


openheart Systematic review of validated case definitions to identify hypertensive disorders of pregnancy in administrative healthcare databases

Amy Johnston ^{1,2}, Sonia R Dancey,³ Victrine Tseung,² Becky Skidmore,⁴ Peter Tanuseputro,^{5,6,7} Graeme N Smith,⁸ Thais Coutinho,^{1,9,10,11} Jodi D Edwards^{1,2,6}

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/openhrt-2022-002151>).

To cite: Johnston A, Dancey SR, Tseung V, *et al*. Systematic review of validated case definitions to identify hypertensive disorders of pregnancy in administrative healthcare databases. *Open Heart* 2023;**10**:e002151. doi:10.1136/openhrt-2022-002151

Received 22 September 2022
Accepted 16 June 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Amy Johnston; Amy.Johnston@uottawa.ca

ABSTRACT

Background Administrative data are frequently used to study cardiovascular disease (CVD) risk in women with hypertensive disorders of pregnancy (HDP). Little is known about the validity of case-finding definitions (CFDs, eg, disease classification codes/algorithms) designed to identify HDP in administrative databases.

Methods A systematic review of the literature. We searched MEDLINE, Embase, CINAHL, Web of Science and grey literature sources for eligible studies. Two independent reviewers screened articles for eligibility and extracted data. Quality of reporting was assessed using checklists; risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool, adapted for administrative studies. Findings were summarised descriptively.

Results Twenty-six studies were included; most (62%) validated CFDs for a variety of maternal and/or neonatal outcomes. Six studies (24%) reported reference standard definitions for all HDP definitions validated; seven reported all 2×2 table values for ≥1 CFD or they were calculable. Most CFDs (n=83; 58%) identified HDP with high specificity (ie, ≥98%); however, sensitivity varied widely (3%–100%). CFDs validated for any maternal hypertensive disorder had the highest median sensitivity (91%, range: 15%–97%). Quality of reporting was generally poor, and all studies were at unclear or high risk of bias on ≥1 QUADAS-2 domain.

Conclusions Even validated CFDs are subject to bias. Researchers should choose the CFD(s) that best align with their research objective, while considering the relative importance of high sensitivity, specificity, negative predictive value and/or positive predictive value, and important characteristics of the validation studies from which they were derived (eg, study prevalence of HDP, spectrum of disease studied, methodological rigour, quality of reporting and risk of bias). Higher quality validation studies on this topic are urgently needed.

PROSPERO registration number CRD42021239113.

INTRODUCTION

Hypertensive disorders of pregnancy (HDP) are multisystemic conditions¹ that affect

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Little is known about the validity of case-finding definitions (CFDs, eg, disease classification codes/algorithms) designed to identify hypertensive disorders of pregnancy (HDP) in administrative databases.

WHAT THIS STUDY ADDS

⇒ We found that most validated CFDs identify HDP with high specificity (ie, ≥98%); however, their sensitivity varies widely—especially for those designed to identify specific HDP subtypes.

⇒ Even CFDs with the highest sensitivity and specificity are subject to substantial bias.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Researchers should choose the CFD(s) that best aligns with their research objective, while considering the relative importance of high sensitivity, specificity, negative predictive value and/or positive predictive value and important characteristics of the validation studies from which they were derived (eg, study prevalence of HDP, spectrum of disease, methodological rigour).

5%–10% of all pregnancies.² Although there is no international consensus on what defines HDP,³ these conditions can be broadly characterised by elevated blood pressure (BP), typically defined as ≥140 mm Hg systolic (sBP) and/or ≥90 mm Hg diastolic (dBP), diagnosed pre-pregnancy or at <20 weeks' gestation (chronic hypertension in pregnancy), at ≥20 weeks' gestation (de novo hypertension diagnosed during pregnancy; that is, gestational hypertension (GH), preeclampsia) or during the postpartum period (postpartum hypertension).

The most common HDP subtypes are GH and preeclampsia, with GH the more common of the two diagnoses.^{4,5} In addition

to distinguishing between GH and preeclampsia subtypes, HDPs are often further differentiated based on symptom severity and time of diagnosis, specifically: preeclampsia with or without severe features, pre-term preeclampsia, eclampsia (preeclampsia with seizures), and haemolysis, elevated liver enzyme levels and a low platelet count (HELLP) syndrome.^{6,7} Although most women with HDP return to their pre-pregnancy (ie, normotensive) state during the postpartum period, all women with a history of these conditions are at an increased risk for premature cardiovascular disease (CVD) and CVD-related mortality.^{8–11}

Both descriptive (ie, prevalence, incidence) and CVD risk estimates for women with a history of HDP are frequently derived using administrative healthcare data¹² due to the availability of large sample sizes, long follow-up times and relatively little loss to follow-up.^{13–16} However, these data often have less clinical and sociodemographic detail than desired, and those that are available are prone to error and misclassification, threatening the accuracy of research findings.¹⁵

The ability to correctly identify women with a history of HDP from administrative databases is critical to generate accurate estimates of HDP disease burden and CVD risk. Indeed, the reliability of descriptive and CVD effect estimates could be drastically impacted if there is low confidence that HDP exposures have been classified correctly. Despite this, little is known about the validity of case-finding definitions (CFDs, eg, International Classification of Disease (ICD) codes, code combinations) used to identify these conditions in healthcare administrative databases.

Objective

To systematically identify and summarise available evidence on the validity of CFDs in identifying HDP from healthcare administrative databases.

METHODS

This review was registered a priori (PROSPERO CRD42021239113) and was conducted and reported in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses Diagnostic Test Accuracy (PRISMA-DTA) checklists,^{17,18} adapted for reviews of administrative validation studies (online supplemental tables 1 and 2). Updates and amendments made to our registered protocol are detailed in online supplemental 2. We use the term women throughout this manuscript; however, we recognise that this is a gendered term that may not apply to all currently or previously pregnant people.

Literature search

An experienced medical information specialist developed and tested the search strategy through an iterative process in consultation with the primary author. The strategy was peer-reviewed by another senior information specialist using the Peer Review of Electronic

Search Strategies (PRESS) Checklist.¹ Any suggested edits from PRESS were reviewed and incorporated where appropriate. Using the multifile option and deduplication tool in OVID, we searched Ovid MEDLINE ALL, including Epub Ahead of Print, In-Process and Other Non-Indexed Citations, and Embase Classic+Embase. We also searched CINAHL (Ebsco) and Web of Science (Core Collection). All searches were performed on 6 February 2021 and updated on 24 January 2023. Strategies utilised a combination of controlled vocabulary (eg, ‘Hypertension, Pregnancy-Induced’, ‘International Classification of Diseases’, ‘Validation Study’) and keywords (eg, ‘pre-eclampsia’, ‘classification’, ‘gold standard’). Vocabulary and syntax were adjusted across databases. There were no language or vocabulary restrictions on searches; however, where possible, editorials or letters were removed from the results. Records identified by searches were downloaded to EndNote V.9.3.3 (Clarivate Analytics) and deduplicated. A grey literature search of pertinent websites and databases was performed based on relevant sites identified in CADTH’s Grey Matters checklist.² We also screened included study reference lists and articles included in a relevant systematic review.¹⁹ ‘Cited by’ searches were also conducted in PubMed, Google Scholar and Web of Science. Publishing organisations were contacted for potentially eligible reports unavailable through institutional access. The initial search strategy, 2023 update and PRISMA Statement literature search extension (PRISMA-S) checklist²⁰ are provided in online supplemental 2 table 3.

Study eligibility

Eligibility criteria are summarised in online supplemental table 4. Briefly, studies were included if authors validated the accuracy of ≥ 1 CFD in identifying any HDP in an administrative database. No studies were excluded based on quality or outcome(s) reported. To reduce the possibility of publication bias, we included data from published and grey literature sources and did not exclude studies based on language or location of conduct.

Study selection and data extraction

EndNote software (Clarivate Analytics, 2020) was used to manage bibliographic records. DistillerSR review software (Evidence Partners, 2021) was used to manage data screening and extraction.

Titles and abstracts were independently screened in for eligibility by three reviewers (AJ, SRD, VT). The full text of any record deemed potentially eligible by two reviewers was then assessed for full eligibility. All data were extracted from included studies in duplicate (AJ, SRD) into standardised abstraction forms that were first piloted on a small number of articles and revised as necessary prior to full data extraction. Extracted data included: study characteristics (author, publication year, country of conduct); population characteristics (eg, participant age, race); definition of each HDP validated; reference standard definition; characteristics of the index CFD (eg,

specific disease classification codes/code combinations that comprised the definition) and individual measures of performance for each CFD reported (ie, 2×2 contingency table values—true positives (TP), true negatives (TN), false positives (FP), false negatives (FN)), sensitivity (SN), specificity (SP)), positive and negative predictive values (PPV and NPV), and any other reported measure of agreement. Authors were not contacted to obtain missing or unclear information. All disagreements were resolved through consensus discussion and, if necessary, with the input of senior authors (JED, TC).

Reporting quality and risk of bias assessments

Quality of reporting and risk of bias (ROB) assessments were performed on studies in which all 2×2 table values for ≥1 CFD were available (ie, reported directly or calculable) (online supplemental table 5). Quality of reporting was assessed using checklists developed for this review; ROB assessed using a modified version of the QUADAS-2 tool. Targeted evaluations of five key biases were also completed. All assessments were first piloted on a single article by two reviewers (AJ, SRD) and questions revised as necessary. The remaining studies were fully assessed by the primary author; results were independently reviewed for accuracy and consistency by another (SRD). All disagreements were resolved through consensus discussion. Detailed methods are provided in the online supplemental 3B,C and tables 6–9.

Data synthesis

All study findings are presented in tables and figures and summarised descriptively. Performance metrics and analytic measures of diagnostic accuracy reported for all CFDs are summarised separately by type of analysis (primary, sensitivity or subgroup). Primary analyses include CFDs validated as part of a primary outcome/analysis. Sensitivity analyses include CFDs validated using alternate coding strategies. In subgroup analyses, authors reported CFD validation statistics separately according to participant characteristics.

When not reported, and whenever possible, 2×2 table values, validation statistics and kappa coefficients along with corresponding 95% exact binomial CIs²¹ were calculated if information was missing. When all 2×2 values were available, we calculated two measures of HDP prevalence for validation cohorts (primary CFDs only): (1) prevalence based on the reference standard (pretest or 'true' prevalence) and (2) prevalence based on the index CFD (post-test or 'apparent' prevalence). We also calculated prevalence-adjusted PPVs for all estimates reported by studies with a pretest prevalence of HDP ≥3% higher than would be expected in the general North American obstetric population. All calculations were made using the `epi.tests` and `epi.kappa` functions (`epiR` package) as well as the `biconf` function (`Hmisc` package) in RStudio (V.1.4.1103, R Foundation for Statistical Computing, 2021) as well as MedCalc's Diagnostic Test Evaluation Calculator (MedCalc Software, 2022).²²

We summarised the SN and SP (95% CIs) of all CFDs with a complete set of 2×2 values in forest plots to visually represent these associations (RevMan V.5.4, The Cochrane Collaboration, 2020). Due to high heterogeneity with respect to study design, study participants, CFD characteristics and reference standards, meta-analyses were not performed.

To guide researchers presently pursuing HDP research using administrative databases, we summarised the best currently available CFDs for HDP (ie, those with an SP ≥97% in combination with a PPV of ≥70%).

RESULTS

Study characteristics

The literature search yielded 5341 titles and abstracts. Thirty records met the eligibility criteria, corresponding to a total of 26 unique validation studies (n=3 records^{23–25} had four published companion papers^{26–29}) (figure 1). Reasons for excluding articles after full-text screening are provided in online supplemental table 10.

Of the 26 included validation studies, 25 reported at least one validation statistic of interest. Study characteristics, participant eligibility criteria and sources of study funding for these studies are summarised in online supplemental tables 11 and 12, respectively.

Nearly half of the 25 studies reporting ≥1 validation statistic of interest were conducted in the USA (n=12, 48%).^{24 30–41} Of the 13 others, six were conducted in Australia,^{25 42–46} three in Canada^{47–49} and two each in Denmark^{50 51} and France.^{23 52} Publication years ranged from 1998 to 2021; n=7 (28%) were published within the past 5 years. Nineteen studies (76%) included participants from multiple sites (eg, state-wide population-based studies), whereas four studies^{31 33–35} included participants from a single hospital.

Sampling methods used to build validation cohorts followed three general approaches (online supplemental table 13). The most commonly used approach involved selecting all (n=8^{23 34 35 37 39 48 50 51}) or a simple/stratified/multi-stage sample (n=11^{24 25 36 38 40 42 43 45 47 49 52}) of administrative records for women hospitalised for obstetrical delivery during their respective study periods. One study⁴⁰ oversampled records with (a) higher rates of hospital readmission and (b) caesarean deliveries to increase the number of high-risk patients. Four studies^{37 49–51} included all, or a random/stratified sample of, participants included in existing research cohorts (Danish National Birth Cohort,⁵⁰ Odense Child Cohort,⁵¹ OaK study,⁴⁹ antidepressant safety studies³⁷). Six studies selected records for all, or a random sample of, women who were exclusively code positive for HDP (24%^{31–33 41 44 46}).

Only nine studies^{25 31 35 36 44 46 49–51} were specifically designed to validate ≥1 CFD for HDP; most (64%) validated CFDs for a variety of maternal and/or neonatal outcomes. Most studies (n=16, 64%) validated CFDs for ≥2 HDP definitions; two studies^{25 35} each validated CFDs designed to identify eight different HDP definitions.

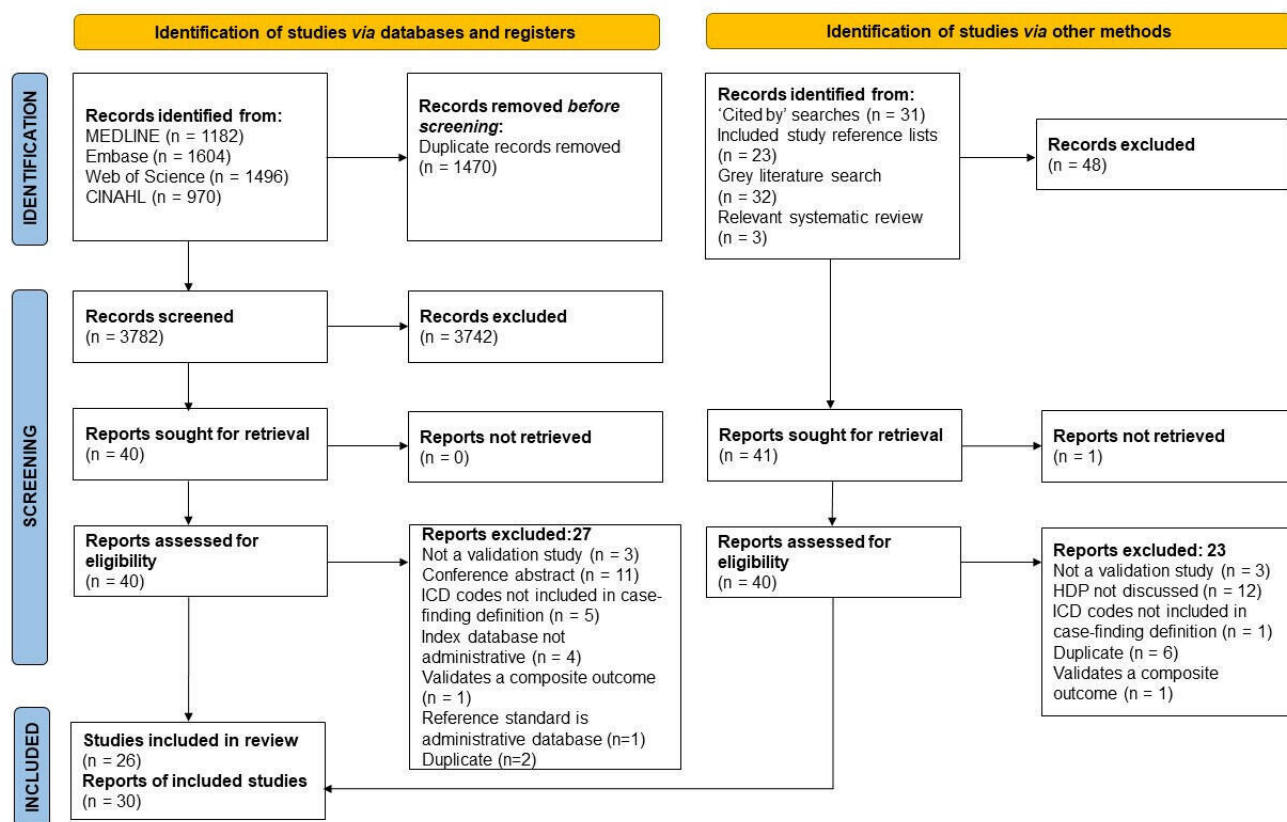


Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram summarising the literature search and selection process. HDP, hypertensive disorders of pregnancy; ICD, International Classification of Disease.

Administrative databases

Included CFDs were validated in 24 unique administrative databases (online supplemental table 11). Two studies each validated their CFDs in the Medicaid Analytic eXtract (MAX),^{32 37} the Danish National Patient Registry,^{50 51} the Victorian Perinatal Data Collection database^{42 44} and the Canadian Institute for Health Information's Discharge Abstract Database (CIHI-DAD).^{47 48} Three studies^{36 38 44} validated the same CFD in two different databases.

Reference standards

Most studies used medical record review/abstraction as their reference standard. Three^{39 46 48} used information obtained from a separate, linked, clinical database/registry. Seven studies^{31 32 34 36 38 41 49} reported that ≥ 1 physician abstracted and/or reviewed medical charts and made a final decision on the HDP diagnosis. In three studies,^{40 43 47} medical records were re-coded by coding specialists and re-abstractions compared with HDP codes identified in an administrative database.

Only six studies^{25 31 40 41 49 51} (24%) reported reference standard definitions for all HDPs validated. An additional two studies^{37 50} only reported reference standard definitions for some HDP definitions tested (online supplemental table 14). In general, HDPs were defined heterogeneously with respect to BP threshold considerations, number of elevated BP measurements required for a diagnosis, as well as timing of HDP occurrence.

Specifically, Luef *et al*⁵¹ defined GH as the onset of elevated BP at >22 weeks' gestation, whereas Roberts *et al*²⁵ and Labgold *et al*³⁵ defined GH as elevated BP diagnosed at >20 weeks' gestation. Further, Luef *et al*⁵¹ reported that ≥ 3 elevated BP measurements were required for a GH diagnosis, whereas Labgold *et al*³⁵ indicated that two instances of elevated BP were sufficient. Finally, Shachkina⁴⁹ reported that a $\text{dBP} \geq 90$ mm Hg alone was sufficient to diagnose preeclampsia, while three other studies^{25 37 41} reported that an $\text{sBP} \geq 140$ or $\text{dBP} \geq 90$ mm Hg was required.

Administrative case-finding definitions

We identified 143 unique CFDs for HDP: 67 reported as part of primary analyses (online supplemental table 15), 29 as part of sensitivity analyses (online supplemental table 16) and 47 as part of a subgroup analysis (online supplemental table 17). The two most validated CFDs were designed to identify preeclampsia ($n=36$; 25%), preeclampsia subtypes (mild, moderate, serious, severe, serious, superimposed, unspecified; $n=27$) and any maternal hypertensive disorder (ie, various hypertensive conditions in pregnancy, whether diagnosed prior to or after 20 weeks' gestation; $n=25$). Few CFDs were validated for HELLP syndrome ($n=6$; 4%) and only one CFD each was validated for mild or unspecified preeclampsia,³¹ moderate to severe preeclampsia⁵² and serious preeclampsia⁵⁰ (online supplemental figure 1).

In eight studies,^{24 31–33 37 38 44 47} CFDs were comprised of ≥ 1 ICD-9 code; however, most ($n=17$; 68%)^{23–25 34 35 39–43 45 46 48–52} validated CFDs comprised solely of ≥ 1 ICD-10 codes. One study³⁶ used Hospital International Classification of Diseases Adapted (HICDA) codes to identify HDP cases in a ‘historical’ administrative cohort. In four studies,^{25 42 48 51} the codes that comprised validated CFDs were not explicitly reported.

Most studies validated CFDs applied exclusively to a hospital discharge abstract for obstetrical delivery (online supplemental tables 15–17). However, four^{23 36 37 50} also considered administrative codes recorded during specific pregnancy and postpartum periods (ie, 10 months prior to and 2 months after delivery; after 140 gestational days; within 30 days post-delivery). Three CFDs reported by Chomistek *et al*⁴¹ incorporated information about treating physician specialty (obstetrics/gynaecology, pathology) and two CFDs validated for preeclampsia required a minimum of two occurrences of a code (‘claim’) in an administrative database to indicate a potential case. Goueslard *et al*⁵² and Tawfik *et al*³⁹ validated CFDs using infant and maternal records. No studies examined the validity of CFDs in identifying HDP in women with multiple gestations or for postpartum hypertension.

Participant characteristics

Fourteen studies (56%) reported any participant socio-demographic information (online supplemental table 18). Among these, only ten^{24 25 31 33 35 36 40 41 43 49} provided summary statistics for maternal age, though two additional studies^{23 40} reported the age range of eligible participants (10–55 years,⁴⁰ 14–50 years²³). One study²⁴ reported participant marital status (69% married) and six^{24 31–33 35 37} reported participants’ medical insurer, two^{32 37} of which limited inclusion to individuals insured through Medicaid. Only eight studies (32%) summarised participant pregnancy and birth characteristics, including information about the number/proportion with risk factors for, and characteristics associated with, HDP (eg, gestational age, birth weight, preexisting hypertension, multiple gestation, previous HDP) (online supplemental table 19).

Only 11 studies^{24 31 33 35–37 39–41 49 50} (44%) reported any information about participant race, ethnicity, or national origin. In studies that validated ≥ 1 CFD for HDP as a primary aim, black and non-Hispanic black women were the most represented racial group ($n=2646$ participants; 39%). In studies that validated CFDs for >1 maternal morbidity, white and non-Hispanic white women were the highest represented racial group ($n=4476$ participants), comprising 63% of the total population.

Synthesis of results

Median SP reported for CFDs across all analyses and HDP diagnoses were consistently very high at $\geq 97\%$. PPV estimates were less consistently high, with median estimates ranging from 55% to 96% (online supplemental table 20). Median SN varied widely, from 3% for CFDs validated for mild to moderate preeclampsia to 91% for

those validated for any maternal hypertensive disorder. In most cases, median SN for CFDs validated as part of sensitivity and/or subgroup analyses were higher than those reported as part of a primary analysis.

Primary analyses

Validation statistics for the 67 CFDs reported in primary analyses are provided in online supplemental table 15. Half ($n=33$) identified HDPs with an SP of $\geq 99\%$, while just over 60% ($n=41$) identified these conditions with a PPV of $\geq 70\%$. Most CFDs were validated for preeclampsia ($n=15$ for any preeclampsia; $n=11$ for preeclampsia subtypes). Those validated for hypertension diagnosed during pregnancy had the highest median SN (88%, IQR: 25%; range: 49%–100%), followed by CFDs validated for any maternal hypertensive disorder and preeclampsia (73%, IQR: 11%; range: 49%–96% and 73%, IQR: 36%; range: 29%–100%, respectively). All 2x2 table values were reported, or calculable, for 23 primary CFDs (34%; $n=7$ studies^{23 35 36 39 49–51}). The SN, SP and corresponding 95% CIs of these ‘comprehensively reported’ CFDs are summarised in figure 2A.

As illustrated in figure 2A, comprehensively reported primary CFDs identified all HDP diagnoses with very high SP. Only one study³⁶ reported an SP estimate of $<90\%$ for two CFDs validated for GH and preeclampsia. Nearly all definitions ($n=20$, 87%) had a PPV of $\geq 70\%$, 13 of which had a PPV of $>80\%$ in combination with high SP.

Pre-test and post-test prevalence estimates for the 23 CFDs shown in figure 2A are summarised in table 1. In two studies,^{35 51} the pretest study prevalence of HDP (ie, prevalence according to the reference standard) was markedly higher than authors reported would have been expected in their target obstetrics population. Specifically, Leuf *et al*⁵¹ reported that preeclampsia impacts approximately 3% of all pregnancies; however, the true prevalence of HDP in their validation cohort was 8%. Further, while Labgold *et al*³⁵ reported that HDP complicates up to 10% of all pregnancies, the true prevalence of GH in their validation cohort was 20%, and the true prevalence of any maternal hypertensive disorder, 37%. There were also marked discrepancies between pretest and post-test prevalence estimates associated with at least one CFD in five studies.^{35 36 49–51} For example, Klemmensen *et al*⁵⁰ reported a pretest prevalence of 6% for hypertension diagnosed during pregnancy; however, according to their CFD, the study prevalence of these conditions was 3%. In another validation study,³⁵ the true prevalence of mild to moderate preeclampsia was 7%, however, according to their CFD, the prevalence of preeclampsia was 0.2%.

High study prevalence of HDP had a marked impact on PPV estimates reported for several CFDs (online supplemental table 21). Specifically, PPVs reported for nine CFDs ($n=4$ studies^{35 36 39 51}) were validated using study cohorts with a $\geq 3\%$ higher study prevalence of HDP than would be expected the general North American obstetrics population. These estimates were an average of 31% higher than their corresponding prevalence-adjusted

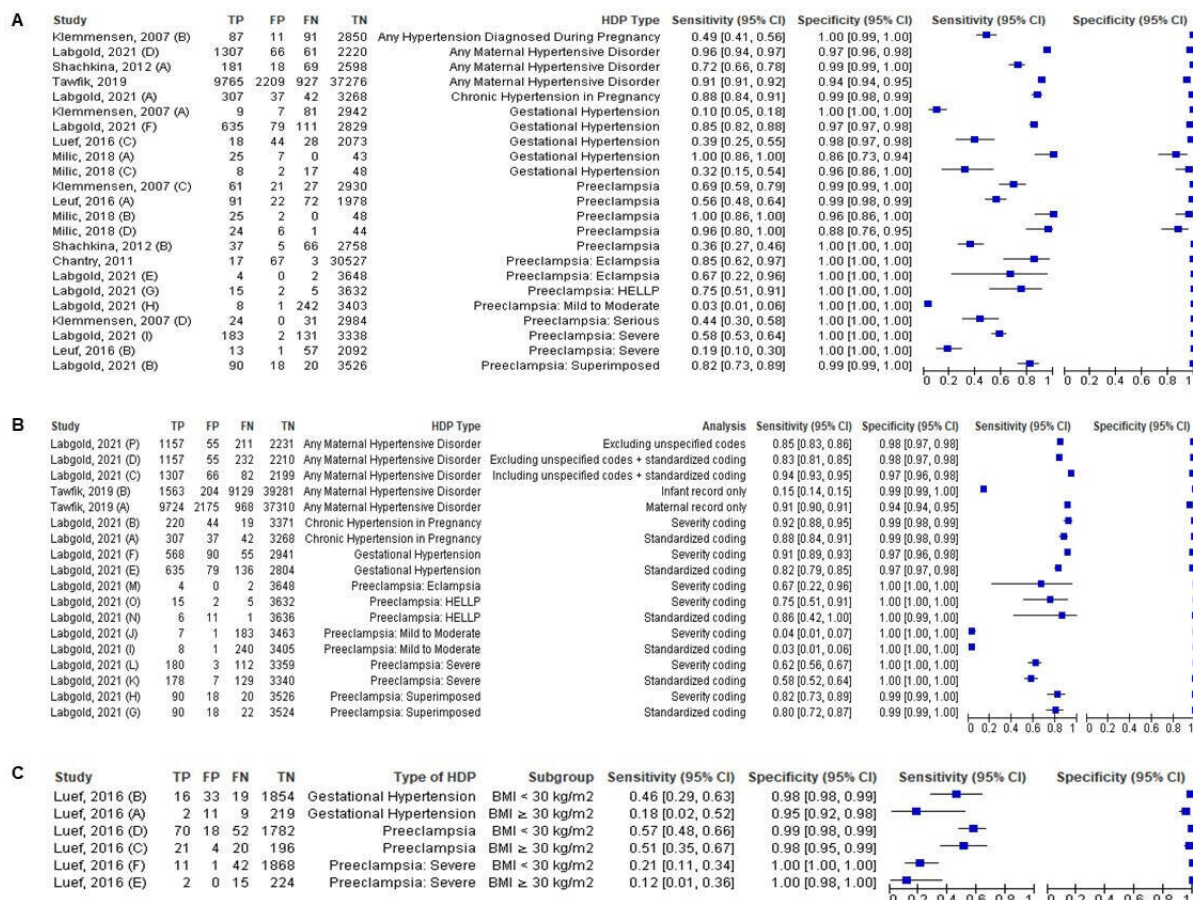


Figure 2 Forest plot showing estimates of sensitivity, specificity and corresponding 95% CIs for 23 administrative case-finding definitions (primary analyses—A), 18 administrative case-finding definitions (sensitivity analyses—B) and 6 administrative case-finding definitions (subgroup analyses—C) for hypertensive disorders of pregnancy (HDP) extracted from seven studies reporting all 2x2 contingency table values. Note: Letters in brackets behind each study's year of publication are meant to distinguish between different case-finding definitions reported by the same study. For gestational hypertension and preeclampsia Milic, 2018: (A) and (B)=results of their 'contemporary cohort' and Milic, 2018 (C) and (D)=results of their 'historic cohort'. Unspecified codes=International Classification of Diseases (ICD) codes for HDP that do not fully identify the diagnosis (eg, ICD-10-CM O14.90=unspecified pre-eclampsia, unspecified trimester); standardised coding=validation parameters are presented for HDP diagnoses that met the American College of Obstetricians and Gynecologists 2020 criteria, regardless of whether it was recorded in the medical record; severity coding=validation parameters are presented for the most severe recorded hypertensive outcome recorded. BMI, body mass index; FP, false positives; FN, false negatives; HELLP, haemolysis, elevated liver enzymes and low platelets syndrome; TP, true positives; TN, true negatives.

estimates. In one study, the PPV of a CFD validated for preeclampsia³⁶ decreased from 78% to 27% after applying a prevalence adjustment.

Sensitivity analyses

Four studies^{35 37 39 41} validated alternative coding strategies for a total of 29 CFDs (online supplemental table 16). Although Goueslard *et al*⁵² reported exploring alternative coding strategies, no results were reported.

The majority of sensitivity analyses were conducted by Labgold *et al*³⁵ (55% of all those reported). Strategies reported by this study included testing the validity of CFDs that used: (a) standardised coding (using 2020 American College of Obstetricians and Gynecologists criteria as the reference standard, even if an HDP diagnosis was not explicitly recorded in the medical record), (b) severity coding (validation parameters for CFDs limited to the

most severe HDP recorded in the medical record), as well as those that (c) excluded unspecified codes for HDP (eg, ICD-10 O10.9 'unspecified pre-existing hypertension complicating pregnancy').

Only two studies^{35 39} reported all 2x2 table values for the CFDs validated (n=18 CFDs; figure 2B). The SN, SP and corresponding 95% CIs of these comprehensively reported CFDs are summarised in figure 2B.

Comprehensively reported CFDs validated as part of sensitivity analyses identified HDPs with very high SP (>94%); however, SN values were quite variable, ranging from 3.2% (95% CI: 1.4% to 6.3%) for mild to moderate preeclampsia³⁵ to 94.1% (95% CI: 92.7% to 95.3%) for any maternal hypertensive disorder³⁵ (figure 2B). Compared with those that employed a standardised coding approach, SN estimates for CFDs that used severity

Table 1 Pre-test and post-test prevalence estimates for 23 administrative case-finding definitions for HDP (primary analyses only)

Author, year (country)	Author-reported prevalence of HDP in target obstetrics population	Type of hypertensive disorder of pregnancy case definition validated*	Pretest prevalence†	Post-test prevalence‡
Chantry <i>et al</i> , ²³ 2020 (France)	Not reported	Eclampsia	0.1%	0.3%
Klemmensen <i>et al</i> , ⁵⁰ 2007 (Denmark)	Prevalence of preeclampsia: 2%–6% of all pregnancies	Any pregnancy-induced hypertension	6%	3%
		Gestational hypertension	3%	1%
		Preeclampsia	3%	3%
		Serious preeclampsia	2%	1%
Luef <i>et al</i> , ⁵¹ 2016 (Denmark)	Prevalence of preeclampsia: approximately 3% of all pregnancies	Gestational hypertension	2%	3%
		Preeclampsia	8%	5%
		Severe preeclampsia	3%	1%
Milic <i>et al</i> , ³⁶ 2018 (USA)	Not reported	Gestational hypertension—historical	33%	13%
		Gestational hypertension—contemporary	33%	43%
		Preeclampsia—historical	33%	40%
		Preeclampsia—contemporary	33%	36%
Tawfik <i>et al</i> , ³⁹ 2019 (USA)	Not reported	Any maternal hypertensive disorder	21%	24%
Labgold <i>et al</i> , ³⁵ 2021 (USA)	Prevalence of hypertensive disorders during pregnancy: up to 10% of all pregnancies	Chronic hypertension	10%	9%
		Any maternal hypertensive disorder	37%	38%
		Gestational hypertension	20%	20%
		Superimposed preeclampsia	3%	3%
		Mild to moderate preeclampsia	7%	0.2%
		Severe preeclampsia	9%	5%
		Eclampsia	0.2%	0.1%
		HELLP syndrome	0.5%	0.5%
Shachkina, ⁴⁹ 2012 (Canada)	Prevalence of hypertension in pregnancy: approximately 5% of all pregnancies (11% of first pregnancies). Prevalence of preeclampsia: 3%–10% of all pregnancies§	Any maternal hypertensive disorder	9%	7%
		Preeclampsia	4%	1%

Note: Historical refers to the cohort that gave birth between 1976 and 1982; contemporary refers to the cohort that gave birth between 2012 and 2015.

*Reference standard definitions for each diagnosis validated can be found in the online supplemental table S12.

†Pretest prevalence=prevalence estimate generated according to the reference standard: $(TP+FN)/(TP+FP+FN+TN)$ ⁵⁵; the 'true' prevalence of HDP in the study.

‡Post-test prevalence=prevalence estimate generated according to the administrative CFD: $(TP+FP)/(TP+FP+FN+TN)$ ⁵⁵; the prevalence of HDP in the study according to the case-finding definition.

§Author conducted sample size calculations based on a preeclampsia prevalence of 5%.

CFD, case-finding definition; FN, false negatives; FP, false positives; HDP, hypertensive disorders of pregnancy; TN, true negatives; TP, true positives.

coding were higher for all HDP subtypes with the exception of HELLP syndrome. One CFD validated by Tawfik *et al*³⁹ identified any maternal hypertensive disorder in infant records with very poor SN (15%, 95% CI: 14% to 15%). Labgold *et al*³⁵ reported that including unspecified codes in CFDs validated for any maternal hypertensive disorder improved SN and NPV compared with those that excluded these codes; however, SPs and PPVs were similar. For more specific HDP diagnoses, the use of unspecified codes did not add substantial value.

Subgroup analyses

Five studies^{25 35 38 50 51} provided validation statistics for 47 CFDs reported separately according to participant

characteristics (eg, body mass index (BMI) group, prenatal care utilisation, maternal age, smoking status, parity, previous HDP diagnosis). Similar to primary and sensitivity analyses, the SP with which CFDs identified HDP were very high (all >95%) across all participant subgroups (online supplemental table 16). However, SNs were quite variable, ranging from 3.2% (95% CI: 1.3% to 6.4%) for a CFD validated for mild to moderate preeclampsia in Medicaid-insured women³⁵ to 97.2% (95% CI: 94.9% to 98.6%) for a CFD validated for any maternal hypertensive disorder in women who delivered in 2018.³⁵ Only one study⁵¹ reported all 2×2 table values for ≥1 CFD (n=6 comprehensively reported CFDs, figure 2C).

Across HDP subtypes, the SN with which CFDs identified HDP was lower for multiparous women compared with first-time mothers (56% vs 75%⁵⁰ for preeclampsia; 58% vs 80%²⁵ for any hypertension diagnosed during pregnancy, respectively) (online supplemental table 17). There were also marked differences in CFD sensitivity based on maternal age (74.7% for women <35 years vs 40.2% for women aged ≥35 years),²⁵ smoking status (75.4% for non-smokers vs 52.2% for smokers) and history of HDP (71.6% for no previous HDP vs 53.8% for previous HDP). Further, the SNs with which CFDs identified GH and preeclampsia were consistently higher for women in lower BMI categories. Specifically, Klemmensen *et al*⁵⁰ reported their CFD for preeclampsia was 77% sensitive in identifying the condition in women with a BMI <25 kg/m²; however, it was only 53% sensitive in identifying preeclampsia in women with a BMI ≥25 kg/m². As illustrated in figure 2C, the SNs with which CFDs identified women with GH and two preeclampsia subtypes were consistently lower for women with a BMI ≥30 kg/m² compared with those with a BMI <30 kg/m².⁵¹

Quality of reporting and ROB assessments

No studies cited the use of a published reporting guideline (eg, Standards for Reporting of Diagnostic Accuracy 2015⁵³). Further, ≥4 measures of diagnostic accuracy, along with corresponding 95% CIs, were reported or calculable for just 55% (n=78) of the 143 CFDs validated

for HDP. Only seven studies^{23 35 36 39 49–51} reported all 2×2 table values for ≥1 CFD, or all such values were calculable, and were formally assessed for quality of reporting and ROB. All other studies (n=19^{24 25 30–34 37 38 40–48 52}) were considered of limited overall value due to critically poor reporting and were not formally assessed for reporting quality or ROB.

The most reported (or calculable) validation statistic was PPV (92% of CFDs), followed by SN (81%), SP (73%) and NPV (55%). In six studies,^{31 32 37 38 41 44} only presumed cases of HDP identified by ≥1 CFD were checked for accuracy against a reference standard, thus, only TP and FP values were reported for these CFDs, and only PPVs were reported or calculable.

Quality of reporting

Quality of methodology reporting was poor to moderate (figure 3A) and the quality of results reporting generally poor (figure 3B); however, two studies^{35 49} were judged to be of ‘good’ overall reporting quality, one of which was published in 2012⁴⁹ and the other in 2021.⁴⁹ Study-specific results are provided in online supplemental tables 22 and 23.

ROB and study applicability

The results of targeted bias assessments (spectrum, incorporation, prevalence-related, review, selective outcome reporting) are presented in online supplemental figure 2.

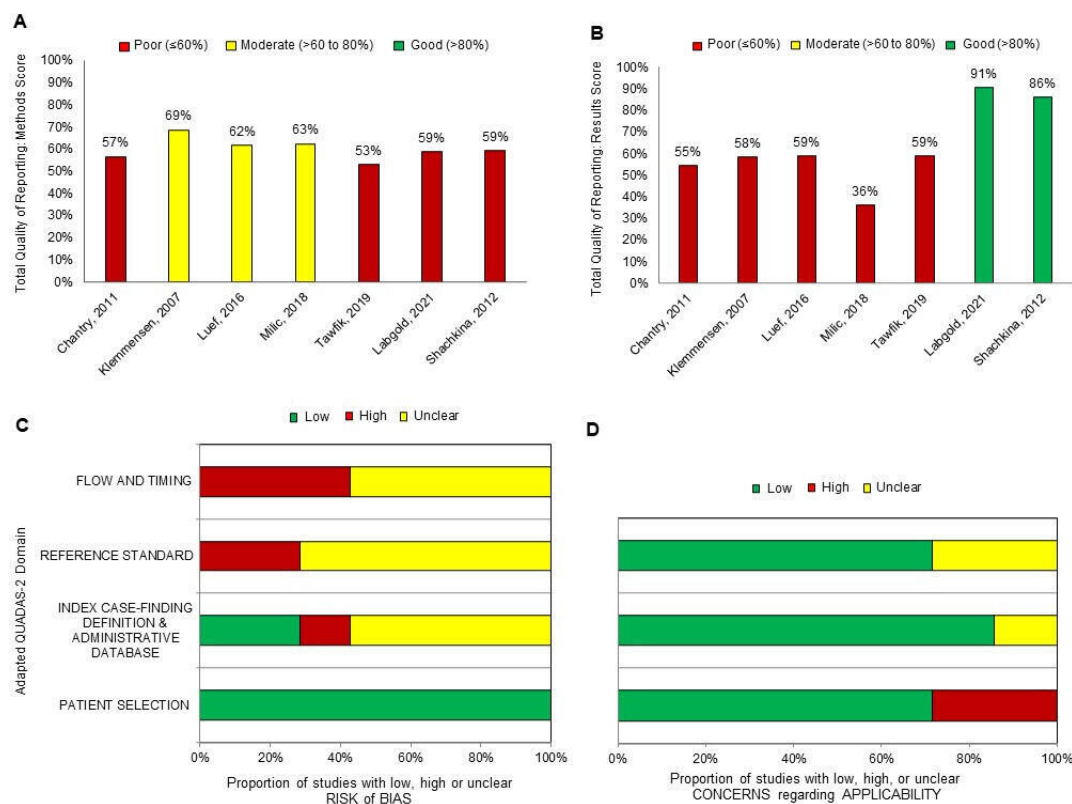


Figure 3 Results of the methodology (A) and results (B) reporting quality assessments, as well as the adapted QUADAS-2 (C), and applicability (D) assessment results for seven studies that reported all 2×2 contingency table values (true positives, true negatives, false positives, false negatives) for at least one hypertensive disorder of pregnancy administrative case-finding definition.

Most studies assessed^{35 36 39 49 50} (n=5; 71%) were judged to be at low ROB for selective outcome reporting and incorporation bias (n=4; 57%).^{23 36 49 51} Prevalence-related bias was of concern (ie, high ROB) in four studies^{35 36 39 51} as the prevalence of ≥ 1 HDP in each validation cohort was higher than would be expected in the general obstetrics population. Four studies^{23 35 36 39} were also judged to be at risk for spectrum bias, as participants were recruited from a high-risk setting. One unpublished Canadian study⁴⁹ was judged to be at low risk for all biases assessed. Study-specific results are provided in online supplemental table 24.

No studies were judged to be at low ROB in all four QUADAS-2 domains, though all studies were judged to be at low ROB for patient selection (figure 3C). Across the other three domains, most were judged to be at unclear ROB due to poor study reporting. Three studies^{23 39 51} were judged to be at high ROB related to study flow and timing concerns; two of which^{23 51} were also judged to be at high ROB due to concerns related to the reference standard. Domain-specific signalling questions and study-specific ROB judgements are summarised in online supplemental tables 25 and 26.

There were no concerns that participants with HDP in included studies matched our research question; however, two studies assessed did not validate CFDs for HDP as a primary aim, raising some concern about their direct applicability to this review (figure 3D). In one study,³⁹ we were unclear about whether the CFDs validated, and/or their interpretation by study authors, differed from our research question. We were also unclear about whether the target condition as defined by the reference standard met our research question in two studies.^{23 39} Study-specific applicability judgements are summarised in online supplemental table 27.

Best currently available administrative CFDs for HDP

The characteristics of all CFDs with an SP $\geq 97\%$, in combination with a (reported or prevalence-adjusted) PPV of $\geq 70\%$, along with corresponding study reporting and bias assessment judgements, are summarised in table 2. These 11 CFDs, which validated eight different HDP definitions, represent the best currently available CFDs for identifying HDP in administrative data. Disease classification codes comprising these definitions are summarised in online supplemental table 28.

DISCUSSION

We identified and synthesised available evidence on the validity of CFDs designed to identify HDP in administrative data. Most validated CFDs were applied exclusively to discharge abstracts for delivery hospitalisation, and no studies validated CFDs for HDP in women with multiple gestations or for a postpartum hypertension diagnosis. While SP estimates were consistently very high across all validated CFDs, SN estimates varied widely, especially for those designed to identify specific HDP subtypes. We also found clear evidence that CFDs identify HDP diagnoses

in women with additional CVD risk factors (eg, elevated BMI, multiparity, smokers, previous HDP diagnosis) with markedly lower SN compared with women without these additional risk factors at the time of delivery. Overall, quality of study reporting was poor to moderate, and we noted substantial heterogeneity across included studies with respect to participant and study design characteristics, reference standard definitions, CFD characteristics, and validation statistics reported.

Findings in light of previous research

Ideally, all CFDs validated for HDP would identify these conditions with high SN and SP. However, there is always a trade-off between these measures such that higher SP will usually mean lower SN and vice versa. Our finding that most CFDs for HDP had very high SP, but poor to moderate SN, is a trade-off consistent with the findings of previous validation studies reporting on other maternal conditions.^{25 34 42} These findings may be explained by a few factors. First, some women may only receive a diagnostic code for HDP if an anti-hypertensive medication is prescribed³⁰ or if they have more severe forms of HDP (eg, eclampsia) that have a substantial impact on patient management.²⁵ Conversely, conditions with more mild features (eg, chronic hypertension in pregnancy, GH) may not be documented in hospital discharge summaries if they do not impact the current hospital admission.²⁵ Further, because clinical coders cannot infer diagnoses, if HDPs are not explicitly recorded in the medical record, they are unlikely to be recorded in the administrative record.⁴⁵ Finally, as noted in a similar review of CFDs for non-pregnancy-related hypertension,⁵⁴ physicians may be less likely to bill for HDP management in the presence of ≥ 1 comorbid condition.

As highlighted by our ROB and applicability assessments, participant characteristics and sampling strategies were of particular concern across included studies, leading us to scrutinise their impact on the validation statistics reported. Specifically, few studies randomly sampled participants from the general (or an objectively generalisable) obstetric population, which would have resulted in pretest prevalence estimates reflective of the general obstetrical population (the preferred approach).⁵⁵ Instead, several authors selected participant records with at least one ICD code for HDP, and either randomly selected records that were code negative for HDP or included all women that were code negative for HDP but code positive for any number of other maternal morbidities (the TN reference). Finally, at least one study³⁶ selected an arbitrary ratio of participants with and without HDP, resulting in a pretest prevalence of 33% for both GH and preeclampsia in their validation cohort.

Given the rarity of certain HDP subtypes, some included studies recruited women from high-risk settings (eg, high-risk hospitals) to increase the likelihood of including participants with maternal morbidity. While identifying participants from these settings allowed authors to examine the validity of CFDs for rarer HDP subtypes, this came at the expense of falsely inflated PPVs (as illustrated by our prevalence-adjusted

Table 2 Characteristics of the best currently available case-finding definitions designed to identify hypertensive disorders of pregnancy in administrative data

Author, year (country)	HDP definition validated	SN (%) (95% CI)	SP (%) (95% CI)	Pretest HDP prevalence*	PPV (95% CI)†	Reporting quality: methods/results	Key bias assessment judgements	Adapted QUADAS-2 risk of bias judgements†††
Labgold <i>et al.</i> , ³⁵ 2021 (USA)	cHTN in pregnancy	88 (84 to 91)	99 (99 to 99)	10%	81 (75 to 85)‡	Poor/good	Incorporation: unclear risk	D1: low risk
	Any HDP§	96 (94 to 97)	97 (96 to 98)	37%	79 (74 to 82)¶		Prevalence: high risk	D2: unclear risk
	GH	85 (82 to 88)	97 (97 to 98)	20%	70 (65 to 75)**		Spectrum: high risk	D3: unclear risk
	SPE	82 (73 to 89)	100 (99 to 100)	3%	83 (75 to 90)		Review: unclear risk	D4: unclear risk
	Eclampsia	67 (22 to 96)	100 (100 to 100)	0.2%	100 (40 to 100)			
Shachkina, ⁴⁹ 2012 (Canada)	HELLP	75 (51 to 91)	100 (100 to 100)	0.5%	88 (64 to 99)			
	Any HDPT‡	72 (67 to 78)	99 (99 to 100)	9%	91 (86 to 94)	Poor/good	Nil	D1: low risk
	PE‡‡	36 (27 to 46)	100 (100 to 100)	4%	88 (74 to 94)			D2: low risk
Klemmensen <i>et al.</i> , ⁵⁰ 2007 (Denmark)	De novo HTN in pregnancy§§	49 (41 to 56)	100 (99 to 100)	6%	89 (81 to 94)	Moderate/poor	Review: unclear risk	D1: low risk
	PE¶¶	69 (59 to 79)	99 (99 to 100)	3%	74 (64 to 83)		Incorporation: unclear risk	D2: unclear risk
Luef <i>et al.</i> , ⁵¹ 2016 (Denmark)	PE***	56 (48 to 64)	99 (98 to 99)	8%	81 (72 to 87)	Moderate/poor	Selective reporting: high risk	D1: low risk
							Prevalence: high risk	D2: high risk
							Review: unclear risk	D3: high risk
							Spectrum: unclear risk	D4: high risk

*Prevalence estimate generated according to the reference standard: (TP+FN)/(TP+FP+FN+TN)⁵⁵; that is, the 'true' prevalence of hypertensive disorders of pregnancy in the study.
†Estimated prevalence of HDP in the general North American obstetrical population based on recently published estimates. Adjusted estimates are indicated.
‡Adjusted based on a population-based prevalence estimate of 5%.
§Includes ICD-10 codes for pre-existing and essential hypertension, chronic hypertension with superimposed preeclampsia and hypertension diagnosed during pregnancy.
¶Adjusted based on a population-based prevalence estimate of 10%.
**Adjusted based on a population-based prevalence estimate of 7%.
‡‡Includes ICD-10-CA codes for chronic hypertension with superimposed preeclampsia and hypertension diagnosed during pregnancy.
‡‡‡Includes ICD-10-CA codes for pre-existing hypertension with preeclampsia, severe preeclampsia and eclampsia.
§§Includes ICD-10 codes for GH, mild to moderate PE, HELLP, unspecified PE and eclampsia.
¶¶Includes ICD-10 codes for mild to moderate, unspecified and severe preeclampsia, HELLP syndrome and eclampsia.
***Includes ICD-10 codes for mild to moderate preeclampsia, unspecified and severe preeclampsia and HELLP syndrome.
†††D1 to D4 refer to the results of each of the 4 domains of the modified QUADAS-2 assessments: D1 - patient selection, D2 - index case-finding definition and administrative database, D3 - reference standard, D4 - flow and timing. Further details can be found in the data supplement.
Any HDP, any maternal hypertensive disorder; cHTN in pregnancy, chronic hypertension in pregnancy; de novo HTN in pregnancy, new hypertension diagnosed during pregnancy; GH, gestational hypertension; HELLP, haemolysis, elevated liver enzymes and low platelets syndrome; PE, preeclampsia; PPV, positive predictive value; SN, sensitivity; SP, specificity; SPE, chronic hypertension with superimposed preeclampsia.

PPVs), driven by largely unrealistic pretest HDP prevalence. Importantly, CFDs may still perform well under these circumstances if the post-test prevalence of disease approximates pretest prevalence.⁵⁵ However, we found that this was not the case for a number of CFDs.

Although PPVs and NPVs are directly impacted by prevalence, SN and SP may also be indirectly impacted if inflated disease prevalence is also accompanied by higher than expected disease severity.⁵⁶ For example, if studies validating CFDs for ‘any preeclampsia’ mostly include women with severe clinical features, the SN and SP associated with the resulting CFDs would be artificially inflated if applied to more general clinical settings where the spectrum of disease may look quite different.¹⁴ Unfortunately, we were unable to fully interpret the reliability of validation statistics reported across studies, as reporting of participant characteristics was poor, and most studies failed to provide adequate descriptions of HDP severity, or report on additional clinical factors associated with HDP severity (eg, participants who experienced a preterm birth), for their validation cohorts.⁵⁷

Choosing the best CFD(s) for HDP

CFDs should always be chosen in light of several important considerations, including (1) applicability to available data sources,^{58–60} (2) validation study design and participant characteristics, (3) methodological rigour and (4) assessments of bias and reporting quality for the studies from which they were derived. The relative importance of high SN, SP, NPV and/or PPV must also be considered, and considerations prioritised based on study objectives. Specifically, if the goal is to estimate the overall burden of HDP (ie, disease surveillance), CFDs with high SN should be prioritised, as this would maximise the number of individuals identified, generating a more accurate estimate of the burden of HDP in the population.^{54 61 62} However, in aetiologic studies, decisions about what validation statistic to prioritise may not be as straightforward.⁵⁰

If researchers investigate the association between HDP and CVD and classify HDP as a binary exposure (ie, present vs absent), in the presence of non-differential misclassification, HDPs identified using a CFD with imperfect SN or SP will normally bias effect estimates toward the null.⁵⁰ However, many argue that CFDs with high SP and/or PPV should be prioritised in aetiologic studies, even at the expense of SN.^{55 61 62} Specifically, while CFDs that identify HDP with low SN will result in a number of true cases being missed, reducing statistical power,⁵⁰ even in circumstances where the SN is high, low SP can also reduce statistical power, leading to overly conservative effect estimates.^{42 61} Furthermore, when conditions are rare, risk estimates generated in the presence of non-differential misclassification will be unbiased when SP is nearly perfect, even when SN is low.³² However, risk differences will be underestimated in such a scenario.³²

Notwithstanding the above considerations, researchers should note that using CFDs with low SN may result in selection bias and lower overall generalisability.⁶³ This is

an especially important consideration within the context of HDP, since many validated CFDs for this clinical population identify women with additional CVD risk factors (eg, elevated BMI, smokers, previous HDP) with markedly lower SN compared with women without these risk factors. Furthermore, even high SN, or near-perfect SP in the presence of non-differential misclassification is not sufficient to guarantee that CVD risk estimates for women with a history of HDP will be accurate or unbiased.⁴⁹ Indeed, many factors can impact the reliability of effect estimates generated using administrative data (eg, misclassification of covariates and confounding), and the combined effect of all such factors should also be considered when results are interpreted.⁴⁹

Although we identified and summarised the best currently available CFDs for HDP based on reported/prevalence-adjusted validation estimates, researchers should carefully consider the appropriateness of their use—and actual performance—if applied to their own data. Importantly, just over half (n=6) of the CFDs presented in [table 2](#) and online supplemental table 28 were validated in a high burden American population who delivered at a single hospital.³⁵ Further, all four studies^{35 49–51} from which the CFDs were derived were judged to be at unclear or high ROB for ≥1 type of key bias and/or QUADAS-2 domain. Ultimately, researchers should choose the CFD(s) that best aligns with their research objective and population of interest, while also considering additional details about study prevalence of HDP, quality of reporting and ROB assessment results for each of these studies from which CFDs were derived.

Considerations for future administrative validation studies for HDP

Higher quality validation studies in this area are urgently needed. Specifically, future validation studies for HDP should (a) follow best practices in study design, conduct and reporting; (b) use appropriate sampling methods to ensure that validation cohorts closely reflect their target obstetrics population and (c) make use of established reporting guidance for administrative validation studies. Given increasing evidence about the importance of dose–response relationships between HDP severity and CVD risk, future validation studies should also focus on improving the SN with which specific HDP subtypes are identified and validating CFDs for HDP recurrence as well as ‘early-’ and ‘late-onset’ HDP.

Strengths and limitations

We used established systematic review methodology specifically tailored for use on administrative validation studies, registered the review protocol a priori, carried out a comprehensive peer-reviewed literature search strategy and employed broad eligibility criteria that were not restrictive of geography, time or HDP definition. Although this review was undertaken to inform population-based research in Ontario, Canada, because we placed no limitation on location of conduct, our

findings will also be relevant to researchers in other jurisdictions. Despite these strengths, our findings should be interpreted in the light of important limitations. First, our ability to directly compare and pool validation statistics was inhibited by poor reporting of study and participant characteristics, heterogeneous CFDs and HDP definitions and diagnostic criteria that differed across jurisdictions and over time. Second, although we undertook a comprehensive, peer-reviewed, literature search, we may be missing eligible studies that have not been indexed in the bibliographic databases we searched. Third, we limited thorough assessments of ROB and reporting quality to a subset of included studies that reported all 2×2 table values for ≥1 CFD of interest or reported sufficient information such that those values were calculable. While we considered studies that did not meet this criterion to be of limited overall value due to critically poor reporting, it is possible that these studies were methodologically rigorous. However, given (a) the reporting recommendation by Benchimol *et al*¹⁴ that these values always be reported in administrative validation studies, (b) that reporting quality can have a substantial impact on a study's internal and external validity and credibility^{57 64} and (c) these studies were also generally poor at reporting other critical information (eg, the ICD codes that comprised the CFDs), (d) that it can be very difficult to carry out any meaningful quality assessments on studies with poor reporting,^{57 65} (e) that studies were not excluded based on reporting quality or ROB and (f) that all relevant validation statistics for all CFDs of interest reported by included studies were extracted and reported, we feel it is unlikely that assessment of quality or ROB for these studies would have impacted the study's conclusions. Finally, we cannot rule out the possibility that CFDs with poor validity remain unpublished and, thus, have not been included in this review.

CONCLUSIONS

Validated CFDs identify HDP with high SP; however, their SNs vary widely and are especially poor for specific HDP diagnoses/subtypes. Critically, researchers should be aware that even validated CFDs are subject to bias and should choose the CFD(s) that best aligns with their research objective(s), while also considering the relative importance of high SN, SP, NPV and/or PPV, and important characteristics of the validation studies (eg, study prevalence of HDP, spectrum of disease, methodological rigour, quality of reporting and ROB) from which they were derived. Until optimal CFDs for HDP are created, researchers are encouraged to use the best currently available CFDs summarised in this review, but carefully manage expectations about their performance when applied to their own data.

Author affiliations

¹School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada

²Brain and Heart Nexus Research Program, University of Ottawa Heart Institute, Ottawa, Ontario, Canada

³School of Medicine, University of Ottawa, Ottawa, Ontario, Canada

⁴Independent Information Specialist, Ottawa, Ontario, Canada

⁵Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

⁶ICES, Ottawa, Ontario, Canada

⁷Division of Palliative Care, Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada

⁸Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Kingston Health Sciences Centre, Kingston, Ontario, Canada

⁹Division of Cardiology, University of Ottawa Heart Institute, Ottawa, Ontario, Canada

¹⁰Canadian Women's Heart Health Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada

¹¹Division of Cardiac Prevention and Rehabilitation, University of Ottawa Heart Institute, Ottawa, Ontario, Canada

Twitter Amy Johnston @amydjohn

Acknowledgements We wish to thank Sanaz Johnston for her assistance in screening non-English language articles for eligibility. We also wish to thank Stephanie Metcalfe, MSc (Public Health Agency of Canada), Robyn Hocking, MLIS (Health Library, Health Canada), Hugo Crites (Geographic, Statistical and Government Information Centre, University of Ottawa Library) and Sarah Visintini, MLIS (Berkman Library, University of Ottawa Heart Institute) for their assistance in locating the full text versions of included grey literature.

Contributors AJ conceived and designed the study, participated in data acquisition, performed the analyses, interpreted the findings, drafted the manuscript, and acts as guarantor of this work. BS, TC and JDE assisted with study design; SRD, VT and BS participated in data acquisition, and BS drafted the literature review section of the manuscript. All authors reviewed and revised the draft version for important intellectual context and approved the final version to be published.

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

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

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

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

ORCID iD

Amy Johnston <http://orcid.org/0000-0003-1934-7867>

REFERENCES

- 1 Roberts JM, Rich-Edwards JW, McElrath TF, *et al*. Subtypes of preeclampsia: recognition and determining clinical usefulness. *Hypertension* 2021;77:1430–41.
- 2 Magee LA, Pels A, Helewa M, *et al*. Canadian hypertensive disorders of pregnancy working G. Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy: executive summary. *JOGC* 2014;36:416–41.

- 3 Wójtowicz A, Zembala-Szczerba M, Babczyk D, *et al.* Early- and late-onset preeclampsia: a comprehensive cohort study of laboratory and clinical findings according to the new ISHHP criteria. *Int J Hypertens* 2019;2019:4108271.
- 4 Riise HKR, Sulo G, Tell GS, *et al.* Association between gestational hypertension and risk of cardiovascular disease among 617 589 Norwegian women. *J Am Heart Assoc* 2018;7:e008337.
- 5 Melamed N, Ray JG, Hladunewich M, *et al.* Gestational hypertension and preeclampsia: are they the same disease. *J Obstet Gynaecol Can* 2014;36:642–7.
- 6 Butalia S, Audibert F, Côté A-M, *et al.* Hypertension Canada's 2018 guidelines for the management of hypertension in pregnancy. *Can J Cardiol* 2018;34:526–31.
- 7 Forest J-C, Girouard J, Massé J, *et al.* Early occurrence of metabolic syndrome after hypertension in pregnancy. *Obstet Gynecol* 2005;105:1373–80.
- 8 Yagel S, Cohen SM, Goldman-Wohl D. An integrated model of preeclampsia: a multifaceted syndrome of the maternal cardiovascular-Placental-fetal array. *Am J Obstet Gynecol* 2022;226:S963–72.
- 9 Wilson BJ, Watson MS, Prescott GJ, *et al.* Hypertensive diseases of pregnancy and risk of hypertension and stroke in later life: results from cohort study. *BMJ* 2003;326:845.
- 10 Kestenbaum B, Seliger SL, Easterling TR, *et al.* Cardiovascular and thromboembolic events following hypertensive pregnancy. *Am J Kidney Dis* 2003;42:982–9.
- 11 Ray JG, Vermeulen MJ, Schull MJ, *et al.* Cardiovascular health after maternal placental syndromes (CHAMPS): population-based retrospective cohort study. *Lancet* 2005;366:1797–803.
- 12 Johnston A, Tseung V, Dancy SR, *et al.* Use of race, ethnicity, and national origin in studies assessing cardiovascular risk in women with a history of hypertensive disorders of pregnancy. *CJC Open* 2021;3:S102–17.
- 13 Harron K, Dibben C, Boyd J, *et al.* Challenges in administrative data linkage for research. *Big Data & Society* 2017;4:205395171774567.
- 14 Benchimol EI, Manuel DG, To T, *et al.* Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–9.
- 15 Ehrenstein V, Nielsen H, Pedersen AB, *et al.* Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clin Epidemiol* 2017;9:245–50.
- 16 Mazzali C, Paganoni AM, Ieva F, *et al.* Methodological issues on the use of administrative data in healthcare research: the case of heart failure hospitalizations in Lombardy region. *BMC Health Serv Res* 2016;16:234.
- 17 Salameh J-P, Bossuyt PM, McGrath TA, *et al.* Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ* 2020;370:m2632.
- 18 McInnes MDF, Moher D, Thoms BD, *et al.* Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;319:388–96.
- 19 Lain SJ, Hadfield RM, Raynes-Greenow CH, *et al.* Quality of data in perinatal population health databases: a systematic review. *Med Care* 2012;50:e7–20.
- 20 Rethlefsen ML, Kirtley S, Waffenschmidt S, *et al.* PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev* 2021;10:39.
- 21 Collett D. *Modelling Binary Data*. Boca Raton, Florida: Chapman & Hall/CRC, 1999.
- 22 Diagnostic test evaluation Calculator 20.104 ed. MedCalc Software Ltd; 2022.
- 23 Chantry AA, Deneux-Tharaux C, Cans C, *et al.* Hospital discharge data can be used for monitoring procedures and intensive care related to severe maternal morbidity. *J Clin Epidemiol* 2011;64:1014–22.
- 24 Lydon-Rochelle MT, Holt VL, Cárdenas V, *et al.* The reporting of pre-existing maternal medical conditions and complications of pregnancy on birth certificates and in hospital discharge data. *Am J Obstet Gynecol* 2005;193:125–34.
- 25 Roberts CL, Bell JC, Ford JB, *et al.* The accuracy of reporting of the hypertensive disorders of pregnancy in population health data. *Hypertens Pregnancy* 2008;27:285–97.
- 26 Chantry AA, Deneux-Tharaux C, Bal G, *et al.* Le programme de Médicalisation Du Système D'Information (PMSI) – Processus de production des Données, Validité et sources D'Erreurs Dans le Domaine de la Morbidité Maternelle Sévère. *Revue d'Épidémiologie et de Santé Publique* 2012;60:177–88.
- 27 Hadfield RM, Lain SJ, Cameron CA, *et al.* The prevalence of maternal medical conditions during pregnancy and a validation of their reporting in hospital discharge data. *Aust N Z J Obstet Gynaecol* 2008;48:78–82.
- 28 Lydon-Rochelle MT, Holt VL, Nelson JC, *et al.* Accuracy of reporting maternal in-hospital diagnoses and intrapartum procedures in Washington state linked birth records. *Paediatr Perinat Epidemiol* 2005;19:460–71.
- 29 Roberts CL, Cameron CA, Bell JC, *et al.* Measuring maternal morbidity in routinely collected health data: development and validation of a maternal morbidity outcome indicator. *Med Care* 2008;46:786–94.
- 30 Chen L, Shortreed SM, Easterling T, *et al.* Identifying hypertension in pregnancy using electronic medical records: the importance of blood pressure values. *Pregnancy Hypertension* 2020;19:112–8.
- 31 Geller SE, Ahmed S, Brown ML, *et al.* International classification of diseases–9Th revision coding for preeclampsia: how accurate is it. *Am J Obstet Gynecol* 2004;190:1629–33.
- 32 He M, Huybrechts KF, Dejene SZ, *et al.* Validation of algorithms to identify adverse perinatal outcomes in the Medicaid analytic extract database. *Pharmacoepidemiol Drug Saf* 2020;29:419–26.
- 33 Himes KP, Bodnar LM. Validation of criteria to identify severe maternal morbidity. *Paediatr Perinat Epidemiol* 2020;34:408–15.
- 34 Korst LM, Gregory KD, Gornbein JA. Elective primary Caesarean delivery: accuracy of administrative data. *Paediatr Perinat Epidemiol* 2004;18:112–9.
- 35 Labgold K, Stanhope KK, Joseph NT, *et al.* Validation of hypertensive disorders during pregnancy: ICD-10 codes in a high-burden southeastern United States hospital. *Epidemiology* 2021;32:591–7.
- 36 Milic NM, Codsí E, Butler Tobah YS, *et al.* Electronic algorithm is superior to hospital discharge codes for diagnoses of hypertensive disorders of pregnancy in historical cohorts. *Mayo Clin Proc* 2018;93:1707–19.
- 37 Palmsten K, Huybrechts KF, Kowal MK, *et al.* Validity of maternal and infant outcomes within nationwide Medicaid data. *Pharmacoepidemiol Drug Saf* 2014;23:646–55.
- 38 Sigakis MJG, Leffert LR, Mirzakhani H, *et al.* The validity of discharge billing codes reflecting severe maternal morbidity. *Anesth Analg* 2016;123:731–8.
- 39 Tawfik DS, Gould JB, Profit J. Perinatal risk factors and outcome coding in clinical and administrative databases. *Pediatrics* 2019;143:e20181487.
- 40 Yasmeen S, Romano PS, Schembri ME, *et al.* Accuracy of obstetric diagnoses and procedures in hospital discharge data. *Am J Obstet Gynecol* 2006;194:992–1001.
- 41 Chomistek AK, Phiri K, Doherty MC, *et al.* Development and validation of ICD-10-CM-based algorithms for date of last Menstrual period, pregnancy outcomes, and infant outcomes. *Drug Saf* 2023;46:515.
- 42 Davey M-A, Sloan M-L, Palma S, *et al.* Methodological processes in validating and analysing the quality of population-based data: a case study using the Victorian perinatal data collection. *Health Inf Manag* 2013;42:12–9.
- 43 Henderson T, Shephard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. *Med Care* 2006;44:1011–9.
- 44 Riley M, Halliday J. The accuracy of eclampsia cases reported to the Victorian inpatient minimum database and the perinatal data collection unit. *Health Inf Manag* 1998;28:13–5.
- 45 Taylor LK, Travis S, Pym M, *et al.* How useful are hospital morbidity data for monitoring conditions occurring in the perinatal period. *Aust N Z J Obstet Gynaecol* 2005;45:36–41.
- 46 Thornton C, Makris A, Ogle R, *et al.* Generic obstetric database systems are unreliable for reporting the hypertensive disorders of pregnancy. *Aust N Z J Obstet Gynaecol* 2004;44:505–9.
- 47 An evaluation of the quality of obstetric/neonatal discharge abstract data by reabstraction of medical charts. Ottawa Health Canada; 2003.
- 48 Joseph KS, Fahey J, Canadian Perinatal Surveillance System. Validation of perinatal data in the discharge abstract database of the Canadian Institute for health information. *Chronic Dis Can* 2009;29:96–100.
- 49 Shachkina S. *Validity of Administrative Database for Reporting Pre-Eclampsia*. Ottawa, ON: University of Ottawa, 2012.
- 50 Klemmensen AK, Olsen SF, Osterdal ML, *et al.* Validity of preeclampsia-related diagnoses recorded in a national hospital registry and in a postpartum interview of the women. *Am J Epidemiol* 2007;166:117–24.
- 51 Luef BM, Andersen LB, Renåult KM, *et al.* Validation of hospital discharge diagnoses for hypertensive disorders of pregnancy. *Acta Obstet Gynecol Scand* 2016;95:1288–94.

- 52 Goueslard K, Cottenet J, Benzenine E, *et al.* Validation study: evaluation of the metrological quality of French hospital data for perinatal algorithms. *BMJ Open* 2020;10:e035218.
- 53 Cohen JF, Korevaar DA, Altman DG, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
- 54 Pace R, Peters T, Rahme E, *et al.* Validity of health administrative database definitions for hypertension: a systematic review. *Can J Cardiol* 2017;33:1052–9.
- 55 Widdifield J, Labrecque J, Lix L, *et al.* Systematic review and critical appraisal of validation studies to identify rheumatic diseases in health administrative databases. *Arthritis Care Res (Hoboken)* 2013;65:1490–503.
- 56 Whiting P, Rutjes AWS, Reitsma JB, *et al.* Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189–202.
- 57 Schmidt RL, Factor RE. Understanding sources of bias in diagnostic accuracy studies. *Arch Pathol Lab Med* 2013;137:558–65.
- 58 Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. *BMJ* 2010;341:bmj.c4226.
- 59 Benchimol EI. Epidemiology and health administrative data: focus on methodology and transparency. *Inflamm Bowel Dis* 2014;20:1780–1.
- 60 Chan J, Mack DR, Manuel DG, *et al.* Validation of an algorithm to identify children with biopsy-proven celiac disease from within health administrative data: an assessment of health services utilization patterns in Ontario, Canada. *PLoS One* 2017;12:e0180338.
- 61 Osterdal M, Klemmensen A, Tabor A, *et al.* Respond to "identifying women with hypertension during pregnancy". *Am J Epidemiol* 2007;166:128–9.
- 62 Tu K, Campbell NR, Chen ZL, *et al.* Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007;1:e18–26.
- 63 Rector TS, Wickstrom SL, Shah M, *et al.* Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. *Health Serv Res* 2004;39:1839–57.
- 64 Tsakiridis I, Arvanitaki A, Zintzaras E. Assessing the reporting quality of systematic reviews of observational studies in preeclampsia. *Arch Gynecol Obstet* 2019;299:689–94.
- 65 Jensen PN, Johnson K, Floyd J, *et al.* A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:141–7.