Completeness and correctness of cerebral palsy diagnoses in two health registers: implications for estimating prevalence

SANDRA JULSEN HOLLUNG 1 | TORSTEIN VIK 2 | ROBERT WIJK 3 | INGER JOHANNE BAKKEN 4 | GURO L ANDERSEN 1

1 The Cerebral Palsy Register of Norway, Vestfold Hospital Trust, Tønsberg; 2 Norwegian University of Science and Technology, Trondheim; 3 The Norwegian Patient Register, The Norwegian Directorate of Health, Trondheim; 4 Norwegian Institute of Public Health, Oslo, Norway.

AIM To assess completeness and correctness of cerebral palsy (CP) diagnoses in the Cerebral Palsy Register of Norway (CPRN) and the Norwegian Patient Register (NPR), and to estimate CP prevalence.

METHOD Among 747 883 Norwegian residents born from 1996 to 2007, 2231 had a diagnosis of CP in the NPR while 1441 were registered in the CPRN. Children registered in the CPRN were considered to have a valid CP diagnosis. For those with a diagnosis of CP only in the NPR, two paediatricians reviewed the hospital records. The prevalence rate of CP with 95% confidence intervals (CI) was calculated on the basis of the combined data sets.

RESULTS One thousand three hundred and ninety-eight children were registered with a diagnosis of CP in both registers, 43 children were only registered in the CPRN, and 824 only in the NPR. The review of hospital records revealed that 464 (59.5%) had CP. Thus, the NPR was 98% complete, and for 86% the diagnosis was correct. The completeness of the CPRN was 76%, while the diagnosis was considered correct for all children (100%). The resulting prevalence of CP was 2.5 (95% CI 2.4–2.7) per 1000.

INTERPRETATION To gain accurate estimates of prevalence rates of CP, it is essential to combine data sources and to validate register data.

The birth prevalence of cerebral palsy (CP) is considered to be a potential indicator of the quality of perinatal care, while population-based prevalence rates provide important information for health care providers and society. In recently published Norwegian studies, the prevalence of CP has varied significantly. A study using information from the Norwegian Social Insurance Scheme reported a birth prevalence of 1.8 per 1000 among individuals born from 1967 to 2002. Surén et al. found a population-based prevalence of 3.0 per 1000 Norwegian residents born from 1999 to 2010, using information extracted from the Norwegian Patient Register (NPR). The NPR is a compulsory national administrative health register, established in 1997. The NPR includes person-identifiable data from 2008 onwards. It contains structured data on all patients treated by the national specialist health services, including individual-level demographic, administrative, and clinical data. In a third study, using information collected by the Cerebral Palsy Register of Norway (CPRN), Andersen et al. found the birth prevalence of CP to be 2.1 per 1000 for children born from 1996 to 1998. The CPRN is a consent-based national medical quality register established in 2006. This register contains clinical data on individual children born from 1996 onwards. Dedicated specialists from each of the 21 habilitation centres record data at three points in time: at diagnosis, and at ages 5 and 15 to 17 years. A paediatrician/paediatric neurologist is responsible for determining the CP diagnosis using the ‘Decision tree for cerebral palsy’ and ‘Classification tree of CP subtypes’ guidelines developed by the Surveillance of Cerebral Palsy in Europe (SCPE). The age recommended for confirmation of the diagnosis is 5 years old. All children with CP in Norway have the right to be diagnosed and treated at one of the habilitation centres. Finally, the CPRN receives summative, anonymized information on the total number of patients with CP per birth year from each habilitation centre. This information is used to estimate the prevalence of CP in Norway.

Variation in the completeness and correctness of data sources used to identify children with CP is most probably a major cause of variability in prevalence. For instance, using information from the Norwegian Social Insurance Scheme will probably underestimate the prevalence, because not all children with CP receive social benefits. Indeed, in a previous study, we found that about 60% of children in the CPRN born from 1996 to 2003 were recorded with CP in the Norwegian Social Insurance Scheme. Furthermore, it is reasonable to assume that the
prevalence reported by the CPRN, which relies upon sum-
mative reports provided by local paediatric habilitation
centres, may be underestimated. Possibly, some CP sub-
types could be more consistently reported to the CPRN
since registration requires an informed consent, increasing
the risk for selection bias. Finally, the prevalence of CP
based upon information in the NPR may be overestimated
because, in regular hospital care, specialists other than pa-
ediatricians/paediatric neurologists might record a CP diag-
nosis code in the hospital record without being aware of
the strict definition of the disorder, or a diagnosis code
may be set on suspicion.

A difference in the population-based prevalence of 1 per
1000 (i.e. 2 per 1000 vs 3 per 1000) represents significant
differences in absolute numbers of people in need of spe-
cial care. In Norway, with a population of 5 million, the
estimated number varies from 10 000 to 15 000 individuals
on the basis of these prevalences. Moreover, imprecise es-
imates of prevalence may lead to inaccurate conclusions
about the assessment of perinatal care, international com-
parisons, and the study of time trends.

Thus, the aim of this study was to assess the complete-
ness and correctness of the CPRN and the NPR, and to
use the combined information to obtain an accurate esti-
mate of the prevalence of CP.

METHOD
Study population and design
This register-based study included 747 883 Norwegian
residents born from 1996 to 2007 and 699 927 live births
in Norway during the same years. In all, 2231 children had
a main or secondary G80.0 to G80.9 (G80) diagnosis code
from the International Statistical Classification of Diseases
and Related Health Problems 10th revision (ICD-10) of
‘Cerebral palsy’ in the NPR, while 1441 children were
recorded with an ICD-10 G80 and SCPE CP subtype7 in
the CPRN. The registers were linked using the 11-digit
personal identification number unique to each Norwegian
resident. The diagnosis of CP was considered correct for
the 1398 children in both registers, and for the 43 children
only in the CPRN. This decision was based upon the
detailed information recorded in the CPRN. A total of 824
children with a CP diagnosis code in the NPR, but not in
the CPRN, were eligible for hospital record review.

Assessment of the correctness of the NPR
To validate the CP diagnosis codes of the 824 children
only in the NPR, hospital records were reviewed by two
experienced paediatric neurologists according to a prede-
terminated standardized registration form. Nineteen children
from rural hospitals were excluded for practical reasons,
and 25 children were not found in the hospital records.
Thus, 780 (95%) hospital records were reviewed. If the
diagnosis of CP was assessed as correct, the reviewer
recorded the appropriate ICD-10 G80 code, and the date
when a paediatrician had confirmed the diagnosis. If the
diagnosis code was assessed as incorrect, the reviewer
noted the most likely correct ICD-10 code, whether the child had suffered a postneonatal (≥28d after birth) brain
trauma, and, if so, at what age.

Statistical analysis
In line with Hogan and Wagner’s description of the valid-
ity of health registers,11 we defined register completeness
as the proportion of children with a true CP diagnosis
code in the register, according to the combined and valid-
dated data set (i.e. equivalent to sensitivity in studies of
diagnostic tests). Register correctness was defined as the
proportion of children with a CP diagnosis code in the
register that were true cases of CP, according to the same
data set (i.e. equivalent to positive predictive value in stud-
ies of diagnostic tests). The two terms are illustrated in
Table I. The use of both completeness and correctness is
necessary to provide an accurate measure of data validity in
a register.11

The reliability of ICD-10 G80 codes only in the NPR
was evaluated by comparing them with the classification of
subtypes determined by the reviewers using Cohen’s un-
weighted kappa, where a kappa value of 1.0 indicates
complete agreement. Kappa values were interpreted as less
than 0.40 indicating poor, 0.40 to 0.75 intermediate to
good, and greater than 0.75 excellent agreement.12

To assess selection bias in the CPRN, χ² statistics were
used to compare the different CP subtypes among children
with CP only in the NPR with the proportion of children

Table I: Calculation of data accuracy in a health register using complete-
ness and correctness

<table>
<thead>
<tr>
<th>True health status in combined and validated register data seta</th>
<th>CP</th>
<th>No CP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration present</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Registration absent</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>a+b+c+b+d</td>
</tr>
</tbody>
</table>

Completeness = a / (a+c)
Correctness = a / (a+b)

Using the combined and validated data set, completeness was
used to calculate the proportion of children with cerebral palsy (CP)
that should have been registered were present in the register, and
correctness was used to assess the proportion of children present
in the register that were regarded as true cases of cerebral palsy.
aThe combined and validated data set includes all children regis-
tered in the CPRN, and children registered with a CP diagnosis
code only in the NPR.
registered in the CPRN. A p-value below 0.050 was considered statistically significant.

The population-based prevalence of CP was calculated by dividing the number of children having a confirmed diagnosis by the number of children residing in Norway.13 Birth prevalence was calculated by subtracting the number of children either born abroad or with a postneonatal cause from the number of children with a confirmed diagnosis, divided by the number of live births.14 Ninety-five percent confidence intervals (95% CI) were calculated according to Altman et al.15

Statistics Norway13 provided population data, and the Medical Birth Registry of Norway14 provided live birth data.

Statistical analyses were performed using VassarStats (www.vassarstats.net; Poughkeepsie, NY).

**Ethics**

The validation of the CP diagnosis codes in the NPR was conducted under Norwegian Patient Register Regulation §2–4, and did not require patient consent. The NPR linked the two registers, and the CPRN only had access to anonymized, aggregated results. The CPRN is approved by The Norwegian Directorate of Health and The Norwegian Data Protection Authority (08/01067-9/EOL). This study was covered by the CPRN consent form, and did not require specific ethical approval.

**RESULTS**

**Correctness of the NPR**

Review of the 780 hospital records of children only in the NPR revealed that 464 (60%) had a correct CP diagnosis code, whereas 302 (39%) did not have CP and 14 (2%) could not be classified.16 In 412 (89%) of the 464 children with a correct CP diagnosis, the code had been determined by a paediatrician/paediatric neurologist. Adding the 464 children with a confirmed diagnosis to the number of children in both registers (n=1398) and to the number of children only in the CPRN (n=43) resulted in 1905 children with CP. Of the 2231 children with a diagnosis of CP in the NPR, 302 did not have CP, suggesting a correctness of 86%. Regarding CP subtypes, there was acceptable agreement between the ICD-10 codes only in the NPR, and the subtypes classified by the reviewers (κ=0.75) (Table SI, online supporting information).

For the 302 children with an incorrect CP diagnosis code in the NPR, the most common ICD-10 diagnoses were epilepsy (G40), specific developmental disorder of motor function (F82), unspecified mental retardation (F79), and other disorders of brain (G93) (Table SI, online supporting information). Furthermore, the reviewers noted that 43 (14%) children with incorrect CP diagnosis codes were recorded with a postneonatal cause, acquired at a mean age of 5 years 2 months.

**Correctness of the CPRN**

The detailed information provided by paediatricians working in habilitation centres indicates that a diagnosis of CP in the CPRN is correct. However, from time to time, the CP subtype may be revised, even in children who are more than 5 years old. Also, in a very few cases, if a diagnosis of CP is later considered to be incorrect, the information for this child is removed from the register. Thus, at any given point in time, the register may contain a few cases of incorrectly diagnosed CP. However, for all practical purposes, the correctness of the CPRN is considered to be 100%. Regarding potential selection bias, Table II shows that the distribution of CP subtypes did not differ between the 464 children assessed by the reviewers to have CP only in the NPR and those recorded in the CPRN.

**Completeness of the two registers**

Only 43 of the 1905 children with CP were not in the NPR with this diagnosis, indicating a completeness of 98% for the NPR. Among these 43 children, over 60% were born from 1996 to 2001, had unilateral CP, and were registered in the CPRN before 2008, which was the first year NPR held individual-level data. One of the 43 children was deceased before 2008.

In the CPRN, 1441 children were registered with detailed information, suggesting a completeness of 76%. Yet, there was a steady increase in completeness from 61% in 1996 to 1998, to 91% in 2006 to 2007.

**Implications for estimating prevalence**

A total of 1905 children had a confirmed diagnosis of CP as of 1 January 2013, corresponding to a population-based prevalence of 2.5 (95% CI 2.4–2.7) per 1000 among Norwegian residents born from 1996 to 2007. Relying upon the information provided only by the NPR, the prevalence rate would have been 3.0 (95% CI 2.9–3.1) per 1000

<table>
<thead>
<tr>
<th>SCPE*</th>
<th>ICD-10</th>
<th>Only in NPR</th>
<th>CPRN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Spastic</td>
<td>Unilateral</td>
<td>G80.2 hemiplegic</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>Bilateral</td>
<td>G80.1 diplegic</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>G80.0 quadriplegic</td>
<td></td>
<td>82</td>
</tr>
<tr>
<td>Dyskinetic</td>
<td>G80.3 dystonic</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>G80.3 athetoid</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Ataxic</td>
<td>G80.4 ataxic</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>Other</td>
<td>G80.8 other</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>G80.9 unspecified</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>464</td>
</tr>
</tbody>
</table>

* p=0.245 for comparison of the distributions between the SCPE subtypes (i.e. spastic unilateral, spastic bilateral, dyskinetic, ataxic, and other) only in the NPR with the proportions in the CPRN.
residents, whereas relying only upon the summative information in the CPRN (n=1679), the corresponding prevalence would have been 2.2 (95% CI 2.1–2.4).

Owing to the administrative nature of the NPR, data from this registry could not be used to calculate the birth prevalence of CP. Nor were the summative reports collected by the CPRN useful for this purpose. Using only detailed information on children registered in the CPRN (n=118 born abroad; n=78 postneonatal cause) and the number of children with a confirmed diagnosis of CP in this study, the corresponding birth prevalence was 2.4 (95% CI 2.3–2.6) per 1000 among the 699 927 live births in Norway from 1996 to 2007.

DISCUSSION
In this study, we found that almost all children with CP were registered with a G80 diagnosis in the NPR, suggesting nearly 100% completeness. Yet, 10% to 15% of the children in this register with this diagnosis code were incorrect. While completeness of the CPRN was under 80% for the entire study period, it reached 90% for children born in the last two years of the birth cohort. In addition, the CPRN was considered 100% correct, and the results did not suggest selection bias of specific CP subtypes. Finally, the results show the importance of combining data sources to obtain more accurate prevalence estimates.

Strengths and limitations
A strength of this study was that it was population-based, covering children with a diagnosis of CP in two national health registers. Moreover, a paediatrician/paediatric neurologist confirmed the CP diagnoses. This allowed us to gain an understanding of the mechanisms behind correctly and incorrectly classified CP diagnoses. Combining information from the two registers made it more likely that all Norwegian children with CP were included, although it cannot be completely excluded that some children with mild CP may not have been recorded in either register. The lower completeness of the CPRN has been explained by work overload among paediatricians.17 Indeed, 98% of parents invited provided informed consent.5 It is thus reassuring that the completeness of the CPRN increased significantly during later years, and that the comparison of CP subtypes does not indicate selection bias. Regarding the NPR, our results suggest that some of the older children (i.e. born from 1996 to 2001) with mild CP were not registered with a diagnosis of CP. This was probably because the register did not include person-identifiable data before 2008. It seems likely that the completeness of CP diagnosis codes in the NPR will approach 100% in the future. For the correctness of the diagnosis in the CPRN, there is also a theoretical, albeit small, possibility that the diagnoses in some cases were revised, even when the children were more than 5 years old.

Comparison with other studies
The prevalence rates and the distribution of CP subtypes reported in this study are similar to those reported by other CP registers. Although this is the first Scandinavian study to cover an entire country, similar studies based on CP registers in Denmark and Sweden have been performed for smaller geographical areas. These studies also identified children with CP by using information from national patient registers, followed by medical record review.18,19 The Cerebral Palsy Registry in eastern Denmark reported a birth prevalence of 2.1 per 1000 for children born from 1995 to 1998.20 Using multiple sources, a study from southern Sweden reported a population-based prevalence of 2.7 per 1000 residents and a birth prevalence of 2.4 per 1000 live births for children born from 1990 to 1997.17 Furthermore, both registers reported a similar distribution of CP subtypes as in our study. The National Surveillance of Cerebral Palsy in Portugal reported a birth prevalence of 1.9 per 1000 live births in 2006, 1.4 to 1.7 for birth years 2007 to 2010, and less than 1 for later years, with 84% spasticity (Cadete A, personal communication 2015). In a study in the USA, Maenner et al. estimated the prevalence of CP by comparing two different surveys based on parental reporting to identify children with CP.21 They reported population-based prevalence rates of 2.6 and 2.9 per 1000 for 2- to 17-year-olds living in the USA from 2011 to 2013. Lastly, compared with Maenner et al., Kirby et al. reported a slightly higher prevalence of CP of 3.3 per 1000 among 8-year-olds in four US areas in 2006.22 However, they reported that 81% had spastic CP, which is similar to CP subtype rates in Scandinavia and Portugal.

Interpretation of results
The high completeness of the NPR is reasonable, since it is recommended in Norway that children with CP should be seen regularly by specialized health care services. Because specialists other than paediatricians are allowed to record the diagnosis, it is not surprising that some children are incorrectly registered with CP in the NPR. In addition, diagnoses registered in the NPR will not be changed if they are disproved later. Taking these factors into consideration, the proportion of children with an incorrect CP diagnosis code in the NPR seems acceptable.

In the CPRN, the diagnosis can be considered to be correct in close to 100% of the cases, because it is based on strict criteria and is confirmed by a paediatrician/paediatric neurologist when the children are 5 years old. Nonetheless, it is possible that at any point in time there are a few children with an erroneous CP diagnosis in the CPRN. Our experience suggests that this misclassification is unlikely to be present in more than two cases per birth year, and these cases are removed from the CPRN. Thus, for all practical purposes, we consider the diagnosis of CP to be correct in this register. In contrast, the completeness in this register was low for the total birth cohort. We have earlier argued that we consider selection bias of specific CP subtypes in the register to be less likely. This is supported by the findings in the present study. The improvement in completeness during the later years may have two causes. First, it may take time before data submission to a
The following additional material may be found online:

**Table SI**: Cross tabulation of the classification of International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) G80 codes of children registered only in the Norwegian Patient Register (NPR) and assessed not to have CP by two paediatric neurologists during a hospital record review.

**Table SII**: Top 10 International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) diagnosis codes for the 302 children registered only in the Norwegian Patient Register (NPR) and assessed not to have CP by two paediatric neurologists during a hospital record review.

**ACKNOWLEDGEMENTS**

The Norwegian Advisory Unit for Medical Quality Registries funded the connection of the two registers, and the Vestfold Hospital Trust funded the hospital record review. We thank paediatric neurologists Ivar Haagaas and Finn Sommer, who performed the hospital record review, as well as our colleagues at the habilitation centres who provided clinical data to the CPRN. The authors have stated that they had no interests that might be perceived as posing a conflict or bias.

**SUPPORTING INFORMATION**

The following additional material may be found online:

**Table SI**: Cross tabulation of the classification of International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) G80 codes of children registered only in the Norwegian Patient Register (NPR) and classified by two paediatric neurologists during a hospital record review.

**Table SII**: Top 10 International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) diagnosis codes for the 302 children registered only in the Norwegian Patient Register (NPR) and assessed not to have CP by two paediatric neurologists during a hospital record review.