SHORT COMMUNICATION

Drug Saf 2010; 33 (9): 741-750
0114-5916/10/0009-0741/$49.95/0

Identifying Major Congenital Malformations in the UK General Practice Research Database (GPRD)

A Study Reporting on the Sensitivity and Added Value of Photocopied Medical Records and Free Text in the GPRD

Rachel A. Charlton,[1] *John G. Weil*,[2] *Marianne C. Cunnington*[2] and *Corinne S. de Vries*[1]

1 Department of Pharmacy and Pharmacology, University of Bath, Bath, UK
2 Worldwide Epidemiology, GlaxoSmithKline, Harlow, Essex, UK

Abstract

Background: Postmarketing teratogen surveillance is essential and requires a data source that can reliably capture a wide range of congenital malformations. The UK General Practice Research Database (GPRD) may have the potential to be used for this kind of surveillance.

Objective: To assess the extent to which this database can be used to accurately identify major congenital malformations.

Methods: This study was carried out as part of a broader study to compare data on anticonvulsant use and safety in pregnancy between the GPRD and a pregnancy registry. The study period ran from 1 January 1990 until 31 December 2006. Mother-baby pairs where the mother had a record of epilepsy, seizure or convulsion were identified using the GPRD computerized medical records. Infants of mother-baby pairs who had a record of a major congenital malformation were identified. Full photocopied paper medical records were requested from the infant's general practitioner and where this was not possible any data entries consisting of uncoded comments, so-called 'free text', in the electronic GPRD record were requested from the database provider. This additional information was then reviewed in order to determine the extent to which the congenital malformation diagnoses identified via the computerized records could be confirmed or rejected and then classified as being major or minor.

Results: Within the study population of 3869 live mother-baby pairs, 188 potentially major congenital malformations were identified from the GPRD computerized record relating to 161 unique individuals. Using a combination of photocopied medical records and free text it was possible to verify 160 malformations (85.1%) as the malformation indicated by the computerized records; this ranged from 91.7% of those cases verified using photocopied medical records and 77.9% of cases verified using free text. Of the verified

congenital malformations, using a combination of computerized data, photocopied medical records and free text, it was possible to classify 78.1% as being major and 15.0% as minor, and this percentage was found to be the same for those cases reviewed by photocopied records and those where free text was used. The proportions of malformations that could be verified and those that could be classified as major or minor were found to vary by malformation class.

Conclusions: The GPRD can be used to ascertain a wide range of congenital malformations. In many cases, when a malformation is identified in the GPRD via the computerized medical records, the malformation is likely to exist. However, in this study a small proportion of identified cases had to be excluded because they had been coded incorrectly or diagnostically ruled out. Therefore, depending on the congenital malformation of interest, verification of such malformations using photocopied medical records or free text is generally recommended.

Background

The UK General Practice Research Database (GPRD) is the world's largest computerized database of anonymized, longitudinal medical records from primary care,[1] and has the potential, in some circumstances, to aid research into the safety of medicines used during pregnancy.[2] Work has begun to validate the recording of congenital malformations in the GPRD, although to our knowledge thus far it has been limited to a selection of cardiovascular defects[3] and neural tube defects.[4]

This paper forms part of a broader study to assess the potential of the GPRD to act as a pregnancy registry system focusing on a cohort of women with a diagnosis of epilepsy, seizure or convulsion. Here we describe the methodology used for the surveillance of the range of congenital malformations that would be required if the GPRD is to be used to monitor drug safety in pregnancy. We report the extent to which diagnoses identified in the GPRD could be confirmed, rejected or made more specific by using the photocopied paper medical records or where this was not possible any data entries consisting of uncoded comments, so-called 'free text,' in the GPRD, which are not routinely available for research purposes.

Methods

Data Source

The GPRD contains over 44 million person-years of data and currently captures data from approximately 4 million active patients (approximately 7% of the UK population).[5] Virtually all prescriptions, non-drug interventions and referrals issued by general practitioners (GPs) are recorded in the database, as are medical diagnoses, including those relating to pregnancy. The database is managed by the GPRD division at the Medicines and Healthcare products Regulatory Authority (MHRA), who also provide a number of services to allow verification of events and diagnoses identified from the computerized records. For patients still registered at the GP practice it is possible to request anonymized photocopies of the patient's full paper medical record, enabling access to all referrals and outpatient letters and correspondence from consultants and specialists. For all patients, regardless of whether they are still registered with the GP practice, it is possible to request the information recorded by GPs as free text in the electronic GPRD record. GPs have the option of recording free text comments, i.e. uncoded additional information relating to symptoms, more detailed diagnoses, test

results, etc., each time they record a medical code within the database. It is recommended that these verification services are used to assess the sensitivity and specificity of medical codes.[1] All methods of verification are requested via the GPRD group at the MHRA to ensure that both practice and patient confidentiality are maintained.

Study Population

Women were eligible for inclusion if they were, or had been, permanently registered at a GP practice regarded by the MHRA to be contributing data up to research standard. Women were included in the analysis of verification of major congenital malformation codes within the child's record if they had a medical record indicating a live pregnancy outcome between 1 January 1990 and 31 December 2006 and they had a code relating to a diagnosis of epilepsy, seizure or convulsion.

Women were excluded from the cohort if they were not 14–49 years of age at the date of delivery and if they did not have a code indicating a pregnancy in the 280 days leading up to the delivery date. In view of requirements for the broader study, women were also excluded if they were not continuously enrolled in the GPRD for the 4 months before the estimated date of the last menstrual period or if the diagnosis of epilepsy, seizure or convulsion was not recorded before the pregnancy indicator code.

The offspring of women meeting all inclusion criteria were identified where possible (based on having the same family and GP practice numbers, and the child's year and month of birth being equal to the mother's year and month of delivery). Mother-baby pairs were included if the linked child was registered and present in the GPRD 3 months after the mother's pregnancy outcome date. If the child had been registered and died before reaching 3 months of age the mother-baby pair was still included and the need to be registered 3 months after the pregnancy outcome date was no longer required. Women were entitled to contribute more than one mother-baby pair to the study and each unique offspring was considered separately.

Identification of Major Congenital Malformations

To identify individuals with a major congenital malformation, search terms for all congenital malformations were created based on those listed in the *International Classification of Diseases 9th Revision*[6] 'congenital anomalies' chapter (ICD-9 codes 740–759). These search terms contained 'wild cards' to allow for variations in terms, e.g. 'hydroceph*' to account for 'hydroceph-aly' and 'hydroceph-alus'. All medical codes containing these search terms were then selected plus any relevant codes identified by being in their hierarchical vicinity. Children of women in the cohort were then identified in the GPRD if they had any of these codes within their medical record.

As pregnancy registries are primarily concerned with major malformations, the malformations identified were categorized as major or minor based on the information available through the Read and OXMIS codes used in the GPRD, photocopied medical records and free text. This categorization was based on the classification used by the European network of population-based registers for the epidemiological surveillance of congenital anomalies (EUROCAT). Minor defects and malformations associated with prematurity, when isolated, are excluded from EUROCAT reports. In addition, there are some malformations (e.g. hypospadias, hydronephrosis, talipes, syndactyly) that are only classified as major if certain criteria are met.[7,8]

All congenital malformation codes identified within the cohort were reviewed independently by two of the authors (RC, JW). Those relating to a major malformation and those relating to a malformation that could be classed as major, if specific criteria were met, were selected for verification.

Verification of Congenital Malformations

Verification of congenital malformations was carried out by requesting a photocopy of the child's entire medical record for children still registered with the practice. These were then reviewed to ensure that the congenital malformation recorded in the computerized medical record was a true malformation and had been correctly

coded. Secondly, for those malformations that needed to meet specific inclusion criteria in order to be classified as major, information relating to this was identified, where present, enabling the true number of major malformations to be determined. Photocopied medical records were requested via the MHRA, who sent out an initial request and three subsequent reminders to GPs at fortnightly intervals. For children who had transferred out of the practice and for those where the GP did not return photocopied records, all the information recorded in the free text within the child's entire electronic GPRD medical record was obtained and reviewed.[1] One example recorded alongside a code for an oral cleft included "Right cleft lip – admit ***, repair cleft lip under GA, Hare lip palate intact." All information in the photocopied medical records and free text was anonymized by the MHRA before being transferred to the investigators.

Analyses

The percentage of diagnoses identified from the computerized records that could be confirmed as being the congenital malformation recorded was calculated for the photocopied medical records and free text verification methods separately and combined. Of those malformations that were confirmed, the percentage where sufficient information was available to classify as a major or minor congenital malformation was then calculated both overall and for different subgroups of congenital malformations. As we did not request photocopied medical records or free text for those individuals who were not identified as having a major congenital malformation, it was not possible to calculate the specificity of the computerized records.

Results

Verifying the Presence of a Congenital Malformation

The study population consisted of 3869 live births within this cohort of women with a record of epilepsy, seizure or convulsion any time before the pregnancy. 188 potentially major congenital malformations recorded at any age were identified relating to 161 unique individuals. Photocopied medical records were requested for the 123 individuals (76.4%) still registered with the practice. Figure 1 shows the response to the photocopied record requests, with 96 records (78.0%) being returned, relating to 109 unique malformations. In addition, the GPs of 12 individuals (9.8%) replied to explain why they were not enclosing the photocopied records.

For 15 individuals (12.2%) the GP did not respond to the record request and therefore free text was obtained. Free text was also obtained for the ten cases where the GP responded but did not provide additional information on the potential major congenital malformation and for the 38 individuals no longer registered with the practice. In total, free text was therefore obtained for 63 individuals (39.1%) representing 77 major congenital malformations.

Table I along with figures 2 and 3 summarize the percentage of malformations that could be confirmed (i.e. verified) or refuted as a congenital malformation, and the percentage of those confirmed that could be further classified as being major or minor. Using a combination of photocopied medical records and free text it was possible to verify 160 malformations (85.1%) as the malformation indicated by the computerized records. For those where photocopied records were available, the percentage verified was 91.7% (100/109) and for those where the free text was used it was possible to verify 77.9% of malformations (60/77). Nine cases (4.8%) in total were found not to have the malformation recorded; these included seven cases of diagnostic suspicion that were not confirmed on investigation and where computer records were not updated, one case where the GP wrote in response to receiving the request for photocopied medical records to state that the infant did not have a major congenital malformation and one case where an incorrect code had been recorded. Using photocopied medical records it was

1 It is not possible to obtain photocopied records for patients who have transferred out of the practice.

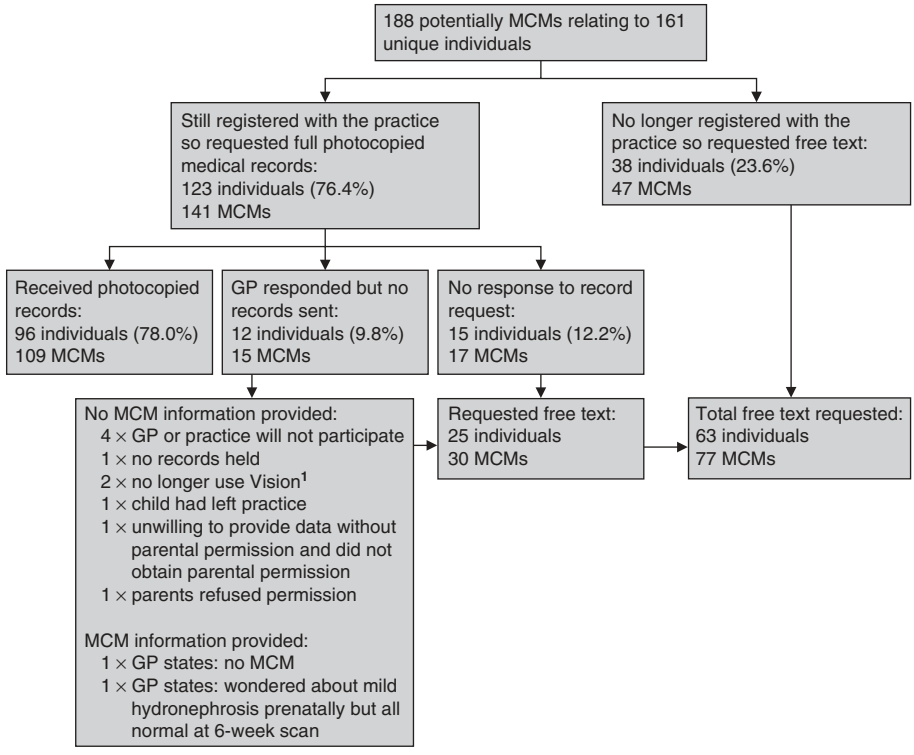

Fig. 1. Methods used for verifying malformations. **1** The general practitioner (GP) practice software that feeds into the General Practice Research Database. **MCM** = major congenital malformation.

possible to either confirm or refute the presence of a malformation in 96.3% of cases (105/109) and in 80.5% of cases (62/77) using the free text.

Table II demonstrates that the verification of the presence of a congenital malformation varied by malformation class. It was possible to verify all CNS, digestive system, chromosomal and fetal valproate syndrome anomalies. Within the majority of other malformation classes, with the exception of talipes, congenital dislocation or dysplasia of the hip and anomalies of the eye, verification was found to be over 80.0%.

Classifying as a Major or Minor Congenital Malformation

Table I shows that of the 160 verified congenital malformations, it was possible to classify 93.1% (149) as being major or minor. This percentage was found to be the same for those cases reviewed by photocopied medical records and those reviewed using free text. For 125 cases (78.1% of those verified), review of the photocopied medical records or free text resulted in the malformation identified on the computerized records being classified as major, whilst 24 cases (15.0%) were classified as minor. For the remaining 6.9% (11/160) of verified malformations there was insufficient information to classify them as being major or minor. Of the 188 congenital malformations initially identified, 125 (66.5%) could be both verified as a malformation and classified as being major, which is what would be of primary interest to pregnancy registries.

The percentage of verified congenital malformations that could be classified as major was found to vary by malformation class (table II). For many classes of malformation the percentage that could be classified as major was ≥80.0%. This was, however, lower for those classes that

Table I. Percentage of congenital malformations that could be verified and, of those, the percentage that could be classified as major or minor

Verification method	No. of malformations	Outcome of verification of congenital malformations identified by medical codes in the GPRD [n (%)]			Verified congenital malformations that could be classified as major or minor (n = 160) [n (%)]		
		yes (a malformation)	no (refute, incorrectly coded)	not enough information	major	minor	not enough information
Photocopied medical records[a]	109	100 (91.7)	5 (4.6)	4 (3.7)	78 (78.0)	15 (15.0)	7 (7.0)
Free text[b]	77	60 (77.9)	2 (2.6)	15 (19.5)	47 (78.3)	9 (15.0)	4 (6.7)
Both methods combined[c]	188[c]	160 (85.1)	9[c] (4.8)	19 (10.1)	125 (78.1)	24 (15.0)	11 (6.9)

a Anonymized photocopies of the patient's full paper medical record, enabling access to all referrals and outpatient letters and correspondence from consultants and specialists.

b Anonymized uncoded additional information, i.e. relating to symptoms, more detailed diagnoses, test results, etc. recorded by the general practitioner in the patient's computerized medical record.

c Includes the two cases where the general practitioner wrote to say that there was no malformation but did not send photocopied medical records and free text was not requested.

GPRD = General Practice Research Database.

included malformations needing to meet specific inclusion criteria (e.g. hydronephrosis, hypospadias, syndactyly, talipes, patent ductus arteriosus) to be classified as being major (e.g. talipes – where cases of a postural origin were excluded from the major malformation count).

Discussion

This study indicates that a wide range of congenital malformations can be identified reliably using the GPRD. The presence or absence of a malformation, regardless of whether major or minor, was confirmed in 85.1% (160/188) of those identified using the computerized records. In infants for whom photocopied medical records were received, it was possible to confirm or refute 96.3% of recorded malformations. The overall verification rate was lowered by the review of some cases being limited to information in the free text, but this resulted largely from the study covering a retrospective 16-year time period. Had the study been carried out in real-time following a drug being launched on the market, or had a review been carried out at frequent intervals, then photocopied medical records would have been available for a larger proportion of individuals, probably resulting in higher verification rates.

At 78.1%, the overall verification and classification of major malformations within this cohort of individuals in the GPRD compares favourably to that found by Cooper et al.[9] using Tennessee Medicaid computerized records. They found that 67.7% of all major congenital malformations identified by birth certificates or patient claims could be confirmed by medical record review and that medical records were available for 98.9% of individuals. Two studies have identified major congenital malformations using the Ingenix Research Data Mart (which contains medical and pharmacy claims data from United Healthcare affiliated health plans) and verified them by medical record abstraction.[10,11] Only one study reported on the availability of medical records and that they were available for 86% of potential major congenital malformation cases.[10] Both studies reported only the number of confirmed major congenital malformations and did not

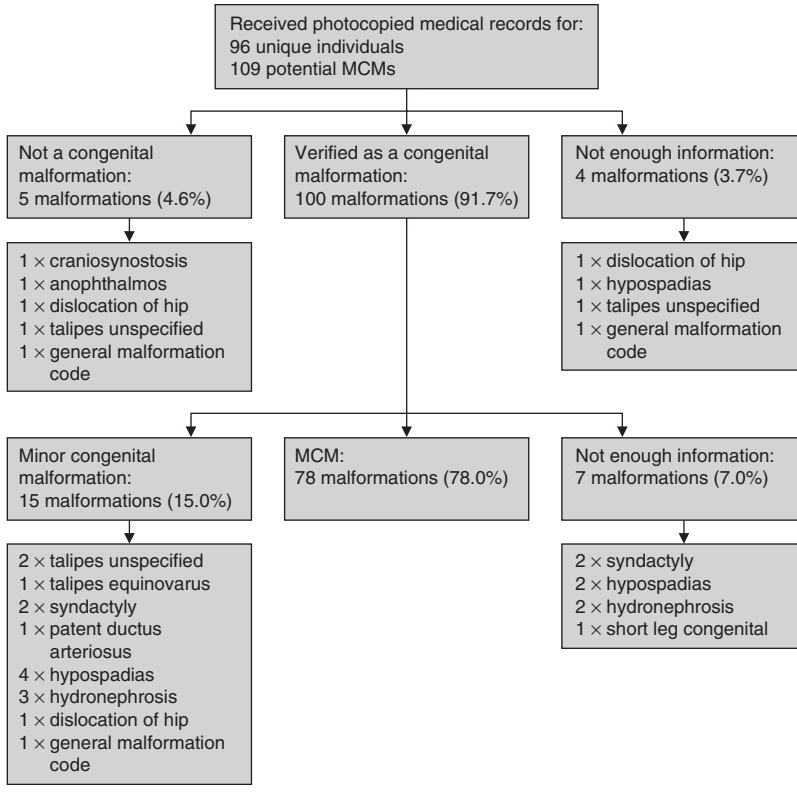

Fig. 2. Percentage of congenital malformations that could be verified using photocopied medical records, and the percentage of those that could be classified as major or minor. Major/minor classification was based on that used by EUROCAT.[7,8] **EUROCAT** = European network of population-based registers for the epidemiological surveillance of congenital anomalies; **MCM** = major congenital malformation.

report the proportion of those identified via the computerized records that could be verified, so direct comparison with our study was not possible.

Verification rates in the GPRD, as with the Tennessee Medicaid records, appeared to vary by the type of defect recorded. Malformations that were compatible with life and required medical intervention or monitoring could be verified relatively easily. For malformations that could be defined as being either major or minor depending on anatomical details regarding the malformation (e.g. hypospadias, hydronephrosis, syndactyly), as well as for malformations that could resolve spontaneously, it proved more difficult to obtain the necessary detail. For instance, the six congenital heart defects where there was insufficient information to confirm the diagnosis consisted of four ventricular septal defects, one patent ductus arteriosus and one case of pulmonary stenosis. We were unable to obtain the photocopied records for any of these cases and no additional information had been recorded in the free text. It is possible that these were less serious, largely asymptomatic defects that did not prompt the GP to record additional information in the free text. For six hypospadias cases (21.4%) there was insufficient information available to classify as major or minor. The EUROCAT guidelines, however, changed in 2005 to include the reporting of all cases of hypospadias as a major malformation and therefore the need for additional detail to distinguish between major and minor types is no longer an issue.

Compared with the photocopied medical records, the free text was less likely to provide sufficient information to confirm or refute the presence

of a malformation. However, if there was evidence that the malformation was present, then using the free text enabled the same percentage of malformations to be classified as major or minor as with the photocopied medical records. With more GP practices scanning patient letters into the free text it is possible that the level of detail recorded in the free text has increased over time.

Within this study we chose to request and review full photocopied medical records in order to obtain as much information as possible about individuals still registered on the database. Previous verification studies of major congenital malformations, however, involved using a medical record abstraction questionnaire, which has also been found to be successful. Wurst et al.,[3] when validating three specific types of congenital heart defect, reported a 94% response rate from GPs to the questionnaire and identified a positive predictive value ≥90% for each defect type. Devine et al.[4] created an algorithm to identify cases of neural tube defects from the computerized records and then sent questionnaires to verify the diagnoses. This study reported a 76.0% response rate to the questionnaire, although a small number were sent to GPs where the patient concerned had already transferred out of the practice. The overall positive predictive value of the algorithm for neural tube defects was 71%, ranging from 47% for spina bifida to 83% for encephalocoele. Questionnaires are, therefore, an additional tool that could be used to verify diagnoses, although when attempting to verify a wide range of malformations the photocopied medical records have the advantage of not requiring the design of a large number of malformation-specific questionnaires.

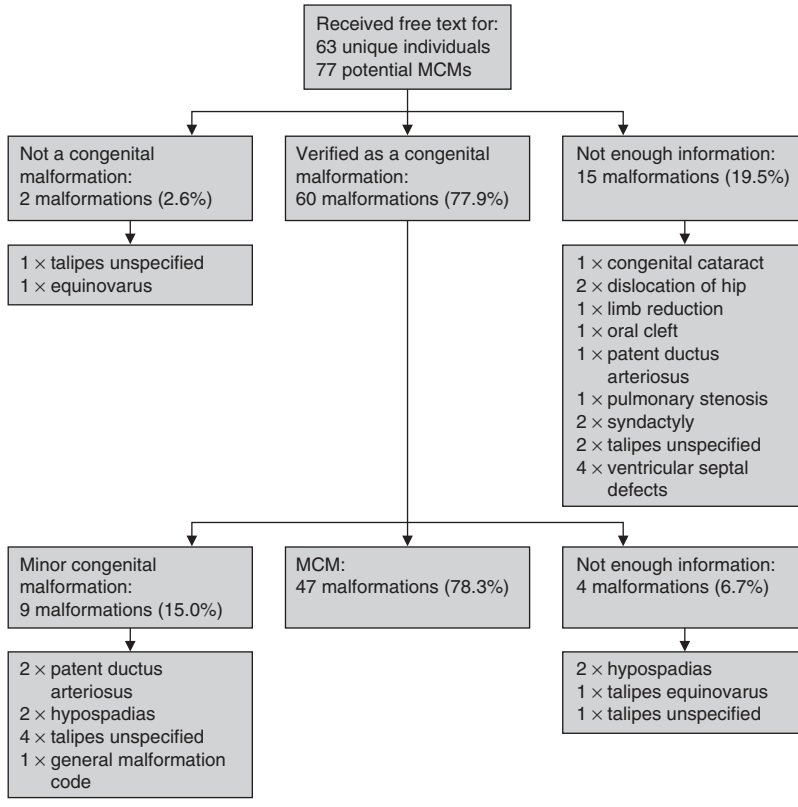

Fig. 3. Percentage of congenital malformations that could be verified using free text, and the percentage of those that could be classified as major or minor. Major/minor classification was based on that used by EUROCAT.[7,8] **EUROCAT** = European network of population-based registers for the epidemiological surveillance of congenital anomalies; **MCM** = major congenital malformation.

Table II. The number of congenital malformations that could be verified and, of those, the number that could be classified as major, by malformation class

Malformation class	No. of malformations	Outcome of verification of congenital malformations identified by medical codes in the GPRD (n)			Of the verified congenital malformations the number that could be classified as major or minor (n)		
		yes (a malformation)	no (refute, incorrectly coded)	not enough information	major	minor	not enough information
CNS	5	5	0	0	5	0	0
Congenital heart defects	50	44	0	6	41	3	0
Orofacial clefts	7	6	0	1	6	0	0
Eye	4	2	1	1	2	0	0
Digestive system	6	6	0	0	6	0	0
Internal urogenital system	23	22	1	0	17	3	2
Hypospadias	28	27	0	1	17	6	4
Talipes	18	12	3	3	4	6	2
Hip dislocation/dysplasia	16	11	2	3	10	1	0
Poly-/syndactyly	10	8	0	2	4	2	2
Limb reduction	6	5	0	1	4	0	1
Musculoskeletal	1	0	1	0	NA	NA	NA
Chromosomal	1	1	0	0	1	NA	NA
Fetal valproate syndrome	4	4	0	0	4	0	0
Other	9	7	1	1	4	3	0
Total	**188**	**160**	**9**	**19**	**125**	**24**	**11**

NA = not applicable; **GPRD** = General Practice Research Database.

This study has demonstrated that in a large number of cases, when a malformation is identified in the GPRD via the computerized medical records, the malformation is likely to exist. There will, however, be a small percentage of identified cases (approximately 5%) that need to be excluded due to being incorrectly coded or diagnostically ruled out. Therefore, making selected information recorded in the free text such as 'excluded' or 'ruled out' routinely available to researchers would be beneficial. For a slightly larger proportion, the malformation identified is present but it may be in a minor form. However, for those who are interested in aetiology, the inclusion of these cases may well be of value.

Conclusions

Postmarketing surveillance programmes are essential to monitor drug safety in pregnancy, and require a data source that can reliably capture cases of congenital malformations. This study has demonstrated the GPRD can be used to ascertain a wide range of congenital malformations. Photocopied medical records and, to a lesser extent, free text have proven valuable sources in carrying out verification of malformations identified through the computerized records. For more severe malformations and those that are easily externally visible, the computerized records were found to be reliable; verification is, however, still recommended, especially for those malformations that can occur with different levels of severity. Further work might assess the accuracy and completeness of the computerized GPRD malformation codes in a different study population and in relation to terminations of pregnancy.

Acknowledgements

The authors wish to thank all the general practitioners who contribute data to the GPRD, particularly those who

responded to the request for photocopied medical records as part of this study.

This study is based in part on data from the Full Feature General Practice Research Database obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. However, the interpretation and conclusions contained in this report are those of the authors alone. The study was funded in part by GlaxoSmithKline and in part by the University of Bath.

Drs Cunnington and Weil are employees of and hold shares in GlaxoSmithKline. Miss Charlton is a past employee of and holds shares in GlaxoSmithKline. Professor de Vries has no conflicts of interest that are relevant to the content of this article.

References

1. Wood L, Martinez C. The General Practice Research Database: role in pharmacovigilance. Drug Saf 2004; 27 (12): 871-81
2. Charlton RA, Cunnington MC, de Vries CS, et al. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. Drug Saf 2008; 31 (1): 39-51
3. Wurst KE, Ephross SA, Loehr J, et al. The utility of the General Practice Research Database to examine selected congenital heart defects: a validation study. Pharmacoepidemiol Drug Saf 2007; 16: 867-77
4. 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: 434-44
5. MHRA. The General Practice Research Database, 2009 [online]. Available from URL: http://www.gprd.com/home/default.asp [Accessed 2010 Mar 1]
6. Hart AC, Hopkins CA, editors. ICD-9-CM professional for pospitals. Volumes 1, 2 and 3. International classification of diseases 9th revision, clinical modification 6th ed. Reston (VA): Ingenix Inc. 2003, St Anthony Publishing/Medicode, 2003
7. EUROCAT. List of minor anomalies for exclusion up to birth year 2004 [online]. Available from URL: http://www.eurocat-network.eu/content/EUROCAT-Old-List-Minor-Anomalies.pdf [Accessed 26 Jan 2010]
8. EUROCAT. Minor anomalies for exclusion (chapter 3.2, guide 1.3), 2005 [online]. Available from URL: http://www.eurocat-network.eu/content/EUROCAT-Guide-1.3.pdf [Accessed 26 Jan 2010]
9. Cooper WO, Hernandez-Diaz S, Gideon P, et al. Positive predictive value of computerised records for major congenital malformations. Pharmacoepidemiol Drug Saf 2008; 17: 455-60
10. Cole JA, Modell JG, Haight BR, et al. Bupropion in pregnancy and the prevalence of congenital malformations. Pharmacoepidemiol Drug Saf 2007; 16: 474-84
11. Cole JA, Ephross SA, Cosmatos IS, et al. Paroxetine in the first trimester and the prevalence of congenital malformations. Pharmacoepidemiol Drug Saf 2007; 16: 1075-85

Correspondence: *Miss Rachel A. Charlton,* Department of Pharmacy and Pharmacology, University of Bath, Claverton Down, Bath BA2 7AY, UK.
E-mail: r.a.charlton@bath.ac.uk