for the same test.

In Australia, a Medicare review taskforce was formed to ensure current clinical practice aligned with evidence and best practice.[21] Iron panel tests were found to be ordered over three times higher than standalone ferritin tests by general practitioners in diagnosing iron deficiency. Our results provide empirical evidence demonstrating the resultant diagnostic errors due to serum iron tests among children. The Medicare taskforce made recommendations to relabel tests in clinically relevant terms, such that a clinician would order ‘iron deficiency studies’ containing ferritin, rather than the

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignment Supérieur (ABES).



current iron panel test with all iron biomarkers. Such an approach may reduce cognitive errors by removing confounding factors out of the decision-making process.

Some limitations should be taken into consideration when interpreting the findings of our study. Our study sample is derived from general practice electronic health records, and the data quality and completeness is therefore largely influenced by general practitioner's use of their software. Diagnostic records are often not complete and many diagnoses, including iron deficiency, are omitted. Although medication prescriptions are well recorded, iron deficiency medication can be obtained without prescription; however the dosage of over the counter iron supplements may not be appropriate for patients with  $<12 \mu\text{g/L}$  serum ferritin, particularly for paediatric patients. We identified the closest iron deficiency related diagnosis and medication records within a cut-off period of 14 days after the test result as our outcome. It is possible that some records were missed due to this time period, although we expect a majority of patients to return to the practice within this timeframe. Nonetheless, it is also possible that the patient may not return to the clinic at all, particularly if their perceived condition has improved. This has additional consequences on one of our outcome criteria: follow up within the next three months. These limitations may have resulted in an underestimate of our outcome measure. Nonetheless, our study objective did not necessitate all of the potential outcomes to be included in the analysis to reach a meaningful conclusion. A sample size large enough to make a meaningful inference from our sample to the population, and no systematic bias in the outcome measure in one category in relation to the other which may have occurred due to missed cases, was required. Based on our final study sample size and the possibility for a systematic bias between our categories being highly unlikely, we are confident both criteria have been achieved and our conclusion is accurate.



Nonetheless, the effect size should be cautiously interpreted, which may have been affected by the missed cases.

Our study also has several strengths. Our outcome was identified through several criteria, in line with the RECORD-PE guidelines: a diagnostic record, medication record, specialist referral, or follow up treatment. Our iron deplete population, having <12 µg/L serum ferritin, requires treatment with medication and follow up.[22] We therefore believe that most patients who are diagnosed and are being treated would be captured with our criteria.

To the best of our knowledge, our study is the first to quantitatively demonstrate the interpretation biases of serum iron test results in general practice, and the resultant error in diagnosing iron deficiency. Iron panel tests are ordered in a similar fashion worldwide in diagnosing iron deficiency, with all four tests being included. Although our results are from Australian general practices within the state of Victoria, it is possible that a similar diagnostic error is occurring in other states/territories and countries.

**Conclusions**

Our study demonstrates cognitive biases in diagnosing iron deficiency in children, which is related to serum iron results and augmented by abnormal haemoglobin test results and female sex. Serum iron tests have consequences for both over- and under-diagnosis, as it appears to lead general practitioners to overlook serum ferritin tests results. In diagnosing iron deficiency, serum iron levels are no longer indicated by clinical guidelines and should not form a basis in establishing iron deficiency status of patients.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

**Funding source:** This study was supported by funding from the Quality Use of Pathology Program, Department of Health, Government of Australia (4-2QFVW4M).

**Competing interests:** The authors have no competing interests relevant to this article to disclose.

**Author contributions:** G.S. conceptualized the study, carried out the statistical analysis, drafted the initial manuscript, and reviewed and revised the manuscript. P.M. and V.I. conceptualized the study, contributed to the drafting of the initial manuscript, provided expertise, and critically reviewed the manuscript for important intellectual content. L.L. provided expertise on the statistical analysis, contributed to the design of the study, and reviewed and revised the manuscript for important intellectual content. A.M. and C.P. developed the data collection instruments, collected data, contributed to the interpretation of the data, and reviewed and revised the manuscript. J.W., A.G., E.W., and D.A. contributed to the design of the study, critically reviewed the manuscript for important intellectual content, and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

**Data Sharing Statement:** Deidentified individual participant data, and ethics approval documents for the collection of data by the data custodians Outcome Health can be made available to researchers who provide a methodologically sound proposal for use in achieving the goals of their approved proposal, on condition of approval of the primary health networks on whose behalf the data is being collected for. Proposals should be submitted to Outcome Health at [admin@outcomehealth.org.au](mailto:admin@outcomehealth.org.au)

REFERENCES

1 Kassebaum NJ, Jasrasaria R, Naghavi M, *et al.* A systematic analysis of global anemia burden from 1990 to 2010. *Blood* 2014;**123**:615–24. doi:10.1182/blood-2013-06-508325

2 Beard JL. Why Iron Deficiency Is Important in Infant Development. *J Nutr* 2008;**138**:2534–6. doi:10.1093/jn/138.12.2534

3 Camaschella C. New insights into iron deficiency and iron deficiency anemia. *Blood Rev* 2017;**31**:225–33. doi:10.1016/j.blre.2017.02.004

4 Soppi ET. Iron deficiency without anemia - a clinical challenge. *Clin Case Rep* 2018;**6**:1082–6. doi:10.1002/ccr3.1529

5 Pasricha S-RS, Flecknoe-Brown SC, Allen KJ, *et al.* Diagnosis and management of iron deficiency anaemia: a clinical update. *Med J Aust* 2010;**193**:525–32.

6 Kell DB, Pretorius E. Serum ferritin is an important inflammatory disease marker, as it is mainly a leakage product from damaged cells. *Metallomics* 2014;**6**:748–73. doi:10.1039/C3MT00347G

7 Cook JD. Diagnosis and management of iron-deficiency anaemia. *Best Pract Res Clin Haematol* 2005;**18**:319–32. doi:10.1016/j.beha.2004.08.022

8 Laposata M. Errors in clinical laboratory test selection and result interpretation: commonly unrecognized mistakes as a cause of poor patient outcome. *Diagnosis* 2014;**1**:85–87. doi:10.1515/dx-2013-0010

9 Plebani M. Diagnostic Errors and Laboratory Medicine – Causes and Strategies. *EJIFCC* 2015;**26**:7–14.

10 Pearce C, McLeod A, Rinehart N, *et al.* What a Comprehensive, Integrated Data Strategy Looks Like: The Population Level Analysis and Reporting (POLAR) Program. *Stud Health Technol Inform* 2019;**264**:303–7. doi:10.3233/SHTI190232

11 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**. doi:10.1136/bmj.k3532

12 Stram M, Gigliotti T, Hartman D, *et al.* Logical Observation Identifiers Names and Codes for Laboratorians: Potential Solutions and Challenges for Interoperability. *Arch Pathol Lab Med* 2020;**144**:229–39. doi:10.5858/arpa.2018-0477-RA

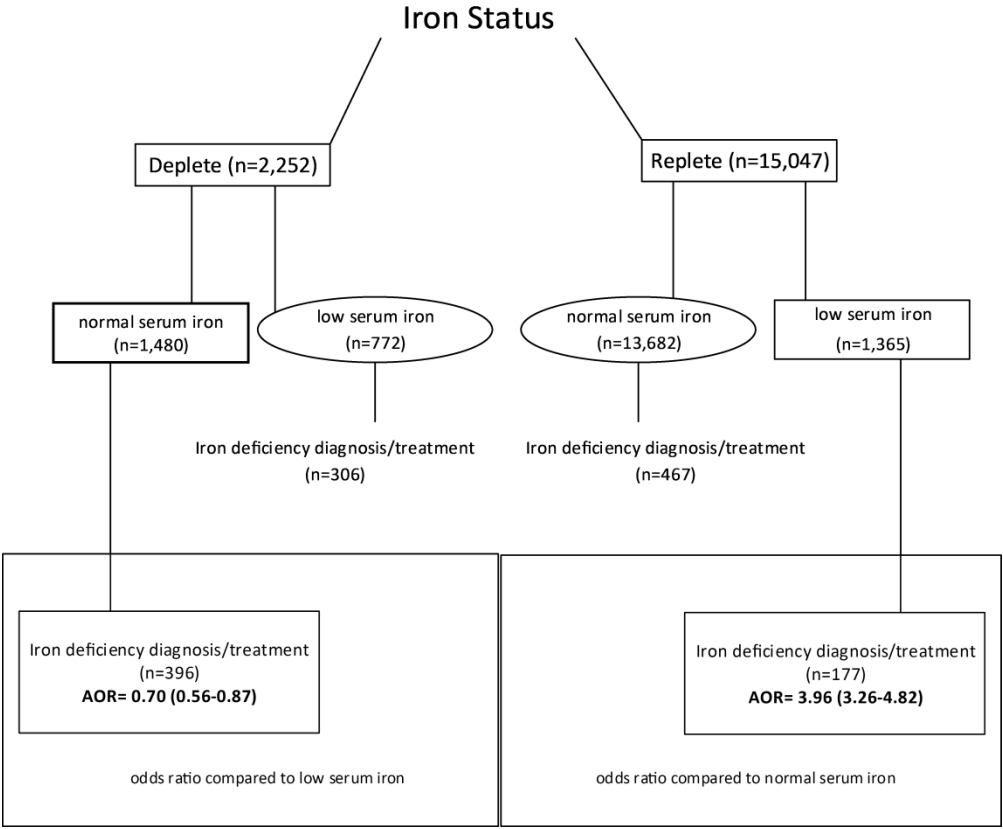
13 Côté RA, Robboy S. Progress in Medical Information Management: Systematized Nomenclature of Medicine (SNOMED). *JAMA* 1980;**243**:756–62. doi:10.1001/jama.1980.03300340032015

14 WHO. Serum ferritin concentrations for the assessment of iron status and iron deficiency in populations. [https://www.who.int/vmnis/indicators/serum\\_ferritin.pdf](https://www.who.int/vmnis/indicators/serum_ferritin.pdf) (accessed 19 May 2020).

15 De Benoist B, World Health Organization, Centers for Disease Control and Prevention (U.S.). *Worldwide prevalence of anaemia 1993-2005 of: WHO Global Database of anaemia*. Geneva: : World Health Organization 2008. [http://whqlibdoc.who.int/publications/2008/9789241596657\\_eng.pdf](http://whqlibdoc.who.int/publications/2008/9789241596657_eng.pdf) (accessed 11 Sep 2018).

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

- 16 Croskerry P. Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. *Acad Emerg Med* 2002;**9**:1184–204. doi:10.1197/aemj.9.11.1184
- 17 Boyle PJ, Purdon M. The information distortion bias: implications for medical decisions. *Med Educ* 2019;**53**:1077–86. doi:10.1111/medu.13919
- 18 Rushton DH, Barth JH. What is the evidence for gender differences in ferritin and haemoglobin? *Crit Rev Oncol Hematol* 2010;**73**:1–9. doi:10.1016/j.critrevonc.2009.03.010
- 19 Schiff GD, Hasan O, Kim S, *et al*. Diagnostic Error in Medicine: Analysis of 583 Physician-Reported Errors. *Arch Intern Med* 2009;**169**:1881–7. doi:10.1001/archinternmed.2009.333
- 20 Hickner J, Thompson PJ, Wilkinson T, *et al*. Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results. *J Am Board Fam Med* 2014;**27**:268–74. doi:10.3122/jabfm.2014.02.130104
- 21 MBS Review Taskforce. Diagnostic Medicine Clinical Committee Final Report. [https://www1.health.gov.au/internet/main/publishing.nsf/Content/mbs-review-2018-taskforce-reports-cp/\\$File/Diagnostic-Medicine-Clinical-Committee-Final%20Report.pdf](https://www1.health.gov.au/internet/main/publishing.nsf/Content/mbs-review-2018-taskforce-reports-cp/$File/Diagnostic-Medicine-Clinical-Committee-Final%20Report.pdf) (accessed 17 Jan 2020).
- 22 Ozdemir N. Iron deficiency anemia from diagnosis to treatment in children. *Türk Pediatri Arş* 2015;**50**:11–9. doi:10.5152/tpa.2015.2337



Characteristics of the study population. Iron deplete was defined as having a serum ferritin result value of <12 µg/L, while iron replete was defined as having a serum ferritin result value of >30 µg/L. Adjusted odds ratio (AOR) and related 95% confidence intervals present the odds of an iron deficiency diagnosis in the iron deplete population associated with normal serum iron results, and in the iron replete population, diagnosis associated with low serum iron results.

### Accurate diagnosis of iron deficiency in children: an observational study

Gorkem Sezgin<sup>a</sup>, Ling Li<sup>a</sup>, Johanna Westbrook<sup>a</sup>, Elisabeth Wearne<sup>b</sup>, Denise Azar<sup>b</sup>, Adam McLeod<sup>c</sup>, Chris Pearce<sup>c</sup>, Vera Ignjatovic<sup>d,e</sup>, Paul Monagle<sup>d,e,f</sup>, Andrew Georgiou<sup>a</sup>

**Affiliations:** <sup>a</sup> Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia ; <sup>b</sup> Gippsland Primary Health Network, Traralgon, Victoria, Australia; <sup>c</sup> Outcome Health, East Burwood, Victoria, Australia ; <sup>d</sup> Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia; <sup>e</sup> Murdoch Children's Research Institute, Parkville, Victoria, Australia ; and <sup>f</sup> Royal Children's Hospital Melbourne, Parkville, Victoria, Australia

**Address correspondence to:** Gorkem Sezgin, Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia, gorkem.sezgin@mq.edu.au, +61 2 9850 2940

Table 1. Associations between normal iron test results and iron deficiency diagnosis for the iron deplete population (serum ferritin <15 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variables included in the model.

Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.47 (0.41-0.54; <0.001)
Adjusted Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.7 (0.58-0.85; <0.001)
Female sex	1.29 (1.02-1.63; 0.031)
Increasing age	1.02 (1-1.04; 0.047)
Anaemia	1.1 (0.9-1.36; 0.357)
Low MCV	1.22 (0.92-1.62; 0.171)
Low MCH	1.24 (0.97-1.59; 0.089)
Low MCHC	0.75 (0.6-0.93; 0.009)
High RDW	1.18 (0.89-1.56; 0.257)
Increasing serum ferritin value	0.9 (0.88-0.93; <0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width

Table 2. Associations between normal iron test results and iron deficiency diagnosis for the iron deplete population (serum ferritin <20 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variables included in the model.

Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.42 (0.38-0.48; <0.001)
Adjusted Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.72 (0.61-0.86; <0.001)
Female sex	1.29 (1.06-1.57; 0.01)
Increasing age	1.02 (1-1.04; 0.019)
Anaemia	1.19 (0.99-1.42; 0.068)
Low MCV	1.29 (0.99-1.67; 0.061)
Low MCH	1.22 (0.98-1.53; 0.079)
Low MCHC	0.77 (0.63-0.94; 0.011)
High RDW	1.13 (0.87-1.49; 0.36)
Increasing serum ferritin value	0.91 (0.89-0.92; <0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).



Table 3. Associations between low iron test results and iron deficiency diagnosis for the iron replete population (serum ferritin >50 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variables included in the model.

<b>Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	4.6 (3.79-5.57; <0.001)
<b>Adjusted Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	4.02 (3.08-5.26; <0.001)
Female sex	1.59 (1.26-2.02; <0.001)
Increasing age	1.03 (1-1.07; 0.036)
Anaemia	1.81 (1.26-2.6; 0.001)
Low MCV	0.36 (0.09-1.44; 0.149)
Low MCH	0.9 (0.44-1.81; 0.758)
Low MCHC	1.02 (0.59-1.75; 0.954)
High RDW	3.65 (1.43-9.32; 0.007)
Increasing serum ferritin value	1 (1-1.01; 0.003)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width

Table 4. Associations between low iron test results and iron deficiency diagnosis for the iron replete population (serum ferritin >100 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variables included in the model.

<b>Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	2.52 (1.82-3.5; <0.001)
<b>Adjusted Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	2.2 (1.36-3.56; 0.001)
Female sex	1.68 (1.15-2.47; 0.008)
Increasing age	1.01 (0.96-1.07; 0.646)
Anaemia	1.61 (0.85-3.06; 0.147)
Low MCV	0.44 (0.05-3.96; 0.464)
Low MCH	0.8 (0.17-3.76; 0.78)
Low MCHC	1.35 (0.54-3.38; 0.518)
High RDW	2.85 (0.7-11.64; 0.143)
Increasing serum ferritin value	1 (1-1; 0.113)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width





# The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE)

Sinéad M Langan,<sup>1</sup> Sigrún AJ Schmidt,<sup>2</sup> Kevin Wing,<sup>1</sup> Vera Ehrenstein,<sup>2</sup> Stuart G Nicholls,<sup>3,4</sup> Kristian B Filion,<sup>5,6</sup> Olaf Klungel,<sup>7</sup> Irene Petersen,<sup>2,8</sup> Henrik T Sorensen,<sup>2</sup> William G Dixon,<sup>9</sup> Astrid Guttman,<sup>10,11</sup> Katie Harron,<sup>12</sup> Lars G Hemkens,<sup>13</sup> David Moher,<sup>3</sup> Sebastian Schneeweiss,<sup>14</sup> Liam Smeeth,<sup>1</sup> Miriam Sturkenboom,<sup>15</sup> Erik von Elm,<sup>16</sup> Shirley V Wang,<sup>14</sup> Eric I Benchimol<sup>10,17,18</sup>

For numbered affiliations see end of article.

Correspondence to: S M Langan  
sinead.langan@lshtm.ac.uk

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2018;363:k3532  
<http://dx.doi.org/10.1136/bmj.k3532>

Accepted: 30 July 2018

In pharmacoepidemiology, routinely collected data from electronic health records (including primary care databases, registries, and administrative healthcare claims) are a resource for research evaluating the real world effectiveness and safety of medicines. Currently available guidelines for the reporting of research using non-randomised, routinely collected data—specifically the REporting of studies Conducted using OBservational Routinely collected health Data (RECORD) and the Strengthening the Reporting of OBservational studies in Epidemiology (STROBE) statements—do not capture the complexity of pharmacoepidemiological research. We have therefore extended the RECORD statement to include reporting guidelines specific to pharmacoepidemiological research (RECORD-PE). This article includes the RECORD-PE checklist (also available on [www.record-statement.org](http://www.record-statement.org)) and

explains each checklist item with examples of good reporting. We anticipate that increasing use of the RECORD-PE guidelines by researchers and endorsement and adherence by journal editors will improve the standards of reporting of pharmacoepidemiological research undertaken using routinely collected data. This improved transparency will benefit the research community, patient care, and ultimately improve public health.

Routinely collected health data are a by-product of the daily operations of healthcare systems, collected independently of specific a priori research questions.<sup>1,2</sup> A broad range of sources (eg, disease registries, health administrative data, quality/safety surveillance databases, electronic health records, and pharmacy data) contain routinely collected data and have both drug exposure and clinical outcomes that are of potential use in pharmacoepidemiology.<sup>3,4</sup>

In pharmacoepidemiology, routinely collected health data are a broadly accepted, necessary, and cost effective resource widely used for evaluating the real world effectiveness and safety of medicines. Studies conducted with routinely collected data are necessary for many reasons. Clinical trials might not be available, or ethical, and could have limitations owing to restrictive inclusion and exclusion criteria. Primary data collection could be costly or infeasible, have limited statistical power to detect safety events, or have durations that prevent the assessment of long term safety outcomes. In many cases, routinely collected health data can be used to provide timely answers and reduce waste in biomedical research when analysing important and novel healthcare issues. The use of routinely collected health data not only leverages existing investment but also could reduce the need for additional investment in de novo data collection.<sup>5,6</sup> Research based on real world evidence, such as routinely collected data, has been conducted on health system planning and evaluation, drug utilisation, comparative drug effectiveness,

## SUMMARY POINTS

- The RECORD reporting guidelines represent the current best practice standard to ensure the clarity and completeness of reporting of non-interventional research using routinely collected health data
- The RECORD-PE statement was derived by use of rigorous methodology and endorsed by the International Society for Pharmacoepidemiology. It is intended to act as a guideline to improve the reporting of pharmacoepidemiological research undertaken using routinely collected health data
- The 15 item checklist should be used in parallel with the RECORD and STROBE guidelines to ensure transparent reporting of pharmacoepidemiology studies using routinely collected health data

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

BMJ: first published as 10.1136/bmj.k3532 on 30 July 2018. Copyright: 2018 BMJ Publishing Group Ltd. All rights reserved. No reuse allowed without permission. See <http://www.bmj.com/permissions> for more information. This article is protected by copyright, for all rights reserved. See <http://www.bmj.com/permissions> for more information. This article is protected by copyright, for all rights reserved.

## RESEARCH METHODS AND REPORTING

epidemiological surveillance, and postmarketing drug surveillance (phase IV studies).<sup>7-9</sup>

Although routinely collected health data are commonly used in pharmacoepidemiological research, these studies are often suboptimally reported.<sup>10-12</sup> Reporting guidelines have been developed for a range of study designs, and represent a minimum standard or items that should be reported in academic manuscripts.<sup>13 14</sup> The main purposes of reporting guidelines are to ensure that readers can easily determine the research question, the methodology used, and the study findings; facilitate understanding of study strengths and limitations, specifically providing insight regarding possible biases; and facilitate replication. Reporting guidelines can also indirectly improve the quality of research by indicating which items to address during study design.<sup>15 16</sup>

The RECORD (REporting of studies Conducted using Observational Routinely collected Data) guideline represents the current best practice standard for the reporting of research using non-randomised routinely collected health data. The guideline was the product of an international collaboration focused on improving the reporting of observational studies using routinely collected data.<sup>1 17</sup> RECORD consists of a checklist of 13 items that supplement or modify the earlier best practice guideline, STROBE (STrengthening the Reporting of OBservational studies in Epidemiology), which focused on the reporting of observational studies.<sup>1 18</sup> The RECORD statement was informed by a systematic review that highlighted major deficiencies in the reporting of studies using routinely collected health data.<sup>11</sup> Since its publication, RECORD has been endorsed by more than 20 major journals (for more information, see [www.record-statement.org](http://www.record-statement.org)).

However, the methodological complexity of pharmacoepidemiological research means that certain reporting requirements are beyond the scope of either RECORD or STROBE. Here, we aimed to extend the RECORD statement to include reporting guidelines specific to pharmacoepidemiological research—that is, the reporting of research focusing on the uses and effects of drugs.<sup>19</sup> This initiative is complementary to existing guidance in the field that mainly focuses on methods for doing (instead of reporting) pharmacoepidemiological research and evaluating the quality of published papers.<sup>20 21</sup> We welcome global community engagement in this endeavour and comments from interested parties by email as these guidelines will be updated periodically.

## The RECORD for Pharmacoepidemiology (RECORD-PE) checklist

### Creation and development of the checklist

We convened a group of international experts in pharmacoepidemiology, routinely collected health data research, reporting guidelines, journalology (that is, the science of publication practices), the joint International Society for Pharmacoepidemiology/International Society for Pharmacoeconomics and Outcomes Research, consensus paper on reporting

requirements to make database studies reproducible,<sup>22</sup> and knowledge users to adapt RECORD for non-interventional pharmacoepidemiological research (RECORD-PE), as presented below. Draft items to be considered in a RECORD-PE checklist were proposed by authors and considered during regular teleconferences and electronic communication, resulting in a draft checklist. A face-to-face meeting was then held in Montreal, Canada, on 25 August 2017. At this meeting, attendees voted on the inclusion of proposed statements and the appropriate wording of these statements, using the approach previously described for the creation of the RECORD statement.<sup>17</sup> Items were included in the checklist if more than 80% of participants agreed on the concept, wording, and message of the item. The draft manuscript and checklist were subsequently revised and circulated to all authors and the RECORD steering committee for comment and approval. It was also circulated to the members of the International Society for Pharmacoepidemiology for comment after completion of the draft and revised accordingly.

### RECORD-PE checklist items

Table 1 shows the complete RECORD-PE checklist, which is organised according to standard manuscript sections and follows the conventions set out in STROBE (and subsequently RECORD).<sup>1 18</sup> The checklist consists of 15 additional items, of which 13 focus on the methods section. Because this checklist is an extension of RECORD, which in turn is an extension of available STROBE items, the statements specific to pharmacoepidemiology are presented next to corresponding STROBE and RECORD checklist items. STROBE additionally has specific checklists for study designs including cohort, cross sectional, and case-control studies. For RECORD and RECORD-PE, we have extended the general STROBE checklist.<sup>10 18</sup> Authors will be expected to consider each checklist item when drafting their manuscript and include items in their manuscript submissions. Below, we provide explanatory text for each RECORD-PE checklist item, organised by manuscript section, and provide a glossary in supplementary table 1.

For content sufficiently covered under STROBE or RECORD, no additional items are provided, although explanatory text regarding particular aspects that might be more pertinent to pharmacoepidemiological studies is provided. All relevant explanations are presented under the respective RECORD-PE item or article section. The RECORD-PE statement is intended for use only in reporting on pharmacoepidemiology studies conducted with routinely collected health data, and represents a minimum standard of reporting for such research in published papers. Such studies include investigation of the use, effectiveness, and safety of drugs or drug eluting devices (eg, drug eluting stents) used in clinical practice. In addition to the widely accepted uses of routinely collected health data for pharmacoepidemiology, in recent years, the concept has emerged that cohort studies of

Table 1 | The RECORD statement for pharmacoepidemiology (RECORD-PE) checklist of items, extended from the STROBE and RECORD statements,<sup>1 18</sup> which should be reported in non-interventional pharmacoepidemiological studies using routinely collected health data

Item No	STROBE items	RECORD items	RECORD-PE items	Page No
Title and abstract				
1	(a) Indicate the study's design with a commonly used term in the title or the abstract. (b) Provide in the abstract an informative and balanced summary of what was done and what was found.	1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. 1.2: If applicable, the geographical region and timeframe within which the study took place should be reported in the title or abstract. 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	—	
Introduction				
Background rationale				
2	Explain the scientific background and rationale for the investigation being reported.	—	—	
Objectives				
3	State specific objectives, including any prespecified hypotheses.	—	—	
Methods				
Study design				
4	Present key elements of study design early in the paper.	—	4.a: Include details of the specific study design (and its features) and report the use of multiple designs if used. 4.b: The use of a diagram(s) is recommended to illustrate key aspects of the study design(s), including exposure, washout, lag and observation periods, and covariate definitions as relevant.	
Setting				
5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.	—	—	
Participants				
6	(a) Cohort study—give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study—give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross sectional study—give the eligibility criteria, and the sources and methods of selection of participants. (b) Cohort study—for matched studies, give matching criteria and number of exposed and unexposed. Case-control study—for matched studies, give matching criteria and the number of controls per case.	6.1: The methods of study population selection (such as codes or algorithms used to identify participants) should be listed in detail. If this is not possible, an explanation should be provided. 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1.a: Describe the study entry criteria and the order in which these criteria were applied to identify the study population. Specify whether only users with a specific indication were included and whether patients were allowed to enter the study population once or if multiple entries were permitted. See explanatory document for guidance related to matched designs.	
Variables				
7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	7.1.a: Describe how the drug exposure definition was developed. 7.1.b: Specify the data sources from which drug exposure information for individuals was obtained. 7.1.c: Describe the time window(s) during which an individual is considered exposed to the drug(s). The rationale for selecting a particular time window should be provided. The extent of potential left truncation or left censoring should be specified. 7.1.d: Justify how events are attributed to current, prior, ever, or cumulative drug exposure. 7.1.e: When examining drug dose and risk attribution, describe how current, historical or time on therapy are considered. 7.1.f: Use of any comparator groups should be outlined and justified. 7.1.g: Outline the approach used to handle individuals with more than one relevant drug exposure during the study period.	
Data sources/measurement				
8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.	—	8.a: Describe the healthcare system and mechanisms for generating the drug exposure records. Specify the care setting in which the drug(s) of interest was prescribed.	

## RESEARCH METHODS AND REPORTING

Table 1 | Continued

Item No	STROBE items	RECORD items	RECORD-PE items	Page No
Bias				
9	Describe any efforts to address potential sources of bias.	—	—	
Study size				
10	Explain how the study size was arrived at.	—	—	
Quantitative variables				
11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why.	—	—	
Statistical methods				
12	(a) Describe all statistical methods, including those used to control for confounding. (b) Describe any methods used to examine subgroups and interactions. (c) Explain how missing data were addressed. (d) Cohort study—if applicable, explain how loss to follow-up was addressed. Case-control study—if applicable, explain how matching of cases and controls was addressed. Cross sectional study—if applicable, describe analytical methods taking account of sampling strategy. (e) Describe any sensitivity analyses.	—	12.1.a: Describe the methods used to evaluate whether the assumptions have been met. 12.1.b: Describe and justify the use of multiple designs, design features, or analytical approaches.	
Data access and cleaning methods				
12	—	12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. 12.2: Authors should provide information on the data cleaning methods used in the study.	—	
Linkage				
12	—	12.3: State whether the study included person level, institutional level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	—	
Results				
Participants				
13	(a) Report the numbers of individuals at each stage of the study (eg, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed). (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram.	13.1: Describe in detail the selection of the individuals included in the study (that is, study population selection) including filtering based on data quality, data availability, and linkage. The selection of included individuals can be described in the text or by means of the study flow diagram.	—	
Descriptive data				
14	(a) Give characteristics of study participants (eg, demographic, clinical, social) and information on exposures and potential confounders. (b) Indicate the number of participants with missing data for each variable of interest. (c) Cohort study—summarise follow-up time (eg, average and total amount).	—	—	
Outcome data				
15	Cohort study—report numbers of outcome events or summary measures over time. Case-control study—report numbers in each exposure category, or summary measures of exposure. Cross sectional study—report numbers of outcome events or summary measures.	—	—	
Main results				
16	(a) Give unadjusted estimates and, if applicable, confounder adjusted estimates and their precision (eg, 95% confidence intervals). Make clear which confounders were adjusted for and why they were included. (b) Report category boundaries when continuous variables are categorised. (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.	—	—	



Table 1 | Continued

Item No	STROBE items	RECORD items	RECORD-PE items	Page No
Other analyses				
17	Report other analyses done—eg, analyses of subgroups and interactions, and sensitivity analyses.	—	—	
Discussion				
Key results				
18	Summarise key results with reference to study objectives.	—	—	
Limitations				
19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	19.1.a: Describe the degree to which the chosen database(s) adequately captures the drug exposure(s) of interest.	
Interpretation				
20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	—	20.a: Discuss the potential for confounding by indication, contraindication or disease severity or selection bias (healthy adherer/sick stopper) as alternative explanations for the study findings when relevant.	
Generalisability				
21	Discuss the generalisability (external validity) of the study results.	—	—	
Other information				
Funding				
22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.	—	—	
Accessibility of protocol, raw data, and programming code				
22	—	22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	—	

RECORD=reporting of studies conducted using observational routinely collected data; RECORD-PE=RECORD for pharmacoepidemiological research; STROBE=strengthening the reporting of observational studies in epidemiology. This table can be downloaded as a separate document in the web appendix; page numbers can be added electronically to the PDF document.

interventions using such data could also be considered attempts to emulate a target trial of the intervention of interest. This concept has been considered helpful within epidemiology and pharmacoepidemiology.<sup>23 24</sup> Routinely collected health data can also help studies with baseline randomisation or pragmatic trials, because the data are frequently collected as part of routine care or health system administration; however, guidance for the reporting of pragmatic trials or trials using these data is beyond the scope of RECORD-PE.

Keywords and medical subject headings terms

The STROBE and RECORD statements do not address the use of specific keywords or medical subject headings (MeSH) terms to identify studies using routinely collected health data.<sup>1 25</sup> There are currently no specific MeSH terms to identify these studies, and researchers use a range of search terms to identify these studies, which is a clear limitation in terms of undertaking systematic reviews and meta-epidemiological research and highlights a need for future research focus.<sup>11</sup>

Title and abstract

No items specific to the RECORD-PE guidelines are needed in addition to the STROBE and RECORD items for the title and abstract. STROBE guidelines advise For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

that an abstract should provide “an informative and balanced summary of what was done and found.”<sup>18</sup> Providing such a summary in the abstract highlights that the study is a pharmacoepidemiological study, and details the research question, the approach taken, and the study findings. Because screening of titles and abstracts is a key step in knowledge syntheses (eg, scoping reviews, systematic reviews), clarity in wording the title and abstract will facilitate appropriate reuse of research findings, thus reducing the waste of research resources. In describing the conduct of a study using routinely collected health data, the RECORD guideline items recommend reporting the type of data used and the name of the database(s), including highlighting whether linked data were used; these specifications are also directly relevant to pharmacoepidemiological studies.<sup>1</sup>

Introduction

No items specific to the RECORD-PE guidelines are needed in addition to the STROBE and RECORD items for the introduction section. The STROBE guidelines advise authors to detail “specific objectives, including any prespecified hypotheses.”<sup>18</sup> The RECORD explanatory paper further recommends that authors be explicit about whether analyses were exploratory or confirmatory, post hoc or prespecified, or a mixture

## RESEARCH METHODS AND REPORTING

of these characteristics. Authors should highlight how interested readers can access the study protocol. These recommendations are needed to enable stakeholders to interpret pharmacoepidemiological studies.

### Methods (study design)

#### RECORD-PE item 4.a

Include details of the specific study design (and its features) and report the use of multiple designs if used.

#### Explanation

STROBE recommends that researchers present key elements of the study design early in the paper.<sup>18</sup> Because routinely collected health data are typically collected in advance of undertaking a study, researchers can theoretically use a range of study designs (eg, self controlled case-series, cohort or case-control studies) or design features (eg, new user designs) depending on the research question. Use of a range of study designs within papers was not directly considered by RECORD. Two aspects of pharmacoepidemiological research warrant an extension to the STROBE statement. Firstly, researchers in the field often use specific study design features (eg, the active comparator new user design) not covered by existing STROBE guidance. Secondly, it is common to use more than one such design or design feature in one publication. Readers should be able to determine which study designs or design features were used. This information will facilitate those readers interested in replicating the methods used.

Study authors should describe such study designs or design features with as much detail as is necessary to make clear to readers what the design involved. If using multiple study designs or design features, authors should comment on which was used for the primary analysis. Authors also should comment on and justify deviations from any study protocol or explicitly state that there was no changes from the protocol.

#### Examples

*Specific design feature (active comparator new user design)*—"A new-user cohort design was used to compare patients initiating dabigatran or rivaroxaban at standard doses for treatment of non valvular AF [atrial fibrillation]. We identified all patients with any inpatient or outpatient diagnoses of AF or atrial flutter, based on *International Classification of Diseases, Ninth Revision (ICD-9)*, coding, who filled their first prescription for either drug from November 4, 2011, when rivaroxaban was approved for AF in the United States, through June 30, 2014. Patients were excluded if they had less than 6 months of enrolment in Medicare Parts A, B, and D, were younger than 65 years, had received prior treatment with warfarin or any NOAC [novel oral anticoagulant], resided in a skilled nursing facility or nursing home, or were receiving hospice care on the date of their cohort-qualifying prescription (index date) . . . Because our purpose was to directly compare dabigatran and rivaroxaban, we did not include a warfarin-treated cohort."<sup>26</sup>

*Specific design feature (interrupted time series analysis)*—"To estimate trend changes in antibiotic prescribing over time, we used segmented linear regression analysis of interrupted time series data, a common quasi-experimental method to assess trend changes after clearly defined events. Separately for each birth week cohort, we estimated the 1-year risk of redeeming at least 1 prescription for an antibiotic, with subanalysis for amoxicillin and penicillin V."<sup>27</sup>

*Specific design feature (drug utilisation/evaluation of the effectiveness of risk minimisation interventions)*—"Following a centralised authorisation within the European Union, on 1 August 2011, dabigatran was marketed in two doses (either 150 or 110 mg bid) for stroke prevention in patients with NVAF [non-valvular atrial fibrillation] and having one or more stroke risk factors. . . . Following early post-marketing reports of bleedings, cautionary recommendations were issued by regulatory authorities. In the safety update from the European Medicines Agency (EMA) (18 November 2011), it was recommended that low doses should be prescribed to elderly patients. Also, this update emphasised the importance of monitoring of renal function, in particular in patients over 75 years. The impact of this safety update, as well as the features of the framework previously described, is assessed in the case study described below."<sup>28</sup>

#### RECORD-PE item 4.b

The use of a diagram(s) is recommended to illustrate key aspects of the study design(s), including exposure, washout, lag and observation periods, and covariate definitions as relevant.

#### Explanation

We recommend the inclusion of a diagram or figure that illustrates the overall study design or timeline for patient inclusion (including key study aspects such as prescription start and end, risk periods, exposed periods, unexposed periods, grace periods, induction periods, and washout periods). Exposure periods are complex in pharmacoepidemiology and are often difficult for readers to follow; this item was not specifically recommended by RECORD. If more than one type of design or analysis is included in the study, a diagram for each is recommended. This allows potentially complex analysis designs, including multiple at-risk periods between or within patients, to be visually summarised in a way that can prevent misinterpretation of paragraphs of text describing the design and implementation.

#### Examples

*Illustration of exposure assessment periods in self controlled case series studies*—a paper by Douglas and colleagues<sup>29</sup> relates to the use of orlistat and risk of acute liver injury and contains a figure describing a typical timeline for a patient in the study (fig 1). The study used a self controlled case series design, and the diagram provides an example of the distribution of unexposed and exposed periods for one patient

(baseline, pretreatment, and multiple time periods of orlistat exposure) and highlights the liver injury risk periods.

*Illustration of exposure assessment periods in cohort studies*—this study by Kim and colleagues<sup>30</sup> on tolclizumab use and the risk for cardiovascular events describes an active comparator new user comparing two biological medicines to treat rheumatoid arthritis (fig 2). The figure shows two key inclusion criteria (use of  $\geq 1$  biological medicine and a diagnosis of rheumatoid arthritis without prior exposure to the specific drug), and the follow-up periods for the two exposure groups, including washout windows. Time periods are clearly marked and censoring events described.

**Methods (setting)**

**RECORD-PE items**

No additional RECORD-PE items are needed to broaden the existing STROBE items.

*Explanation*

As discussed in the RECORD explanatory document, readers need to understand both the reasons and context of data collection to be able assess the potential for information bias, for example, were the data collected for clinical care or billing purposes. Readers also should be able to determine whether the population in the database represents the source population in order to evaluate the generalisability of findings.

**Methods (participants)**

**RECORD-PE item 6.1.a**

Describe the study entry criteria and the order in which these criteria were applied to identify the study population. Specify whether only users with a specific indication were included and whether patients were allowed to enter the study population once or if multiple entries were permitted. See below for guidance related to matched designs.

*Explanation*

When patients are included in a study on the basis of their exposure status to one (or more) drugs, there are likely to be several ways of defining the entry criteria; hence an extension to RECORD is required. We refer here to three levels of population hierarchy described in detail in the RECORD explanatory text—namely, the source population, database population, and study population. In many Scandinavian databases, the source population and database population might represent the same individuals because they comprise the entire population of the specific country.<sup>1</sup> The database population is derived from the study population and contains people who meet eligibility criteria (eg, in the case of primary care databases, they are in primary care practices and have not opted out of inclusion in the database). It is important to provide details of the inclusion and exclusion criteria applied to identify the study population, which includes clearly specification of how exposure status and other

eligibility criteria are defined. Authors should also be clear whether the exclusion criteria are applied before or after selection of study entry dates. Reporting these details would greatly enhance study reproducibility and ability to evaluate the relevance and validity of findings.

A detailed description of matching procedures should be provided. For control sampling, the time axis on which the risk set or incidence density sampling was conducted should also be reported. The procedure for handling cases without eligible controls should be explained (eg, loosening of matching criteria, exclusion). A description should further include whether frequency or individual matching was used, whether matching was done with or without replacement, and the algorithm used (eg, greedy v nearest neighbour matching).

*Examples of selection of the population*

A study by van Staa and colleagues included users of oral corticosteroids who were defined as “permanently registered patients aged 18 years or older who received one or more prescriptions for oral corticosteroids during the period of time from the enrolment date in their practice in GPRD up to the end of the study (December 1997).”<sup>31</sup>

Shin and colleagues report: “The index date for cases was defined by the day of follow-up on which hospital admission occurred. For each case, up to 10 controls were randomly selected using risk set sampling, with controls matched on sex, age ( $\pm 1$  year), cohort entry date ( $\pm 90$  days), and follow-up duration; for one case, the age caliper was widened to 2 years to identify an eligible control.”<sup>32</sup>

**Methods (variables)**

**RECORD-PE item 7.1.a**

Describe how the drug exposure definition was developed.

*Explanation*

Authors should specify clearly how drug exposure code lists were obtained. This information could include which dictionary was searched (eg, the Anatomical Therapeutic Chemical (ATC) classification, or database or country specific codes such as the National Drug Codes in the US)<sup>33</sup> and how these dictionaries or data sources were searched (eg, automated or manual), as well as which drug substance name(s) and what route of administration was used for the search and which ATC classification level was applied. This level of detail allows readers to interpret the completeness and veracity of the exposure definition and permits replication of study findings, and goes beyond the detail of RECORD.

*Example*

*Defining the drug exposure definition*—“The main exposures evaluated were first-trimester exposure to any antidepressants (medications with Anatomical Therapeutic Chemical Classification [ATC] codes

## RESEARCH METHODS AND REPORTING

beginning with N06A) and selective serotonin reuptake inhibitors (SSRIs; medications with ATC codes beginning with N06AB). Exposure was defined according to two sources of information: maternal self-reports (available for offspring born between 1996 and 2012) and dispensation records (available for both parents of offspring born between 2006 and 2012). Information about maternal *self-reported* medication use during the first trimester of pregnancy came from the Medical Birth Register . . . Information about medication use based on *dispensation* records came from the Prescribed Drug Register, which covers all medication dispensations and accompanying prescriptions written in Sweden since July 2005.”<sup>34</sup>

**RECORD-PE item 7.1.b**

Specify the data sources from which drug exposure information for individuals was obtained.

*Explanation*

Authors should make clear what the data source is and whether the electronic records represent issued prescriptions from electronic health records or redeemed prescriptions. Readers also need to understand whether a database contains information on reimbursed prescriptions, out-of-network dispensations, drugs directly dispensed by healthcare providers (samples), or over-the-counter drug use and the completeness of these variables.

*Example*

*Specifying the data sources for drug exposure identification*—“In Denmark, the study population included users of OHAs [oral hypoglycaemic agents] identified in the Aarhus University Prescription Database (AUPD). The database’s catchment area covers the North and Central Regions of Denmark (hereafter referred to as ‘northern Denmark’), with a combined population in mid-2010 of 1.8 million persons, which is about one-third of the Danish population. AUPD captures reimbursed prescriptions redeemed in the regions’ outpatient pharmacies since 1998. In the UK, OHA users were identified from the General Practice Research Database (GPRD), currently also known as the Clinical Practice Research Datalink.”<sup>35</sup>

**RECORD-PE item 7.1.c**

Describe the time window(s) during which an individual is considered exposed to the drug(s). The rationale for selecting a particular time window should be provided. The extent of potential left truncation or left censoring should be specified.

*Explanation*

The time of exposure can be, for instance, the number of days after the start of a first prescription (see also recommendation for figure in RECORD-PE item 4). The number of days may be derived from the number of tablets prescribed, the number of recorded refills, or the number of tablets taken per day for the stated or assumed indication. Frequently, for researchers

using routinely collected health data, there might be no access to data regarding the instructions for taking the medicine. Readers should be able to know whether duration was assumed, based on usual prescribing or directly ascertained from instructions. A description of specific variables generated in association with drug exposure also should be provided. These could include variables capturing information related to dosage information or the total number of prescriptions redeemed within a defined period. Examples include duration, cumulative dose, and recency (that is, current, new, recent, former use). Authors should specify whether only initiators or both initiators and prevalent users are included. They should clearly outline whether they included new users and treatment naive new users by defining the required washout period before a patient is categorised as new user (eg, a new user, by contrast to a treatment naive new user, could be a re-initiator). Authors should detail assumptions about the prescribed daily dose (if not recorded), the duration of prescription coverage, and the length of grace periods used in defining characteristics such as switching, discontinuation, persistence, and adherence.<sup>36</sup>

To account for variation in refill behaviour, refills that are sufficiently close together usually are considered to represent continuous use. A drug is often considered discontinued, in the absence of a new refill, if a prespecified time interval goes past the assumed expiration date supplied in a given prescription (based on the estimated number of days prescribed). This issue could be important in systems such as electronic health records, which have variable coding dates. Readers may want to consult recently published recommendations on how to compute such durations of exposure and how to report methodology.<sup>37</sup> The definition of the exposed period also can be used to assess the outcome of discontinuation. Because prescription or redemption records are an imperfect measure of true drug intake, the algorithms and assumptions used by authors to define exposed time should be reported. Different definitions of the exposed period can be examined by researchers in sensitivity analyses, which should be reported in the manuscript or appendices.

In routinely collected health data, issues of left or right truncation and censoring might also affect the definition of drug exposure and outcome data, and might result in important misclassification and bias—hence, these issues should be reported in publications of routinely collected health data.<sup>38</sup> For example, right truncation could exist in an electronic health records setting because patients transfer between practices or within administrative systems if they become ineligible for insurance coverage; decisions around these aspects should be made clear to the reader.

*Examples of defining exposure time windows*

Paterno and colleagues report: “Exposure was defined as at least one filled prescription for lithium during the first trimester (first 90 days after the date of the



last menstrual period). The primary reference group included women with no lithium or lamotrigine dispensings during the 3 months before the start of pregnancy or during the first trimester. The criterion of no dispensing during the 3 months before the start of pregnancy was imposed to avoid misclassifying as unexposed women who still had medications from an earlier filling available at the start of pregnancy.”<sup>39</sup>

Larivée and colleagues report: “As is true with most healthcare databases, data are left truncated, resulting in the incomplete capture of medical history and previous use of medications. This issue is particularly important in insurance databases, where no information is available outside of the coverage period, and databases such as US Medicare, which only cover patients aged 65 years or older. This truncation is partially mitigated in the CPRD [Clinical Practice Research Datalink] by the transfer of patient records from one practice to another when patients change practices, but such transfers are only feasible between practices that use the same software and it is not possible to link patient records across practices.”<sup>40</sup>

**RECORD-PE item 7.1.d**

Justify how events are attributed to current, prior, ever, or cumulative drug exposure.

*Explanation*

In pharmacoepidemiological studies, it is common to compare rates of adverse events between two drug groups or two or more periods. The adverse event rate is defined as the number of adverse events divided by the total time at risk for a given exposure. It is important to consider and report transparently how the time at risk is defined. The definition of time at risk depends on the pharmacokinetic properties of the drug, the nature of the endpoint of interest, patient related factors, and the plausible hypothesis about the induction period linking the drug and the endpoint. Outcomes can be attributed to drug exposure anywhere along the spectrum from “currently exposed” to “ever exposed” when a binary exposure variable is considered. Another risk attribution model within this spectrum is “on drug plus a lag window.” In this model, an event can be attributed to treatment for a given time period beyond drug discontinuation, thereby allowing time for the drug to continue having a residual effect within the body, or for delayed presentation of the outcome. Different risk attribution models can lead to different conclusions based on the same data. This issue has led to guidelines for rheumatology biological medicine registers, suggesting that research groups use similar risk attribution models when addressing the same research question, to increase comparability of findings.<sup>41</sup> When defining the risk attribution model, researchers may also consider the possibility of protopathic bias, for example, starting a drug to treat early symptoms of the undiagnosed outcome.<sup>42</sup> If protopathic bias is possible, authors should describe it in their manuscript.

*Example*

*Describing how events are attributed to drug exposure—* [In the] “statistical analysis TB [tuberculosis] cases were attributed to anti-TNF [tumour necrosis factor] therapy using two different models: “on drug” (if the patient was actively receiving that drug at the time of diagnosis) and “most recent drug.”<sup>43</sup>

**RECORD-PE item 7.1.e**

When examining drug dose and risk attribution, describe how current treatment, historical treatment, or time on treatment are considered.

*Explanation*

The risk of an adverse event could be influenced by current or historical treatment. Therefore, researchers should look at how current and past drug exposure are considered in analyses. Modelling only current use, either as a binary variable or as current dose, assumes that previous use has no effect on the outcome of interest. Recent use, such as exposure in the past 30 days, allows historical exposure to be considered, but assumes that exposure 29 days ago was important but exposure 31 days ago was not. Selection of an appropriate risk window varies according to the research question and the biological mechanism through which the exposure might lead to the outcome. For example, historical drug exposure is unlikely to influence a hypersensitivity reaction today, while drug exposure months or years ago could contribute to current risk of malignancy.<sup>44</sup> Complex models, such as the weighted cumulative exposure model, allow history of drug use to be modelled flexibly up to the time point when risk is assessed.<sup>45</sup> Although no model is perfect, researchers should consider and report how past exposure was taken into consideration. The approach for handling individuals who were exposed to more than one of the drug exposures of interest during the study period also should be outlined and authors will want to report their approach to time varying confounding. Authors might also want to directly address the issue of depletion of susceptible people or healthy adherer bias.<sup>46 47</sup>

*Examples describing how current and historical exposures are considered*

Movahedi and colleagues report: “Because of uncertainty about mechanisms linking glucocorticoid (GC) exposure and diabetes mellitus (DM), we fitted 7 conventional models, each using a different representation of time-varying GC exposure . . . Models 5 and 6 used continuous time-varying measures of cumulative dose until a given time point, either in the last year or since study entry, respectively. Model 7 categorized cumulative dose since cohort entry, with cutoff points (based on quartiles) at 0, 960, 3,055, and 7,300 mg PED [prednisolone equivalent dose].”<sup>48</sup>

Larivée and colleagues report: “The aim of this study was to describe the challenge of studying the risk of VTE [venous thromboembolism] among first-time users of drospirenone-containing COCs [combined oral

## RESEARCH METHODS AND REPORTING

contraceptives] in a healthcare database and assess the risk among first-time users and restarters . . . The first-time user cohort included all women aged 16-45 years who received a first ever prescription of drospirenone- or levonorgestrel-containing COCs between May 2002 and March 2015. The restarter cohort included those who were restarting a COC after a period of non-use of  $\geq 6$  months.”<sup>40</sup>

**RECORD-PE item 7.1.f**

Use of any comparator groups should be outlined and justified.

*Explanation*

Confounding by indication has also been called an “intractable” bias in epidemiology,<sup>49</sup> because the choice of treatment is guided foremost by the risk of a particular outcome; hence this item has particular relevance to pharmacoepidemiology. This bias might lead to strong confounding, perhaps greater than that arising from associations due to underlying common causes. Moreover, the degree of confounding by indication is difficult to assess, because it is based on an expected prognosis, and that expectation is formed in the mind of an individual health professional dealing with an individual patient.

An appropriate choice of a comparator treatment is key to reducing confounding by indication or severity (see RECORD-PE item 4.a). If there is no comparator group or cohort, authors should state why. Clear description of the use and justification of comparator groups is needed for the assessment of the potential for confounding by indication or severity. Comparators might include alternative drug exposures for the same indication, differing time windows for the same drug exposures, use of historical comparators, unexposed periods, or unexposed individuals. In the absence of randomisation, confounding (by indication) deserves close attention. Therefore, researchers can use more than one comparison group and make inferences based on whether the estimate of association changes in response to better control of confounding (eg, whether an odds ratio based on an active comparator differs with or without adjustment for confounders), and these analyses should be reported in the published paper. Historical active comparator groups can be assembled from routinely collected health data for single arm studies or when a contemporaneous active comparator is not available; any of these approaches should be clearly reported.

*Examples of consideration of comparator drugs*

In a study examining the association between use of antidepressants and pregnancy/offspring outcomes, Sujan and colleagues dealt with confounding by indication by using exposures during non-relevant gestational periods: “To explore whether intrauterine exposure was specifically associated with outcomes over and above maternal depression treatment around the time of pregnancy, associations for maternal first-trimester antidepressant dispensations were compared

with associations for dispensations before pregnancy, while adjusting for measured pregnancy covariates, maternal covariates, and paternal covariates . . . Additionally, the fit of models that included separate parameters for before-pregnancy dispensations and first trimester dispensations were compared with models that included 1 parameter for both dispensation windows. Paternal first trimester antidepressant dispensations were used as a negative control to further explore the role of familial confounding.”<sup>34</sup>

Filion and colleagues report: “Our primary reference group was patients receiving treatment with combinations of oral antidiabetic drugs. With guidelines recommending that incretin-based drugs be used as second-line or third-line therapy, the use of this reference group both reduced potential confounding by indication and provided a clinically relevant treatment comparison.”<sup>50</sup>

**RECORD-PE item 7.1.g**

Outline the approach used to handle individuals with more than one relevant drug exposure during the study period.

*Explanation*

In a cohort study comparing the incidence of an adverse reaction between two or more drug exposures, the method of handling individuals who receive multiple drugs at the start of their exposure period (or drug 1 initially, followed by drug 2) should be described to enable readers to interpret findings (see also RECORD-PE items 7.1.d and 7.1.e). Some studies exclude patients who, according to the prescription or dispensing record, receive more than one treatment at the same time at cohort entry, since attribution of risk is difficult. Censoring is used most often if more than one treatment is experienced during follow-up (see below). Alternatively, exposure to more than one treatment (eg, while switching from an old treatment to a novel treatment) can be handled through time varying exposure, whereby each patient’s person time is segmented, based on the dispensing record, with appropriate methods for handling time dependent confounding (eg, marginal structural models, g estimation). The approach taken by authors should be reported transparently, including defining risk attribution models and lag periods.

*Examples of handling multiple drug exposures*

Use of time varying exposures was reported clearly by Xue and colleagues in an international pharmacovigilance study of women with postmenopausal osteoporosis treated with denosumab: “because a large proportion of new Prolia users may have been previously treated with a bisphosphonate, a new-user design which mitigates biases associated with previous treatments, if adopted, will be based on a very small number of patients. Also, patients with osteoporosis tend to switch treatments over time, so an open-cohort design combined with an ‘as treated’ analysis was selected to account for time-varying medication exposure.”<sup>51</sup>

Wong and colleagues report: “Based on age within five years, sex, and calendar year at use, we matched one clarithromycin user with one or two amoxicillin users. In both groups we excluded patients who had been prescribed clarithromycin up to four years before the date of first antibiotic prescription during the observation period. However, amoxicillin users could be classified as using clarithromycin at a later date. The observation period commenced from the date of the first antibiotic prescription (index date) and ended at the earliest occurrence of the outcome, death, subsequent use of clarithromycin or amoxicillin, or end of study (31 December 2012).”<sup>52</sup>

**Methods (data sources)**  
**RECORD-PE item 8.a**

Describe the healthcare system and mechanisms for generating the drug exposure records. Specify the care setting in which the drug(s) of interest was prescribed.

*Explanation*

The type of healthcare system, the characteristics of patients for whom drug data are available, and the extent to which patients are reimbursed for prescription drugs could affect the likelihood of drug use, and the likelihood of a record of drug use being included in the study data—for example, formulary restrictions could preclude the use of drugs. Understanding this context will be important for interpretation of generalisability or understanding the limitations in availability of drugs in different settings. For example, although Canada has a government funded, universal healthcare system, some provinces reimburse all prescription drug costs, whereas other provinces only cover drug costs in specific age groups or in people with low incomes who are receiving social assistance. In this second group of provinces, supplemental private insurance might be common among non-covered groups, and therefore drug records could be incomplete in provincial health administrative data. This missing information could result in partial ascertainment, because complete prescription records are available only for certain patients. Left truncation also could create bias if public insurance coverage is available only for older patients. Therefore, characteristics of the health system and context of drug data collection should be provided.

*Examples describing the healthcare system within which drugs were prescribed*

Larivée and colleagues report: “Restarters of COCs [drospirenone-containing combined oral contraceptives] can also be misclassified as first-time users in UK databases as oral contraceptives are commonly prescribed at family planning clinics (i.e., community contraception clinics, genitourinary medicine clinics, sexual health clinics). In England, approximately 7.9% of women aged under 16 attended a family planning clinic from 2009 to 2010 and 21.5% of women aged 16–19 years visited a family planning clinic from 2008

to 2009. The CPRD [Clinical Practice Research Datalink] only captures prescriptions issued by the general practitioner, and the availability of oral contraceptives at family planning clinics makes the identification of first-time users difficult. To attempt to overcome this issue, we applied several exclusion criteria, such as the exclusion of all women with previous prescriptions for hormonal contraception issued by the general practitioner and those with diagnostic codes indicating previous use of hormonal contraception. In addition, we excluded all women with a diagnostic or referral code indicating previous visits to a family planning clinic any time before cohort entry.”<sup>40</sup>

Khan and colleagues report: “Using unique patient identifiers, stroke patients identified in the registry were linked to the Ontario Drug Benefits Database, which contains information on antihypertensive drug prescriptions, including the quantity and dates of drugs dispensed as well as the number of days supplied from each prescription, for patients ≥65 years of age. Residents may fill prescriptions at any outpatient pharmacy in Ontario with a maximum copayment of \$6.11 (Canadian) for each prescription after a yearly \$100 (Canadian) deductible. Low income seniors have a \$2 (Canadian) maximum copayment with no yearly deductible. Using postal codes, patients in the registry are also linked to data from the 2006 Canada Census to determine median neighborhood income.”<sup>53</sup>

**Methods (bias)**  
**RECORD-PE items**

No items specific to the RECORD-PE guidelines are needed in addition to the RECORD and STROBE items.

*Explanation*

Biased studies are characterised by systematic error in observed associations, and readers need to understand the approaches taken to manage bias in order for them to judge whether the results are biased. Several potential biases are likely to arise within pharmacoepidemiological studies, and could be more prominent when routinely collected data are used. Recent papers describing use of triangulation could be helpful in discussions about bias.<sup>54</sup> The ROBINS-I (risk of bias in non-randomised studies of interventions) tool might also help to focus discussions of bias.<sup>55</sup> We list below some potential sources of bias in pharmacoepidemiological studies that should be reported.

- Confounding in pharmacoepidemiological analyses might be addressed by design or analytical approaches.<sup>56</sup> Examples of designs or design features include the use of self controlled case series, instrumental variables, regression discontinuity design, and active comparators. Examples of analytical approaches include the use of multivariable regression analysis or propensity scores, although these approaches will not guarantee absence of confounding. The study design or analytical approach used to address confounding should be

## RESEARCH METHODS AND REPORTING

reported, and the authors should note in the discussion the extent to which these methods potentially addressed or failed to address the risk of confounding. If more than one method was used, authors should make clear which approach was the main analysis and which were sensitivity analyses.

- For example, in studies applying the propensity score methodology to handle baseline confounding, the method of propensity score estimation should be reported (eg, logistic regression). These scores can be used to adjust for baseline confounding by several methods: propensity score matching, propensity score stratification, covariate adjustment, and inverse probability of treatment weighting.<sup>57</sup> The specific approach (or approaches) used should be clearly described, together with any attempts to assess the similarities of the resulting treated and untreated groups for each baseline variable.<sup>58 59</sup> If investigators have used trimming approaches, they should discuss the resulting number of excluded participants. In particular, high dimensional proxy adjustment based on propensity score methods has been reported to reduce residual confounding in studies using claims data and if this approach is used, it should be described.<sup>60</sup> Lists of empirically identified potential confounders should be reported in online appendices.<sup>39</sup> If other approaches (such as instrumental variables) were used, these should be clearly described in the publication with similar detail to those outlined for propensity scores.<sup>61</sup>
- Of particular relevance when considering confounding in studies of drug treatment is the type of treatment effect that the non-interventional study is attempting to measure. Types include the intention-to-treat effect (the comparative effect of being assigned to treatment strategies at baseline carried forward, regardless of whether study individuals adhere to the specific treatment) and the as-treated effect (the comparative effect of a drug while it is actually used). In cases in which observational studies based on routinely collected data are designed to emulate a hypothetical or real trial, authors should clearly specify any relevant existing or hypothetical trial that is being emulated. For studies that allow individuals to switch drug treatments as part of the analysis, the role of potential time varying confounders should be considered and reported in the text, along with details of any complex statistical methods applied (such as inverse probability weighting of marginal structural models). For example, in their study of the effect of aspirin on cardiovascular mortality,<sup>62</sup> Cook and colleagues include a directed acyclical graph of the hypothetical relation between aspirin use, cardiovascular death, and intervening

cardiovascular events to show the role of non-fatal cardiovascular events as potential time varying confounders or intermediate steps in the association between aspirin use and cardiovascular mortality.

- Confounding by health status (healthy initiator bias or depletion of susceptibles) can be dealt with at the design stage by aligning the start of observation with treatment initiation (the active comparator new user design; see RECORD-PE item 4.a). The reasoning behind the decision to use such a design feature as well as the extent to which selection bias was or was not addressed should be discussed.
- Information and selection bias due to potential misclassification of drug exposure by prescription or redemption records can be dealt with in a sensitivity analysis that includes different definitions of exposures (eg, when different algorithms are used to define duration of prescribing episodes<sup>63</sup>). An alternative approach is to include only people with more than one prescription or redemption over a given period (eg, within six months), because those individuals with just one prescription or redemption might never have used the treatment. Use of different washout periods to define new episodes of treatment could also affect the interpretation of data. Each issue should be clearly described and discussed in reports of pharmacoepidemiological studies based on routinely collected health data.

### Methods (statistical methods)

#### RECORD-PE item 12a

Describe the methods used to evaluate whether the assumptions have been met.

#### Explanation

In reporting all study designs, authors should discuss whether the underlying study assumptions have been met. A failure of the assumptions being met could undermine the methods used. Determining whether the methods used were appropriate, given the data, is important for readers to understand whether the resulting analysis requires further consideration. A pharmacoepidemiological example is the use of self controlled studies, in which participants act as their own controls, and include the case crossover design and self controlled case series studies.<sup>64</sup> When applying the self controlled case series method, several assumptions must be met to obtain valid and unbiased estimates,<sup>65</sup> for instance, exposure to the drug of interest must not be influenced by a previous outcome event.<sup>66 67</sup> Authors should specify clearly how the assumptions of self controlled or other study designs were evaluated. Detailed guidance on the conduct and reporting of self controlled case series are beyond the scope of these guidelines, but are currently being developed by the SCOPE (self controlled crossover observational



pharmacoepidemiology) initiative.<sup>68</sup> All reports should explain any assumptions that were not evaluated or not met. Discussions should also address the possibility of time related bias (eg, immortal time bias),<sup>69</sup> if these are likely to be a problem.

*Example on testing study assumptions*

Wilson and colleagues report: “We graphed the number of combined endpoint events in the days before and after vaccination. In the self-controlled case series model, the date of vaccination serves as the index date for exposure for each patient. Previous studies have identified that children are at increased risk for systemic reactions at different times from 5–14 days after vaccination . . . Because *a priori* we did not know with certainty the time period following vaccination for which there would be an increased risk of our combined endpoint, we modified the standard self-controlled case series approach by looking for an elevation in risk during each post-vaccination day up to day 17 . . . We then classified days 20–28 as unexposed, establishing a washout period in between the exposed and unexposed periods . . . When multiple events occurred to a given individual, the first occurrence of the composite outcome in the post-vaccination period was used (e.g., someone attending the ER [emergency department] who was then admitted would have one event counted in that period). The relative incidence rate of the composite endpoint during the exposed period compared with the unexposed period was analyzed using a fixed effects Poisson regression model. This model included a term for exposure period and a term for patient, thereby allowing each individual to serve as his or her own control and accounting for intra-individual correlation. An offset term was also included to account for the differing durations of the exposed and unexposed periods.”<sup>70</sup>

**RECORD-PE item 12.b**

Describe and justify the use of multiple designs, design features, or analytical approaches.

*Explanation*

As discussed in RECORD-PE item 4.a, the use of multiple designs or design features in the same report is a commonly used strategy in pharmacoepidemiological studies to assess the potential for bias and residual confounding. If authors have used multiple approaches to analysis, these should be clearly outlined for readers to assess strengths and limitations. The authors also should state clearly how they approached reproducibility across different databases, including issues such as variability in coding and healthcare systems. If authors used a common data model<sup>71</sup> (see the second example<sup>72</sup> below) to analyse data across different data sources, they should describe this and specify which common data model they used. If any data pooling across data sources was done, the approaches used should be described.

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

*Examples describing each design, design feature, or analytical approach*

Wong and colleagues report: “We used Poisson regression to estimate the rate ratios for clarithromycin users compared with amoxicillin users during current, recent, and past use . . . For the self-controlled case series analysis, we estimated incidence rate ratios using conditional Poisson regression, comparing the rate of events during risk windows with the rate during baseline periods . . . we also performed a post hoc case crossover analysis, which is not vulnerable to this limitation of the self controlled case series. The case crossover design is applied for studies investigating the association between transient drug use and outcome with abrupt time of onset. We estimated odds ratios using conditional logistic regression, comparing drug use before the event (current period) with that at other earlier control periods within patients.”<sup>52</sup>

But and colleagues report: “The individual-level data from the five cohorts were standardised by each research partner locally using the common data model. We then conducted centralised analyses by uploading the unified data to a server at Statistics Denmark, where for each cohort we constructed the individual-level dataset to assess insulin exposure and other variables in exactly the same way. We employed a semi-aggregate level approach to combine the datasets, which were tabulated by cancer site as the number of cancer cases and person-years aggregated by categorical variables.”<sup>72</sup>

**Methods (data access and cleaning methods)**

**RECORD-PE items**

No items specific to the RECORD-PE guidelines are needed in addition to the RECORD items.

*Explanation*

RECORD states that “authors should provide information on the data cleaning methods used in the study.”<sup>1</sup> This information is particularly important for pharmacoepidemiological studies, because the preparation of drug exposure data is complex and reflects serial assumptions that are typically not disclosed. Therefore, data cleaning extends substantially beyond the removal of outlying values. When data require preparation for analysis (eg, conversion of raw prescription data to exposed and unexposed episodes of person time), authors should be transparent about the steps undertaken in cleaning the data. These steps might include decisions on deriving start and stop dates; and assumptions made when instructions on administration instructions provide for flexibility (eg, prescriptions as needed), when prescriptions are overlapping, and when clinically implausible values are encountered.

**Results**

**RECORD-PE items**

For the results section, no items specific to RECORD-PE are needed in addition to the earlier STROBE and RECORD recommendations.

## RESEARCH METHODS AND REPORTING

*Explanation*

STROBE guidelines recommend that researchers report the number of individuals included at each stage of the study, including reasons for exclusion.<sup>18</sup> The RECORD guidelines further underscore the importance of reporting how results were filtered based on data quality, availability, and linkage.<sup>1</sup> Use of a flow diagram to illustrate the selection of the study population is encouraged by both STROBE and RECORD—note that this diagram is distinct from the study design diagram discussed in RECORD-PE item 4.b. A high level of transparency is equally important in pharmacoepidemiological studies in which additional eligibility criteria are often used (eg, based on indications for use, washout periods, and lag periods), adding further complexity to the selection process. Researchers should report the number of participants included at all stages of the study, including the analysis stage and for analyses performed to assess different objectives (eg, subgroup and sensitivity analyses).

Pharmacoepidemiological studies that examine adverse drug events or reactions should report whether and how researchers assessed or validated the outcome on the individual case level (eg, through record review by a specialist blinded to the exposure(s) under study, in order to try and rule out other more likely causes of the event). This process should be clear from the text, a table, or a flowchart describing how many events were considered to be caused by the study drug(s) after record review. Presenting numbers of potential cases that lacked sufficient data to be classified as non-cases or definite cases (and were assigned a final status such as possible or uncertain) is also encouraged. In an article by Kaye and colleagues on the risk of liver injury associated with use of oral antimicrobials, figure 1 provides a good example.<sup>73</sup> Clear delineation of the selection process facilitates critical appraisal, applicability, and reproducibility of the study findings.

Regarding results from descriptive analyses, STROBE recommends that authors present detailed data on the distribution of demographic, clinical, and social variables, including the number of participants with missing data. Missing data are frequently encountered in pharmacoepidemiological studies based on routinely collected data. In studies using routinely collected data, we might not know whether there is non-recorded or unmeasured information on diagnoses, symptoms, and management. Cohort studies also should provide summary measures of follow-up time. For studies based on routinely collected data, RECORD does not include additional items.<sup>1</sup> However, in terms of clinical variables, pharmacoepidemiological studies should report the distribution of indications for the drug of interest. It is also advised that authors summarise person time on and off drug exposure, including the sensitivity of “at risk” periods to different definitions of risk attribution if appropriate. Furthermore, in the case of time varying variables, which are increasingly used in pharmacoepidemiology, authors should

consider reporting the number and characteristics of individuals with time varying data.

RECORD-PE supports the STROBE recommendations for presentation of outcome data, main results, and other analyses.<sup>18</sup> Thus, researchers should report the number of events or summary measures of outcomes (or exposures in case-control studies), unadjusted and adjusted estimates and their precision, confounder variables adjusted for, category boundaries when continuous variables are categorised, absolute measures of risk for a meaningful time period (if relevant), and other analyses performed (including subgroup, interaction, and sensitivity analyses). Authors should present the results determined using the different approaches, which could include conventional methods and more complex approaches.

If multiple approaches have been used in an attempt to account for confounding (eg, matching and adjustment), the results of all methods should be presented and any differences discussed. It is advisable to present descriptive results showing covariate distribution (number and percentages) in exposure groups before propensity score matching, as well as the distribution after propensity score matching if appropriate. Authors should explicitly state whether an analysis was prespecified or post hoc. Researchers also are advised to report in detail the results of analyses used to explore and handle missing data, which are frequently encountered in pharmacoepidemiological studies based on routinely collected data.

**Discussion (limitations)****RECORD-PE item 19.1.a**

Describe the degree to which the chosen database(s) adequately captures the drug exposure(s) of interest.

*Explanation*

Authors should report whether the drug exposure in question could be ascertained by an alternative source, if not fully captured in the database used for the study. Some of the explanation outlined in RECORD-PE item 8.a is also relevant here. An additional issue relates to whether a drug exposure of interest could have been obtained over the counter, and whether such use is captured by the data source.<sup>74</sup> If not, authors may want to discuss the likely extent of misclassification. A similar issue is that if patients are admitted to hospital for extended periods and the database does not capture in-hospital dispensing of drug treatments, misclassification also might occur.<sup>69 75</sup> Authors should also discuss whether the database is likely to have had information on diagnoses, symptoms, and management, and discuss the implications for study findings.

*Examples on adequacy of capture of drug exposure in database*

Weinstein and colleagues report: “This analysis was restricted to prescription use of paracetamol and ibuprofen, and it is unknown whether these results would generalize to non-prescription

exposures. There are several reasons for a GP [general practitioner] to prescribe these medications in the CPRD [Clinical Practice Research Datalink], including record keeping and giving the patient access to the medication at a lower cost because the patient qualifies for free filling of prescriptions. In addition, those using these medications chronically may need larger quantities than typically available over the counter. Thus, it is likely that, by relying on prescriptions, we skewed our study population toward elderly subjects with chronic conditions who may also be at the low end of the economic spectrum.”<sup>76</sup>

Suissa reports: “In our illustration, the naive approach that does not account for the immeasurable hospitalized time during the 30-day period prior to the index date estimated a significant 40 percent reduction in mortality associated with a prescription of inhaled corticosteroids during this period. However, there were 806 cases (deaths) that had been hospitalized during this same 30-day period and that were considered unexposed by this analysis since they did not receive a prescription. These cases had spent 16.2 out of the 30 days in the hospital, time during which they could not receive outpatient prescriptions, compared with 8.8 days for the corresponding 253 such controls. In fact, 190 of these 806 cases (24 percent) had spent the entire 30-day period in the hospital, compared with seven of the 253 controls (3 percent), and could not possibly have received any prescription at all.”<sup>69</sup>

**Discussion (interpretation)**

**RECORD-PE item 20.a**

Discuss the potential for confounding by indication, contraindication, or disease severity or for selection bias (healthy adherer or sick stopper) as alternative explanations for the study findings when relevant.

*Explanation*

As discussed in the methods section, confounding by indication is a major issue in interpreting the findings of pharmacoepidemiological studies, beyond studies of routinely collected data in general. Particularly in the case of insurance or billing data (such as health administrative data), identification of study participants, drug exposures, confounders, and outcomes are based on coded data. Little or no information might be available to describe the indication for drug treatment, personal preferences and values of the patient and prescriber, any potential contraindications to use, or disease severity, all of which could confound the association between the drug and the outcome of interest. Even in clinical data (such as data from electronic health records), the indication or contraindication for drugs might not be recorded or might be contained in free text fields, and not accessible to investigators using these data for pharmacoepidemiological research. Important confounding variables thus might not be available for investigators or readers of the research report.

Therefore, to the best of their ability, authors should report such potential confounding in the routinely collected health data and how the confounding was managed overall.

We recommend the inclusion of a clear statement in the conclusions (or other discussion section) to explain whether the results could be explained by confounding by indication. Such a statement would help reduce misguided decision making and increase the trustworthiness of the evidence and its interpretation. This statement could report any post hoc analyses designed to evaluate the robustness of the finding and alternative explanations—for instance, to evaluate whether patients in different exposure groups were likely to have been prescribed the drugs for similar conditions. If moving of an individual to another practice or insurance database cannot be tracked, authors should clearly outline whether that individual’s person time is unique (that is, only considered from the new registration date) and discuss the effect of missing past exposures or events on the study findings. A histogram of people leaving for reasons other than death by year of age also might be useful, as well as the age distribution of people entering the population.

Further, as per RECORD-PE item 12.1.b, authors should include explicit consideration of results from different approaches when they have used multiple designs, design features, or analytical approaches. This information is especially pertinent if such efforts yield inconsistent results and thus guidance on interpretation is needed.

*Examples on alternative explanation for findings*

Sujan and colleagues report: “The findings from the present study should be considered in light of several limitations. First, and most important, observational designs such as these cannot fully rule out all sources of confounding. In particular, like other register-based approaches, this study could not comprehensively assess maternal depression or its severity, nor could it compare different antidepressant treatment regimens. Thus, associations could have been influenced by confounding by antidepressant indication . . . the study used multiple designs to address this limitation, each of which could help rule out some but not all sources of confounding, to provide complementary evidence. For example, sibling comparisons ruled out all stable confounders (e.g., chronic maternal depression), but that design may not have been able to account for confounding from maternal depression that varied across pregnancies. Thus, the within-family associations with preterm birth may plausibly be driven by unmeasured time-varying maternal depression rather than by antidepressant use.”<sup>34</sup>

Filion and colleagues report: “Our study was designed to examine the impact of drug formulary restrictions on the validity of pharmacoepidemiologic studies using the example of fluticasone/salmeterol combination therapy. We found that the implementation of these restrictions had a profound effect on drug utilization,



## RESEARCH METHODS AND REPORTING

with the policy resulting in an important decrease in the rates of prescription and of new use of fluticasone/salmeterol. These prescription changes resulted in channeling and confounding by indication, with new users of fluticasone/salmeterol having a significantly higher crude rate of hospitalization for respiratory causes during the restricted period (crude HR [hazard ratio] = 1.41, 95%CI [confidence interval] = 1.32, 1.51) because of the presence of more severe underlying respiratory disease. Adjustment for potential confounders attenuated and reversed the association, with new users during the restricted period having a significantly lower rate of hospitalization for respiratory causes compared with those during the liberal period (fully adjusted HR = 0.78, 95%CI = 0.73, 0.83). These results suggest that drug formulary restrictions can result in substantial and unexpected confounding by indication that threatens the validity of study results. These results also suggest that adjusting for patient demographic and clinical characteristics is insufficient to account for channeling because of formulary restrictions. Consequently, such restrictions must be considered in the design and analysis of pharmacoepidemiologic studies.<sup>77</sup>

Schneeweiss and colleagues report: "Aprotinin rather than aminocaproic acid was used in sicker patients, and the modest reduction in the relative mortality estimates after the control of confounding by covariates is consistent with the hypothesis of confounding on the basis of indication. Multivariate analyses resulted in weaker associations between aprotinin and death than those reported in unadjusted analyses (unadjusted relative risk, 1.83; adjusted relative risk, 1.64). Matching according to propensity score permitted us to control for an additional 10 covariates in a highly selected cohort, which further reduced the relative-risk estimate.

"Our analyses were adjusted for some, but not all, covariates typically included in risk-prediction scores for patients undergoing CABG [coronary artery bypass grafting]. However, we adjusted for many covariates not typically included, and controlling for proxies of confounders results in control of the confounders themselves if the proxies capture the relations with the true confounding variable, exposure, and outcomes. Our joint adjustment for 41 characteristics before CABG was performed resulted in the prediction of in-hospital death that is as good as that from widely accepted clinical risk-prediction models for patients undergoing CABG. Prediction was almost identical for patients receiving aprotinin and for those receiving aminocaproic acid."<sup>78</sup>

### Discussion of the RECORD-PE checklist

The complete and accurate reporting of research is an ethical requirement endorsed by leading declarations and recommendations internationally.<sup>79 80</sup> The RECORD-PE guidelines have been developed to meet an identified need and are designed to improve the reporting of pharmacoepidemiological studies undertaken using routinely collected data. They are

an extension of the STROBE and RECORD statements and should be used in conjunction with the existing guidelines.<sup>1 18 81</sup> RECORD-PE represents a minimum standard of reporting and complements a recent set of comprehensive methodological and reporting items created with the aim of making pharmacoepidemiological research more reproducible.<sup>22</sup> Better reporting is a prerequisite of replication, but replication requires much more detail. However, RECORD-PE also guides authors on the transparency of reporting and helps readers to understand strengths and potential limitations of the work. Therefore, RECORD-PE represents a minimum standard of reporting for pharmacoepidemiological studies undertaken using routinely collected health data.

### Limitations

We consulted widely in the creation of these guidelines, including international experts in pharmacoepidemiology, journalology, editors, and policy makers. Despite wide consultation, we may have missed some key points. In addition, members of our guideline development working committee were primarily from western Europe and North America.

Pharmacoepidemiology is a rapidly changing field with frequent new methodological developments, reflecting the growth of big data, the development of integrated or distributed data systems, and the innovative approaches being developed to reduce bias associated with the use of non-randomised data to assess drug effects. Increasing collaborative approaches across traditional geographical and data source boundaries are leading to new discoveries for patient benefit in pharmacovigilance and pharmacoepidemiology to overcome problems such as sample sizes insufficient to detect less common safety signals. We have included some recent developments briefly in this article, for example, the use of common data models. We recognise that these guidelines will need to be updated to encompass new developments.

The RECORD-PE guidelines are an extension of the STROBE guidelines for non-interventional research, hence the focus of these guidelines is largely on non-interventional research in pharmacoepidemiology. There is much discussion in the literature about whether the term "observational" should be used as opposed to the term "non-interventional studies," because all studies involve observation. For RECORD-PE, we have continued with the term "observational" in the title because RECORD-PE is an extension of RECORD which, in turn, is an extension of the STROBE guidance.<sup>1 18</sup> We have briefly mentioned the use of pragmatic trials using routinely collected health data in the RECORD-PE guidelines; however, the increasing use of pragmatic randomised controlled trials (and particularly the development of registry based trials and trials within cohorts<sup>82 83</sup>) will likely necessitate expansion of the currently available guidance with both RECORD and CONSORT as guiding documents.



Conclusions of the RECORD-PE checklist

The RECORD-PE statement aims to extend existing STROBE and RECORD guidelines providing guidance for the reporting of pharmacoepidemiological studies using routinely collected data. It aims to enable readers to understand what was planned, what was done, and what was found in the research. This essential information is critical for users of research to optimally interpret the findings, including their strengths and limitations. Poor research reporting hampers the use of research findings and is an important component of research waste.<sup>84</sup> We anticipate that with increasing use of the RECORD-PE guidelines by researchers and endorsement and adherence by journal editors, the reporting of pharmacoepidemiological research undertaken using routinely collected health data will improve. The improved transparency and accuracy will benefit the research community, and ultimately improve patient care.

AUTHOR AFFILIATIONS

- <sup>1</sup>Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
- <sup>2</sup>Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark
- <sup>3</sup>Ottawa Hospital Research Institute, Ottawa, ON, Canada;
- <sup>4</sup>School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada
- <sup>5</sup>Departments of Medicine and of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada
- <sup>6</sup>Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, QC, Canada
- <sup>7</sup>Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, Netherlands
- <sup>8</sup>Department of Primary Care and Population Health, University College London, London, UK
- <sup>9</sup>Arthritis Research UK Centre for Epidemiology, Division of Musculoskeletal and Dermatological Sciences, School of Biological Sciences, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK
- <sup>10</sup>Institute for Clinical Evaluative Sciences, Toronto, ON, Canada
- <sup>11</sup>Hospital for Sick Children, Department of Paediatrics, University of Toronto, Toronto, ON, Canada
- <sup>12</sup>ICH Population, Policy, and Practice Programme, University College London, Great Ormond Street Institute of Child Health, London, UK
- <sup>13</sup>Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland
- <sup>14</sup>Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
- <sup>15</sup>Julius Global Health, University Medical Center Utrecht, Utrecht, Netherlands
- <sup>16</sup>Cochrane Switzerland, Institute of Social and Preventive Medicine, University of Lausanne, Lausanne, Switzerland
- <sup>17</sup>Department of Pediatrics and School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada
- <sup>18</sup>Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada

**Contributors:** SML and EIB conceived and designed the study, carried out the statistical analysis, had full access to all the data in the study, take responsibility for the integrity of the data and the accuracy of the data analysis, and are guarantors. All authors acquired data, analysed and interpreted data, drafted the manuscript, and critically revised the manuscript for important intellectual content. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

**Funding:** We gratefully acknowledge funding from the International Society for Pharmacoepidemiology as well as comments received from its membership. SML is supported by a Wellcome Senior Clinical fellowship in Science (205039/Z/16/Z); EIB is supported by a New Investigator Award from the Canadian Institutes of Health Research, Canadian Association of Gastroenterology, and Crohn's and Colitis Canada; EIB is also supported by the career enhancement programme of the Canadian Child Health Clinician Scientist Programme; KBF is supported by a salary support award from the Fonds de recherche du Québec—Santé (FRQS; Quebec Foundation for Health Research); HTS is supported by the Programme for Clinical Research Infrastructure (PROCRIN) Programme.

**Competing interests:** All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/col\\_disclosure.pdf](http://www.icmje.org/col_disclosure.pdf) and declare: main support from the International Society for Pharmacoepidemiology 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.

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

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

1 Benchimol EI, Smeeth L, Guttman A, et al, RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12:e1001885. doi:10.1371/journal.pmed.1001885

2 Spasoff R. *Epidemiologic methods for health policy*. Oxford University Press, 1999.

3 Mor A, Petersen I, Sørensen HT, Thomsen RW. Metformin and other glucose-lowering drug initiation and rates of community-based antibiotic use and hospital-treated infections in patients with type 2 diabetes: a Danish nationwide population-based cohort study. *BMJ Open* 2016;6:e011523. doi:10.1136/bmjopen-2016-011523

4 Nørgaard M, Johnsen SP. How can the research potential of the clinical quality databases be maximized? The Danish experience. *J Intern Med* 2016;279:132-40. doi:10.1111/joim.12437

5 Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Current use of routinely collected health data to complement randomized controlled trials: a meta-epidemiological survey. *CMAJ Open* 2016;4:E132-40. doi:10.9778/cmajo.20150036

6 Moher D, Glasziou P, Chalmers I, et al. Increasing value and reducing waste in biomedical research: who's listening?. *Lancet* 2016;387:1573-86. doi:10.1016/S0140-6736(15)00307-4

7 Canadian Institutes of Health Research. *Canada's strategy for patient-oriented research. Improving health outcomes through evidence-informed care*. Canadian Institutes of Health Research, 2011.

8 Hoffman S, Podgurski A. The use and misuse of biomedical data: is bigger really better? *Am J Law Med* 2013;39:497-538. doi:10.1177/009885881303900401

9 Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Aff (Millwood)* 2014;33:1139-47. doi:10.1377/hlthaff.2014.0048

10 Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821-9. doi:10.1016/j.jclinepi.2010.10.006

11 Hemkens LG, Benchimol EI, Langan SM, et al. The reporting of studies using routinely collected health data was often insufficient. *J Clin Epidemiol* 2016;79:104-11. doi:10.1016/j.jclinepi.2016.06.005

12 Wang SV, Verpillat P, Rassen JA, Patrick A, Garry EM, Bartels DB. Transparency and reproducibility of observational cohort studies using large healthcare databases. *Clin Pharmacol Ther* 2016;99:325-32. doi:10.1002/cpt.329

13 Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Lancet* 2008;371:1149-50. doi:10.1016/S0140-6736(08)60505-X

14 Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869. doi:10.1136/bmj.c869

15 Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263-7.

## RESEARCH METHODS AND REPORTING

- 16 Enhancing the QUALITY and Transparency Of health Research (EQUATOR) Network Library 2015. [www.equator-network.org/library/](http://www.equator-network.org/library/). Accessed March 7 2015.
- 17 Nicholls SG, Quach P, von Elm E, et al. The REporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) Statement: methods for arriving at consensus and developing reporting guidelines. *PLoS One* 2015;10:e0125620. doi:10.1371/journal.pone.0125620
- 18 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JPSTROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology* 2007;18:800-4. doi:10.1097/EDE.0b013e3181577654
- 19 Strom B. *Pharmacoepidemiology*. 1.What is pharmacoepidemiology? Wiley, 2005: 3-16.
- 20 Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE Initiative. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010;16:467-71.
- 21 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology. 2011. [www.encepp.eu/standards\\_and\\_guidances/methodologicalGuide.shtml](http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml).
- 22 Wang SV, Schneeweiss S, Berger ML, et al, joint ISPE-ISPOR Special Task Force on Real World Evidence in Health Care Decision Making. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf* 2017;26:1018-32. doi:10.1002/pds.4295
- 23 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758-64. doi:10.1093/aje/kwv254
- 24 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. doi:10.1136/bmj.i4919
- 25 Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;4:e297doi:10.1371/journal.pmed.0040297.
- 26 Graham DJ, Reichman ME, Wernecke M, et al. Stroke, bleeding, and mortality risks in elderly Medicare beneficiaries treated with dabigatran or rivaroxaban for nonvalvular atrial fibrillation. *JAMA Intern Med* 2016;176:1662-71. doi:10.1001/jamainternmed.2016.5954
- 27 Kinlaw AC, Stürmer T, Lund JL, et al. Trends in antibiotic use by birth season and birth year. *Pediatrics* 2017;140:e20170441. doi:10.1542/peds.2017-0441
- 28 Nyeland ME, Laursen MV, Callréus T. Evaluating the effectiveness of risk minimisation measures: the application of a conceptual framework to Danish real-world dabigatran data. *Pharmacoepidemiol Drug Saf* 2017;26:607-14. doi:10.1002/pds.4203
- 29 Douglas IJ, Langham J, Bhaskaran K, Brauer R, Smeeth L. Orlistat and the risk of acute liver injury: self controlled case series study in UK Clinical Practice Research Datalink. *BMJ* 2013;346:f1936. doi:10.1136/bmj.f1936
- 30 Kim SC, Solomon DH, Rogers JR, et al. Cardiovascular safety of tocilizumab versus tumor necrosis factor inhibitors in patients with rheumatoid arthritis: a multi-database cohort study. *Arthritis Rheumatol* 2017;69:1154-64. doi:10.1002/art.40084
- 31 van Staa TP, Leufkens HG, Abenham L, Begaud B, Zhang B, Cooper C. Use of oral corticosteroids in the United Kingdom. *QJM* 2000;93:105-11. doi:10.1093/qjmed/93.2.105
- 32 Shin JY, Eberg M, Ernst P, Filion KB. Statin potency and the risk of hospitalization for community-acquired pneumonia. *Br J Clin Pharmacol* 2017;83:1319-27. doi:10.1111/bcp.13208
- 33 Huerta C, Abbing-Karahagopian V, Requena G, et al. Exposure to benzodiazepines (anxiolytics, hypnotics and related drugs) in seven European electronic healthcare databases: a cross-national descriptive study from the PROTECT-EU Project. *Pharmacoepidemiol Drug Saf* 2016;25(suppl 1):56-65. doi:10.1002/pds.3825
- 34 Suján AC, Rickert ME, Öberg AS, et al. Associations of maternal antidepressant use during the first trimester of pregnancy with preterm birth, small for gestational age, autism spectrum disorder, and attention-deficit/hyperactivity disorder in offspring. *JAMA* 2017;317:1553-62. doi:10.1001/jama.2017.3413
- 35 Ehrenstein V, Hernandez RK, Ulrichsen SP, et al. Rosiglitazone use and post-discontinuation glycaemic control in two European countries, 2000-2010. *BMJ Open* 2013;3:e003424. doi:10.1136/bmjopen-2013-003424
- 36 Vrijens B, Vincze G, Kristanto P, Urquhart J, Burnier M. Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories. *BMJ* 2008;336:1114-7. doi:10.1136/bmj.39553.670231.25
- 37 Pye SR, Sheppard T, Joseph RM, et al. Assumptions made when preparing drug exposure data for analysis have an impact on results: An unreported step in pharmacoepidemiology studies. *Pharmacoepidemiol Drug Saf* 2018;27:781-8. doi:10.1002/pds.4440
- 38 Vandenbroucke J, Pearce N. Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology? *Am J Epidemiol* 2015;182:826-33. doi:10.1093/aje/kwv225
- 39 Patomo E, Huybrechts KF, Bateman BT, et al. Lithium use in pregnancy and the risk of cardiac malformations. *N Engl J Med* 2017;376:2245-54. doi:10.1056/NEJMoa1612222
- 40 Larivée N, Suissa S, Coulombe J, Tagalakis V, Filion KB. Drospirenone-containing oral contraceptive pills and the risk of venous thromboembolism: an assessment of risk in first-time users and restarters. *Drug Saf* 2017;40:583-96. doi:10.1007/s40264-017-0525-2
- 41 Dixon WG, Carmona L, Finckh A, et al. EULAR points to consider when establishing, analysing and reporting safety data of biologics registers in rheumatology. *Ann Rheum Dis* 2010;69:1596-602. doi:10.1136/ard.2009.125526
- 42 Horwitz RJ, Feinstein AR. The problem of "protopathic bias" in case-control studies. *Am J Med* 1980;68:255-8. doi:10.1016/0002-9343(80)90363-0
- 43 Dixon WG, Hyrich KL, Watson KD, et al, BSRBR Control Centre ConsortiumBSR Biologics Register. Drug-specific risk of tuberculosis in patients with rheumatoid arthritis treated with anti-TNF therapy: results from the British Society for Rheumatology Biologics Register (BSRBR). *Ann Rheum Dis* 2010;69:522-8. doi:10.1136/ard.2009.118935
- 44 Pottegård A, Friis S, Stürmer T, et al. Considerations for pharmacoepidemiological studies of drug-cancer associations. *Basic Clin Pharmacol Toxicol* 2018;122:451-9. doi:10.1111/bcpt.12946
- 45 Abrahamowicz M, Beauchamp ME, Sylvestre MP. Comparison of alternative models for linking drug exposure with adverse effects. *Stat Med* 2012;31:1014-30. doi:10.1002/sim.4343
- 46 Moride Y, Abenham L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research [correction in: *J Clin Epidemiol* 2004;57:111]. *J Clin Epidemiol* 1994;47:731-7.
- 47 Guess HA. Behavior of the exposure odds ratio in a case-control study when the hazard function is not constant over time. *J Clin Epidemiol* 1989;42:1179-84. doi:10.1016/0895-4356(89)90116-9
- 48 Movahedi M, Beauchamp ME, Abrahamowicz M, et al. Risk of incident diabetes mellitus associated with the dosage and duration of oral glucocorticoid therapy in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2016;68:1089-98.
- 49 Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983;2:267-71. doi:10.1002/sim.4780020222
- 50 Filion KB, Azoulay L, Platt RW, et al, CNODES Investigators. A multicenter observational study of incretin-based drugs and heart failure. *N Engl J Med* 2016;374:1145-54. doi:10.1056/NEJMoa1506115
- 51 Xue F, Ma H, Stehman-Breen C, et al, Denosumab Global Safety Assessment Team. Design and methods of a postmarketing pharmacoepidemiology study assessing long-term safety of Prolia (denosumab) for the treatment of postmenopausal osteoporosis. *Pharmacoepidemiol Drug Saf* 2013;22:1107-14.
- 52 Wong AY, Root A, Douglas IJ, et al. Cardiovascular outcomes associated with use of clarithromycin: population based study. *BMJ* 2016;352:h6926. doi:10.1136/bmj.h6926
- 53 Khan NA, Yun L, Humphries K, Kapral M. Antihypertensive drug use and adherence after stroke: are there sex differences? *Stroke* 2010;41:1445-9. doi:10.1161/STROKEAHA.110.579375
- 54 Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;45:1866-86.
- 55 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. doi:10.1136/bmj.i4919
- 56 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323-37. doi:10.1016/j.jclinepi.2004.10.012
- 57 Williamson EJ, Forbes A. Introduction to propensity scores. *Respirology* 2014;19:625-35. doi:10.1111/resp.12312
- 58 Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009;28:3083-107. doi:10.1002/sim.3697
- 59 Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *Am J Epidemiol* 2005;161:891-8. doi:10.1093/aje/kwi106
- 60 Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009;20:512-22. doi:10.1097/EDE.0b013e3181a663cc

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

61 Swanson SA, Hernán MA. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013;24:370-4. doi:10.1097/EDE.0b013e31828d0590

62 Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. *Am J Epidemiol* 2002;155:1045-53. doi:10.1093/aje/k111045

63 Størring H, Pottegård A, Hallas J. Refining estimates of prescription durations by using observed covariates in pharmacoepidemiological databases: an application of the reverse waiting time distribution. *Pharmacoepidemiol Drug Saf* 2017;26:900-8. doi:10.1002/pds.4216

64 Gault N, Castañeda-Sanabria J, De Rycke Y, Guillo S, Foulon S, Tubach F. Self-controlled designs in pharmacoepidemiology involving electronic healthcare databases: a systematic review. *BMC Med Res Methodol* 2017;17:25. doi:10.1186/s12874-016-0278-0

65 Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ* 2016;354:i4515. doi:10.1136/bmj.i4515

66 Whitaker H. The self controlled case series method. *BMJ* 2008;337:a1069. doi:10.1136/bmj.a1069

67 Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;25:1768-97. doi:10.1002/sim.2302

68 Cadarette S, Tadrous M, Delaney J, et al. Control yourself: guidance for the application and reporting of self-controlled study designs in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2017;26(suppl 2):3.

69 Suissa S. Immeasurable time bias in observational studies of drug effects on mortality. *Am J Epidemiol* 2008;168:329-35. doi:10.1093/aje/kwn135

70 Wilson K, Hawken S, Kwong JC, et al. Adverse events following 12 and 18 month vaccinations: a population-based, self-controlled case series analysis. *PLoS One* 2011;6:e27897. doi:10.1371/journal.pone.0027897

71 Reisinger SJ, Ryan PB, O'Hara DJ, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. *J Am Med Inform Assoc* 2010;17:652-62. doi:10.1136/jamia.2009.002477

72 But A, De Bruin ML, Bazelier MT, et al. Cancer risk among insulin users: comparing analogues with human insulin in the CARING five-country cohort study. *Diabetologia* 2017;60:1691-703. doi:10.1007/s00125-017-4312-5

73 Kaye JA, Castellsague J, Bui CL, et al. Risk of acute liver injury associated with the use of moxifloxacin and other oral antimicrobials: a retrospective, population-based cohort study. *Pharmacotherapy* 2014;34:336-49. doi:10.1002/phar.1367

74 Schmidt M, Hallas J, Friis S. Potential of prescription registries to capture individual-level use of aspirin and other nonsteroidal anti-inflammatory drugs in Denmark: trends in utilization 1999-2012. *Clin Epidemiol* 2014;6:155-68. doi:10.2147/CLEP.S59156

75 Suissa S. Co-morbidity in COPD: the effects of cardiovascular drug therapies. *Respiration* 2010;80:3-7. doi:10.1159/000315387

76 Weinstein RB, Ryan P, Berlin JA, et al. Channeling in the use of nonprescription paracetamol and ibuprofen in an electronic medical records database: evidence and implications. *Drug Saf* 2017;40:1279-92. doi:10.1007/s40264-017-0581-7

77 Filion KB, Eberg M, Ernst P. Confounding by drug formulary restriction in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Saf* 2016;25:278-86. doi:10.1002/pds.3923

78 Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008;358:771-83. doi:10.1056/NEJMoa0707571

79 Nicholls SG, Langan SM, Benchimol EI. Reporting and transparency in big data: the nexus of ethics and methodology. In: Mittelstadt BD, Floridi L, eds. *The ethics of biomedical big data*. Springer International Publishing, 2016:339-65. doi:10.1007/978-3-319-33525-4\_15.

80 Nicholls SG, Langan SM, Benchimol EI, Moher D. Reporting transparency: making the ethical mandate explicit. *BMC Med* 2016;14:44. doi:10.1186/s12916-016-0587-5

81 Etminan M. Reporting guidelines for pharmacoepidemiological studies are urgently needed. *BMJ* 2014;349:g5511. doi:10.1136/bmj.g5511

82 Mathes T, Buehn S, Prengel P, Pieper D. Registry-based randomized controlled trials merged the strength of randomized controlled trails and observational studies and give rise to more pragmatic trials. *J Clin Epidemiol* 2018;93:120-7. doi:10.1016/j.jclinepi.2017.09.017

83 Mc Cord KA, Al-Shahi Salman R, Treweek S, et al. Routinely collected data for randomized trials: promises, barriers, and implications. *Trials* 2018;19:29. doi:10.1186/s13063-017-2394-5

84 Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267-76. doi:10.1016/S0140-6736(13)62228-X

Supplementary table 1: Glossary of terms  
Supplementary table 2: RECORD-PE checklist

# BMJ Open

## Influence of serum iron test results on the diagnosis of iron deficiency in children: a retrospective observational study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-046865.R1
Article Type:	Original research
Date Submitted by the Author:	17-Mar-2021
Complete List of Authors:	Sezgin, Gorkem; Macquarie University Faculty of Medicine and Health Sciences, Centre for Health Systems and Safety Research, Australian Institute of Health Innovation Li, Ling; Macquarie University, Westbrook, Johanna; Macquarie University Faculty of Medicine and Health Sciences, Australian Institute of Health Innovation Wearne, Elisabeth; Gippsland Primary Health Network Azar, Denise; Gippsland Primary Health Network McLeod, Adam; Outcome Health Pearce, Christopher; Melbourne East General Practice Network, General Practice Ignjatovic, Vera; The University of Melbourne, Department of Paediatrics; Murdoch Childrens Research Institute Monagle, Paul; University of Melbourne, Paediatrics; Murdoch Childrens Research Institute Georgiou, A; Macquarie University, Australian Institute of Health Innovation
<b>Primary Subject Heading</b>:	Paediatrics
Secondary Subject Heading:	General practice / Family practice, Evidence based practice, Haematology (incl blood transfusion)
Keywords:	CHEMICAL PATHOLOGY, Anaemia < HAEMATOLOGY, Paediatric pathology < PAEDIATRICS, Clinical chemistry < PATHOLOGY, Paediatric pathology < PATHOLOGY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).



## Influence of serum iron test results on the diagnosis of iron deficiency in children: a retrospective observational study

Gorkem Sezgin<sup>a</sup>, Ling Li<sup>a</sup>, Johanna Westbrook<sup>a</sup>, Elisabeth Wearne<sup>b</sup>, Denise Azar<sup>b</sup>, Adam McLeod<sup>c</sup>, Chris Pearce<sup>c</sup>, Vera Ignjatovic<sup>d,e</sup>, Paul Monagle<sup>d,e,f</sup>, Andrew Georgiou<sup>a</sup>

**Affiliations:** <sup>a</sup> Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia ; <sup>b</sup> Gippsland Primary Health Network, Traralgon, Victoria, Australia; <sup>c</sup> Outcome Health, East Burwood, Victoria, Australia ; <sup>d</sup> Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia; <sup>e</sup> Murdoch Children's Research Institute, Parkville, Victoria, Australia ; and <sup>f</sup> Royal Children's Hospital Melbourne, Parkville, Victoria, Australia

**Address correspondence to:** Gorkem Sezgin, Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia, gorkem.sezgin@mq.edu.au, +61 2 9850 2940

**Word Count:** 2688



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**ABSTRACT**

Background and objective: Serum iron results are not indicative of iron deficiency yet may be incorrectly used to diagnose iron deficiency instead of serum ferritin results. Our objective was to determine the association between serum iron test results and iron deficiency diagnosis in children by general practitioners.

Design, setting, patients, main outcome measures: A retrospective observational study of 14,187 children aged 1 to 18 years with serum ferritin and serum iron test results from 137 general practices in Victoria, Australia, between 2008 and 2018. Generalised estimating equation models calculating odds ratios were used to determine the association between serum iron test results (main exposure measure) and iron deficiency diagnosis (outcome measure) in two population groups: 1) iron deplete population, defined as having a serum ferritin <12 µg/L if aged less than five years, and <15 µg/L if aged five years or over; and 2) iron replete population, defined as having a serum ferritin >30 µg/L.

Results: 3,484 tests were iron deplete and 15,528 were iron replete. Iron deplete children were less likely to be diagnosed with iron deficiency if they had normal serum iron levels (AOR: 0.73; 95% CI 0.57-0.96). Iron replete children had greater odds of an iron deficiency diagnosis if they had low serum iron results (AOR: 2.59; 95%CI 1.72-3.89). Other contributors to an iron deficiency diagnosis were female sex and having anaemia.

Conclusion: Serum ferritin alone remains the best means of diagnosing iron deficiency. Reliance on serum iron test results by general practitioners is leading to significant over- and under-diagnosis of iron deficiency in children.

**STRENGTHS AND LIMITATIONS OF THIS STUDY**

Our study utilises general practice electronic health record data which contains a large sample size to draw our analysis from

Our study contains data from 137 general practices across a large region

Through general practice data, our study is able to correlate diagnostic procedures associated with pathology test orders, and to an extent, treatment thereafter.

General practice data is limited to information recorded in the electronic health software, which is not always complete for all entry fields.

For peer review only

INTRODUCTION

Iron deficiency is the most common nutritional deficiency worldwide and the highest contributing factor to anaemia. The global prevalence of anaemia was estimated as 32.9% in 2010, with the highest burden among young children.[1] Missed diagnosis of iron deficiency can be debilitating for young children, with studies suggesting a link between iron deficiency anaemia and cognitive disorders.[2] Iron deficiency occurs some time before patients become anaemic, marked by a process in which red cells become microcytic and hypochromic.[3] Therefore, anaemia can be prevented with early diagnosis and treatment of iron deficiency.

Diagnosing iron deficiency can be challenging: patients may present with non-specific clinical symptoms, and interpretation of laboratory tests in diagnosing iron deficiency can be difficult.[4] Low serum ferritin levels (a measure of iron stores in the body) are clinically indicative of iron deficiency.[5] In general practice serum ferritin is often ordered as part of an iron panel test, which also contains serum iron, transferrin, and total iron binding capacity. The provision of these accompanying tests can complicate the interpretation of results for diagnosing iron deficiency, particularly due to the presence of serum iron (a measure of circulating iron), which is not clinically indicative of iron deficiency. The utility of serum iron in clinical diagnosis is limited, particularly in general practice. Serum iron and transferrin saturation levels, when interpreted together, provide an indication of iron use by tissue.[6] However, their use in diagnosing iron deficiency is not supported by current guidelines.[6]

Research focusing on diagnostic errors among paediatric patients is limited, particularly in general practice. Failure to order appropriate laboratory tests and incorrect interpretation of test results commonly go unrecognized,[7] and are a

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

considerable cause of diagnostic errors and adverse outcomes to patients.[8] Our aim was to determine whether serum iron tests results (which are often ordered alongside serum ferritin) were mistakenly used in the diagnosis of iron deficiency for paediatric patients in general practice. We therefore assessed the association between serum iron test results and the diagnosis of iron deficiency.

## METHODS

We conducted a retrospective observational study of the association between serum iron test results for children, and their subsequent diagnosis of iron deficiency. We obtained our data from general practice electronic health records from 137 practices in Victoria, Australia.[9]. This study was conducted and reported according to Reporting of Studies Conducted Using Observational Routinely Collected Health Data Statement for Pharmacoepidemiology (RECORD-PE) guidelines.[10]

### Study population and variables

Our sample population consisted of paediatric patients aged between one and 18 years with iron panel test results. We included laboratory tests containing both serum iron and serum ferritin test results between 2008 and 2018.

Serum ferritin, serum iron, transferrin saturation (TSAT), C-reactive protein (CRP) tests, and red blood cell indices were identified from laboratory test records by searching for related terms in Logical Observation Identifiers Names and Codes (LOINC) [11] and test name fields. We determined ages by subtracting the years between test request date and patient year of birth.

Diagnostic records of iron deficiency were identified by searching SNOMED (Systematized Nomenclature of Medicine) [12] terms for “iron” related to iron deficiency, and additionally, text fields for “iron”, “ID”, and “IDA” related to iron deficiency. Specialist referrals were identified by searching for the terms “referral” and “specialist”. Medications used in the treatment of iron deficiency were identified by searching for commercial and generic names of iron supplements in the field record of medications prescribed by the general practitioner. Due to limitations related to electronic health records, the association between test results and related diagnoses, referrals, and medication are not recorded. The closest iron deficiency diagnosis, referral, and medication within a cut-off of 14 days after the test request was considered as being associated with the test result. Treatment for iron deficiency is recommended to be followed up with a repeat test within a three month period. We therefore identified test requests with a follow up serum ferritin test within a 90 day period, which may indicate that the patient is being treated.

**Inclusion/Exclusion criteria**

To ensure our study population contained only new cases, we excluded cases where a serum ferritin or serum iron test was ordered within the last 12 months. Prior serum ferritin results may influence the diagnoses of iron deficiency due to a variety of reasons, including: continuation of care to assess the effect of any prior treatments, or re-testing to confirm results. Although re-test interval is recommended as three months, our criteria was purposefully wide to ensure no potential known-cases were included. Likewise, we excluded patients who had a past iron medication or a past diagnosis for iron deficiency or anaemia at any time.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

Presence of inflammation can complicate the diagnosis of iron deficiency, as serum ferritin levels rise under this condition [13], and can mask an underlying iron deficiency. Where inflammation is suspected, clinicians may either delay testing for iron deficiency until inflammation is resolved, or order a CRP test, an inflammation marker, and interpret ferritin results in context of the CRP result.[14] We excluded test results accompanied by high CRP results, as indicated by the pathology provider's report or having a result of  $>10$  mg/L to account for possible inflammatory diseases.

Thalassemia can often be mistaken for iron deficiency. We therefore excluded patients who had a known thalassemia diagnosis. Only test results containing both serum iron and serum ferritin results were included.

### Statistical analyses

We set out to determine association between serum iron test results and iron deficiency diagnoses in two population groups: (i) in an iron deplete population group; and (ii) in an iron replete population group. Iron deplete was defined as having low serum ferritin levels: denoted as the serum ferritin result value being less than 12  $\mu\text{g/L}$  for children under the age of five, and less than 15  $\mu\text{g/L}$  for children five years of age and over.[6] A patient was considered as iron replete if their serum ferritin result was either denoted as normal or high by the pathology provider, or had a ferritin result value above 30  $\mu\text{g/L}$ . [15] Results for other thresholds ( $<15$   $\mu\text{g/L}$  and  $<20$   $\mu\text{g/L}$  for iron deplete;  $>50$   $\mu\text{g/L}$  and  $>100$   $\mu\text{g/L}$  serum ferritin for iron replete) can be found in the supplementary document.

In our iron deplete population group, we tested if having normal serum iron results decreased the odds of a diagnosis; which indicates under-diagnosis. In our iron replete



population group, we tested if having low serum iron results increased the odds of an iron deficiency diagnosis; which indicates over-diagnosis (Figure 1). For the purposes of our study, the term ‘diagnosis’ comprises a diagnostic record, referral, medication, or treatment related to iron deficiency; identified as described under the study population and variables heading within the methods section.

Generalised estimating equations were formed to model the probability of an iron deficiency diagnosis associated with serum iron test results, fitted as a binomial distribution with a logit link function and robust standard error estimations. Clusters by general practice and patient were formed, with exchangeable within group correlation. The model was used to report odds ratios, p-values, and 95% confidence intervals: with iron deficiency diagnosis as the binary outcome variable (diagnosed; not diagnosed). Serum iron test result was our main exposure variable of interest, as a binary variable (pathology provider noted low serum iron levels; pathology provider noted normal or high serum iron levels). Covariates included in our model were age (continuous), sex (binary), serum ferritin result value (continuous), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), and mean corpuscular haemoglobin concentration (MCHC) values reported by pathology provider (binary variable as: low result value; normal or high result value), as well as red cell distribution width (RDW); binary variable as: high result value; normal or low result value) and TSAT (binary variable as: low result value; normal or high result value). We also included anaemia status as a covariate, identified as having a haemoglobin concentration below 12 g/dL (120 g/L) for females, and 13 g/dL (130 g/L) for males. Our covariates were selected due to their relevance to the diagnosis of iron deficiency. Female sex and younger age have high prevalence of iron deficiency.[16,17] Red cell indices are relevant to the diagnosis of iron deficiency anaemia, and serum transferrin

is ordered alongside serum ferritin and can be an indicator of tissue iron absorption (if interpreted in consideration of serum iron levels), although its use in the diagnosis of iron deficiency is not recommended.[6] All statistical analyses were conducted using Stata MP 16.0 (StataCorp, TX, USA).

### Patient and public involvement

Patients were not involved in this study.

### Ethics approval

Approval to use the de-identified general practice electronic health record data in research was obtained by the data custodians, Outcome Health, from the RACGP National Research and Evaluation Ethics Committee (NREEC 17–008). Ethics approval for the use of the de-identified data was obtained from Macquarie University Human Research Ethics Committee (5201700872).

## RESULTS

A total of 17,929 test results from 14,187 patients were included in our analyses. Females accounted for 59.7% of the test results, and the median age was 14 years (10–17 years).

A total of 3,484 test results with low serum ferritin results (i.e. iron deplete) were identified (Figure 1). There were 2,518 test results with normal serum iron, of which 483 were diagnosed with iron deficiency. Our analysis revealed a decrease in the odds of an iron deficiency diagnosis by 0.73 (95% CI 0.56–0.96; p-value 0.022) when serum iron results were normal (Table 1). Having a low MCHC result (AOR 0.77;

95%CI 0.62-0.97; p-value 0.027), female sex (AOR 1.40; 95%CI 1.09-1.81; p-value 0.009) and increasing serum ferritin result value (AOR 0.91; 95%CI 0.88-0.94; p-value 0.001) also decreased the odds of a diagnosis.

Table 1. Associations between normal iron test results and iron deficiency diagnosis for the iron deplete population. Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between serum iron result and diagnosis, with all co-variates included in the model.

Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.46 (0.40-0.52; <0.001)
Adjusted Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.73 (0.57-0.96; 0.022)
Female sex	1.41 (1.09 -1.81; 0.009)
Increasing age	1.01 (0.99-1.03; 0.526)
Anaemia	1.19 (0.96-1.48; 0.115)
Low MCV	1.22 (0.92-1.63; 0.172)
Low MCH	1.15 (0.89-1.49; 0.272)
Low MCHC	0.77 (0.62-0.97; 0.027)
High RDW	1.23 (0.92-1.63; 0.158)
Low TSAT	1.03 (0.81-1.32; 0.797)
Increasing serum ferritin value	0.91 (0.88-0.94; <0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width; TSAT: transferrin saturation

A total of 15,528 test results had normal or high serum ferritin levels (i.e. iron replete). Low serum iron results accompanied 1,470 of these tests, with a further 170 iron deficiency diagnoses recorded among these tests (Figure 1). We observed a statistically significant increase in the odds of iron deficiency diagnosis with low serum iron results by 2.59 (95% CI 1.72-3.89; p-value <0.001) (Table 2). Female sex (AOR 1.62; 95%CI 1.34-1.96; p-value<0.001), having anaemia (AOR 1.81; 95%CI 1.40-2.35; p-value<0.001), and having low TSAT (AOR 1.84; 95%CI 1.23-2.75; p-value 0.003) were also associated with increased odds a diagnosis.

Table 2. Associations between low iron test results and iron deficiency diagnosis for the iron replete population. Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between serum iron result and diagnosis, with all co-variables included in the model.

<b>Odds Ratio (95% CI ; p-value)</b>	
Low serum iron result	4.57 (3.94-5.30; <0.001)
<b>Adjusted Odds Ratio (95% CI ; p-value)</b>	
Low serum iron result	2.59 (1.72-3.89; <0.001)
Female sex	1.62 (1.34-1.96; <0.001)
Increasing age	1.04 (1.01-1.06; 0.001)
Anaemia	1.81 (1.40-2.35; <0.001)
Low MCV	0.76 (0.36-1.63; 0.486)
Low MCH	0.94 (0.58-1.51; 0.792)
Low MCHC	0.94 (0.64-1.39; 0.757)
High RDW	1.79 (0.92-3.49; 0.087)
Low TSAT	1.84 (1.23-2.75; 0.003)
Increasing serum ferritin value	1.004 (1.002-1.01; <0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width; TSAT: transferrin saturation

## DISCUSSION

Our study identified both under and over diagnosis of iron deficiency among children treated in general practice related to the interpretation of serum iron results. Children without iron deficiency were over two times more likely to be erroneously diagnosed (over-diagnosed) when they had low serum iron results. Children with iron deficiency were 0.7 times less likely to be diagnosed (under-diagnosed) when they had normal serum iron levels.

Our study demonstrates serum iron results being erroneously used in the diagnosis of iron deficiency, consequently leading to diagnostic errors. Our results suggest a misconception of the changed role of iron levels in diagnosing iron deficiency; serum iron being confused as the biomarker of iron deficiency where in fact, serum ferritin

is. Serum iron tests performed as part of a panel act as a distractor in interpreting the test.

We also observed an increase in the likelihood of an iron deficiency diagnosis with low haemoglobin results. Low haemoglobin levels are indicative of anaemia, and the most common cause of anaemia is iron deficiency.[18] Based on Croskerry's [19] study on biases in decision making, our observation may be a result of 'availability and non-availability' bias; defined as a tendency to make a judgement based on a thought coming more readily to mind due to its frequency. Alternatively, it may be a result of 'information distortion bias', defined by Boyle and Purdon [20] as using information to support of a course of action more so than is warranted by the available information. Anaemia is frequently caused by iron deficiency and may have consequently led to the misdiagnosis. We also observed females being more likely to be diagnosed with iron deficiency. Females are known to have a high incidence of iron deficiency in the general population, predominantly thought to be a result of menses.[21] Similar to haemoglobin, female sex could also be resulting in bias in interpretation of the iron panel tests due to its stronger association with the disease.

Laboratory testing related errors are one of the highest contributing factors to diagnostic errors, as demonstrated by Schiff *et al.*'s [22] study. Schiff *et al.* asked clinicians across the US to report three cases of diagnostic errors, and describe their cause, significance, and frequency. They found that laboratory testing related errors, including erroneous interpretation of test results, was the highest contributor to diagnostic errors: with 44% of the errors accounted to it. In Hickner *et al.*'s survey [23], doctors reported uncertainty in test result interpretation 8.3% of the time. Some

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).

of the reported challenges related to nomenclature of the tests, such as different tests with the same name, or different names for the same test.

In Australia, a Medicare review taskforce was formed to ensure current clinical practice aligned with evidence and best practice.[24] Iron panel tests were found to be ordered over three times higher than standalone ferritin tests by general practitioners in diagnosing iron deficiency. Our results provide empirical evidence demonstrating the resultant diagnostic errors due to serum iron tests among children. The Medicare taskforce made recommendations to relabel tests in clinically relevant terms, such that a clinician would order 'iron deficiency studies' containing serum ferritin, rather than the current iron panel test with all iron biomarkers. Such an approach may reduce cognitive errors by removing confounding factors out of the decision-making process. In the diagnosis of iron deficiency, this may reduce burden of disease by ensuring patients with iron deficiency are correctly and timely diagnosed, potentially preventing anaemia and cognitive impairment. Additionally, correct use of serum ferritin to diagnose iron deficiency may reduce the number of patients being misclassified as having iron deficiency based on serum iron results, potentially reducing unnecessary treatment, and promoting correct diagnosis by ruling out iron deficiency.

Some limitations should be taken into consideration when interpreting the findings of our study. Our study sample is derived from general practice electronic health records, and the data quality and completeness is therefore largely influenced by general practitioner's use of their software. The reason for ordering of iron panel tests is not known, and may have an impact on the diagnosis. Nonetheless, serum ferritin is a test used in checking the iron status of a patient, and low levels, particularly  $<12/15 \mu\text{g/L}$



should be treated. Diagnostic records are often not complete and many diagnoses, including iron deficiency, are omitted. Certain diseases may impact the interpretation of serum ferritin and the diagnosis of iron deficiency, including thalassemia and inflammatory diseases. We accounted for a number of these potential confounders, although it is likely that other unknown factors (diseases, clinical investigations, patient history) may have impacted the interpretation of test results and the diagnosis of iron deficiency. Although medication prescriptions are well recorded, iron deficiency medication can be obtained without prescription; however the dosage of over the counter iron supplements may not be appropriate for patients with  $<12/15$   $\mu\text{g/L}$  serum ferritin, particularly for paediatric patients. We identified the closest iron deficiency related diagnosis and medication records within a cut-off period of 14 days after the test result as our outcome. It is possible that some records were missed due to this time period, although we expect a majority of patients to return to the practice within this timeframe. Nonetheless, it is also possible that the patient may not return to the clinic at all, particularly if their perceived condition has improved. This has additional consequences on one of our outcome criteria: follow up within the next three months. These limitations may have resulted in an underestimate of our outcome measure. Nonetheless, our study objective did not necessitate all of the potential outcomes to be included in the analysis to reach a meaningful conclusion. A sample size large enough to make a meaningful inference from our sample to the population, and no systematic bias in the outcome measure in one category in relation to the other which may have occurred due to missed cases, was required. Based on our final study sample size and the possibility for a systematic bias between our categories being highly unlikely, we are confident both criteria have been achieved and our conclusion

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).

is accurate. Nonetheless, the effect size should be cautiously interpreted, which may have been affected by the missed cases.

Our study also has several strengths. Our outcome was identified through several criteria, in line with the RECORD-PE guidelines: a diagnostic record, medication record, specialist referral, or follow up treatment. Our iron deplete population, having  $<12/15$   $\mu\text{g/L}$  serum ferritin, requires treatment with medication and follow up.[25] We therefore believe that most patients who are diagnosed and are being treated would be captured with our criteria.

To the best of our knowledge, our study is the first to quantitatively demonstrate the interpretation biases of serum iron test results in general practice, and the resultant error in diagnosing iron deficiency. Iron panel tests are ordered in a similar fashion worldwide in diagnosing iron deficiency, with all four tests being included. Although our results are from Australian general practices within the state of Victoria, it is possible that a similar diagnostic error is occurring in other states/territories and countries where iron panel tests are ordered in the investigation of iron deficiency instead of standalone serum ferritin.

## Conclusions

Our study demonstrates cognitive biases in diagnosing iron deficiency in children, which is related to serum iron results and augmented by abnormal haemoglobin test results and female sex. Serum iron tests have consequences for both over- and under-diagnosis, as it appears to lead general practitioners to overlook serum ferritin test results. In diagnosing iron deficiency, serum iron levels are no longer indicated by clinical guidelines and should not form a basis in establishing iron deficiency status of patients.

**Funding source:** This study was supported by funding from the Quality Use of Pathology Program, Department of Health, Government of Australia (4-2QFVW4M).

**Competing interests:** The authors have no competing interests relevant to this article to disclose.

**Author contributions:** G.S. conceptualized the study, carried out the statistical analysis, drafted the initial manuscript, and reviewed and revised the manuscript. P.M. and V.I. conceptualized the study, contributed to the drafting of the initial manuscript, provided expertise, and critically reviewed the manuscript for important intellectual content. L.L. provided expertise on the statistical analysis, contributed to the design of the study, and reviewed and revised the manuscript for important intellectual content. A.M. and C.P. developed the data collection instruments, collected data, contributed to the interpretation of the data, and reviewed and revised the manuscript. J.W., A.G., E.W., and D.A. contributed to the design of the study, critically reviewed the manuscript for important intellectual content, and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

**Data Sharing Statement:** Deidentified individual participant data, and ethics approval documents for the collection of data by the data custodians Outcome Health can be made available to researchers who provide a methodologically sound proposal for use in achieving the goals of their approved proposal, on condition of approval of the primary health networks on whose behalf the data is being collected for. Proposals should be submitted to Outcome Health at [admin@outcomehealth.org.au](mailto:admin@outcomehealth.org.au)

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

## Figure Captions

Figure 1. Characteristics of the study population. Iron deplete was defined as having a serum ferritin result value of  $<12 \mu\text{g/L}$ , while iron replete was defined as having a serum ferritin result value of  $>30 \mu\text{g/L}$ . Adjusted odds ratio (AOR) and related 95% confidence intervals present the odds of an iron deficiency diagnosis in the iron deplete population associated with normal serum iron results, and in the iron replete population, diagnosis associated with low serum iron results.

For peer review only

REFERENCES

1 Kassebaum NJ, Jasrasaria R, Naghavi M, *et al.* A systematic analysis of global anemia burden from 1990 to 2010. *Blood* 2014;**123**:615–24. doi:10.1182/blood-2013-06-508325

2 Beard JL. Why Iron Deficiency Is Important in Infant Development. *J Nutr* 2008;**138**:2534–6. doi:10.1093/jn/138.12.2534

3 Camaschella C. New insights into iron deficiency and iron deficiency anemia. *Blood Rev* 2017;**31**:225–33. doi:10.1016/j.blre.2017.02.004

4 Soppi ET. Iron deficiency without anemia - a clinical challenge. *Clin Case Rep* 2018;**6**:1082–6. doi:10.1002/ccr3.1529

5 Pasricha S-RS, Flecknoe-Brown SC, Allen KJ, *et al.* Diagnosis and management of iron deficiency anaemia: a clinical update. *Med J Aust* 2010;**193**:525–32.

6 Geneva: World Health Organization (WHO). WHO guideline on use of ferritin concentrations to assess iron status in individuals and populations. 2020.

7 Laposata M. Errors in clinical laboratory test selection and result interpretation: commonly unrecognized mistakes as a cause of poor patient outcome. *Diagnosis* 2014;**1**:85–7. doi:10.1515/dx-2013-0010

8 Plebani M. Diagnostic Errors and Laboratory Medicine – Causes and Strategies. *EJIFCC* 2015;**26**:7–14.

9 Pearce C, McLeod A, Rinehart N, *et al.* What a Comprehensive, Integrated Data Strategy Looks Like: The Population Level Analysis and Reporting (POLAR) Program. *Stud Health Technol Inform* 2019;**264**:303–7. doi:10.3233/SHTI190232

10 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**. doi:10.1136/bmj.k3532

11 Stram M, Gigliotti T, Hartman D, *et al.* Logical Observation Identifiers Names and Codes for Laboratorians: Potential Solutions and Challenges for Interoperability. *Arch Pathol Lab Med* 2020;**144**:229–39. doi:10.5858/arpa.2018-0477-RA

12 Côté RA, Robboy S. Progress in Medical Information Management: Systematized Nomenclature of Medicine (SNOMED). *JAMA* 1980;**243**:756–62. doi:10.1001/jama.1980.03300340032015

13 Kell DB, Pretorius E. Serum ferritin is an important inflammatory disease marker, as it is mainly a leakage product from damaged cells. *Metallomics* 2014;**6**:748–73. doi:10.1039/C3MT00347G

14 Cook JD. Diagnosis and management of iron-deficiency anaemia. *Best Pract Res Clin Haematol* 2005;**18**:319–32. doi:10.1016/j.beha.2004.08.022

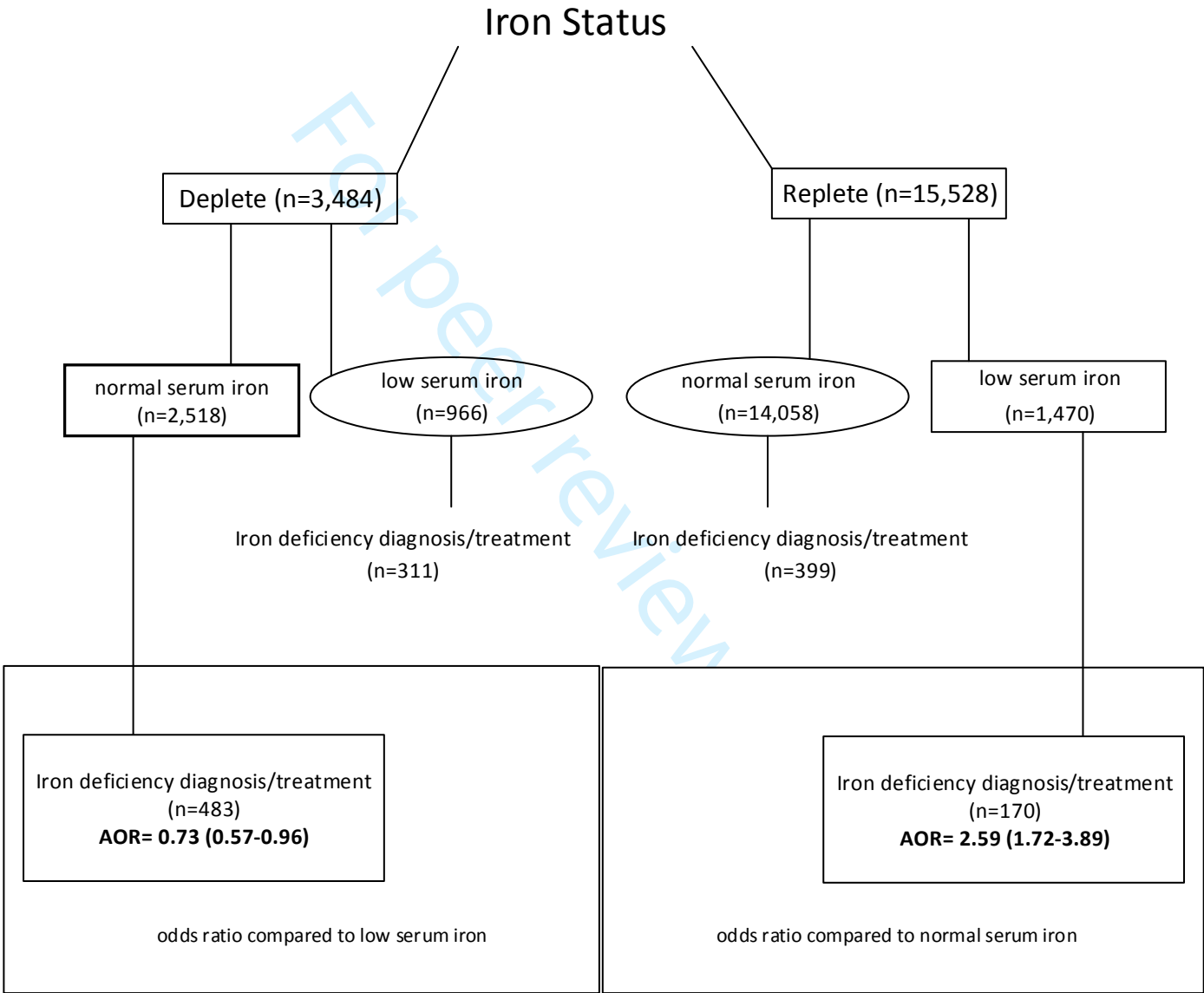
15 WHO. Serum ferritin concentrations for the assessment of iron status and iron deficiency in populations. [https://www.who.int/vmnis/indicators/serum\\_ferritin.pdf](https://www.who.int/vmnis/indicators/serum_ferritin.pdf) (accessed 19 May 2020).

16 Percy L, Mansour D, Fraser I. Iron deficiency and iron deficiency anaemia in women. *Best Pract Res Clin Obstet Gynaecol* 2017;**40**:55–67. doi:10.1016/j.bpobgyn.2016.09.007

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignment Supérieur (ABES).

- 17 Parkin PC, Borkhoff CM. Practical tips for paediatricians: Assessment and management of young children with iron deficiency. *Paediatr Child Health* 2018;**23**:433–4. doi:10.1093/pch/pxy045
- 18 De Benoist B, World Health Organization, Centers for Disease Control and Prevention (U.S.). *Worldwide prevalence of anaemia 1993-2005 of: WHO Global Database of anaemia*. Geneva: : World Health Organization 2008. [http://whqlibdoc.who.int/publications/2008/9789241596657\\_eng.pdf](http://whqlibdoc.who.int/publications/2008/9789241596657_eng.pdf) (accessed 11 Sep 2018).
- 19 Croskerry P. Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. *Acad Emerg Med* 2002;**9**:1184–204. doi:10.1197/aemj.9.11.1184
- 20 Boyle PJ, Purdon M. The information distortion bias: implications for medical decisions. *Med Educ* 2019;**53**:1077–86. doi:10.1111/medu.13919
- 21 Rushton DH, Barth JH. What is the evidence for gender differences in ferritin and haemoglobin? *Crit Rev Oncol Hematol* 2010;**73**:1–9. doi:10.1016/j.critrevonc.2009.03.010
- 22 Schiff GD, Hasan O, Kim S, *et al*. Diagnostic Error in Medicine: Analysis of 583 Physician-Reported Errors. *Arch Intern Med* 2009;**169**:1881–7. doi:10.1001/archinternmed.2009.333
- 23 Hickner J, Thompson PJ, Wilkinson T, *et al*. Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results. *J Am Board Fam Med* 2014;**27**:268–74. doi:10.3122/jabfm.2014.02.130104
- 24 MBS Review Taskforce. Diagnostic Medicine Clinical Committee Final Report. [https://www1.health.gov.au/internet/main/publishing.nsf/Content/mbs-review-2018-taskforce-reports-cp/\\$File/Diagnostic-Medicine-Clinical-Committee-Final%20Report.pdf](https://www1.health.gov.au/internet/main/publishing.nsf/Content/mbs-review-2018-taskforce-reports-cp/$File/Diagnostic-Medicine-Clinical-Committee-Final%20Report.pdf) (accessed 17 Jan 2020).
- 25 Ozdemir N. Iron deficiency anemia from diagnosis to treatment in children. *Türk Pediatri Arş* 2015;**50**:11–9. doi:10.5152/tpa.2015.2337





## **Influence of serum iron test results on the diagnosis of iron deficiency in children: a retrospective observational study**

Gorkem Sezgin<sup>a</sup>, Ling Li<sup>a</sup>, Johanna Westbrook<sup>a</sup>, Elisabeth Wearne<sup>b</sup>, Denise Azar<sup>b</sup>, Adam McLeod<sup>c</sup>, Chris Pearce<sup>c</sup>, Vera Ignjatovic<sup>d,e</sup>, Paul Monagle<sup>d,e,f</sup>, Andrew Georgiou<sup>a</sup>

**Affiliations:** <sup>a</sup> Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia ; <sup>b</sup> Gippsland Primary Health Network, Traralgon, Victoria, Australia; <sup>c</sup> Outcome Health, East Burwood, Victoria, Australia ; <sup>d</sup> Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia; <sup>e</sup> Murdoch Children's Research Institute, Parkville, Victoria, Australia ; and <sup>f</sup> Royal Children's Hospital Melbourne, Parkville, Victoria, Australia

**Address correspondence to:** Gorkem Sezgin, Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Ryde, New South Wales, Australia, gorkem.sezgin@mq.edu.au, +61 2 9850 2940

Table 1. Associations between normal iron test results and iron deficiency diagnosis for the iron deplete population (serum ferritin <15 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variates included in the model.

Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.46 (0.41-0.53; <0.001)
Adjusted Odds Ratio (95% CI ; p-value)	
Normal serum iron result	0.77 (0.59-0.998; 0.048)
Female sex	1.39 (1.09-1.78; 0.009)
Increasing age	1.01 (0.99-1.03; 0.257)
Anaemia	1.15 (0.93-1.43; 0.187)
Low MCV	1.23 (0.93-1.63; 0.148)
Low MCH	1.19 (0.92-1.54; 0.175)
Low MCHC	0.78 (0.62-0.97; 0.029)
High RDW	1.24 (0.94-1.64; 0.124)
Low TSAT	1.05 (0.82-1.34; 0.696)
Increasing serum ferritin value	0.91 (0.88-0.93; <0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width; TSAT: transferrin saturation

Table 2. Associations between normal iron test results and iron deficiency diagnosis for the iron deplete population (serum ferritin <20 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variates included in the model.

<b>Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	0.42 (0.37-0.47; <0.001)
<b>Adjusted Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	0.77 (0.61-0.99; 0.038)
Female sex	1.31 (1.07-1.61; 0.01)
Increasing age	1.02 (1-1.03; 0.059)
Anaemia	1.20 (0.99-1.44; 0.063)
Low MCV	1.32 (1.01-1.72; 0.04)
Low MCH	1.14 (0.90-1.43; 0.268)
Low MCHC	0.81 (0.66-1.00; 0.05)
High RDW	1.18 (0.91-1.54; 0.219)
Low TSAT	1.01 (0.81-1.26; 0.945)
Increasing serum ferritin value	0.91 (0.89-0.92; <0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width; TSAT: transferrin saturation

Table 3. Associations between low iron test results and iron deficiency diagnosis for the iron replete population (serum ferritin >50 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variables included in the model.

Odds Ratio (95% CI ; p-value)	
Normal serum iron result	4.47 (3.66-5.45; <0.001)
Adjusted Odds Ratio (95% CI ; p-value)	
Normal serum iron result	2.56 (1.51-4.36; 0.001)
Female sex	1.65 (1.29-2.12; <0.001)
Increasing age	1.05 (1.02-1.08; 0.001)
Anaemia	2.24 (1.56-3.22; <0.001)
Low MCV	0.58 (0.18-1.83; 0.354)
Low MCH	0.99 (0.51-1.94; 0.984)
Low MCHC	0.78 (0.44-1.36; 0.377)
High RDW	2.33 (0.92-5.95; 0.076)
Low TSAT	1.90 (1.12-3.21; 0.017)
Increasing serum ferritin value	1.004 (1.002-1.01; <0.001)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width; TSAT: transferrin saturation

Table 4. Associations between low iron test results and iron deficiency diagnosis for the iron replete population (serum ferritin >100 µg/L). Odds ratios present the binary association between the indicated variable and iron deficiency diagnosis. Adjusted odds ratios present the multivariate association between iron result and diagnosis, with all co-variates included in the model.

<b>Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	2.58 (1.86-3.58; <0.001)
<b>Adjusted Odds Ratio (95% CI ; p-value)</b>	
Normal serum iron result	1.60 (0.64-3.99; 0.318)
Female sex	1.60 (1.08-2.38; 0.019)
Increasing age	1.02 (0.98-1.07; 0.344)
Anaemia	2.15 (1.18-3.91; 0.012)
Low MCV	0.31 (0.04-2.53; 0.274)
Low MCH	0.99 (0.26-3.83; 0.989)
Low MCHC	0.76 (0.27-2.11; 0.6)
High RDW	3.63 (0.99-13.22; 0.051)
Low TSAT	1.75 (0.71-4.34; 0.227)
Increasing serum ferritin value	1.002 (1.0001-1.004; 0.04)

MCV: mean corpuscular volume; MCH: mean corpuscular haemoglobin; MCHC: mean corpuscular haemoglobin concentration; RDW: red cell distribution width; TSAT: transferrin saturation



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(a) Indicate the study’s design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(a) Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5-7
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	
		(b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed	Figure 1
		Case-control study—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-8
Bias	9	Describe any efforts to address potential sources of bias	6-7
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-9
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	8-9
		(b) Describe any methods used to examine subgroups and interactions	8
		(c) Explain how missing data were addressed	8
		(d) Cohort study—If applicable, explain how loss to follow-up was addressed	8
		Case-control study—If applicable, explain how matching of cases and controls was addressed	
		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(e) Describe any sensitivity analyses	8

Continued on next page

<b>Results</b>			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9
		(b) Indicate number of participants with missing data for each variable of interest	N/A
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	Figure 1
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 1 / Table 2
		(b) Report category boundaries when continuous variables were categorized	8
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-13
Generalisability	21	Discuss the generalisability (external validity) of the study results	15
<b>Other information</b>			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	16

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).