

# Validity and reliability of the guidelines of the Surveillance of Cerebral Palsy in Europe for the classification of cerebral palsy

Mary Gainsborough\* MRCPCH, United Bristol Healthcare Trust, Bristol;

Geraldine Surman BSc MSc, 4 Child, National Perinatal Epidemiology Unit, University of Oxford, UK;

Giovanna Maestri MD, Registre des handicaps de l'Enfant et Observatoire Perinatal, Grenoble, France;

Allan Colver MD FRCPCH, Northumbria Healthcare NHS Trust and Newcastle University, UK;

Christine Cans MD PhD, University Hospital, Grenoble, France, on behalf of the Surveillance of Cerebral Palsy in Europe collaborative group.

*\*Correspondence to first author at 4 Child, National Perinatal Epidemiology Unit, University of Oxford, Old Road Campus, Headington, Oxford OX3 7LF, UK.  
E-mail: m.gainsborough@doctors.org.uk*

DOI: 10.1111/j.1469-8749.2008.03141.x

The validity and reliability of the guidelines of the Surveillance of Cerebral Palsy in Europe (SCPE) for the classification of cerebral palsy (CP) were tested by administering 10 written case vignettes via an interactive web-based link to 30 SCPE partners. There was a moderately good level of agreement ( $\kappa=0.59$ ) about inclusion as a CP case on the SCPE database. Classification by CP subtype differed in two main areas: assigning spastic versus dyskinetic and judgement of distribution of spastic involvement. Agreement on Gross Motor Function Classification System (GMFCS) level was less good than reported in previous studies. Twenty respondents repeated the test 5 months later and there was good repeatability for case inclusion ( $\kappa=0.72$ ) but considerable variation in assignment of CP subtype and GMFCS level. There is a need for further collaborative work and training to improve harmonization of the classification of CP, including examination, application of SCPE guidelines, and register coding.

In 2000 the Surveillance of Cerebral Palsy in Europe (SCPE) reported the collaboration between cerebral palsy (CP) surveys and registers in 14 centres in eight countries.<sup>1</sup> Consensus was reached on definition, inclusion/exclusion criteria, and classification of CP (see Appendix Ia and b, supporting information published online). The subtypes of CP are applied in a hierarchical manner using the predominant clinical features in the following categories: spastic (bilateral, unilateral), dyskinetic (dystonic, choreoathetoid, unclassifiable), ataxic, and unclassifiable. The purpose of the collaboration was to improve harmonization of data for monitoring rates of CP and to provide a framework for research and service planning. Subsequently, a central database was established and the prevalence of CP and characteristics of more than 6000 children with CP have been reported.<sup>2</sup> However, the clinical validity and reliability of the guidelines have not been adequately assessed.

The interobserver reliability of the SCPE guidelines and algorithm was evaluated in the Netherlands.<sup>3</sup> Medical records of 57 children registered as having CP at a rehabilitation centre were reviewed by two medical students. Each case was classified as CP or non-CP with good agreement. The reviewers also had fair agreement about type of motor disorder and topographical distribution. The limitations were the small number of observers and the use of medical students who are not generally involved in such decisions in clinical practice. However, the study showed the potential for the SCPE guidelines to assist classification.

In order for SCPE data to be reliable the following criteria must be achieved: (1) those examining children must make reproducible observations, (2) examiners must record their observations accurately and systematically, (3) coders in each register must interpret the observations accurately, (4) coders must apply the guidelines precisely, and (5) the SCPE guidelines must be unambiguous.

For registers where the examiner also does the coding, point (3) does not apply.

This study aimed to test the reliability of points (3) to (5). It was decided to use written case vignettes so that all participants across the collaboration could be given the same information. Therefore, the study tests the coders' ability to extract and interpret clinical information, and their ability to apply the guidelines.

## METHOD

Patient records were the basis for 10 written case vignettes, each of 300 to 400 words written in English. An example is given in Appendix II (supporting information published online). The vignettes included details of birth and early developmental history, surgery or procedures, neurology as described in the clinical records, and functional ability at age 5 years. This reflects the sort of detail provided to local registers, although some use a standardized form to collect information.<sup>4</sup> The cases were chosen in order to include a variety of aetiologies and clinical findings. The cases included a child born in a different country and a postneonatal case (after a near-drowning episode at age 26mo).

None of the respondents was involved in the preparation of the case histories. Respondents were not informed that the task would be repeated and there was no facility to print

or review their original answers. The number of cases was chosen so that the task took less than 1 hour to complete, and most of the results are descriptive.

The case histories were presented as an interactive task on the SCPE website in March 2003. The task was repeated in July 2003 to the same respondents.

Respondents were asked: (1) to use the SCPE algorithm and indicate whether they would include each child as a CP case on the SCPE database (i.e. does the child have a clinical pattern fitting the criteria for CP at age 5 years and do they meet the other criteria?); (2) to classify CP type according to the SCPE algorithm (a link was provided); (3) to assess function using the Gross Motor Function Classification System (GMFCS)<sup>5</sup> at age 5 years (a link was provided); and (4) to enter comments about each case.

#### STATISTICAL ANALYSIS

Results were automatically collated in an Excel database and statistical analysis was performed using SPSS (version 11.5). Agreement was expressed as the highest percentage of respondents selecting the same category. Intraobserver agreement was expressed as the percentage selecting the same response on the second occasion.

Chance-corrected agreement was measured using Cohen's kappa statistic ( $\kappa$ ) and categorized as follows: a negative value is worse than chance; 0 to 0.20 poor; 0.21 to 0.40 fair; 0.41 to 0.60 moderate; 0.61 to 0.80 substantial; and more than 0.80 excellent. However, the kappa results could only reflect the interobserver agreement with the selected cases, and applying the results more generally would need to be treated with caution. However, with multiple raters and options it was not appropriate to calculate a kappa statistic for classification of CP subtype.

#### Results

Thirty respondents undertook the first exercise and 20 completed the repeat exercise. Eight countries and all 14 centres were represented with a mean of 2.1 respondents per centre (range 1–5). Respondents' professions are shown in Table I. Six respondents gave answers in only nine out of the 10 cases.

#### INTEROBSERVER RELIABILITY

##### Case conclusion

There was a moderately good level of agreement on whether to include a child as a CP case with a  $\kappa=0.59$  and mean agreement 89.3% (range 80–100%). One case with a poorer level of agreement was a child who had an acquired injury at 26 months; another case with a poorer level of agreement was a child born in a different country. Excluding these two cases, the mean agreement was 94.6% (range 80–100%) and  $\kappa=0.59$ .

Although SCPE guidelines suggest inclusion of all post-natal cases, many respondents cited different policies used in individual registers, showing a lack of rigorous application of SCPE guidelines. Other areas of some disagreement were about inclusion of children with a minimal motor disorder or with any suggestion of progression. Some of the features interpreted as progression may have reflected changing abilities with age and increased spasticity. However, the majority excluded the case of a child with a clearly evolving motor disorder and loss of skills.

#### CP subtype

Classification of CP subtypes is shown in Table II. Selection between spastic and dyskinetic subtypes was a particular area of contention.

Concerning the distribution of spastic signs, there was unanimity in labelling child 4, who had a severe motor disorder, as 'spastic bilateral' but in some of the milder cases there was disagreement about whether to classify children with subtle problems on the less affected side as bilateral or unilateral, e.g. cases 1 and 3.

For dyskinetic cases, most respondents decided between 'dyskinetic dystonic' and 'dyskinetic choreoathetotic', but with great variability, reflecting some difficulty isolating and interpreting the predominant features.

#### GMFCS level

Agreement over GMFCS level was very variable, ranging from 48 to 95% (mean 78%) for the most frequently selected level (Table III). There was, however, a broad clustering of responses around two levels in most cases. The best agreement (96%) was in case 5, a very severely affected child. In other children with some independent mobility, respondents differed in their interpretation of function. For example, some regarded use of a gait trainer frame as an example of walking with an assistive mobility device while others did not, which resulted in assigning GMFCS Level IV or Level V respectively. Another child, case 3, who could walk between furniture and used a Kaye walker for longer distances was assigned Level II by some and Level III by others.

#### INTRAOBSERVER RELIABILITY

Intraobserver agreement about inclusion/exclusion as cases of CP was substantial with an overall  $\kappa=0.72$  and mean agreement 90.3% (range 82.4–100%). However, individual kappas varied widely from –0.25 to 1.0.

For CP subtype, intraobserver agreement could only be assessed when the case was included as CP on both occasions. Looking at broad CP subtypes ('spastic bilateral', 'spastic unilateral', 'dyskinetic', 'ataxic', and 'unclassifiable'), there was 61.1 to 100% agreement (mean 74.4%). In some cases, altered responses tended to be between 'spastic bilateral' and 'spastic unilateral' (cases 1, 3, and 6) whereas in other cases the changes were between 'spastic bilateral' and 'dyskinetic' subtypes (cases 7 and 8). Within the 'dyskinetic' subtype, there were substantial changes between 'dyskinetic dystonic', 'dyskinetic choreoathetotic' and 'dyskinetic unclassifiable' in 7 out of 16 pairs of responses to case 8.

Intraobserver agreement for GMFCS level was even less good, ranging from 40 to 100% (mean 69.3%). However, the

**Table I: Professions of respondents**

23 Clinical doctors
Neuropaediatricians 14
Paediatricians 5
General practitioner 1
Rehabilitation doctors 2
Geneticist 1
3 Non-clinical doctors
Epidemiologists 2
Public health doctor 1
Physiotherapists 2
Nurses 2

differences were always by one level only, and were less marked for cases generally thought to have the most functional limitation, i.e. GMFCS Level V.

#### QUALITATIVE COMMENTS

Respondents commented that they had difficulty classifying because of limited information and the inability to examine the child, particularly where neurological features appeared to be mixed, although the guidelines clearly state to classify as 'spastic' in mixed forms, and in general according to 'predominant' signs. Terms such as 'spastic hemiplegia' and 'diplegia' persisted in the comments although SCPE use 'bilateral' and 'unilateral' for distribution of spasticity. Some respondents were not happy to record 'bilateral' involvement if the sides were not equally affected.

#### Discussion

The SCPE collaboration and website provided a unique platform to conduct a study of this kind at a small cost with many participants. Even though the number of cases was small in order to keep the task manageable for respondents, the number of participants completing the second round reduced.

The study found an encouraging level of agreement in assessing CP 'caseness' for use in epidemiological registers but more variability regarding description of CP subtype and GMFCS level.

The participants, all partners in SCPE, were familiar with and in agreement with the SCPE guidelines and algorithm. Respondents would have varied in their familiarity with the guidelines which, although straightforward to apply, require change to previous practice. In addition, the study was set up to test classification for entry onto CP population-based registers rather than for individual clinical description. This may

have lead to different use of categories from clinical practice. However, the study is one of very few comparing different practitioners' use of the label 'CP' and its subtypes.

As the cases were presented as text descriptions, they could be presented to many participants simultaneously. Interpretation of neurological terms and limited information reflects real practice for those centres where cases are assessed from forms completed by local clinicians or entries made in case notes. Interpretation may have been further complicated by the use of English language only. An alternative means of testing would have been through video recordings, although these would still require substantial accompanying text to give background details and to explain signs that are felt rather than visualized, such as tone, power, and dynamic catch. Web-based video material would also have presented problems over consent and child protection. Even direct examination of children presents difficulties with standardization, as children get tired, cooperation levels vary, and clinical signs fluctuate.

The SCPE classification of subtype requires an assessment of the predominant motor disorder. There is not scope to record other coexisting neurological features, although in practice many children present with mixed features. Rosenbaum et al.<sup>6</sup> proposed that children continue to be classified by the predominant type of tone or movement abnormality, (categorized as spasticity, dystonia, choreoathetosis, or ataxia), but that any additional tone or movement abnormalities present should be listed as secondary types, as well as recording the anatomical distribution of features. In this study, some of the disagreement between respondents in assigning subtypes appears to have arisen because of different opinions about which motor features were predominant.

**Table II: Classification of cerebral palsy subtype by respondents**

	Case No.									
	1	2	3	4	5	6	7	8	9	10
CP subtype										
Spastic bilateral	12	26	24	29	21	4	16	3		5
Spastic unilateral	16		3			19				
Dyskinetic dystonic		2	1			1	10	9		
Dyskinetic choreoathetotic							1	14		
Dyskinetic nonclassifiable					2		1	3	1	
Ataxic					1					
Non-classifiable							1		2	1
<b>Total respondents</b>	<b>28</b>	<b>28</b>	<b>28</b>	<b>29</b>	<b>24</b>	<b>24</b>	<b>29</b>	<b>29</b>	<b>3</b>	<b>6</b>

**Table III: Gross Motor Function Classification System (GMFCS) levels classified by respondents**

GMFCS	Case No.							
	1	2	3	4	5	6	7	8
Level I	22							
Level II	4	22	8					
Level III		4	18	5		2		
Level IV			1	23	1	17	2	7
Level V					22	4	25	20
Not done		1					1	
<b>Total respondents</b>	<b>26</b>	<b>27</b>	<b>27</b>	<b>28</b>	<b>23</b>	<b>23</b>	<b>28</b>	<b>27</b>

Analysis of respondents by profession presents difficulties, as the numbers vary from many neuropaediatricians to fewer physiotherapists and public health doctors. No particular pattern was apparent. However, there was a trend towards particular subtypes being used more frequently by particular individuals, suggesting that individual leanings are more influential than profession as such. These leanings may be affected by exposure to mentors, literature, training material, and specific clinical experience.

The inclusion of 'born out-of-area' and postneonatal cases complicates the presentation of our results. In retrospect it might have been better to test the response to these issues separately from the clinical classification of the cases involved.

In this study, the GMFCS was applied based on limited information; however, the case vignettes were prepared with a view to a GMFCS level being applied and key pieces of information about the child's motor function at age 5 years were included. The poor interobserver agreement may reflect lack of sufficient detail or a tendency of respondents to interpret pieces of information differently. Some respondents had not used the GMFCS before, although training has not been shown to influence reliability.<sup>7</sup> Time constraints may have led to respondents focusing on walking alone rather than sitting ability, transferring, managing stairs, mode of transport used for longer distances, etc. If the GMFCS is to be applied retrospectively by registers, it is essential to collect the necessary clinical information. Reliability ratings are greater if the GMFCS is applied by professionals who see the child regularly and performed by direct observation.<sup>7</sup> However, good reliability has been shown when applying the GMFCS from written records<sup>8</sup> but this was with just two observers. It would be preferable for registers to ensure that the clinician responsible for the child applies the GMFCS directly.

Intraobserver variability for GMFCS level was even greater than interobserver variability. This may reflect practice, exposure to literature regarding the GMFCS in the intervening period, and reduced time spent on the task the second time round.

In summary, the main reasons for inter- and intraobserver variation in this study appear to be: (1) differences between and within individuals in interpretation of clinical information; (2) differences in practice between centres, including some adherence to local guidelines instead of strict application of the SCPE guidelines; and (3) limitations of classifying text-based cases.

Further work is needed to promote a shared understanding and use of the words and definitions used to describe the neurological and functional features of CP. An English version of a Reference and Training Manual (R&TM), developed by SCPE, was presented at the Annual Meeting of the European Academy of Childhood Disability in 2003.<sup>9</sup> This is in the format of a video and text interactive CD-ROM. The video material presents children with typical neurological features and different CP subtypes. The R&TM CD-ROM is available in German, French, Spanish, Italian, Dutch, Lithuanian, and Slovenian. Further details are available on the SCPE website ([http://www-rheop.ujf-grenoble.fr/scpe2/site\\_scpe/index.php](http://www-rheop.ujf-grenoble.fr/scpe2/site_scpe/index.php)). A further piece of work has been to devise a single data

collection form to use across registers to strengthen the quality and consistency of data supplied to the database and this is also available on the SCPE website. Incorporation of regular training and audit into individual registers' standard practice would improve adherence to SCPE guidelines.

*Accepted for publication 25th February 2008.*

#### *Acknowledgments*

This work was supported by European Commission funds: DGXII-BIOMED2-Contrat N°BMH4-983701 and DGXII-FP5-Contrat N° QLG5-CT-2001-30133. The authors thank Pat Yudkin for statistical advice, Ann Johnson for comments, and all respondents for their time and cooperation.

SCPE participants: C Cans, PS Jouk (Grenoble, FR), C Arnaud (Toulouse, FR), V McManus, D Coghlan (Cork, IE), J Parkes, H Dolk (Belfast, UK), P Uvebrant, E Beckung (Göteborg, SW), O Hensey, V Dowding (Dublin, IE), S Jarvis, A Colver (Newcastle, UK), J Kurinczuk, G Surman (Oxford, UK), I Krägeloh-Mann, U Petruch (Tübingen, DE), P Pharoah, MJ Platt (Liverpool, UK), M Topp, P Udall, (DIKE, Copenhagen, DK), MG Torrioli, G Schirripa (Rome, IT), M Wichers (Arnhem, NL), G Andersen, A Meberg (Tonsberg, NO), M Bottos, A Colonna (Bologna, IT)

#### *Supporting Information*

The following supporting information is available for this article online:

**Appendix Ia and Ib:** Surveillance of Cerebral Palsy in Europe (SCPE) decision tree for cerebral palsy (CP)

**Appendix II:** Example of case vignette

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author of the article.

#### **References**

1. Surveillance of Cerebral Palsy in Europe. Surveillance of Cerebral Palsy in Europe: a collaboration of cerebral palsy surveys and registers. *Dev Med Child Neurol* 2000; **42**: 816–24.
2. Surveillance of Cerebral Palsy in Europe. Prevalence and characteristics of children with cerebral palsy in Europe. *Dev Med Child Neurol* 2002; **44**: 633–40.
3. Gorter J, Van der Kraats C, Bulten R, Ketelaar M. Interobserver reliability of the Surveillance of Cerebral Palsy in Europe algorithm for classification of cerebral palsy. Poster presented at the annual meeting of the European Academy of Childhood Disability Oslo, Norway 2003.
4. Evans PM, Alberman E. Recording motor defects of children with cerebral palsy. *Dev Med Child Neurol* 1985; **27**: 404–06.
5. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. The Gross Motor Function Classification System for cerebral palsy. *Dev Med Child Neurol* 1997; **39**: 214–23.
6. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M. The Definition and Classification of Cerebral Palsy. *Dev Med Child Neurol* 2007; **49**(Suppl. 109): 8–14.
7. Morris C, Kurinczuk JJ, Fitzpatrick R, Rosenbaum PL. Who best to make the assessment? Professionals' and families' classifications of gross motor function in cerebral palsy are highly consistent. *Arch Dis Child* 2006; **91**: 675–79.
8. Wood E, Rosenbaum P. The Gross Motor Function Classification System for cerebral palsy: a study of reliability and stability over time. *Dev Med Child Neurol* 2000; **42**: 292–96.
9. Krageloh-Mann I, Petruch U, Gainsborough M, Cans C. The Reference and Training Manual (RTM) of the Surveillance of Cerebral Palsy in Europe (SCPE) – a video and text-based interactive CD-ROM. Invited speaker presentation at the annual meeting of the European Academy of Childhood Disability Oslo, Norway 2003.