Received: 1 September 2020 | Accepted: 15 April 2021

DOI: 10.1002/pds.5251

ORIGINAL ARTICLE

WILEY

The identification and validity of congenital malformation diagnoses in UK electronic health records: A systematic review

Maria Peppa | Caroline Minassian | Punam Mangtani | Sara L. Thomas

Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

Correspondence
Maria Peppa, Faculty of Epidemiology and Population Health, Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
Email: m.peppa@ucl.ac.uk

Funding information
National Institute for Health Research Health Protection Research Unit in Immunisation at the London School of Hygiene and Tropical Medicine in partnership with Public Health England.

Abstract

Purpose: To describe the methods used to identify and validate congenital malformation diagnoses recorded in UK electronic health records, and the results of validation studies.

Methods: Medline and Embase were searched for publications between 1987 and 2019 that involved identifying congenital malformations from UK electronic health records using diagnostic codes. The methods and code-lists used to identify congenital malformations, and the methods and results of validations, were examined.

Results: We retrieved 54 eligible studies; 36 identified congenital malformations from primary care data and 18 from secondary care data alone or in combination with birth and/or death records. Identification in secondary care data relied on codes from the 'Q' chapter for congenital malformations in ICD-10. In contrast, studies using primary care data frequently used additional codes outside of the 'P' chapter for congenital malformation diagnoses in Read, although the exact codes used were not always clear. Eight studies validated diagnoses identified in primary care data. The positive predictive value was highest (80%–100%) for congenital malformations overall, major malformations, and heart defects although the validity of the reference standard used was often uncertain. It was lowest for neural tube defects (71%) and developmental hip dysplasia (56%).

Conclusions: Studies identifying congenital malformations from primary care data provided limited details about the methods used. The few validation studies were limited to diagnoses recorded in primary care. Further assessments of all measures of validity in both data sources and of other malformation subgroups are needed, using robust reference standards and adhering to reporting guidelines.

KEYWORDS
congenital abnormalities, electronic health records, pharmacoepidemiology, systematic review, United Kingdom, validation study

1 | INTRODUCTION

Large databases of anonymised, routinely collected electronic health records (EHR) are increasingly being used for post-licensure safety studies of drugs and vaccines given in pregnancy.[1,2] UK primary care databases, such as the Clinical Practice Research Datalink (CPRD), contain information recorded in general practice for a representative sample of the population.[3] Secondary care databases, such as Hospital Episode Statistics Admitted Patient Care (HES-APC), capture information on all patients admitted to NHS hospitals in England.[4] These and other administrative data such as death certificates can be provided to researchers pre-linked to maximise the ascertainment of safety outcomes.[3,4]

Congenital malformations (CMs) are rare but important outcomes in safety studies. Therefore, the methods used to identify CMs in EHR are of interest. In primary care, clinical data including diagnoses and procedures are currently coded using the hierarchical Read coding system; some practices previously used Oxford Medical Information System (OXMIS) codes.[3] In secondary care, diagnoses and procedures are coded using the International Classification of Diseases 10th Revision (ICD-10) and the Classification of Interventions and Procedures 4th Revision (OPCS-4), respectively.[4] Death certificates are also coded using ICD-10.[5] By developing a list of relevant codes ('code-lists') researchers can search patient records in EHR data to identify CM diagnoses.[6]

Chapters 'P' in Read and 'Q' in ICD-10 are dedicated to CM codes. However, clinical coders may use codes from other chapters to record CMs. For example, congenital pulmonary valve malformations could be recorded using the Read code 'pulmonary valve anomalies' from the CM 'P' chapter or the less specific Read code 'pulmonary valve disorders' from the 'G' chapter on circulatory system diseases. Inclusion of broad-ranging codes from outside dedicated CM chapters enables more complete capture of CMs but also increases the risk of capturing non-congenital conditions.

To aid code-list development, researchers can refer to published guidelines which define codes for particular conditions or provide relevant key terms which can be used to identify equivalent codes. The 'European Surveillance of Congenital Anomalies' network (EUROCAT), for example, publishes the modified ICD-10 codes used by registries to code and classify CMs identified in the first year of life.[7] However, such guidelines are not used to encode CMs in routine clinical practice.

The heterogeneity of CMs and lack of a standardised algorithm for their identification in EHR complicate their ascertainment and may result in a variety of approaches, affecting replicability across studies and validity of the findings. Previous reviews have noted the good validity of diagnostic codes in UK EHR overall.[8,9] However, to date, no study has systematically assessed the validity of code-lists developed to identify CMs in these data.

This systematic review aimed to inform future safety studies by describing the methods used to identify individuals with CMs from UK EHR and by summarising the results of associated validations.

Key Points

- Post-licensure safety studies of drugs and vaccines in pregnancy have relied on primary care data.
- More extensive use of linked data, such as linked primary and secondary care data, would allow fuller ascertainment of malformations.
- The positive predictive value for congenital malformations overall, major malformations and heart defects is high (80%–100%) in primary care data.
- Further validation studies are needed of other congenital malformation subgroups in primary care data and in secondary care data.
- There is a need for fuller reporting of the methods and code-lists used to identify congenital malformations in studies using electronic health records.

2 | METHODS

We carried out a systematic review of studies that involved identifying CMs in UK EHR. The systematic review was registered with PROSPERO (registration number: CRD42017037168).

2.1 | Eligibility criteria

Studies were included if they used diagnostic codes to identify CMs in UK EHR and were published after 1987 (the year CPRD was established).[3] As our main objective was to review the methods used to identify CMs from large electronic datasets, case series and case reports were excluded. We further excluded: studies of populations sourced solely from tertiary care settings or specialist registries as we considered the differences in casemix and coding practices in these settings would limit the generalisability of results; studies restricted to CMs that are frequently excluded from safety studies (e.g., chromosomal abnormalities); conference abstracts due to insufficient methodological detail to meet the objectives of this review.

Validations of diagnostic codes for CMs, reported within the EHR studies identified, were included if they used either of the following methods: (1) comparison of individual diagnoses to a reference standard such as an external database, anonymised free-text (entered by the GP alongside electronic coding), additional information from GPs via questionnaires, or provision of anonymised paper records such as hospital letters, or (2) comparison of prevalence rates with rates derived from external population-based data. Validation studies of individual diagnoses had to report ≥1 validity measurement (sensitivity, specificity, positive predictive value (PPV) or negative predictive value (NPV)), or provide data that allowed their calculation.

2.2 | Search strategy and study selection

Medline and Embase were searched up to 20 September 2019 for English-language publications using a wide-ranging search strategy that included keywords and subject headings for 'congenital malformations' and 'electronic health records' (Supporting Information Methods). This was supplemented with a manual search of the bibliographies of three of the main UK primary care EHR databases (CPRD, The Health Improvement Network [THIN] and QResearch) and the Boston Collaborative Drug Surveillance Programme.[10-13] Reference lists of relevant reviews and eligible studies were also searched.

The titles and abstracts of studies identified in the search were screened by MP to determine eligibility for full-text review. A sample of abstracts were also screened by ST and PM to establish consistency, with any differences resolved through discussion.

2.3 | Data extraction

Using a standardised form, details of the methods used to identify CMs were extracted from each study, including the specific CMs identified with any exclusions or subsequent classification into subgroups, any externally developed guidelines used to inform case definitions, and details of the identification process (any code-lists or methods used to develop them, and whether a computerised search using code-lists was conducted versus manual review of records to identify possible malformations). Code-lists not included in publications but publicly available elsewhere were identified or requested from authors.

When studies examined the diagnostic validity of recorded CMs, we extracted data on: the main conditions validated, the proportion of identified diagnoses that authors chose to validate and the rationale for their selection, the reference standard used, the response rate for any information requests and the reasons for non-responses, and the reported results. When validation comprised prevalence comparisons with external data-sources, we extracted information about the external data-source, the period of comparison and the results.

2.4 | Quality assessment

A modified version of the Quality Assessment of Diagnostic Studies tool (QUADAS-2) was used to assess the quality of validation studies.[14] Six areas from the tool applicable to EHR were adapted for use: enrolment of patients, patient exclusions, blinding to the reference standard results, the validity of the reference standard, consistent use of the reference standard, and inclusion of patients in the analysis (Table S1).

2.5 | Data synthesis and analysis

Studies were stratified by healthcare setting (primary or secondary care) and then by type of CM. The methods used to identify CMs, and the methods and results of validation studies, were summarised for each stratum. When studies presented validation measures without 95% confidence intervals (CIs), these were calculated using the Wilson method. As there were very few validation studies within strata, with diverse methodologies, formal assessment of between-study heterogeneity and summary estimates of validation measures were not attempted.[15] Instead, heterogeneity in validity estimates within strata was investigated using χ^2 tests, using the same approach as that in a recent systematic review of dementia diagnoses in UK EHR.[16]

3 | RESULTS

We retrieved 54 eligible studies that identified CMs from UK EHR (Figure 1). Thirty-six identified CMs from primary care data, with most assessing drug safety during pregnancy. The majority used standalone CPRD ($n = 21$) and examined all major malformations ($n = 19$).[17-52] One study also used linked HES data for validation (discussed further below).[52]

The remaining 18 studies identified CMs from secondary care data, with most assessing health service delivery, surgical outcomes or disease trends.[53-70] Almost all used HES ($n = 16$) and the most commonly identified CMs were heart ($n = 4$) and orofacial defects ($n = 4$). Eight studies also used linked ($n = 6$)[53-57,67] or unlinked ($n = 1$)[58] death records and/or linked birth data ($n = 2$).[67,68]

3.1 | Identifying malformations in primary care data

Published guidelines were often used in the 36 primary care studies to develop case definitions for CMs and any anatomical subgroups (Table 1). Of the 22 studies that identified 'any' or 'major' malformations, 91% ($n = 20$) consulted guidelines, with most using EUROCAT ($n = 16$).[17,19-32,36] Four studies indicated that they also sought clinical input alongside the guidelines (Table 1).[19-21,48]

Once defined, the methods used to identify CMs of interest from EHR were rarely specified. Twenty-three studies were known to have used Read and/or OXMIS code-lists because these were publicly available ($n = 11$)[25-32,36,49,52] or could be obtained from authors ($n = 12$)[23,24,33-35,37,38,40,43,48,50,51] and a further three studies were considered likely to have used code-lists based on their content and communication with authors (Table 2).[39,45,46] Of these, only three specified using computerised searches to apply their code-lists to the data.[33-35] Two studies did not use a code-list but manually reviewed electronic records.[19,22] For the remaining eight studies, the use of a code-list was uncertain.[17,18,20,21,41,42,44,47] Among all 36 studies, three also described examining anonymised free-text in maternal records to identify evidence of CMs among pregnancy losses and/or terminations (Table 1).[18,22,23]

Code-list development was described briefly in 17 studies (Table 1). Most relied on relating relevant ICD codes, often determined using published guidelines such as EUROCAT, to equivalent Read/OXMIS codes ($n = 16$; Table 1).[23-35,37,38,48] However, only three studies described how this was achieved:[23,24,48] by creating text search terms using ICD code descriptions and applying these to Read/OXMIS dictionaries to identify relevant codes[23,24] or by examining all Read/OXMIS codes in patients' records and including those thought to correspond to relevant ICD codes.[48]

We examined the codes used to identify CMs in the 24 studies that definitely or probably used Read code-lists.[23-36,39,40,43,45,46,48-52] For 16 of these, only broad initial code-lists were available, including codes that were not part of the final CM case definition (Table 2).[23-36,40,48] Another three only provided the subset of codes identified in the study population rather than all codes used to search

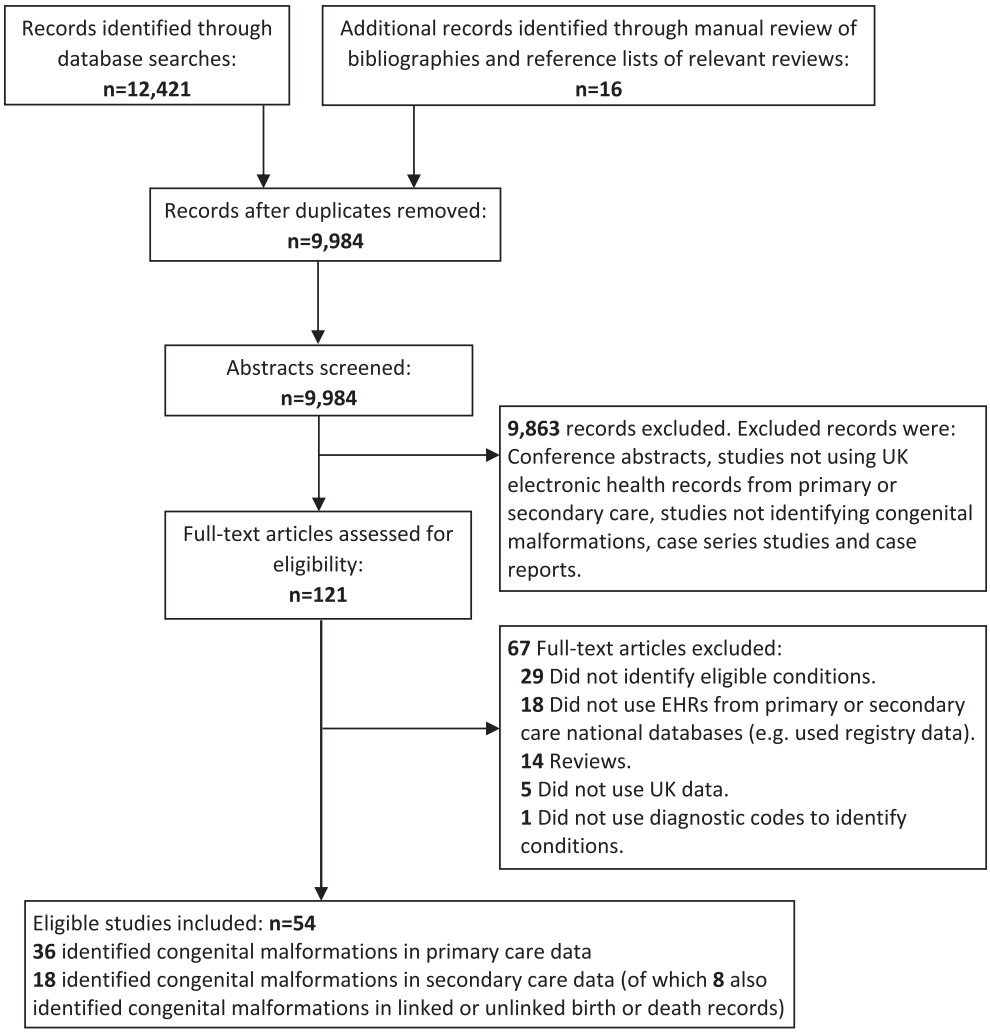

FIGURE 1 Identification of eligible studies

records.[39,45,46] Specific, detailed code-lists were available for just five studies (Table 2).[43,49-52]

All 24 studies considered 'P' chapter Read codes when identifying CMs. Although rarely detailed in the published methods, most also considered to a varying extent codes from other chapters (Table 2). Most frequently considered were codes related to procedures ($n = 20$; e.g., 'repair of cleft lip operations') and diagnoses ($n = 19$; e.g., 'mitral stenosis' from the 'G' chapter on circulatory system diseases; Table 2). Studies also frequently considered codes indicating a personal history of CMs ($n = 14$), and administrative codes for care transfers, monitoring or counselling of individuals with CMs ($n = 12$; e.g., 'transfer of care from paediatric congenital heart services'). Codes relating to testing or screening for CMs (e.g., 'screening for congenital eye anomaly') or conditions identified on examination (e.g., 'observed on examination—pigeon chest') were least frequent ($n = 2$).

3.2 | Identifying malformations in secondary care data

All 18 studies identifying CMs in secondary care and/or birth and death certificate data used publicly available code-lists, although none specified whether computerised or manual searches were carried out (Table 3). All ICD-10 codes were from the 'Q' chapter for CMs. Half of the studies also used OPCS-4 codes, often because procedures were relevant to the study question. The methods used to develop ICD-10 code-lists were publicly available for six studies which examined CMs as a subset of chronic or life-limiting conditions.[53,54,67-70] The methods described were therefore not specific to CMs although two studies did use EUROCAT guidelines to define and identify these.[53,54]

3.3 | Validation of codes used to record malformations in primary care

Eight studies validated CMs identified in primary care data by estimating the PPV of coded diagnoses (Table 4).[17,22,24,38,39,46,49,52] Although all studies examined postnatal diagnoses, only two of these included CMs identified in the antenatal period from maternal records (one of which estimated separate PPVs for postnatal and antenatal diagnoses). One study used linked HES data as the reference standard to validate CMs identified in CPRD, but most relied exclusively on GP questionnaires ($n = 4$) or further primary care data including anonymised free-text and complete paper records ($n = 3$).

TABLE 1 Studies identifying congenital malformations using primary care data

Author	Study aim	Data source	Identification period[a]	Population	Methods used to identify evidence of malformations	Classifications used for malformations and exclusions	Guidelines used to define, classify or exclude malformations	Code-list development described?
Any malformations								
Cea-Soriano[17]	Assess the safety of non-insulin antidiabetic drugs in pregnancy	THIN	1995–2013	Live-born infants	*Read codes* were identified in infant records anytime in the identification period. Maternal records were also searched for codes during pregnancy (methods not described)	*Classified* into subgroups *Excluded* minor and genetic conditions and birth marks in safety assessments	EUROCAT	No
Baril[18]	Assess the safety of the human papillomavirus vaccine in pregnancy	CPRD	2008–2011	Live-born infants Terminations Miscarriages Stillbirths	*Medcodes*[b] were identified in infant clinical or referral files in the first 12 weeks of life and entity types were examined *Free-text* in the maternal record was examined for other pregnancy outcomes	*Classified* as major or minor	Metropolitan Atlanta Congenital Defects Program	No
Ruigomez[19]	Assess the safety of antacids in pregnancy	CPRD	1991–1997	Live-born infants Terminations Stillbirths	*Manual review* of infant records in the first year of life by two physicians. Maternal records were also thought to be examined (methods not described)	*Classified* as major or minor and into subgroups *Excluded* genetic conditions	EUROCAT, with physician input[c]	N/A (Authors communicated that no code-list was developed)
Major malformations								
Petersen[20]	Assess anticonvulsant safety in pregnancy	THIN	1995–2014	Live-born infants	*Read codes from the 'P Chapter'* were identified in infant records	*Excluded* minor conditions	EUROCAT, with GP input	Only identified 'P' codes, unclear if code-list used
Petersen[21]	Assess the safety of antipsychotics in pregnancy	CPRD and THIN	1995–2012	Live-born singletons	*Read codes from the 'P Chapter'* were identified in infant records during the first year of life. Codes were also identified from maternal records during pregnancy	*Excluded* minor conditions and Down syndrome	EUROCAT, with GP input[d]	Only identified 'P' codes in the infant records, unclear if code-list used
Charlton[22]	Assess the safety of inhaled corticosteroids in pregnancy	CPRD	2000–2010	Live-born singletons Terminations Stillbirths	*Manual review* of clinical, referral and test files of infants during the identification period to identify Read codes *Free-text* in the maternal record that was associated with other pregnancy outcomes was searched for evidence	*Classified* into subgroups *Excluded* minor conditions and syndrome-related defects in infants with syndromes	EUROCAT	N/A (Authors communicated that no code-list was developed)

(Continues)

TABLE 1 (Continued)

Author	Study aim	Data source	Identification period[a]	Population	Methods used to identify evidence of malformations	Classifications used for malformations and exclusions	Guidelines used to define, classify or exclude malformations	Code-list development described?
Charlton[23]	Assess the safety of anticonvulsants in pregnancy	CPRD	1990–2006	Live-born infants Terminations Stillbirths Neonatal deaths	*Read and OXMIS* codes were identified in clinical, referral and test files of infants during the identification period *Free-text* in the maternal record was searched. It was examined from 2 months before until 4 months after a termination or 6 months after a stillbirth or neonatal death	*Classified* into subgroups *Excluded* minor conditions, genetic conditions and those not plausibly drug-induced	EUROCAT	Created search terms using ICD-9 codes 740-759. Read codes containing the terms were identified
Charlton[24]	Assess the identification of major malformations in CPRD	CPRD	1990–2006	Live-born infants	*Read and OXMIS* codes were identified in clinical, referral and test files of infants during the identification period	*Classified* into subgroups *Excluded* minor conditions	EUROCAT	Created search terms using ICD-9 codes 740-759. Read codes containing the terms were identified
Dhalwani[25]	Assess nicotine replacement therapy in pregnancy	THIN	2001–2012	Live-born infants	*Read codes* were identified in the medical file of infants during the identification period	*Classified* into subgroups *Excluded* minor defects and those due to known teratogens	EUROCAT	Relevant ICD-10 codes were used to identify equivalent Read codes
Ban[26-30]	Assess the risks in pregnancy of: 1. Depression and therapy 2. Inflammatory bowel disease and therapy 3. Anxiolytics/hypnotics 4. Anti-epileptics 5. Coeliac disease	THIN	1990–2009 1990–2010 1990–2010 1990–2013 1990–2013	Live-born singletons	*Read codes* were identified from infant records during the identification period	*Classified* into subgroups *Excluded* minor and genetic conditions and those due to known teratogens	EUROCAT	Relevant ICD-10 codes were used to identify equivalent Read codes
Sokal[31,32]	Compare prevalence of MCMs in THIN with: 1. EUROCAT 2. Other population-based data	THIN	1990–2010	Live-born singletons	*Read codes* were identified in infant records during the identification period	*Classified* into subgroups *Excluded* minor conditions	EUROCAT	Relevant ICD-10 codes were used to identify equivalent Read codes
Vasilakis-Scaramozza[33-35]	Assess the risk of: 1. Asthma treatment 2. Depression 3. Hypertension in pregnancy	CPRD	1991–2002	Live-born singletons Terminations Stillbirths	*Read codes* were identified in infant records during the identification period. *Cause of death* for stillbirths and terminations was checked for evidence[e]	*Classified* into subgroups *Excluded* minor and genetic conditions and those associated with prematurity among preterm births	Centers for Disease Control and Prevention guidelines	ICD-9 codes 740-7599 were used to identify equivalent Read codes
Tata[36]	Assess asthma and treatment in pregnancy	THIN	1988–2004	Live-born infants	*Read codes* were identified in infant records during the identification period.	*Classified* into subgroups *Excluded* minor conditions	EUROCAT	No

TABLE 1 (Continued)

Author	Study aim	Data source	Identification period[a]	Population	Methods used to identify evidence of malformations	Classifications used for malformations and exclusions	Guidelines used to define, classify or exclude malformations	Code-list development described?
Jick[37]	Assess the safety of antifungals in pregnancy	CPRD	Not specified	Live-born infants	*Manual review* of infant records for evidence at birth. If a malformation was suspected, relevant paper records were requested from the GP for further information	*Classified* into subgroups	Defined as those needing surgery or treatment	ICD-8 codes 7400-7590 were used to identify equivalent OXMIS codes[f]
Jick[38]	Assess the safety of anticonvulsants in pregnancy	CPRD	1988–1993	Live-born infants	*OXMIS codes* were identified in infant records around the time of birth	*Excluded* minor conditions, hypospadias, hernias, and those that could not be drug-induced	N/S	ICD-8 codes 7400-7590 were used to identify equivalent OXMIS codes
Neural tube defects								
Devine[39]	Examine the validity of neural tube defects recorded in CPRD	CPRD	1987–2004	Live-born infants Terminations Stillbirths Miscarriages	*Read and OXMIS codes* for anencephaly, encephalocele, spina bifida and meningocele were identified in infant records in the first year of life. The first code was included. Codes on January 1st were excluded unless within 30 days of birth *Read and OXMIS codes* within 210 days of a pregnancy record were identified in maternal data. The first code was included. Codes were excluded if: within 60 days of the first code, if on January 1st and not within 30 days of a pregnancy record, or, if within 180 days of a code in the infant	*Classified* anencephaly, cephaloceles, meningoceles and spina bifida separately for some analyses	N/S	No
Tata[40]	Assess the risks of celiac disease in pregnancy	CPRD	1987–2002	Live-born infants	*Read codes* for meningocele, meningomyelocele, spina bifida and hydrocephalus were identified in infant records during the identification period	None described	N/S	No
Lawrenson[41]	Estimate prevalence and incidence of renal failure and replacement therapy in those with neural tube defects	CPRD	Prior to 1997	Patients aged 10–69	*Diagnostic codes* for neural tube defects were identified in the identification period (methods were not described further)	None described	N/S	No

(Continues)

TABLE 1 (Continued)

Author	Study aim	Data source	Identification period[a]	Population	Methods used to identify evidence of malformations	Classifications used for malformations and exclusions	Guidelines used to define, classify or exclude malformations	Code-list development described?
Lawrenson[42]	Estimate mortality and prevalence rates of neural tube defects	CPRD	Prior to 1997	Patients aged 10–69	*Diagnostic codes* for meningocele, meningomyelocele, spina bifida and hydrocephalus were identified in the identification period (methods were not described further)	None described	N/S	No
Orofacial defects								
Chi[43]	Assess the safety of topical corticosteroids in pregnancy	CPRD	2000–2006	Live-born singletons	*Read and OXMIS codes* were identified in clinical files of infants	*Classified* by cleft type *Excluded* syndromic cleft	N/S	No
Heart defects								
Petersen[44]	Assess antidepressant safety in pregnancy	THIN	1990–2011	Live-born singletons	*Read codes* were identified from infant records during the first 5 years of life	*Excluded* Down syndrome	N/S	No
Margulis[45] Hammad[46]	Assess the safety of anti-depressants in pregnancy; Validate specific heart defects	CPRD	1996–2010	Live-born singletons	*Read codes* indicating a heart defect or related procedure were identified from infant records in the first year and first 6 years of life	*Classified* by heart defect *Excluded* genetic conditions and sequences[8]	Published development-based classification system[71]	No
Billett[47]	Estimate the prevalence of comorbidities, health service use and recording of clinical indicators in those with heart defects	Q Research	N/S-2005	All patients	*Read codes* indicating a heart defect or related procedure were identified during the identification period	*Classified* by complexity *Excluded* cardiomyopathies, isolated arrhythmias, isolated dextrocardia, bicuspid aortic valve, mitral valve prolapse, cardiac tumours, Marfan's syndrome	Modified version of a published classification system based on anatomical hierarchy[72]	No
Wurst[48]	Compare the prevalence of heart defects between CPRD, NCAS and EUROCAT	CPRD	2001–2003	Live-born infants	*Read and OXMIS* codes were identified in the infant's first year and first 6 years of life	*Classified* by heart defect *Excluded* minor conditions and vascular defects	EUROCAT and National Congenital Anomaly System guidelines, with input from a paediatric cardiologist	Identified potential Read/OXMIS codes and then selected those equivalent to ICD-9 codes 7450-59, 7460-69, 7470-74
Wurst[49]	Examine the validity of specific heart defects recorded in CPRD	CPRD	1992–2005	Live-born infants	*Read and OXMIS codes* for diagnoses and procedures were identified from infant records during the identification period	*Classified* by heart defect *Excluded* codes synonymous to the conditions of interest but which did not	N/S	Identified only those codes that included the terms: 'ventricular septal defect', 'tetralogy of

TABLE 1 (Continued)

Author	Study aim	Data source	Identification period[a]	Population	Methods used to identify evidence of malformations	Classifications used for malformations and exclusions	Guidelines used to define, classify or exclude malformations	Code-list development described?
						contain the specified terms		Fallot', 'coarctation of the aorta', 'COA', 'VSD', or 'TOF'
Gastroschisis								
Bannister[50]	Assess the incidence of infections in children with gastroschisis	THIN	1990–2013	Live-born infants	*Read codes* were identified from infant records during the first 5 years of life	*Excluded* other major malformations	EUROCAT	No
Genitourinary tract or inguinal region malformations and developmental hip dysplasia								
Perry[51]	Assess the presence of malformations in those with Perthes disease	CPRD	1990–2008	All patients in the study	*Read codes* were identified from records anytime in the identification period	*Classified* into subgroups for analyses	N/S	No
Developmental hip dysplasia								
Broadhurst[52]	Estimate incidence of developmental hip dysplasia	Linked CPRD and HES	1990–2016	All patients aged 1–8 years	*Read codes* were identified from records anytime in the identification period and the first code was considered the first diagnostic evidence	*Excluded* those with neuromuscular conditions, syndromes or traumatic hip dislocation that could result in developmental hip dysplasia	N/S	No

Abbreviations: CPRD, Clinical Practice Research Datalink (formerly the General Practice Research Database); GP, General Practitioner; ICD, International Classification of Diseases; MCM, major congenital malformation; N/A, not applicable; N/S, not specified; OXMIS, Oxford Medical Information System; THIN, The Health Improvement Network.

[a]The period in which congenital malformations were identified was not always explicitly defined by authors and did not always correspond to the study period. When possible, it was estimated from available information.

[b]CPRD medcodes, corresponding to Read codes.

[c]If the classification of identified conditions was uncertain, complete paper records were reviewed for further information.

[d]If the classification of identified conditions was uncertain, free-text from THIN or complete paper records from CPRD were reviewed for further information.

[e]This appeared to involve examination of maternal records but methods were not described further.

[f]The authors communicated that they conducted a manual review but also provided information for a code-list.

[g]These are groups of related malformations typically occurring as a result of a MCM disrupting the development of surrounding tissues.

TABLE 2 Read and Oxford Medical Information System (OXMIS) codes used to identify congenital malformations in primary care data

Study	Used a code-list?	Level of detail in code-list published or provided upon request	Source of code-list *or* reason list not available	Coding system	Types of Read and/or OXMIS codes included in code-lists to identify malformations							
					Diagnosis	Diagnosis from 'P' chapter in Read	Diagnosis from other Read chapters	Procedure	Testing/ screening	History[a]	Observation[b]	Administrative[c]
Any malformations												
Cea-Soriano[17]	N/S	None	No response to request for further information	Read[d]	✓	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Baril[18]	N/S	None	No response to request for further information	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Ruigomez[19]	No	None	No code-list used	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Major malformations												
Petersen[20,21]	N/S	Subset (most common codes identified in the study population)	No response to request for further information	Read	✓	✓	No	No	No	No	No	No
Charlton[22]	No	None	No code-list used	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Charlton[23,24]	Yes	Broad	Author	Read and OXMIS	✓	✓	✓	✓	✓	✓	✓	✓
Tata[36] Sokal[31,32] Ban[26-30] Dhalwani[25]	Yes	Broad	Published[e]	Read	✓	✓	✓	✓	No	✓	No	✓
Vasilakis-Scaramozza[33-35]	Yes	Broad	Author	Read	✓	✓	✓	No	No	No	No	No
Jick[37,38]	Yes	Broad[f]	Author	OXMIS	✓	No	No	N/S	N/S	N/S	N/S	N/S
Neural tube defects												
Devine[39]	Likely	Subset (only codes in the study population that were validated)	Published	Read and OXMIS	✓	✓	✓[g]	✓	N/S	N/S	N/S	N/S
Tata[40]	Yes	Broad	Author	Read and OXMIS	✓	✓	✓	✓	N/S	N/S	No	✓
Lawrenson[41,42]	N/S	None	No access to study material	N/S	✓	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Orofacial defects												
Chi[43]	Yes	Specific	Author	Read and OXMIS	✓	✓	No	✓	No	✓	No	No

TABLE 2 (Continued)

Study	Used a code-list?	Level of detail in code-list published or provided upon request	Source of code-list *or* reason list not available	Coding system	Types of Read and/or OXMIS codes included in code-lists to identify malformations							
					Diagnosis	Diagnosis from 'P' chapter in Read	Diagnosis from other Read chapters	Procedure	Testing/ screening	History[a]	Observation[b]	Administrative[c]
Heart defects												
Petersen[44]	N/S	Subset (most common codes identified in the study population)[h]	No response to request for further information	Read	✓	✓	N/S	N/S	N/S	N/S	N/S	N/S
Margulis[45] Hammad[46]	Likely	Subset (only codes in the study population that were validated)	Published	Read	✓	✓	N/S	✓[d]	N/S	N/S	N/S	N/S
Billet[47]	N/S	None	No access to study material	Read	✓[d]	N/S	N/S	✓[d]	N/S	N/S	N/S	N/S
Wurst[48]	Yes	Broad	Author	Read and OXMIS	✓	✓	✓	No	No	No	No	No
Wurst[49]	Yes	Specific	Published	Read and OXMIS	✓	✓	No	✓	No	No	No	No
Gastroschisis												
Bannister[50]	Yes	Specific	Author	Read	✓	✓	No	✓	No	No	No	No
Genitourinary tract or inguinal region malformations and developmental hip dysplasia												
Perry[51]	Yes	Specific	Author	Read	✓	✓	✓	✓	No	✓	No	No
Developmental hip dysplasia												
Broadhurst[52]	Yes	Specific	Published	Read	✓	✓	✓	✓	No	✓	✓[i]	✓[i]

Note: Studies using the same code-list were grouped.

Abbreviations: N/A, not applicable; N/S, not specified.

[a]Codes for history of congenital malformations (e.g. '*history of cleft palate*').

[b]Codes for congenital malformations observed during examination (e.g. '*observed on examination—pigeon chest*').

[c]Codes for monitoring, counselling or transfers of care related to congenital malformations (e.g. '*transfer of care from paediatric congenital heart service*').

[d]Described in methods.

[e]Published as part of a thesis.

[f]Author noted a manual review was likely used in one study.

[g]Codes for suspected foetal malformations.

[h]Published Read terms only.

[i]Codes used as supportive evidence of condition.

TABLE 3 Studies identifying congenital malformations using secondary care data

Author	Congenital malformations identified	Study Aim	Data-source	Identification Period	Study population	Algorithm to identify congenital malformations of interest	Algorithm to identify subgroups, excluded conditions or those used in sensitivity analyses
Zylbersztejn, 2019[53]	Congenital malformations as a subset of chronic conditions in children	1. Compare preventable deaths in Sweden and England and assess the contribution of risk factors.	HES admissions linked to death certificates	2003-2013	Singleton live-births delivered between 2003-2012	1. Presence of one of the codes below in hospital admission records from the 31st day of life until the second birthday or on the death certificate as any cause of death until the fifth birthday.	1. Malformation subgroups not explored.
Zylbersztejn, 2018[54]		2. Assess the reasons for higher infant mortality in England compared to Sweden				2. Presence of one of the codes below in the birth admission record, hospital admission records until the second birthday or on the death certificate as any cause of death up until the fifth birthday. Codes used in both studies[a]: Q00-07, Q104, Q107, Q11-12, Q130-134, Q138-139, Q14-16, Q188; Q20-26; Q30-34; Q35-37; Q380, Q383-384, Q386-388, Q39, Q402-403, Q408-409, Q41-42, Q431, Q433-437, Q439, Q44-45; Q500, Q51, Q520-522, Q524, Q540-543, Q548-550, Q555, Q56; Q601-602, Q604-606, Q61, Q620-626, Q628, Q630-632, Q638-639, Q64; Q650-652, Q658-659, Q675, Q682, Q71-74, Q750-751, Q753-759, Q761-764, Q77-78, Q790, Q792-796, Q798; Q820-824, Q829, Q85, Q860-862, Q868, Q878, Q891-893, Q897-899; Q90-93, Q952-953, Q97, Q99	2. Sensitivity analysis of severe malformations as defined by the following codes[b]: Q00-07; Q20, Q212-214, Q218-219, Q22-24, Q251-259, Q26, Q282-283, Q289; Q30-34; Q390-394, Q41-45; Q60-64; Q722, Q750, Q752, Q759-762, Q764-767, Q77, Q780-784, Q788-795, Q799; Q81, Q871, Q873-874, Q877-879; Q909, Q913-914, Q917, Q928, Q93, Q950, Q969, Q97-98, Q992, Q998-999
Dimopoulos, 2019[55]	Heart defects	Assess transplant survival rates and the capacity for such procedures.	HES admissions linked to death certificates	1997-2015	All	Presence of any 'Q2xx.x' code with a further procedure code indicating a heart or heart-lung transplant (OPCS-4 codes K01-02).	Stratified heart defects by complexity, using guidelines from 32nd Bethesda conference
Kempny, 2017[56]	Heart defects	Examine surgical volume and mortality in patients.	HES admissions linked to death certificates	1997-2015	All	Presence of any 'Q2x.x' code and evidence of a cardiac surgery (except heart or heart-lung transplants).	Excluded CABG and mitral valve surgery in sensitivity analyses
Singhal, 2014[57]	Heart defects; Spina Bifida	Assess the relationship between self-harm/suicide and chronic illnesses.	HES admissions linked to death certificates	1999-2011	All	Presence of any Q20-24 code (heart defects) or a Q05 code (spina bifida).	-

TABLE 3 (Continued)

Author	Congenital malformations identified	Study Aim	Data-source	Identification Period	Study population	Algorithm to identify congenital malformations of interest	Algorithm to identify subgroups, excluded conditions or those used in sensitivity analyses
Billett, 2007[58]	Heart defects	Explore trends in admissions, procedures and patient mortality.	HES admissions, Death certificates (unlinked) from England and Wales	1995-2004 1994-2003	All	Any Q20-28 code in the primary diagnosis field. Any ICD-9 code for 745-747 for underlying cause of death until 2001 and any ICD-10 Q20-28 code thereafter.	-
Fitzsimons, 2017[59]	Orofacial clefts	1. Examine grommet insertion practices in cleft patients.	HES admissions	1. 1997-2011	1. Live-births from 1997 to 2005	Codes used in all studies: Presence of any diagnostic Q35, Q36, or Q37 code as well as a procedure code for primary cleft repair (OPCS-4 codes F031 or F291).	1. Clefts were grouped by type using repair codes (F03, F29, F30, F32) and the available diagnosis code.
Fitzsimons, 2014[60]		2. Examine hospital admissions for dental treatment in cleft.		2. 1997-2011	2. Live-births from 1997 to 2003		2. Clefts were grouped by type using repair codes (F03, F29, F30, F32) and the available diagnosis code.
Fitzsimons, 2013[61]		3. Explore hospital admissions and length of stay for cleft patients.		3. 1997-2011	3. Live-births from 1997 to 2008		3. Clefts were grouped by type using repair codes (F03, F29, F32) and the available diagnosis code.
Fitzsimons, 2012[62]		4. Assess changes in cleft patient care following service changes.		4. 1997-2009	4. Live-births from 1997 to 2008		4. Clefts were not grouped by type. All 4 studies identified those with additional malformations for separate analyses or exclusion using any of the following codes in any field: D821; Q00-07; Q16; Q18; Q20-28; Q380; Q75; Q86-87; Q90-93; Q95-99
Broadhurst, 2019[52]	Developmental hip dysplasia	Estimate incidence of developmental hip dysplasia	Identified cases in CPRD and then used linked HES admissions data for validation	1990-2016	All patients aged 1-8 years	Linked HES admissions data in the 2 years either side of the initial diagnosis in CPRD were searched for supportive evidence in the following order (strongest to weakest evidence): A) A specific diagnosis based on the presence of a Q650-656 code or evidence of a specific procedure (OPCS-4 codes X221-225, X228-229). B) The presence of a related diagnostic code: M244 or R294.	Diagnoses excluded: G80, G800-804, G808-809, Q743, Q796, Q90, Q900-902, Q909, Q980-985, Q824, Q718, Q728, Q999, Q916, G711, G819, Q929, Q824, Q931, Q773 Procedures excluded: W651-655, W658-659, W661-664, W668-669, W671-679

(Continues)

TABLE 3 (Continued)

Author	Congenital malformations identified	Study Aim	Data-source	Identification Period	Study population	Algorithm to identify congenital malformations of interest	Algorithm to identify subgroups, excluded conditions or those used in sensitivity analyses
						C) Evidence of a hospital admission within 6 months. If none of the above were found, CPRD was searched for supportive evidence (≥3 orthopaedic hospital attendances, ≥2 diagnostic codes or related codes such as those for a clicking hip) in the 2 years after the initial diagnosis.	
McAllister, 2018[63]	Developmental hip dysplasia	Assess the risk of surgery for this condition following an intervention.	Scottish Morbidity Record	1997-2014	Live-births from 1997-2013	Presence of any Q650-659 code with a further procedure code (OPCS-4 codes T202, T205, W134, W144, W164, W169, W281, X221–229, W65-66).	-
Dharmasena, 2017[64]	Anophthalmia; Microphthalmia; Malformations of orbit; Agenesis of lacrimal apparatus	Estimate hospital admission trends and incidence of eye malformations.	HES admissions	1999-2011 1990-1998	All <1 years old	Presence of any of the following codes in the birth record or subsequent hospital admission record: Q110-Q111 (anophthalmia), Q112 (microphthalmia), Q107 (congenital malformations of orbit), Q104 (agenesis of lacrimal apparatus). The ICD-9 code 7431 for microphthalmia was used in further analyses that examined years prior to 1995.	
-							
Lansdale, 2017[65]	Pyloric stenosis	Examine surgical outcomes for infantile hypertrophic pyloric stenosis.	HES admissions & Patient Episode Data for Wales	2002-2011	All	Presence of a procedure code for pyloric stenosis (OPCS-4 code G401) that occurred between the 1st day of life and the 1st birthday, as well as a Q400 diagnostic code.	-
Wilkinson, 2017[66]	Hypospadias	Estimate the frequency of re-operations and complications following repair of hypospadias.	HES admissions	1999-2009	Boys <16 years old	Presence of a procedure code for primary repair of hypospadias (OPCS-4 code M731), with or without a diagnostic Q54 code. Post-surgery admissions for were identified from Q540-543 or Q548-549 codes or codes relating to surgical complications or revisions.	Excluded those with disorders of sexual differentiation recorded with the following codes: Q560-564; Q640-641; E250, E258-259, E345

TABLE 3 (Continued)

Author	Congenital malformations identified	Study Aim	Data-source	Identification Period	Study population	Algorithm to identify congenital malformations of interest	Algorithm to identify subgroups, excluded conditions or those used in sensitivity analyses
Jarvis and Fraser, 2018[67]	Life-limiting or life-threatening malformations[c]	1. Compare the identification of these conditions in inpatient data and death records.	1. HES admissions linked to death certificates & Scottish Birth, Morbidity and Death Records.	1. 2001-2015 2003-2014	1. Patients aged 0-25 years with a death record	1. Searched English and Scottish hospital admission data and Scottish birth records for the codes below, except for Q445 and Q748. Death records were searched for codes among the underlying causes of death. If this was not related to a life-limiting condition, then contributing causes of death were checked.	-
Jarvis, 2017[68]		2. Assess clinical stability in those with life-limiting conditions.	2. Scottish Birth & Morbidity Records	2. 2003-2014	2. Patients aged 0-25 years	2. Searched Scottish birth records and hospital admissions data for the codes below except for Q445 and Q748.	
Fraser, 2014[69]		3. Estimate the prevalence of life-limiting conditions.	3. HES admissions	3. 2009-2010	3. Patients aged 0-40 years	3. Searched hospital admissions data for the codes below.	
Fraser, 2012[70]		4. Estimate the prevalence of life-limiting conditions.	4. HES admissions	4. 2000-2010	4. Patients aged 0-19 years	4. Searched hospital admissions data for the codes below. Codes used in all studies: Q000, Q01, Q031, Q039-040, Q042-044, Q046, Q049, Q070, Q200, Q203-204, Q206, Q208, Q213, Q218, Q220-221, Q224-226, Q230, Q232, Q234, Q239, Q254, Q256, Q262, Q264, Q268, Q282, Q321, Q336, Q396, Q410, Q419, Q437, Q442, Q445, Q447, Q601, Q606, Q614, Q619, Q642, Q743, Q748, Q750, Q772-774, Q780, Q785, Q792-793, Q804, Q81, Q821, Q824, Q858, Q860, Q870-872, Q878, Q91, Q920-921, Q924, Q927-928, Q932-935, Q938, Q952.	

Abbreviations: HES, Hospital Episode Statistics; ICD, International Classification of Diseases; OPCS-4, 4th Revision of the Classification of Interventions and Procedures.
[a]Codes were a subset of those developed by Hardelid et al. to identify chronic conditions requiring medical follow-up for more than a year in half or more of cases.
[b]Codes were a subset of those developed by Feudtner et al. to identify conditions likely to last at least a year and involve multiple organ systems or require tertiary care.
[c]Codes for malformations were a subset of those used to identify all life-limiting or threatening conditions.

TABLE 4 Summary of studies performing validations of congenital malformation diagnoses identified in primary care data

Author	Malformations validated	Study population	Validation method	Method summary	No. cases identified for validation	Response rate (received/requested)	Reasons for non-receipt of requested information from GPs	Positive Predictive Value (95% CI)
Cea-Soriano[17]	Any	Live-births	Manual review of free-text and electronic records	Authors: Identified supporting evidence from free-text, comments associated with specialist referrals in the year before or after the first diagnosis, tests or procedures, and repeated records of malformations or symptoms	788[a]	N/A	N/A	81% (78–84)
Charlton[22]	Major	Live-births Stillbirths Terminations	GP questionnaire Manual review of free-text and electronic records	GP confirmed: Diagnosis Authors: Identified supporting evidence (e.g., surgery codes)	622[b]	88% (127/145)[c]	N/S	86% (83–89)
Charlton[24]	Major or could potentially be classed as major under certain criteria	Live-births	Request for and review of complete paper records and free-text	Authors: Requested complete paper records for those registered with a practice and free-text if they could not be provided. Diagnoses were confirmed and those with sufficient information to be classified as major or minor were identified	188[b]	78% (96/123)[d]	15 No response 4 Refused participation 1 No records available 2 Practice left CPRD 1 Transferred out 2 No parental permission 2 No malformation[e]	Combined: 85% (79–90) Records only: 92% (84–96) Free-text only: 78% (67–86)
Jick[38]	Major	Live-births	GP questionnaire	GP confirmed: Diagnosis and diagnosis date	16[b]	N/S	N/S	100% (76–100)[f]
Hammad[46]	Heart	Live-births	GP questionnaire	GP confirmed: Diagnosis and diagnosis date, type of exam used to determine diagnosis, information used to confirm the diagnosis	888[a]	81% (719/888)	N/S	1996–2010:[g] 93% (91–95) 2006–2010: 94% (91–97)

TABLE 4 (Continued)

Author	Malformations validated	Study population	Validation method	Method summary	No. cases identified for validation	Response rate (received/requested)	Reasons for non-receipt of requested information from GPs	Positive Predictive Value (95% CI)
Wurst[49]	Heart VSD TOF COA	Live-births	GP questionnaire	GP confirmed: Diagnosis and diagnosis date, age at diagnosis, reason diagnosis suspected, type of doctor that made diagnosis, diagnostic tests and results, referrals to cardiology, surgery, other heart defects and VSD type	200[b] 104 VSD 72 TOF 24 COA	94% (187/200)	N/S	94% (89–96) 95% (88–98) 90% (80–96) 100% (81–100)
Devine[39]	Neural Tube Anencephaly Encephalocele Meningocele Spina bifida	Live-births Stillbirths Terminations Miscarriages	GP questionnaire	GP confirmed: Diagnosis and diagnosis date, the type of exam used to determine the diagnosis and the information used to confirm it. As this study also identified neural tube defects encoded in the maternal record around the time of pregnancy, GPs were asked to confirm if the diagnosis related to a condition in the mother or her offspring for such cases	217[b]	76% (165/217)	31 No response 18 Transferred out 3 Data entry errors[h]	71% (63–78)[g] 81% (68–89) 83% (36–99) 64% (36–86) 47% (36–58)
Broadhurst[52]	Developmental hip dysplasia	Children with a first diagnosis between 1 and 8 years	Review of linked HES records and CPRD records	Authors: Identified the initial diagnosis in CPRD and searched for supportive evidence in linked HES data in the 2 years either side of this or in CPRD in the 2 years after. Results were stratified by the strength of the supportive evidence	754[a]	N/A	N/A	Using most specific supportive evidence 34% (30–37) Using any available supportive evidence 56% (53–60)

(Continues)

TABLE 4 (Continued)

Author	Malformations validated	Study population	Validation method	Method summary	No. cases identified for validation	Response rate (received/requested)	Reasons for non-receipt of requested information from GPs	Positive Predictive Value (95% CI)
				(from strongest to weakest): A. Specific diagnostic or procedural code in HES B. Non-specific code in HES (e.g., 'clicking hip') C. Hospital admission within 6 months D. No related codes in HES but other coded supportive evidence available in CPRD (≥2 diagnostic codes, ≥3 orthopaedic follow-up visits, other supportive evidence such as Read codes for a clicking hip)				

Abbreviations: CI, confidence interval; COA, coarctation of aorta; GP, general practitioner; N/A, not applicable; N/S, not specified; PPV, positive predictive value; TOF, tetralogy of fallot; VSD, ventricular septal defect.

[a]Number of individuals with malformation(s) identified for validation.

[b]Number of malformations identified for validation.

[c]Did not specify if questionnaires were requested by individual or by malformation.

[d]Received records for 96 patients corresponding to 109 malformations.

[e]These were included in the PPV calculation as individuals whose record of a major malformation diagnosis was not validated.

[f]95% CI calculated on the assumption that 16 questionnaires were sent to GPs and responses received.

[g]Calculated additional PPVs, including PPVs for each code (not shown).

[h]Did not include these when calculating PPV.

TABLE 5 Summary of quality assessment of validation studies

Criterion	Risk of bias, by study								
	Cea-Soriano[17] (Any malformations)	Charlton[24] (Major malformations)	Charlton[22] (Major malformations)	Jick[38] (Major malformations)	Devine[39] (Neural Tube Defects)	Wurst[49] (Heart defects)	Hammad[46] (Heart defects)	Broadhurst[52] (Developmental Hip Dysplasia)
Was a consecutive or random sample of patients enrolled?	Low	Low	Low	Low	Low	Low	Low	Low
Did the study avoid inappropriate exclusions?	Low	Low	Low	Low	Low	Low	Low	Low
Were the index test results interpreted without knowledge of the results of the reference standard?	Low	Low	Low	Low	Low	Low	Low	Low
Is the reference standard likely to correctly classify the target condition?	Uncertain	Uncertain	Uncertain	Uncertain	Uncertain	Uncertain	Low	High
Did all patients receive the same reference standard?	Uncertain	Low	High	Low	Low	Low	Low	Low
Were all patients included in the analysis?	Low	Low	Low	Low	Low	Low	Low	Low

Although the quality of most studies was assessed as good, the validity of the reference standards was often uncertain as details of the type of information used by GPs to confirm diagnoses (e.g., hospital letters) were seldom described (Table 5; Table S2). The length of time either side of the recorded diagnosis date during which supportive evidence was sought was also seldom reported; just one study described searching for all types of supportive evidence within a defined time window for all potential cases (Table 4).[52] Studies also differed in the methods used to calculate PPV; some calculated a conservative PPV by including in the denominator the total number of individuals for whom information was requested whilst others included only those for whom requested information was returned.

Estimates of PPV were generally high (Figure 2). The one study of any CM (major or minor) reported a PPV of 81%[17] and three studies examining major malformations reported estimates from 85% to 100% with no evidence of between-study heterogeneity ($n = 3$; χ^2 test $= 2.7$; $p = 0.3$).[22,24,38] However, one study demonstrated that the PPV varied according to the reference standard used, being lower when validated against free-text review compared to complete paper record review (78% vs. 92%, χ^2 test $= 7.2$; $p = 0.007$).[24] Similarly, both studies of heart defects reported PPVs ≥90% (with the lowest lower confidence limit being 81%).[46,49]

Estimates of PPV were lower for neural tube defects, the majority of which were identified antenatally from maternal records; overall PPV was 71% and was similar for antenatal and postnatal diagnoses (69% vs. 75%). The PPV varied from 47% to 83% for neural tube defect subtypes.[39] Only 34% of patients with a developmental hip dysplasia diagnosis in CPRD also had specific evidence of a diagnosis or procedure in HES in-patient data within 2 years, increasing to 56% when less specific supportive evidence was considered.[52] The reference standard in this study was, however, considered suboptimal (Table S2).

Validation based on comparing prevalence in primary care data with external data-sources suggested prevalence was comparable to or higher than in national or regional population-based registries (-Table S3). Between 1990 and 2009, the prevalence ratio between THIN and UK EUROCAT data for major CMs diagnosed in the first year of life and defined according to EUROCAT guidelines was 1.18 (95%CI, 1.16–1.20)[31] and for heart defects was 1.31 (95%CI, 1.26–1.35). Prevalence was comparable for nervous system malformations (PR = 1.06; 95%CI, 0.98–1.14). Similar results were seen in studies comparing the prevalence of such conditions in the first year of life in CPRD with population-based registries.[39,48] Compared to EUROCAT, THIN appeared to under-ascertain urinary, orofacial, digestive and abdominal malformations, and rare conditions such as Ebstein's anomaly.[31] However, the prevalence of conditions increased in both THIN and CPRD when follow-up was extended beyond the first birthday.[31,48]

3.4 | Validation of codes used to record malformations in secondary care and other UK EHR

No formal validation studies of CM diagnoses identified in secondary care data were retrieved, although some authors described evidence suggesting good validity for these data. For example, between 2000 and

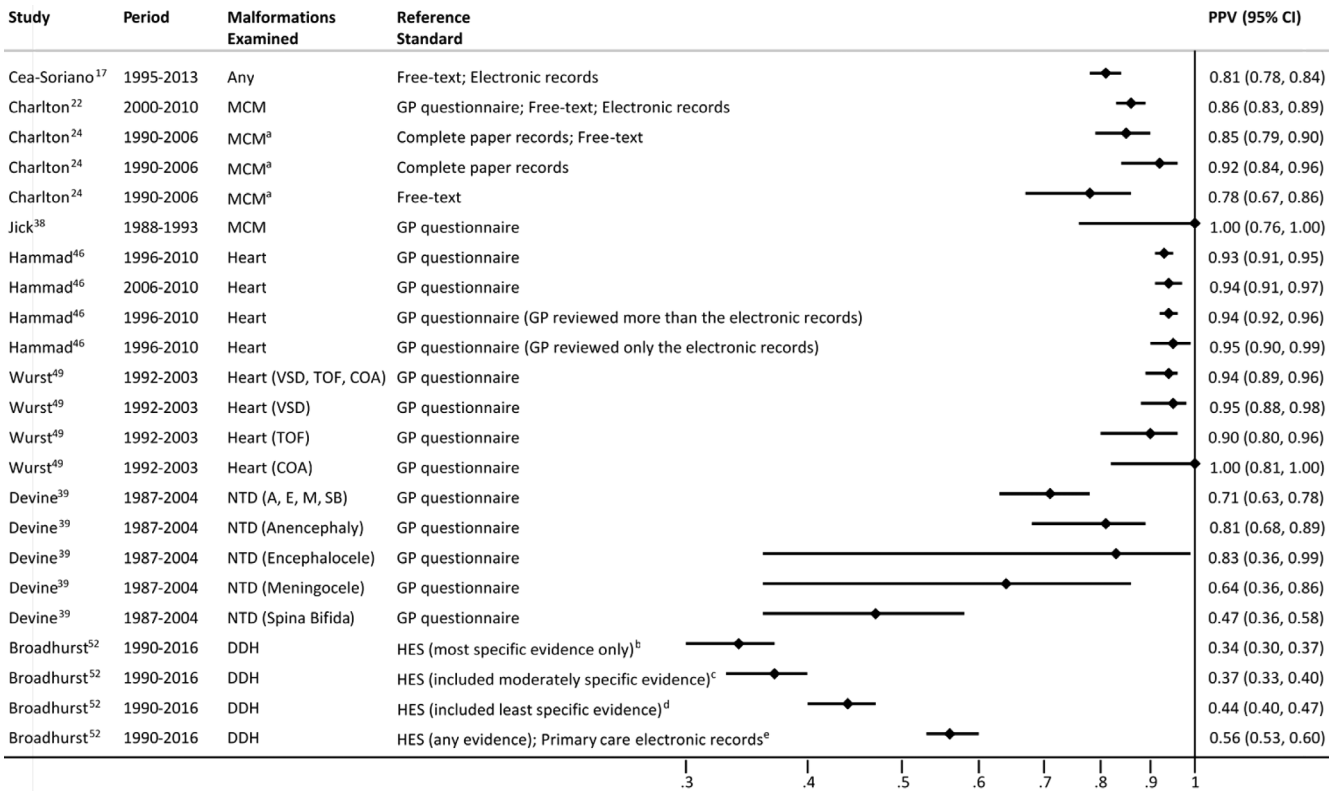

FIGURE 2 Individual-study estimates for PPV of congenital malformations in primary care data. [a]Included conditions that could potentially be classified as major under certain criteria. [b]Specific diagnostic or procedural codes in HES. [c]Searched for specific diagnostic/procedural codes or codes that were likely to relate to developmental hip dysplasia (e.g., 'clicking hip'). [d]Searched for specific or related codes or evidence of a hospital admission within 6 months of the diagnosis in the primary care record. [e]Searched for any evidence in HES (as defined previously) or supportive evidence anywhere in the electronic primary care record such as multiple records of the diagnosis, regular orthopaedic hospital attendances etc. Abbreviations: A, anencephaly; CI, confidence interval; COA, coarctation of aorta; DDH, developmental hip dysplasia; E, encephalocele; GP, general practitioner; HES, hospital episode statistics; M, meningocele; MCM, major congenital malformation; NTD, neural tube defects; PPV, positive predictive value; SB, spina bifida; TOF, tetralogy of fallot; VSD, ventricular septal defect

2009, 85% of almost 9000 individuals identified from a UK orofacial cleft registry could be linked to HES, with >92% concordance of diagnoses.[59-62] Another study noted 98% concordance between a specialist surgical database for developmental hip dysplasia and surgical interventions for this condition in Scottish admissions data.[63]

One study identified deceased infants with a life-limiting CM recorded in in-patient HES data ($n = 6823$) or linked Scottish birth and in-patient data ($n = 555$).[67] Completeness of recording of the CM in death certificates was 80% and 89%, respectively.

4 | DISCUSSION

We retrieved 54 studies, 36 of which used UK primary care data and 18 of which used UK secondary care data (sometimes supplemented with birth or death records) to identify CMs. Completeness of reporting and methodologies differed between these two subgroups. Primary care studies frequently used published guidelines to develop case definitions and Read code-lists for CMs of interest. Code-lists were not always publicly available; those that were demonstrated that although the Read system 'P' chapter is dedicated to CM diagnoses, codes from other diagnostic chapters and codes for procedures, medical histories or administrative tasks were also used to a varying extent, usually without describing the criteria for their inclusion.

In contrast, all secondary care studies made the diagnostic codes used to identify CMs publicly available, but few described using published guidelines to inform case definitions and code-lists. All studies restricted ICD-10 diagnostic codes to the CM 'Q' chapter, with half also using OPCS-4 procedure codes. Clinical coders in hospital settings, unlike primary care, are required to encode definitive diagnoses for reimbursement purposes.[4] This could explain why researchers relied on 'Q' chapter CM codes in these data, but considered Read codes beyond the 'P' chapter in primary care data.

We identified only eight validation studies. All CM diagnoses undergoing validation were identified from primary care data and all studies assessed PPV. No validation studies of CMs identified from secondary care data were retrieved and no study linked to secondary care data for the purpose of assessing completeness of recording of CMs in primary care records. Although Read code-lists for CMs varied across studies, few examined the validity of individual codes and none considered the validity of 'P' chapter versus non-'P' chapter codes. Just three malformation subgroups were validated, with insufficient

studies to obtain robust PPV summary estimates. The PPV was high for CMs overall, for major CMs, and for heart defects, ranging from 80% to 100%.[17,23,24,38,46,49] These estimates are in line with previous findings of a median PPV of 89% for 183 different diagnoses in CPRD.[8] The overall PPV for neural tube defects was slightly lower, and for developmental hip dysplasia was markedly lower, at 56%.[39,52]

These PPV results should be interpreted with caution. Validation studies assume that the reference standard has perfect sensitivity and specificity for the diagnosis of interest. However, the robustness of the reference standard was considered uncertain in six of the eight validation studies.[17,22,24,38,39,49] Studies using GP questionnaires, anonymised free-text or anonymised paper records as the reference rarely described what information was considered sufficient to confirm diagnoses, and this may have varied across studies and reference sources. Indeed, one study estimated a lower PPV using free-text review compared to paper record review; free-text may less completely capture evidence for CMs and thus underestimate PPV. Use of a different EHR dataset as the reference standard may also be imperfect. Using in-patient HES as the reference, the PPV for hip dysplasia recorded in CPRD was low.[39,52] This may be explained by the inclusion of CPRD diagnoses made before the earliest available HES data, and those treated with non-surgical interventions, which would not be captured in in-patient data.[73]

Estimates of diagnostic validity may also depend on the severity of the CM and the length of the time window chosen to search for supportive evidence. Major CMs with severe phenotypes requiring intensive clinical management are more likely to be encoded in the patient record than CMs with milder phenotypes, potentially increasing both sensitivity estimates for EHR diagnoses and their PPV (due to higher levels of diagnostic recording in the reference data source). Searches for supportive evidence conducted over longer periods (e.g., throughout childhood) may also yield higher PPVs than shorter searches (e.g., in the year after initial diagnosis), particularly for milder phenotypes where clinical contact is infrequent. Validation studies rarely defined the length of time used for these searches, however, and so the potential impact of this on PPV was uncertain.

There may also be genuine differences in the PPV among subpopulations assessed using different reference standards. Full paper records may not be provided if the patient has died or left the practice, or has extensive case notes that are time-consuming for practices to process. Such patients could have a higher or lower probability of a genuine CM compared to those with available paper records. Other concerns include differences across studies in the denominators used to calculate PPV, and studies conducted over different time periods. PPV varies with the prevalence of the condition in the population, and prevalence of some CMs may have changed over time, for example, due to increased ascertainment following improved screening methods. PPV estimates obtained in older studies may not therefore be generalisable to current data.

Finally, although some CMs are detected antenatally and are associated with pregnancy outcomes such as terminations, validation studies focused heavily on postnatal diagnoses. GPs are often the first health professional seen by pregnant women but most antenatal checks do not occur in the GP surgery and so antenatal diagnoses of CMs in primary care data may be incompletely captured.[74] Furthermore, pregnancy outcomes such as terminations have been shown to be under-recorded in these data.[75] This may explain why studies rarely examined antenatal diagnoses. A further challenge is distinguishing whether antenatally recorded CMs relate to the mother or offspring. The validation study of neural tube defects found that 37% of antenatally recorded spina bifida diagnoses related to the mother, not the offspring, lowering the overall PPV for neural tube defects.[37]

The prevalence of CMs identified in the first year of life from primary care databases was comparable to or higher than in registries and increased with longer follow-up.[31,39,48] This could reflect limited follow-up and voluntary notification systems for some registries.[7,76] Higher prevalence in primary care data may also reflect imperfect PPV for CM codes. Using Read codes from outside the 'P' chapter may increase the ascertainment of CMs but decrease the PPV if these codes are also used to record non-congenital conditions. The influence of the codes used is also pertinent to studies of secondary care data. A very recent study, published after we completed the search for this review, identified CMs in UK hospitalisation data using three different ICD-10 code-lists.[77] The prevalence of CMs was shown to depend markedly on the code-list applied, ranging from 1.8% to 4.1%.[77]

4.1 | Strengths and limitations

To our knowledge, our study is the first to examine systematically the methods used by researchers to identify individuals with CMs recorded in UK EHR and the results of associated validation studies. Our comprehensive search strategy is likely to have captured the majority of relevant studies. Our review also brings together the ICD-10 code-lists used to identify CMs in secondary care data. The Read code-lists used in primary care studies were not all publicly available, but our efforts to contact authors enabled us to obtain additional lists.

Our findings have some limitations. First, it was not possible to obtain detailed Read code-lists for all the studies of primary care data. Available code-lists were often broad in nature and included codes that were not part of the final case definition. Therefore, our summary of the Read codes used to identify CMs was incomplete and it was unclear in some studies which codes were ultimately used.

In addition, the small number of validation studies identified were insufficient to allow formal assessment of the risk of publication bias in this review. Validations of diagnoses in EHR are often a minor component of the main study. Finally, validations estimating low PPV could be less likely to be included in publications than those estimating higher PPV. If so, the PPVs reported in this review may be a biased subset of the PPVs that have been carried out.

5 | CONCLUSIONS AND RECOMMENDATIONS

This review provides a detailed summary of the methods and code-lists used to identify CMs from UK EHR, which will help guide future

safety studies. It has highlighted the need to increase the methodological detail provided by such studies and the sharing of Read code-lists, in line with the 2015 reporting guidelines for studies using observational routinely collected health data.[78]

Further validation studies are also needed for a range of malformation subgroups, and estimating all measures of validity (sensitivity, specificity and NPV). Validation of diagnoses in hospital data are required to address information gaps and to assess the robustness of studies using these data. In addition, UK general practices are currently switching from use of Read codes to the SNOMED coding system.[79] Work will therefore be needed to assess the validity of equivalent SNOMED codes to those currently used for CMs.

Finally, more extensive use of linked primary and secondary care data could enable an understanding of the completeness of each data source for different CMs and allow for fuller ascertainment of these conditions. The gradual increase in the electronic recording of coded diagnoses in linked outpatient hospital data will enable their future use in malformation research, improving ascertainment of those who do not require in-patient care.

ACKNOWLEDGEMENTS

We thank Dr. Helen I. McDonald for her valuable insight during the preparation of this manuscript and for sharing her expertise around clinical coding practices in hospital settings. The content of this paper has not been published by any other journal or at any conference and is not under consideration for publication by any other journal. This work was supported by the National Institute for Health Research Health Protection Research Unit in Immunisation at the London School of Hygiene and Tropical Medicine in partnership with Public Health England (grant number EPIDZD03). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The funding agency was not involved in the study design, data collection, data management, analysis or interpretation. There was no involvement of the funding agency in the preparation, review or approval of the manuscript for submission for publication.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICS STATEMENT

This study is a systematic review of previously published studies and it was therefore not necessary to apply for ethical approval.

AUTHOR CONTRIBUTIONS

Study concept and design: All authors; Acquisition of data: Maria Peppa; Analysis and interpretation of data: All authors; Drafting the manuscript: Maria Peppa, Sara L. Thomas; Critical revision of the manuscript for important intellectual content: All authors; Statistical analysis: Maria Peppa, Sara L. Thomas; Obtained funding: Sara L. Thomas, Punam Mangtani.

ORCID

Maria Peppa https://orcid.org/0000-0002-4181-0638
Caroline Minassian https://orcid.org/0000-0001-9406-1928
Punam Mangtani https://orcid.org/0000-0001-7074-2999

REFERENCES

1. Charlton RA, Neville AJ, Jordan S, et al. Healthcare databases in Europe for studying medicine use and safety during pregnancy. *Pharmacoepidemiol Drug Saf*. 2014;23(6):586-594.
2. Donegan K, King B, Bryan P. Safety of pertussis vaccination in pregnant women in UK: observational study. *BMJ*. 2014;349:g4219.
3. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836.
4. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data resource profile: Hospital episode statistics admitted patient care (HES APC). (1464-3685 [Electronic]).
5. Office For National Statistics. User guide to mortality statistics; 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatisticsjuly2017. Accessed March 3, 2019.
6. Dave S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf*. 2009;18(8):704-707.
7. Boyd PA, Haeusler M, Barisic I, Loane M, Garne E, Dolk H. Paper 1: the EUROCAT network—organization and processes. *Birth Defects Res A Clin Mol Teratol*. 2011;91(1):S2-S15.
8. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. *Br J Clin Pharmacol*. 2010;69(1):4-14.
9. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *J Public Health (Oxf)*. 2012;34(1):138-148.
10. CPRD Bibliography. 2020. https://www.cprd.com/Bibliography/. Accessed September 20, 2019.
11. QResearch Research Papers. 2020. https://www.qresearch.org/publications/research-papers/. Accessed September 20, 2019.
12. THIN Publications. 2020. https://www.ucl.ac.uk/epidemiology-health-care/research/primary-care-and-population-health/research/thin-database/publications. Accessed September 20, 2019.
13. Boston Collaborative Drug Surveillance Program Publications. 2015. http://www.bu.edu/bcdsp/publications-2/. Accessed September 20, 2019.
14. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
15. Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane Handbook for Systematic Reviews of Interventions Version 6.1 (updated September 2020). Cochrane; 2020. www.training.cochrane.org/handbook.
16. McGuinness LA, Warren-Gash C, Moorhouse LR, Thomas SL. The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: a systematic review. *Pharmacoepidemiol Drug Saf*. 2019;28(2):244-255.
17. Cea-Soriano L, Garcia-Rodriguez LA, Brodovicz KG, Masso Gonzalez E, Bartels DB, Hernandez-Diaz S. Safety of non-insulin glucose-lowering drugs in pregnant women with pre-gestational diabetes: a cohort study. *Diabetes Obes Metab*. 2018;20(7):1642-1651.
18. Baril L, Rosillon D, Willame C, et al. Risk of spontaneous abortion and other pregnancy outcomes in 15-25 year old women exposed to human papillomavirus-16/18 AS04-adjuvanted vaccine in the United Kingdom. *Vaccine*. 2015;33(48):6884-6891.
19. Ruigomez A, Garcia Rodriguez LA, Cattaruzzi C, et al. Use of cimetidine, omeprazole, and ranitidine in pregnant women and pregnancy outcomes. *Am J Epidemiol*. 1999;150(5):476-481.
20. Petersen I, Collings SL, McCrea RL, et al. Antiepileptic drugs prescribed in pregnancy and prevalence of major congenital malformations: comparative prevalence studies. *Clin Epidemiol*. 2017;9:95-103.
21. Petersen I, Sammon CJ, McCrea RL, et al. Risks associated with antipsychotic treatment in pregnancy: comparative cohort studies

based on electronic health records. *Schizophr Res*. 2016;176(2–3): 349-356.
22. Charlton RA, Snowball JM, Nightingale AL, Davis KJ. Safety of fluticasone propionate prescribed for asthma during pregnancy: a UKpopulation-based cohort study. *J Allergy Clin Immunol Pract*. 2015; 3(5):772-779.e773.
23. Charlton RA, Weil JG, Cunnington MC, Ray S, de Vries CS. Comparing the general practice research database and the UKepilepsy and pregnancy register as tools for postmarketing teratogen surveillance: anticonvulsants and the risk of major congenital malformations. *Drug Saf*. 2011;34(2):157-171.
24. Charlton RA, Weil JG, Cunnington MC, de Vries CS. Identifying major congenital malformations in the UKgeneral practice research database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf*. 2010;33(9):741-750.
25. Dhalwani NN, Szatkowski L, Coleman T, Fiaschi L, Tata LJ. Nicotine replacement therapy in pregnancy and major congenital anomalies in offspring. *Pediatrics*. 2015;135(5):859-867.
26. Ban L, Gibson JE, West J, et al. Maternal depression, antidepressant prescriptions, and congenital anomaly risk in offspring: a population-based cohort study. *BJOG*. 2014;121(12):1471-1481.
27. Ban L, Tata LJ, Fiaschi L, Card T. Limited risks of major congenital anomalies in children of mothers with IBD and effects of medications. *Gastroenterology*. 2014;146(1):76-84.
28. Ban L, West J, Gibson JE, et al. First trimester exposure to anxiolytic and hypnotic drugs and the risks of major congenital anomalies: a United Kingdom population-based cohort study. *PLoS One*. 2014;9(6): e100996.
29. Ban L, Fleming KM, Doyle P, et al. Congenital anomalies in children of mothers taking antiepileptic drugs with and without periconceptional high dose folic acid use: a population-based cohort study. *PLoS One*. 2015;10(7):e0131130.
30. Ban L, West J, Abdul Sultan A, Dhalwani NN, Ludvigsson JF, Tata LJ. Limited risks of major congenital anomalies in children of mothers with coeliac disease: a population-based cohort study. *BJOG*. 2015; 122(13):1833-1841.
31. Sokal R, Fleming KM, Tata LJ. Potential of general practice data for congenital anomaly research: comparison with registry data in the United Kingdom. *Birth Defects Res A Clin Mol Teratol*. 2013;97(8): 546-553.
32. Sokal R, Tata LJ, Fleming KM. Sex prevalence of major congenital anomalies in the United Kingdom: a national population-based study and international comparison meta-analysis. *Birth Defects Res A Clin Mol Teratol*. 2014;100(2):79-91.
33. Vasilakis-Scaramozza C, Aschengrau A, Cabral HJ, Jick SS. Asthma drugs and the risk of congenital anomalies. *Pharmacotherapy*. 2013; 33(4):363-368.
34. Vasilakis-Scaramozza C, Aschengrau A, Cabral H, Jick SS. Antidepressant use during early pregnancy and the risk of congenital anomalies. *Pharmacotherapy*. 2013;33(7):693-700.
35. Vasilakis-Scaramozza C, Aschengrau A, Cabral HJ, Jick SS. Antihypertensive drugs and the risk of congenital anomalies. *Pharmacotherapy*. 2013;33(5):476-482.
36. Tata LJ, Lewis SA, McKeever TM, et al. Effect of maternal asthma, exacerbations and asthma medication use on congenital malformations in offspring: a UKpopulation-based study. *Thorax*. 2008; 63(11):981-987.
37. Jick SS. Pregnancy outcomes after maternal exposure to fluconazole. *Pharmacotherapy*. 1999;19(2):221-222.
38. Jick SS, Terris BZ. Anticonvulsants and congenital malformations. *Pharmacotherapy*. 1997;17(3):561-564.
39. Devine S, West SL, Andrews E, et al. Validation of neural tube defects in the full featured—general practice research database. *Pharmacoepidemiol Drug Saf*. 2008;17(5):434-444.
40. Tata LJ, Card TR, Logan RF, Hubbard RB, Smith CJ, West J. Fertility and pregnancy-related events in women with celiac disease: a population-based cohort study. *Gastroenterology*. 2005;128(4):849-855.
41. Lawrenson R, Wyndaele JJ, Vlachonikolis I, Farmer C, Glickman S. Renal failure in patients with neurogenic lower urinary tract dysfunction. *Neuroepidemiology*. 2001;20(2):138-143.
42. Lawrenson R, Wyndaele JJ, Vlachonikolis I, Farmer C, Glickman S. A UKgeneral practice database study of prevalence and mortality of people with neural tube defects. *Clin Rehabil*. 2000;14(6):627-630.
43. Chi CC, Mayon-White RT, Wojnarowska FT. Safety of topical corticosteroids in pregnancy: a population-based cohort study. *J Invest Dermatol*. 2011;131(4):884-891.
44. Petersen I, Evans SJ, Gilbert R, Marston L, Nazareth I. Selective serotonin reuptake inhibitors and congenital heart anomalies: comparative cohort studies of women treated before and during pregnancy and their children. *J Clin Psychiatry*. 2016;77(1):e36-e42.
45. Margulis AV, Abou-Ali A, Strazzeri MM, et al. Use of selective serotonin reuptake inhibitors in pregnancy and cardiac malformations: a propensity-score matched cohort in CPRD. *Pharmacoepidemiol Drug Saf*. 2013;22(9):942-951.
46. Hammad TA, Margulis AV, Ding Y, Strazzeri MM, Epperly H. Determining the predictive value of read codes to identify congenital cardiac malformations in the UKclinical practice research datalink. *Pharmacoepidemiol Drug Saf*. 2013;22(11):1233-1238.
47. Billett J, Cowie MR, Gatzoulis MA, Vonder Muhll IF, Majeed A. Comorbidity, healthcare utilisation and process of care measures in patients with congenital heart disease in the UK: cross-sectional, population-based study with case-control analysis. *Heart*. 2008;94(9): 1194-1199.
48. Wurst KE, Ephross SA, Loehr J, Clark DW, Guess HA. Evaluation of the general practice research database congenital heart defects prevalence: comparison to United Kingdom national systems. *Birth Defects Res A Clin Mol Teratol*. 2007;79(4):309-316.
49. Wurst KE, Ephross SA, Loehr J, Clark DW, Guess HA. The utility of the general practice research database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiol Drug Saf*. 2007; 16(8):867-877.
50. Bannister J, Szatkowski L, Sharkey D, Tan S, Fiaschi L, Ban L. Early life incidence of gastrointestinal and respiratory infections in children with gastroschisis: a cohort study. *J Pediatr Gastroenterol Nutr*. 2018; 67(5):580-585.
51. Perry DC, Bruce CE, Pope D, Dangerfield P, Platt MJ, Hall AJ. Comorbidities in Perthes' disease: a case control study using the general practice research database. *J Bone Joint Surg Br*. 2012;94(12):1684-1689.
52. Broadhurst C, Rhodes AML, Harper P, Perry DC, Clarke NMP, Aarvold A. What is the incidence of late detection of developmental dysplasia of the hip in England?: a 26-year national study of children diagnosed after the age of one. *Bone Joint J*. 2019;101-B(3):281-287.
53. Zylbersztejn A, Gilbert R, Hjern A, Hardelid P. Origins of disparities in preventable child mortality in England and Sweden: a birth cohort study. *Arch Dis Child*. 2020;105(1):53-61.
54. Zylbersztejn A, Gilbert R, Hjern A, Wijlaars L, Hardelid P. Child mortality in England compared with Sweden: a birth cohort study. *Lancet*. 2018;391(10134):2008-2018.
55. Dimopoulos K, Muthiah K, Alonso-Gonzalez R, et al. Heart or heart-lung transplantation for patients with congenital heart disease in England. *Heart*. 2019;105(8):596-602.
56. Kempny A, Dimopoulos K, Uebing A, et al. Outcome of cardiac surgery in patients with congenital heart disease in England between 1997 and 2015. *PLoS One*. 2017;12(6):e0178963.
57. Singhal A, Ross J, Seminog O, Hawton K, Goldacre MJ. Risk of self-harm and suicide in people with specific psychiatric and physical disorders: comparisons between disorders using English national record linkage. *J R Soc Med*. 2014;107(5):194-204.

58. Billett J, Majeed A, Gatzoulis M, Cowie M. Trends in hospital admissions, in-hospital case fatality and population mortality from congenital heart disease in England, 1994 to 2004. *Heart*. 2008;94(3):342-348.
59. Fitzsimons KJ, Copley LP, van der Meulen JH, Panagamuwa C, Deacon SA. Grommet surgery in children with orofacial clefts in England. *Cleft Palate Craniofac J*. 2017;54(1):80-89.
60. Fitzsimons KJ, Copley LP, Smallridge JA, Clark VJ, van der Meulen JH, Deacon SA. Hospital admissions for dental treatment among children with cleft lip and/or palate born between 1997 and 2003: an analysis of hospital episode statistics in England. *Int J Paediatr Dent*. 2014;24(3):200-208.
61. Fitzsimons KJ, Copley LP, Deacon SA, van der Meulen JH. Hospital care of children with a cleft in England. *Arch Dis Child*. 2013;98(12):970-974.
62. Fitzsimons KJ, Mukarram S, Copley LP, Deacon SA, van der Meulen JH. Centralisation of services for children with cleft lip or palate in England: a study of hospital episode statistics. *BMC Health Serv Res*. 2012;12:148.
63. McAllister DA, Morling JR, Fischbacher CM, Reidy M, Murray A, Wood R. Enhanced detection services for developmental dysplasia of the hip in Scottish children, 1997-2013. *Arch Dis Child*. 2018;103(11):1021-1026.
64. Dharmasena A, Keenan T, Goldacre R, Hall N, Goldacre MJ. Trends over time in the incidence of congenital anophthalmia, microphthalmia and orbital malformation in England: database study. *Br J Ophthalmol*. 2017;101(6):735-739.
65. Lansdale N, Al-Khafaji N, Green P, Kenny SE. Population-level surgical outcomes for infantile hypertrophic pyloric stenosis. *J Pediatr Surg*. 2018;53(3):540-544.
66. Wilkinson DJ, Green PA, Beglinger S, et al. Hypospadias surgery in England: Higher volume centres have lower complication rates. *J Pediatr Urol*. 2017;13(5):481.e1-481.e6.
67. Jarvis S, Fraser LK. Comparing routine inpatient data and death records as a means of identifying children and young people with life-limiting conditions. *Palliat Med*. 2018;32(2):543-553.
68. Jarvis S, Parslow RC, Carragher P, Beresford B, Fraser LK. How many children and young people with life-limiting conditions are clinically unstable? A national data linkage study. *Arch Dis Child*. 2017;102(2):131-138.
69. Fraser LK, Lidstone V, Miller M, et al. Patterns of diagnoses among children and young adults with life-limiting conditions: a secondary analysis of a national dataset. *Palliat Med*. 2014;28(6):513-520.
70. Fraser LK, Miller M, Hain R, et al. Rising national prevalence of life-limiting conditions in children in England. *Pediatrics*. 2012;129(4):e923-e929.
71. Ferencz C, Loffredo C, Correa-Villasenor A, Wilson P. Genetic and environmental risk factors of major cardiovascular malformations: the Baltimore-Washington infant study: 1981–1989. *Perspect Pediatr Cardiol*. 1997;5:867-868.
72. Wren C, Richmond S, Donaldson L. Temporal variability in birth prevalence of cardiovascular malformations. *Heart*. 2000;83(4):414-419.
73. Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2005;14(7):443-451.
74. Redshaw M, Heikkila K. Delivered with Care: A National Survey of Women's Experience of Maternity Care 2010; 2010.
75. Minassian C, Williams R, Meeraus WH, Smeeth L, OMR C, Thomas SL. Methods to generate and validate a pregnancy register in the UKclinical practice research Datalink primary care database. *Pharmacoepidemiol Drug Safety*. 2019;28(7):923-933.
76. Bishop C, Small N, Mason D, et al. Improving case ascertainment of congenital anomalies: findings from a prospective birth cohort with detailed primary care record linkage. *BMJ Paediatr Open*. 2017;1(1):e000171.
77. Zylbersztejn A, Verfurden M, Hardelid P, Gilbert R, Wijlaars L. Phenotyping congenital anomalies in administrative hospital records. *Paediatr Perinat Epidemiol*. 2020;34(1):21-28.
78. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.
79. NHS Digital. SNOMED CT Implementation in Primary Care; 2018. https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct/snomed-ct-implementation-in-primary-care. Accessed October 21, 2019.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Peppa M, Minassian C, Mangtani P, Thomas SL. The identification and validity of congenital malformation diagnoses in UK electronic health records: A systematic review. *Pharmacoepidemiol Drug Saf*. 2021;30:875–898. https://doi.org/10.1002/pds.5251