Online Research @ Cardiff

information services
gwasanaethau gwybodaeth

ORIGINAL RESEARCH

Identifying cerebral palsy from routinely-collected data in England and Wales

Bethan Carter [ID] [1]
C Verity Bennett [ID] [1]
Jackie Bethel [ID] [1]
Hywel M Jones [ID] [1]
Ting Wang [ID] [2]
Alison Kemp [ID] [1]

[1]Division of Population Medicine, School of Medicine, Cardiff University, Cardiff, UK; [2]Swansea Medical School, Swansea University, Swansea, UK

Purpose: An observational study using routinely-collected health care data to describe the extent to which children and young people (CYP) with cerebral palsy (CP) can be identified and the prevalence of CP can be estimated.

Patients and methods: Routinely-collected anonymized data, for CYP (aged 0–25 years old between 1 January 2004 and 31 December 2014) were analyzed in two linked datasets, from England and Wales respectively. Datasets included National Health Service; General Practitioner (GP), inpatients, outpatients, and national mortality records. CP was identified using ICD-10 codes G80.0–G83.3 and equivalent Read v2 codes. Ascertainment rates of CP were identified for each data source and compared between countries. Frequency and consistency of coding were investigated, and prevalence of CP estimated.

Results: A total of 7,113 and 5,218 CYP with CP were identified in the English and Welsh datasets respectively. Whilst the majority of CYP with CP would be expected to attend their GP, 65.3% (4,646/7,113) of English and 65.1% (3,396/5,218) of Welsh cases were ascertained from GP datasets. Further cases were identified solely in inpatient datasets (2,410 in England, 1,813 in Wales). Few cases were coded for CP within outpatient datasets. Four character codes that specified CP type were rarely used; one in five health care records were coded both with G80 codes (explicitly CP) and with G81–83 codes (other paralytic syndromes) or equivalent Read codes. Estimated period prevalence of CYP with CP was 2.5–3.4 per 1,000 in England and 2.4–3.2 per 1,000 in Wales.

Conclusion: In England and Wales, coding of CP in routine data is infrequent, inconsistent, non-specific, and difficult to isolate from conditions with similar physical signs. Yet the prevalence estimates of CP were similar to those reported elsewhere. To optimize case recognition we recommend improved coding quality and the use of both primary and secondary care datasets as a minimum.

Keywords: chronic conditions, CPRD, SAIL, coding, data-linkage

Introduction

Routinely-collected health care data within the National Health Service (NHS) offers the opportunity to estimate the prevalence of cerebral palsy (CP) and explore the frequency and consistency of CP coding for this group of children and young people (CYP).

The term CP encompasses a group of permanent but non-progressive disorders of movement and posture development caused by neurological disturbances in early life.[1] These are the most commonly occurring childhood physical disabilities.[2] Presenting in early infancy, most cases are diagnosed by the age of 2 years. CP is a lifelong condition that is associated with multiple co-existing health problems,

Correspondence: Bethan Carter
Division of Population Medicine, School of Medicine, Cardiff University, 3rd Floor Neuadd Meirionnydd, Cardiff, UK
Tel +44 292 068 7765
Email carterbr2@cardiff.ac.uk

including disorders of sensation, cognition, behavior, epilepsy, and musculoskeletal problems.[1] Patients often require a multi-disciplinary range of both primary (General Practitioner [GP]) and secondary (hospital) care.

Prevalence estimates of CP vary between studies, generally falling between 2 and 3 per 1,000 in developed countries,[3] however prevalence of the different CP subtypes is unclear.[4] In England and Wales, it has been estimated that 22,100 children between 3–15 years of age will be living with CP by 2020.[5] Prevalence of CP has been linked to sex,[6,7] ethnicity,[8] and socio-economic status,[9] yet these factors have not been explored at a national level within England and Wales. Previous estimates of CP prevalence in the UK,[10–12] calculated from sources such as local registers or surveys, are regionally restricted, and data from these studies are now more than 20 years old. Despite the complex and ongoing health care needs of CYP with CP, we lack a comprehensive understanding of both the epidemiology and the burden of CP on health care services.

The use of routinely-collected administrative data has been widely promoted[13,14] over the past decade by funders and policy makers on the basis that it can provide a low-cost source of pre-existing, large population data, spanning long periods of time. However, as these data are primarily collected for reimbursement and commissioning purposes, they are potentially problematic from a research perspective as the completeness, consistency, and accuracy of variables recorded that are not pertinent to the primary purpose of the dataset are less likely to be rigorously checked by the data provider.[15,16] In the UK, coding of some chronic conditions (including asthma and COPD) is incentivized by the Quality and Outcomes Framework;[17] but CP is not included in this list and the extent to which the condition is coded may be compromised by the fact that the primary reason for each attendance is due to an associated or unrelated illness and the underlying chronic condition may not be recorded.

Studies of chronic and complex conditions such as congenital anomalies,[18] asthma,[19] and COPD[20] have successfully employed UK administrative data to investigate health care service use by linking routinely-collected data from primary and secondary care settings. A growing body of studies in Australia[21–23] have linked a range of administrative health records (including register data and hospital admissions) to explore CP health care questions, and studies from the USA[24,25] have used administrative records to determine CP health care costs. These studies

did not take account of Primary Care data. Whilst pediatricians in the secondary health care sector are most likely to make the initial diagnosis of CP, most children will attend their GP practice at some time and their underlying medical condition should be coded in routinely-collected primary health care datasets. Thus, UK studies that link primary and secondary health care data have the potential to optimize case ascertainment.

In this study, two sets of linked primary, secondary, and mortality administrative data, one from England and one from Wales, were used to determine how CP cases can be ascertained, to estimate the prevalence of the condition, and to understand the potential strengths and limitations of using routine data to evaluate health care utilization for CYP with CP.

Material and methods
Data

Routinely-collected anonymized data for patients aged 0–25 years old between 1 January 2004 and 31 December 2014, were analyzed in two separately linked datasets for England and for Wales. Each linked dataset included health care records (GP, inpatients, and outpatients) from the NHS, and mortality records (death register).

The English dataset comprised GP records held in the Clinical Practice Research Datalink (CPRD), linked to Hospital Episode Statistics inpatient data from Admitted Patient Care, outpatient data from Outpatients Dataset (OPD), and mortality data (including causes of death) from the Office for National Statistics (ONS). The CPRD dataset is a primary care research database of de-identified patient data from a network of GP practices across the UK. Primary care data are linked to a range of other health-related data to provide a longitudinal, representative UK population health care dataset. The linked dataset included more than 10 million registered patients and represents an estimated 5.34% of the population of England.[26,27]

The Welsh dataset was hosted by Secure Anonymized Information Linkage Databank (SAIL). The SAIL databank anonymously record-links routinely-collected data held in health and social care datasets at the Centre for Improvement in Population Health through E-records Research, Swansea University, United Kingdom which was part of the Farr Institute.[28] For each dataset within the SAIL databank, an individual is assigned an Anonymized Linking Field, based on their names, address, or NHS number, which is used to link back to their medical records hosted in the SAIL databank.[29]

The Welsh dataset included 100% of Welsh hospital records and records from approximately 70% of GP practices. This dataset comprised inpatient data from Patient Electronic Dataset Wales, data from Outpatients Dataset Wales, and ONS (known as the Annual District Death Extract or ADDE in SAIL) linked to GP records from the Wales General Practice Dataset.

Datasets were linked primarily by NHS number (although where this was not possible, a combination of alternative patient identifiable fields were used).[30] Linkage was conducted by NHS Digital and CPRD in England, and SAIL in Wales.[31]

The study was approved as part of a wider project[30] under Section 3 of the Health and Social Care Act. Data were requested from CPRD and SAIL, and access to the data was granted following approval of our study by the Independent Scientific Advisory Committee for CPRD data and the Information Governance Review Panel for SAIL data.

Identifying cases of CP

Each source of health care data for these patients was searched for CP codes from 1 January 1979 and 31 December 2014 to identify all potential cases of CP that were aged 0–25 years at any point during the study period of 1 January 2004 to 31 December 2014. Records were searched from 1 January 1979 as a young person born on 1 January 1979 would turn 25 on 1 January 2004 and any children born after that would be younger and therefore included within the study period.

Cases falling under the definition "Cerebral palsy and other paralytic syndromes" as stated in the ICD-10 were included as cases of possible CP; specifically, those coded in hospital and mortality data using ICD-10 codes G80.0–83.3. Whilst ICD-10 G80.0-G80.9 codes are specific to CP, G81.0–83.3 were also included to account for cases that may have been coded by the extent of limb paralysis experienced rather than the disease-specific codes.

An equivalent list of Read codes, v2 5-character for Welsh GP data and v2 7-character for English GP data, were identified (Table S1). Read codes were derived from a previously published list of Read codes diagnostic for CP,[32] and supplemented by subsequently introduced relevant Read codes, to provide 44 possible Read codes for CP. This list was converted to medcodes that are used in CPRD. The list of Read codes for GP data in the Welsh dataset was created by matching Read v2 5-character codes to the ICD 10 codes of interest (G80.0–G83.3),

based upon individual code definitions. This is due to the Welsh dataset using the 5-character Read codes rather than the 7-character code used in CPRD. Combining the CPRD and SAIL code lists resulted in a list of 61 possible Read codes used to identify cases of CP in GP data.

CP cases were flagged in the data if the patient had been coded as previously mentioned at any time in their patient history records before the age of 25 years. The rest of the population who were aged 0–25 years between 2004 and 2014 and did not have a CP diagnosis code anywhere in their health care or mortality data were flagged as "non-CP cases".

Data cleaning and preparation

In both English and Welsh datasets, individuals were assigned an anonymized patient identifier. Records were excluded within CPRD and SAIL if there were multiple hospital admissions on the same date for the same individual, if the event fell outside age range or time period of interest, or the probabilistic matching was below the adequate threshold. The probabilistic matching was deemed acceptable if the patient identifier was unique to a patient and there was an exact match on NHS number, date of birth, and sex with or without postcode. The overall dataset included 2,122,909 CYP for England and 1,636,252 for Wales. Figures reported here differ marginally to those previously reported,[30] as more rigorous data cleaning was subsequently undertaken for this study.

Data analysis
Case ascertainment

The number and proportion of the total cases of possible CP that were identified from each source of health care data (GP, inpatients, outpatients, and mortality) were compared both within each linked dataset and between England and Wales.

Coding consistency

The proportion of 4-character ICD-10 codes used in hospital and mortality datasets for the study population were compared for patients in England and Wales. To explore the consistency of coding, the different combinations of 3-character ICD codes and equivalent Read codes applied to the same individual in hospital and GP data respectively, were recorded. Three-character ICD-10 code group G80 (G80.0–G80.9), G81 (G81.0–G81.9), G82 (G82.0–G82.5), and G83 (G83.0–G83.3) and equivalent Read codes were

applied in this analysis (Table S1). The proportions of patients with CP that had only one 3-character CP code recorded were compared between data sources and countries and these cases were broken down by age group to identify consistency of coding throughout their records. The proportion of cases that were coded as "cerebral palsy" (ie, G80.0–G80.9 or Read v2 equivalents) were compared to the other code groups (hemiplegia [G81.0–G81.9], paraplegia and tetraplegia [G82.0–G82.5], and other paralytic syndromes [G83.0–G83.3]).

Prevalence of CP

The total number of possible CP cases identified within each linked dataset were used to calculate crude period prevalence rates per 1,000 population at risk within each dataset for each nation between 1 January 2004 and 31 December 2014. The denominator used for the English dataset consisted of CPRD's anonymized list of patients who had linkable GP and secondary care data (if any) of an acceptable standard for research purposes, who were aged 0–25 years at any point during the study period (n=2,122,909 CYP). For the Welsh dataset, the denominator included all CYP within SAIL (anonymized) aged 0–25 years and resident in Wales (and satisfying SAIL's quality assurance checks) at any point during the study period (n=1,636,252). To investigate demographic characteristics of possible CP cases in each nation, prevalence estimates were broken down by sex and Index of Multiple Deprivation (IMD)[33] (or WIMD in Wales)[34] quintiles, and 95% CIs were calculated for each. Deprived areas were defined as those within the 4th or 5th quintiles. Estimated prevalence rates were also calculated after exclusion of cases only ever coded as other paralytic conditions (ie,

ICD-10 G81.0–G83.3 codes and their Read v2 code equivalents).

Results

Ascertainment

Applying the ICD-10 and Read v2 codes for CP and other paralytic conditions, a total of 7,113 patients in the English dataset, and 5,218 patients in the Welsh dataset were identified (Figures 1 and 2). For England 65.3% (4,646/7,113) of cases of possible CP were identified in the GP dataset, 67.2% (3,121/4,646) of which were also identified in inpatient data. However, a further 2,410 cases were identified within inpatient data that were not coded in the GP dataset. A similar pattern was seen in the Welsh dataset (GP identified 65.1% [3,396/5,218], 61.0% [2,073] of whom were within the inpatient dataset and a further 1,813 were identified in inpatient data but not in the GP data). Very few cases were identified from outpatient datasets in either country. Most cases within the mortality dataset were also identified within the inpatient dataset (England 93.7% [148/158]; Wales 97.5% [237/243]).

Despite multiple health care attendances, approximately 30% of CYP were coded only once for possible CP (29.5% 2,095/7,113 in England and 32.3% 1,685/5,218 in Wales). Of these cases, 15.8% (England) and 13.7% (Wales) were less than 1 year old when CP was coded.

### Coding practices

In both hospital and mortality data, the most commonly recorded code was G80.9 "Cerebral Palsy, unspecified" (Table 1). Hence, in a large proportion of cases the type of CP was not coded.
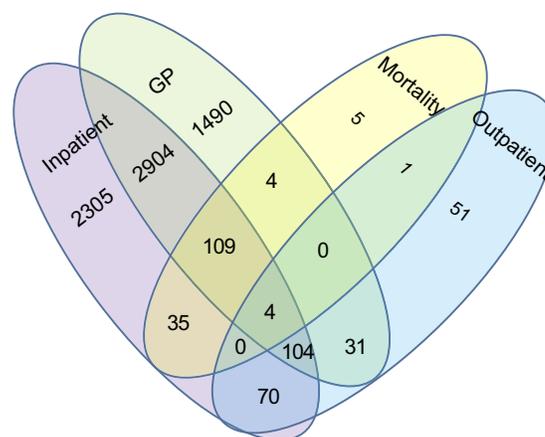
Figure 1 Venn diagram showing the sources of identification of 7,113 CP cases within the English dataset (1979–2014).
Abbreviations: CP, cerebral palsy; GP, General Practitioner.
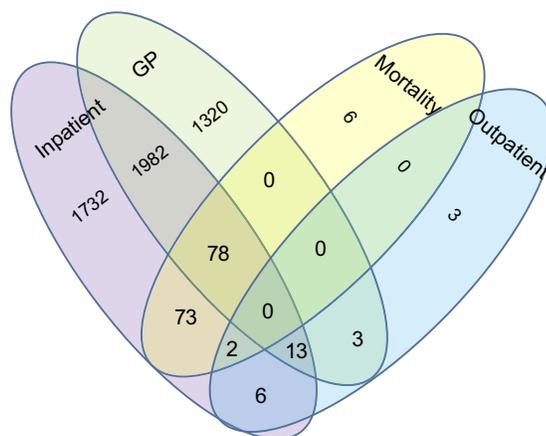
Figure 2 Venn diagram showing the sources of identification of 5,218 CP cases within the Welsh dataset (1979–2014).
Abbreviations: CP, cerebral palsy; GP, General Practitioner.

Overall, 74.5% (5,284/7,113) of cases in England and 75.0% (3,911/5,218) of cases in Wales were specifically identified as "Cerebral palsy" (G80.0–G80.9 codes or an equivalent Read v2 code) at least once (Table 2) during their multiple attendances. Several of these cases were also coded as either "Hemiplegia", "Paraplegia and Tetraplegia", "Other paralytic syndromes" (G81, 82 or 83), or a combination of these. Patterns of multiple coding were similar between the datasets for the two countries and between primary and secondary care. Approximately one in four cases were only ever coded as "other paralytic syndrome" G81–83.

Prevalence

The crude period prevalence for 2004–2014 of CYP (less than 25 years of age) with CP or other paralytic syndromes was 3.4 per 1,000 in England (7,113/2,122,909) and 3.2 per 1,000 (5,218/1,636,252) in Wales. Prevalence was higher in males than females and in the most deprived areas compared to least deprived in both England and Wales (Table 3).

When narrowing case identification to those coded at least once explicitly as CP (ie, G80.0–G80.9 and equivalent Read v2 codes), 5,284 English cases, and 3,911 Welsh cases of CP were identified and period prevalence of CYP with CP reduced to 2.5 per 1,000 (95% CI: 2.4–2.6) for England and 2.4 per 1,000 (95% CI: 2.3–2.5) for Wales.

Discussion

To the best of our knowledge, this is the first study to describe the ascertainment of CYP with CP from data-linkage of routinely-collected primary, secondary care, and mortality data in England and in Wales. We found

that the coding of CP in these records was infrequent, inconsistent, and non-specific. We therefore question the accuracy of CP case identification through coding alone.

It is clear that no one data source provides a comprehensive ascertainment of CP cases. Inclusion of outpatient data added very little to case ascertainment, as the reason for attendance was very rarely coded. Approximately half of the CP cases identified in both English and Welsh inpatient data did not appear with a CP code in the GP data, which suggests a paucity of coding by GPs. Despite the differences in the proportions of data from different sources available for linkage in each country, the pattern of CP case ascertainment remained very similar between England and Wales.

A distinct lack of specificity of codes used across all data sources negates the use of routine data to investigate health care use by type of CP and severity measures are not collected. The most frequently used code in GP data was "Congenital cerebral palsy". Similarly, "Cerebral palsy, unspecified" was most commonly used across secondary care datasets and was particularly heavily used in the mortality data. This lack of precision is possibly because coding is applied to records by administrative staff who never see the patients themselves, and thus rely on the specificity and clarity of the clinicians' notes to code records. Although coding of the primary diagnosis is mandatory,[35] chronic conditions are often contributory factors to an admission and may not always be well-recorded as a secondary diagnosis. Coders in the UK do not receive the same intensity of training as their US counterparts, which may also account for the heavy use of non-specific codes. It has been shown that

Table 1 Percentage of each 4-character code used for the total ICD-10 G80.0–83.3 codes applied to National Health Service contacts in inpatients, outpatients, and mortality data for England and Wales (2004–2014)

ICD-10 code	ICD description	England			Wales		
		Inpatients (%)	Outpatients (%)	Mortality (%)	Inpatients (%)	Outpatients (%)	Mortality (%)
G80.0	Spastic quadriplegic cerebral palsy	7.5	4.2	0.7	8.1	16.7	2.2
G80.1	Spastic diplegic cerebral palsy	7.4	3.6	0.7	8.6	0.0	0.7
G80.2	Spastic hemiplegic cerebral palsy	3.1	0.9	0.0	4.3	0.0	0.0
G80.3	Dyskinetic cerebral palsy	3.1	0.7	1.7	2.7	0.0	0.7
G80.4	Ataxic cerebral palsy	0.3	0.4	0.0	0.5	0.0	0.0
G80.8	Other cerebral palsy	10.2	13.2	1.7	10.3	6.7	2.9
G80.9	Cerebral palsy, unspecified	42.3	69.1	84.3	44.3	53.3	84.7
G81.0	Flaccid hemiplegia	0.1	0.1	0.0	0.1	0.0	0.0
G81.1	Spastic hemiplegia	0.2	0.3	0.0	0.3	0.0	0.0
G81.9	Hemiplegia, unspecified	12.0	6.6	0.0	10.1	20.0	0.0
G82.0	Flaccid paraplegia	0.2	0.0	0.0	0.2	0.0	0.4
G82.1	Spastic paraplegia	0.9	0.1	0.0	0.5	0.0	0.0
G82.2	Paraplegia, unspecified	3.2	0.8	0.7	3.1	0.0	0.0
G82.3	Flaccid tetraplegia	0.1	0.0	0.0	0.0	0.0	0.0
G82.4	Spastic tetraplegia	5.1	0.0	7.6	5.1	3.3	4.0
G82.5	Tetraplegia, unspecified	3.5	0.0	2.7	1.5	0.0	4.4
G83.0	Diplegia of upper limbs	0.1	0.0	0.0	0.1	0.0	0.0
G83.1	Monoplegia of lower limb	0.3	0.0	0.0	0.1	0.0	0.0
G83.2	Monoplegia of upper limb	0.2	0.0	0.0	0.2	0.0	0.0
G83.3	Monoplegia, unspecified	0.1	0.0	0.0	0.0	0.0	0.0
	Total number of codes	**45,391**	*	*	**24,317**	*	*

Note: *Due to disclosure of small numbers that have the potential to identify individual cases we are unable to give the number of codes observed for Outpatient and Mortality data for both England and Wales.

Table 2 Frequency and percentage usage of 3-character ICD-10 codes (or equivalent Read v2 code) per individual CYP with CP in primary and secondary care data

Codes given*	Inpatients and outpatients	General Practitioner
England	**N=5,614**	**N=4,646**
Cerebral palsy (G80)	4,184 (74.5%)	3,524 (75.9%)
• Cerebral palsy (G80) only	3,119 (55.6%)	2,818 (60.7%)
• Cerebral palsy (G80) & Hemiplegia (G81)	500 (8.9%)	407 (8.8%)
• Cerebral palsy (G80) & Paraplegia and Tetraplegia (G82)	497 (8.9%)	257 (5.5%)
• Cerebral palsy (G80) & Other paralytic syndromes (G83)	27 (0.5%)	23 (0.5%)
• Cerebral palsy (G80) and multiple other codes (G81-83)	41 (1.0%)	19 (0.5%)
Other paralytic syndromes (G81–83)	1,430 (25.5%)	1,122 (24.1%)
Wales	**N=3,892**	**N=3,396**
Cerebral palsy (G80)	2,974 (76.4%)	2,520 (74.2%)
• Cerebral palsy (G80) only	2,361 (60.7%)	1,915 (56.5%)
• Cerebral palsy (G80) & Hemiplegia (G81)	289 (7.4%)	400 (11.8%)
• Cerebral palsy (G80) & Paraplegia and Tetraplegia (G82)	281 (7.2%)	168 (5.0%)
• Cerebral palsy (G80) & Other paralytic syndromes (G83)	20 (0.5%)	15 (0.4%)
• Cerebral palsy (G80) and multiple other codes (G81–83)	23 (0.8%)	22 (0.9%)
Other paralytic syndromes (G81-83)	918 (23.6%)	876 (25.8%)

Note: *Table S1 contains equivalent Read codes.

Table 3 Cerebral palsy/other paralytic syndromes (G80.0–G83.3) period prevalence per 1,000 population at risk for 2004–2014 by sex and social deprivation (quintiles of English indices of deprivation 2010 and Welsh index of multiple deprivation [IMD] 2011) within England and Wales. CIs in brackets. NB sex and IMD denominators do not equal the total number of cases in the two datasets due to missing data

Demographic	Prevalence of cerebral palsy or other paralytic conditions (G80–G83)			
	England		Wales	
Sex	**N**	**Rate (95% CI)**	**N**	**Rate (95% CI)**
Male	1,039,850	3.8 (3.7–4.0)	820,532	3.6 (3.5–3.8)
Female	1,083,042	2.9 (2.8–3.0)	815,713	2.7 (2.6–2.9)
IMD				
1 – Least deprived	434,229	3.0 (2.9–3.2)	287,927	3.1 (2.9–3.3)
2	425,447	3.3 (3.1–3.4)	266,765	3.2 (3.0–3.4)
3	404,969	3.3 (3.1–3.5)	273,750	3.4 (3.2–3.6)
4	447,706	3.5 (3.3–3.6)	290,747	3.4 (3.2–3.6)
5 – Most deprived	408,157	3.8 (3.6–3.9)	324,153	3.6 (3.4–3.8)

clinician involvement improves accuracy of coding,[36] thus the imminent NHS adoption of SNOMED CT[37] (a clinical language to be used by clinicians in patient electronic records at the point of contact) has the potential to improve coding in the future. Further, the Community Services Data Set[38] that was introduced from October 2017 should improve the ascertainment of CYP with CP, as these patients are primarily managed by clinicians and therapists within the community setting, which was excluded from routine data collection during the study period.

The fact that almost a third of the cases identified were only ever coded for CP once in their records (despite having multiple NHS contacts) requires further investigation. It is possible that some of these cases are due to erroneous or abandoned diagnoses. Other explanations could include: infrequent coding practice, infrequent service use, or cases of CP of lower severity with associated low morbidity where the condition is not easily apparent or relevant to the clinical consultation. The age range of the child within the study period and study design may also have played a part in this high proportion of single

coded cases. As it was not possible to follow-up the younger CYP in the study period to the age of 25; patients born later in the study period had less chance to be identified as they had fewer years of records to search.

We hypothesized that coding for CP may be inaccurate, and in line with other studies[3,32] we applied a broad selection of codes to include all possible cases of CP (G80–83). It is likely that false positive cases resulted from including G81.0–G83.3 codes in the definition of CP. However, it is clear from our results that cases of CP (coded G80 or Read equivalent during at least one NHS contact) can also be coded at another contact with a G81–G83 code. Hence it is impossible, without more complex algorithms, to confidently separate true cases of CP from other conditions with similar physical symptoms using these data alone. Prevalence estimates for CYP with CP or "other paralytic syndromes" were 3.2 per 1,000 for Wales and 3.4 per 1,000 for England. However, when excluding cases not explicitly coded as CP, estimated prevalence was reduced to 2.4 and 2.5 respectively. This figure exceeds the worldwide estimate of 2.11 per 1,000,[3] and previous studies of UK prevalence using CP register and survey data (2.1 in both South East England,[39] and in Scotland and selected counties in England).[12] However, our lower estimates of prevalence are consistent with that of 2.45 per 1,000 previously published[10] for North East England from regional survey data. All the prevalence rates calculated and reported in this paper are crude rates: no adjustment has been made for age composition or any other demographic or socio-economic factors.

Estimated prevalence rates were higher in males. A finding consistent with previous publications.[40–42] Although measures of deprivation used here are relative within each country, and hence cannot be directly compared between England and Wales, both countries appear to show a trend toward greater CP prevalence in more deprived areas.

There are currently many shortcomings inhibiting the use of routine health care data to quantify and analyze health care utilization of CYP with CP: validation of coding accuracy is strongly recommended to underpin the value of future studies using these data as a research outcome measure. Oskoui et al[43] explored the accuracy of identifying known cases of CP from a regional CP register in a health care administrative dataset, and identified a sensitivity and specificity of the administrative claims data of 65.5% and 99.9% respectively, suggesting that administrative data did not capture the full spectrum of CP. The sensitivity was greater for the more severely affected individuals. We can

hypothesize that this may also be the case in our dataset, but without the ability to analyze cases by severity or validate cases against clinical data or CP register, we cannot be sure.

Other countries with CP registers may be able to continue this work through the validation of codes within routine data against CP register data, however, differences in coding practices across countries, the quality of routinely-collected data, structure of health care services, and varying levels of coder training make it difficult to generalize results such as these across countries.

Conclusion

The use of routinely-collected health care data within England and Wales for the surveillance of CYP with CP presents many challenges. These not only lie in the complexities of creating comprehensive linked datasets, but also in the quality of information currently recorded within those datasets.

Coding of CP in routine primary and secondary care records data needs radical improvement within the UK. The introduction of SNOMED CT and the Children and Young People's Health Services Dataset will hopefully improve case ascertainment in the future, however the introduction of incentives to code for CP would also be prudent to ensure routine data are coded both comprehensively and accurately.

Other information

Study protocols for the wider study are available upon request.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Bax M, Goldstein M, Rosenbaum P, et al. Proposed definition and classification of cerebral palsy, April 2005. *Dev Med Child Neurol*. 2005;47(8):571–576. doi:10.1017/S001216220500112X

2. Krägeloh-Mann I, Cans C. Cerebral palsy update. *Brain Dev*. 2009;31 (7):537–544. doi:10.1016/j.braindev.2009.03.009

3. Oskoui M, Coutinho F, Dykeman J, Jetté N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. *Dev Med Child Neurol*. 2013;55(6):509–519. doi:10.1111/dmcn.12080

4. Graham HK, Rosenbaum P, Paneth N, et al. Cerebral palsy. *Nat Rev Dis Primers*. 2016;2:15082. doi:10.1038/nrdp.2015.82

5. Glinianaia SV, Best KE, Lingam R, Rankin J. Predicting the prevalence of cerebral palsy by severity level in children aged 3 to 15 years across England and Wales by 2020. *Dev Med Child Neurol*. 2017;59 (8):864–870. doi:10.1111/dmcn.13475

6. Westbom L, Hagglund G, Nordmark E. Cerebral palsy in a total population of 4–11 year olds in southern Sweden. Prevalence and distribution according to different CP classification systems. *BMC Pediatr*. 2007;7(1):41. doi:10.1186/1471-2431-7-41

7. Johnson A. Prevalence and characteristics of children with cerebral palsy in Europe. *Dev Med Child Neurol*. 2002;44(9):633–640. doi:10.1017/S0012162201002675

8. Sinha G, Corry P, Subesinghe D, Wild J, Levene MI. Prevalence and type of cerebral palsy in a British ethnic community: the role of consanguinity. *Dev Med Child Neurol*. 1997;39(4):259–262. doi:10.1111/j.1469-8749.1997.tb07422.x

9. Hjern A, Thorngren-Jerneck K. Perinatal complications and socio-economic differences in cerebral palsy in Sweden–a national cohort study. *BMC Pediatr*. 2008;8(1):49. doi:10.1186/1471-2431-8-49

10. Colver A, Gibson M, Hey E, Jarvis S, Mackie P, Richmond S. Increasing rates of cerebral palsy across the severity spectrum in north-east England 1964–1993. *Arch Dis Childhood-Fetal Neonatal Ed*. 2000;83(1):F7–F12. doi:10.1136/fn.83.1.f7

11. Surman G, Newdick H, Johnson A. Cerebral palsy rates among low-birthweight infants fell in the 1990s. *Dev Med Child Neurol*. 2003;45 (7):456–462. doi:10.1111/j.1469-8749.2003.tb00940.x

12. Pharoah P, Cooke T, Johnson M, King R, Mutch L. Epidemiology of cerebral palsy in England and Scotland, 1984–9. *Arch Dis Childhood-Fetal Neonatal Ed*. 1998;79(1):F21–F25. doi:10.1136/fn.79.1.f21

13. Health e-Research Centre. £19m investment into e-health research centres; 2012. https://www.herc.ac.uk/2012/08/02/1178/. Accessed November 9, 2018.

14. Medical Research Council. Research changes lives 2014-2019; 2013. https://mrc.ukri.org/publications/browse/strategic-plan-2014-19/. Accessed November 9, 2018.

15. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *J Public Health (Bangkok)*. 2011;34 (1):138–148. doi:10.1093/pubmed/fdr054

16. Mathur R, Bhaskaran K, Chaturvedi N, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Bangkok)*. 2013;36(4):684–692. doi:10.1093/pubmed/fdt116

17. NHS Digital. Quality Outcomes Framework (QOF). https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/general-practice-data-hub/quality-outcomes-framework-qof. Accessed November 9, 2018.

18. Bishop CF, Small N, Parslow R, Kelly B. Healthcare use for children with complex needs: using routine health data linked to a multiethnic, ongoing birth cohort. *BMJ Open*. 2018;8(3):e018419. doi:10.1136/bmjopen-2017-018419

19. Suruki RY, Daugherty JB, Boudiaf N, Albers FC. The frequency of asthma exacerbations and healthcare utilization in patients with asthma from the UK and USA. *BMC Pulm Med*. 2017;17(1):74. doi:10.1186/s12890-017-0500-9

20. Rothnie KJ, Müllerová H, Thomas SL, et al. Recording of hospitalizations for acute exacerbations of COPD in UK electronic health care records. *Clin Epidemiol*. 2016;8:771–782. doi:10.2147/CLEP.S117867

21. O'Callaghan ME, MacLennan AH, Gibson CS, et al. Epidemiologic associations with cerebral palsy. *Obstetrics Gynecology*. 2011;118 (3):576–582. doi:10.1097/AOG.0b013e31822ad2dc

22. Meehan E, Reid SM, Williams K, et al. Hospital admissions in children with cerebral palsy: a data linkage study. *Dev Med Child Neurol*. 2017;59(5):512–519. doi:10.1111/dmcn.13350

23. Blackmore AM, Bear N, Blair E, et al. Predicting respiratory hospital admissions in young people with cerebral palsy. *Arch Dis Child*. 2018;103:1119–1124. doi:10.1136/archdischild-2017-314346

24. Berry JG, Glader L, Stevenson RD, et al. Associations of coexisting conditions with healthcare spending for children with cerebral palsy. *J Pediatr*. 2018;200:111–117. doi:10.1016/j.jpeds.2018.04.021

25. Kancherla V, Amendah DD, Grosse SD, Yeargin-Allsopp M, Van Naarden Braun K, Van Naarden Braun K. Medical expenditures attributable to cerebral palsy and intellectual disability among med-icaid-enrolled children. *Res Dev Disabil*. 2012;33(3):832–840. doi:10.1016/j.ridd.2011.12.001

26. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827–836. doi:10.1093/ije/dyv098

27. CPRD. https://www.cprd.com. Accessed January 4, 2019.

28. The Farr Institute of Health Informatics Research. http://farrinstitute.org/. Accessed January 4, 2019.

29. Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak*. 2009;9(1):3. doi:10.1186/1472-6947-9-3

30. National Confidential Enquiry into Patient Outcomes and Death. Each and every need: a review of the quality of care provided to patients aged 0–25 years old with chronic neurodisability, using the cerebral palsies as examples of chronic neurodisabling conditions; 2018. www.ncepod.org.uk/2018report1/downloads/EachAndEveryNeed_FullReport.pdf. Accessed November 9, 2018.

31. Jones KH, Ford DV, Jones C, et al. A case study of the Secure Anonymous Information Linkage (SAIL) gateway: a privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform*. 2014;50:196–204. doi:10.1016/j.jbi.2014.01.003

32. Meeraus WH, Petersen I, Gilbert R, Ginsberg SD. Association between antibiotic prescribing in pregnancy and cerebral palsy or epilepsy in children born at term: a cohort study using the health improvement network. *PLoS One*. 2015;10(3):1–14. doi:10.1371/journal.pone.0122034

33. Ministry of Housing, Communities & Local Government. English indices of deprivation 2015; 2015. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015. Accessed November 9, 2018.

34. Welsh Government. Welsh index of multiple deprivation 2014; 2015. http://gov.wales/statistics-and-research/welsh-index-multipledeprivation/?lang=en. Accessed November 9, 2018.

35. Welsh Government. A profile of long-term and chronic conditions in Wales; 2006. http://www.wales.nhs.uk/documents/profile-long_term-chronic-e.pdf. Accessed November 9, 2018.

36. Mahbubani K, Georgiades F, Goh EL, et al. Clinician-directed improvement in the accuracy of hospital clinical coding. *Future Hosp J.* 2018;5(1):47–51. doi:10.7861/futurehosp.5-1-47

37. NHS Digital. https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct. Accessed November 9, 2018.

38. NHS Digital. https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets. Accessed November 15, 2018.

39. Surman G, Bonellie S, Chalmers J, et al. UKCP: a collaborative network of cerebral palsy registers in the United Kingdom. *J Public Health (Bangkok).* 2006;28(2):148–156. doi:10.1093/pubmed/fdi087

40. Bhasin TK, Brocksen S, Avchen RN, Braun KVN. Prevalence of four developmental disabilities among children aged 8 years: metropolitan atlanta developmental disabilities surveillance program, 1996 and 2000. *MMWR. Serveillance Summaries.* 2006;55 (1):1–9.

41. Yeargin-Allsopp M, Braun KVN, Doernberg NS, Benedict RE, Kirby RS, Durkin MS. Prevalence of cerebral palsy in 8-year-old children in three areas of the United States in 2002: a multisite collaboration. *Pediatrics.* 2008;121(3):547–554. doi:10.1542/peds.2007-1270

42. Durkin MS, Maenner MJ, Benedict RE, et al. The role of socio-economic status and perinatal factors in racial disparities in the risk of cerebral palsy. *Dev Med Child Neurol.* 2015;57(9):835–843. doi:10.1111/dmcn.12746

43. Oskoui M, Ng P, Dorais M, et al. Accuracy of administrative claims data for cerebral palsy diagnosis: a retrospective cohort study. *CMAJ Open.* 2017;5(3):E570–E575. doi:10.9778/cmajo.20170013

Supplementary material

Table S1 ICD 10 and corresponding Read v2 codes for cerebral palsy diagnosis

ICD 10 Code	ICD Term	CPRD Read v2 7 Character Code	Read Term	SAIL Read v2 5 Character Code
G80.0–G80.9	**Cerebral Palsy**			
G80	Cerebral Palsy	F23.00	Congenital cerebral palsy	F23.
		F2B.00[a]	Cerebral palsy	F2B.
G80.0	Spastic quadriplegic cerebral palsy	F2B0.00[a]	Spastic quadriplegic cerebral palsy	F2B0.
G80.1	Spastic diplegic cerebral palsy	F230100	Cerebral palsy with spastic diplegia	F2301
		F230111[a]	Spastic diplegic cerebral palsy	"
		F230z00	Congenital diplegia NOS	F230z
		F23y200	Spastic cerebral palsy	F23y2
G80.2	Spastic hemiplegic cerebral palsy	F234.00	Infantile hemiplegia NOS	F234.
		F2B1.00[a]	Spastic hemiplegic cerebral palsy	F2B1.
G80.3	Dyskinetic cerebral palsy	F137.11	Athetoid cerebral palsy	
		F137.12	Athetosis - congenital	
		F137.13[b]	Vogt's disease	
		F137000	Athetoid cerebral palsy	F1370
		F137011	Vogt's disease	"
		F137100	Double athetosis	F1371
		F137111	Congenital athetosis	"
		F23.11	Congenital spastic cerebral palsy	F23.[c]
		F23y300	Dyskinetic cerebral palsy	F23y3
		F23y600[a]	Choreoathetoid cerebral palsy	F23y6
G80.4	Ataxic cerebral palsy	F23y000	Ataxic infantile cerebral palsy	F23y0
G80.8	Other cerebral palsy	F23.12	Infantile cerebral palsy	F23.[c]
		F230.00	Congenital diplegia	F230.
		F230.11	Paraplegia - congenital	"
		F230000	Congenital paraplegia	F2300
		F231.00	Congenital hemiplegia	F231.
		F232.00	Congenital quadriplegia	F232.
		F232.11	Tetraplegia - congenital	"
		F233.00	Congenital monoplegia	F233.
		F233.11	Congenital spastic foot	"
		F23y.00	Other congenital cerebral palsy	F23y.
		F23y100	Flaccid infantile cerebral palsy	F23y1
		F23y400	Ataxic diplegic cerebral palsy	F23y4
		F23y500	Worster-Drought syndrome	F23y5
		F23y511	Congenital suprabulbar paresis	"
		F23yz00	Other infantile cerebral palsy NOS	F23yz
		F2By.00[a]	Other cerebral palsy	F2By.
		Fyu9000	[X]Other infantile cerebral palsy	Fyu90
		G669.00	Cerebral palsy, not congenital or infantile, acute	G669.
G80.9	Cerebral Palsy, unspecified	F23.13	Littles disease	F23.[c]
		F23.14	Cerebral atonia	F23.[c]
		F23z.00	Congenital cerebral palsy NOS	F23z.
		F2Bz.00[a]	Cerebral palsy NOS	F2Bz.

(*Continued*)

Table S1 (Continued).

ICD 10 Code	ICD Term	CPRD Read v2 7 Character Code	Read Term	SAIL Read v2 5 Character Code
G80-83	[X]Cerebral palsy and other paralytic syndromes	Fyu9.00[d]	[X]Cerebral palsy and other paralytic syndromes	Fyu9.
G81.0–G81.9	**Hemiplegia**			
G81.0	Flaccid hemiplegia	F220.00	Flaccid hemiplegia	F220.
G81.1	Spastic hemiplegia	F221.00	Spastic hemiplegia	F221.
G81.9	Hemiplegia unspecified	F22.00	Hemiplegia	F22.
		F222.00	Left hemiplegia	F222.
		F223.00	Right hemiplegia	F223.
		F22z.00	Hemiplegia NOS	F22z.
G82.0–G82.5	**Paraplegia/tetraplegia/ quadriplegia**			
G82.0	Flaccid paraplegia	F241000	Flaccid paraplegia	F2410
G82.1	Spastic paraplegia	F241100	Spastic paraplegia	F2411
G82.2	Paraplegia	F241.00	Paraplegia	F241.
G82.3	Flaccid tetraplegia	F240000	Flaccid tetraplegia	F2400
G82.4	Spastic tetraplegia	F240100	Spastic tetraplegia	F2401
G82.5	Tetraplegia unspecified/quadriplegia NOS	F240.00	Quadriplegia	F240.
		F240.11	Tetraplegia	"
G82–G83	Paraplegia, tetraplegia and other paralytic syndromes	F24.00[e]	Other paralytic symptoms	F24.
G83.0–G83.3	**Monoplegia/diplegia**			
G83.0	Diplegia of upper limb	F242.00	Diplegia of upper limb	F242.
G83.1	Monoplegia of lower limb	F243.00	Monoplegia of lower limb	F243.
G83.2	Monoplegia of upper limb	F244.00	Monplegia of upper limb	F244.
G83.3	Monoplegia unspecified	F245.00	Monoplegia unspecified	F245.

Notes: Clinical Practice Research Datalink (CPRD) uses Read v2 7-character codes to code General Practitioner (GP) data (England); and Secure Anonymized Information Linkage Databank (SAIL) uses Read v2 5-character codes to code GP diagnosis (Wales). "Corresponds to SAIL Read v2 5-character code given in row above. [a]Additional Read v2 7-character codes identified post-Meeraus et al's study period (1990–2010). [b]Read v2 7-character and 5-character code not identified within CPRD or SAIL datasets. [c]SAIL Read v2 5-character code F23. includes the sub codes F23.11, F23.12, F23.13, and F23.14; each of these are given as individual codes for CPRD Read v2 7-character codes. [d]Cerebral palsy and other paralytic syndromes, Read v2 7-character code Fyu9.00 was included as an ICD-10 G80 code. [e]Paraplegia, tetraplegia, and other paralytic syndromes, Read v2 7-character code F24.00 was included as an ICD-10 G82 code.