Development and Evaluation Process of a Pediatric Long-Bone Fracture Classification Proposal

Laurent Audigé[1], James Hunter[2], Annelie M. Weinberg[3], Jay Magidson[4], Theddy Slongo[5]

Abstract

Background and Purpose: A pediatric long-bone fracture classification proposal was developed following the principle of the Müller AO classification long bones for adults and assessed for reliability and accuracy in a series of four pilot agreement studies (classification sessions).

Material and Methods: Six surgeons independently classified 136 radius fractures using digitalized radiographs. Reliability and accuracy were quantified using Kappa and Latent Class Modeling, respectively.

Results: Results from the last two sessions are presented whereby fractures were classified as epiphyseal, metaphyseal or diaphyseal. In session 3, the overall Kappa was 0.78 and pairwise Kappa ranged between 0.74 and 0.95. Surgeons were overall 85.7–99.5% accurate. In session 4, the overall Kappa was 0.98 and pairwise Kappa ranged between 0.95 and 0.99. Surgeons were overall 98.8–99.5% accurate.

Conclusion: The systematic implementation of pilot agreement studies and use of Latent Class Modeling methodology for the development and evaluation of fracture classification systems in orthopedics are encouraged.

Key Words

Fracture classification systems · Orthopedics · Pediatrics · Accuracy · Latent Class Modeling

[1] Manager Methodology, AO Clinical Investigation and Documentation, AO Foundation, Davos Platz, Switzerland,
[2] Consultant Pediatric Orthopedic Surgeon, Queen's Medical Centre, Nottingham, UK,
[3] Department of Pediatric Traumatology, Johannes Gutenberg University, Mainz, Germany,
[4] Director, Statistical Innovations, 375 Concord Avenue, Belmont, Massachusetts, USA,
[5] Consultant Pediatric Orthopedic Surgeon, Department of Pediatric Surgery, University Children's Hospital, Bern, Switzerland.

Received: April 6, 2004; revision accepted: July 13, 2004.

Eur J Trauma 2004;30:248–54
DOI 10.1007/s00068-004-1428-3

Introduction

Fracture classification systems are routinely used in trauma and orthopedics. When classifications are prognostic, they help surgeons make decisions on the most appropriate treatment [1]. They are also useful for documentation, such as registers, and for valid comparison of results between clinical studies. The process of using common classification systems has, however, mostly shown poor interrater reliability [2]; thus, the usefulness of these classifications in practice has been questioned. It is reasonable to require that defined fracture entities can be diagnosed with acceptable reliability and accuracy, before a classification system is promoted for use in practice and research [3–5]. Reliability measures how much repeated ratings of the same fractures agree, while accuracy measures how much ratings correspond with the truth (i.e., cases are correctly classified according to their "true" fracture status).

To classify fractures as accurately as possible in clinical practice, surgeons must perform a thorough analysis of the available information including the patient history, physical examination, diagnostic imaging, and examination under anesthesia at the time of surgery.

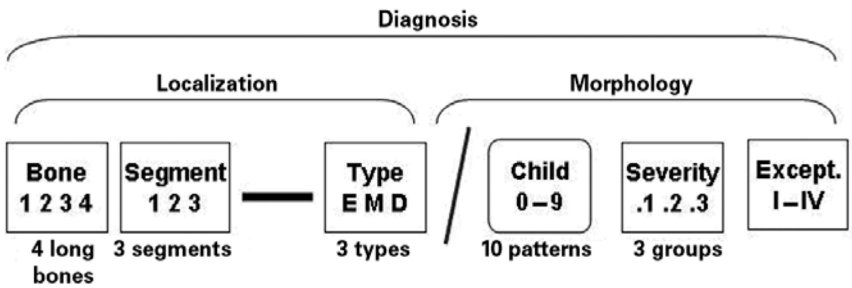

Figure 1. Overall structure of the pediatric long-bone fracture classification.
The current classification proposal is based on the Müller AO classification of fractures–long bones for adults and considers child-specific relevant fracture features. The original unifying principle of the CCF, valid for all fractures, is an anatomic and morphological organization mostly into triads. The anatomy is related to the four long bones and their three segments. It is further described by the fracture type recorded as E, M or D. The morphology of the fracture is documented by a type-specific child code, a severity code as well as an additional code for displacement of specific fractures.

Such analysis is important, since many fracture treatments (e.g., conservative immobilization and minimally invasive surgery) provide no opportunity for direct visualization of fracture sites. This particularly applies to pediatric fractures given that they are mostly treated nonoperatively. Since the truth about fracture patterns cannot be directly recorded, reliability is often assessed as an indirect method to conclude on the validity (i.e., accuracy) of the tested classifications. In the orthopedic literature [2], the Kappa coefficient is commonly used as a "chance-corrected measure of agreement". This coefficient ranges from +1 (complete agreement) through 0 (agreement by chance alone) to < 0 (less agreement than expected by chance). There are several limitations with this methodological approach. Different surgeons may wrongly agree on their diagnosis, i.e., the classification process may be reliable but inaccurate. In addition, the Kappa coefficient is strongly influenced by the distribution of fracture categories within a sample [6–8], and interpretation guidelines remain subjective. For instance, a Kappa coefficient of 0.70 may be interpreted as "substantial" [9], "good" [10], "fair to good" [11] or "fair" [12] agreement depending on the scale used. The meaning attached to these words may vary between individuals. Hence, while using Kappa provides useful initial indicators of reliability, additional or alternative methods are needed to help making better conclusions on the accuracy of a classification process.

Over recent years, analytic methods have been developed to overcome the lack of "gold-standard" diagnostic tools in the evaluation of the characteristics of diagnostic tests [13–15], including a method called Latent Class Modeling [16–18]. Using this latter technique in the evaluation of fracture classifications, the hypothesis is that each fracture belongs to one of several real clinically relevant categories (or classes). However, although these categories can be theoretically defined, they may not be directly observable in practice. In such situation, these classes are said to be "latent" (so the term "latent class"). The analysis aims at identifying the most likely number of these latent classes in the population, given the selected sample of fractures and the agreement data collected among the various raters, such as surgeons. The modeling process assesses how many fracture classes can be reasonably identified in the sample. For each class, the accuracy of classification by each rater is estimated. Each case is allocated to the most probable "latent class", so the analysis provides an estimate of the "true" fracture distribution in the sample.

In this study we explored the use of Latent Class Modeling [16–18] in the context of the evaluation of a proposed pediatric long-bone fracture classification. This is a method for the analysis of classification agreement data to answer the following questions:

(1) How many different classes of fractures are truly present in the study sample?
(2) Are certain individual raters or groups of raters in disagreement with the others?
(3) What are the patterns of disagreements, i.e., do raters disagree with respect to classification of certain classes of fractures or all?
(4) What is the accuracy (proportion of correct classification) by each rater for each class?

Material and Methods

Building on earlier development work on a specific pediatric long-bone fracture classification [19, 20], a revised classification proposal was evaluated using a series of pilot agreement studies. We present here part of the data collected in this context. The current classification proposal is based on the Müller AO classification of fractures–long bones for adults [21] and considers child-specific relevant fracture features (Figure 1). This classification and final results from the last pilot agreement study will be presented in full elsewhere.

However, one of the relevant fracture features is presented here in more detail as an example of our evaluation and analytic approach. Radius fractures were classified as diaphyseal (D), metaphyseal (M) or epiphyseal (E), where M fractures were identified by a square drawn over the anteroposterior radiographic view (Figure 2). This categorization as D, M or E fractures has been proposed earlier [22].

A consecutive series of 136 single radius fractures from pediatric patients (< 16 years old and open bone physis) were collected in 2002 at the University Children's Hospital (Department of Pediatric Surgery) in Bern, Switzerland. Cases were selected irrespective of their treatment or the quality of the radiographs. The size of the series was chosen so that at least ten cases were available for each most common fracture category in the proposed classification. Anteroposterior and lateral standard preoperative radiographs were digitalized and saved in random on a CD for viewing on a personal computer.

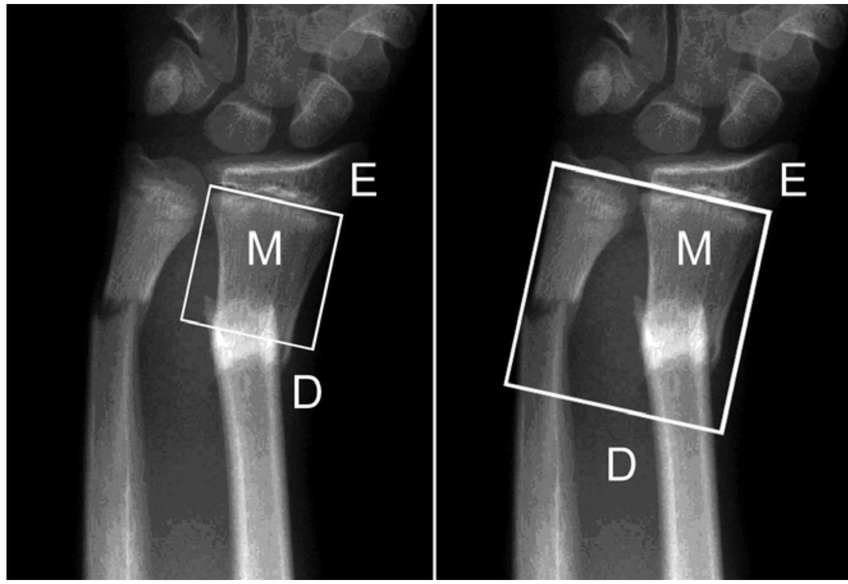

Figure 2. Definitions used for the classification of pediatric radius fractures as D (diaphyseal), M (metaphyseal) and E (epiphyseal) fractures in the context of two consecutive pilot agreement studies.
To distinguish between D, M, and E fractures, a square was applied drawn over the anteroposterior radiographic views (courtesy Th. Slongo). In session 3, the width of the square was that of the radius physis, but this resulted in unreliable and inaccurate classification for some fractures; in the following session, the width was increased to include the ulna as well. This change was considered more clinically relevant and contributed to a dramatic improvement of classification accuracy.

The development and evaluation process involved a series of four pilot agreement studies (called hereafter "classification sessions"), where classification definitions were clarified or modified according to successive agreement results. In this paper, we consider results from the last two classification sessions. In session 3, six pediatric surgeons (called hereafter "raters") from five countries classified fractures, while five pediatric surgeons in one hospital participated in session 4. Only one rater participated in both sessions. The definition used to classify fractures as E, M or D differed between the two sessions, whereby in session 4 the square side length was determined by including the ulna (Figure 2). In addition, in the last session, the classification process was done with the help of a transparency on which a series of colored squares was drawn. Raters were trained prior to the sessions. The classification process was conducted independently by each rater using a specifically developed spreadsheet.

Data from each rater were checked and corrected for inconsistency or data entry error, and then recorded for statistical analysis. For each classification session, a preliminary descriptive analysis was conducted, whereby we identified the most likely distribution of "true" fracture categories in the sample and the percentage of full agreement among raters. A list of the cases for which disagreement occurred was provided to explore reasons for disagreement. Reliability was measured via the overall and category-specific Kappa coefficients, as well as rater pairwise Kappa coefficients using the statistical software Intercooled Stata 8 (Stata Corporation, Texas, USA). The classification accuracy for each category by each rater (percentage of cases correctly classified) was estimated by the technique called "Latent Class Modeling" [18] using the software Latent GOLD® (Statistical Innovations Inc, Belmont, MA, USA) [23, 24].

The most likely distribution of "true" fracture categories in the sample was estimated using two distinct methods. The first estimate was obtained from a consensus classification agreement among the raters. The second estimate was derived from Latent Class Modeling.

Table 1. Distribution of fractures according to fracture type (E/M/D), raw agreement among all raters, overall and category-specific Kappa coefficients in classification sessions 3 and 4. D: diaphysis; E: epiphysis; M: metaphysis.

		Fracture type				Full rater agreement	
		E	M	D	All	n	%
Session 3	Number of fractures	17	51	68	136	94	69
	Kappa	0.74	0.72	0.85	0.78		
Session 4	Number of fractures	15	66	55	136	133	98
	Kappa	0.98	0.98	0.99	0.98		

Table 2. Raters' pairwise Kappa coefficients for classification of pediatric fracture type (E/M/D) in the radius in classification sessions 3 and 4. D: diaphysis; E: epiphysis; M: metaphysis.

	Session 3					Session 4			
Raters	1	2	3	4	5	1	2	3	4
2	0.83					0.99			
3	0.85	0.74				0.99	0.97		
4	0.88	0.79	0.87			0.99	0.97	0.98	
5	0.92	0.83	0.81	0.85		1.00	0.99	0.99	0.99
6	0.95	0.86	0.84	0.87	0.92				
Min.	0.74					0.97			
Max.	0.95					1.00			
Median	0.85					0.99			

Results

The distribution of radial fracture types according to raters' consensus classification is presented in Table 1. Two fractures were classified as E in session 3 (though three and four raters out of six rated them as M, respectively) and as M by all raters in session 4. 13 fractures classified as D by the majority of raters in session 3 were classified as M by all raters in session 4. There was full agreement among raters in 69% of fractures (94/136) in session 3, and 98% of fractures (133/136) in session 4.

The overall Kappa coefficients were 0.78 and 0.98 in session 3 and 4, respectively (Table 1). Category-specific Kappa coefficients ranged 0.72–0.85 between fracture types in session 3, and 0.98–0.99 in session 4. The median of raters' pairwise Kappa coefficients was 0.85 after session 3 (range 0.74–0.95; Table 2). The lowest pairwise Kappa coefficients were between rater 2 and raters 3 and 4 (0.74 and 0.79, respectively), while they were the highest (0.92–0.95) between raters 1, 5 and 6. The median of raters' pairwise Kappa coefficients was 0.99 after session 4 (range 0.97–1.0), which shows an almost perfect agreement between raters.

Latent class analyses results from sessions 3 and 4 are presented in Table 3. The agreement data from session 3 were consistent with three true types of fractures with prevalences of 13%, 39% and 48%, respectively. By considering how raters performed in classifying fractures within each of these three classes, we can make some judgment as which type of fractures these classes truly represent. Raters would classify fractures of class 1 as "E" fractures with probabilities ranging from 58% to 98%. Thus fractures in this class might represent "true" E fractures. Raters 3 and 5 showed misclassification rates of 42% and 25% for these fractures in favor of M fractures. Class 2 fractures might include "true" M fractures since the accuracy of classifying these fractures by raters ranged from 76% to almost 100%. The last class of fractures hence might include "true" D fractures (class 3) with estimated classification accuracies between raters ranging from 80% to 100% (four raters were close to 100%). Rater 2, however, showed a 21% misclassification rate of D fractures. Further analyses (not shown here) showed that raters 3 and 4 behaved similarly in their classification process compared with other raters. Rater 3 showed the lowest accuracy in this classification process. The median (range) of overall classification accuracy between raters was 93% (86–99%). The agreement data from session 4 were consistent with a distribution of fractures in three classes with prevalences of 11% of E fractures, 48% of M fractures and 40% of D fractures in the population of radius fractures. Classification accuracies were around 98–100% for almost all fracture types and raters. Rater 4, the only rater participating in both sessions, overall improved classification accuracy in session 4.

Depending on the profile of responses by the raters, the latent class analysis allocated sampled fractures to one of the identified latent classes, thus estimating the distribution of likely true fracture type in the sample. In our example, this distribution was identical to that obtained by raters' consensus classification.

Discussion

A pediatric long-bone fracture classification proposal was developed using a structured methodological approach [25]. The successive drafts of this classification were assessed by a series of four pilot agreement studies to highlight their weaknesses and identify areas for improvement. For simplicity, we present only results from

Table 3. Accuracy of classification of pediatric fracture type (E/M/D) in the radius in classification sessions 3 and 4: fracture class distribution and probability of correct classification for each fracture class and rater. D: diaphysis; E: epiphysis; M: metaphysis.

Likely fracture type	Session 3 E Class 1	 M Class 2	 D Class 3		Session 4 E Class 1	 M Class 2	 D Class 3
Fracture class distribution[a]	13.1%	38.9%	48.0%		11.2%	48.4%	40.4%
Rater 1				Rater 7			
E	**81.1%**	2.0%	0.1%		**98.1%**	0.1%	0.1%
M	17.9%	**93.8%**	0.2%		1.1%	**99.7%**	0.3%
D	1.0%	4.2%	**99.7%**		0.9%	0.2%	**99.6%**
Rater 2				Rater 8			
E	**86.9%**	5.9%	0.1%		**98.1%**	0.1%	0.1%
M	12.3%	**91.9%**	20.8%		1.1%	**99.8%**	2.1%
D	0.8%	2.2%	**79.2%**		0.9%	0.2%	**97.8%**
Rater 3				Rater 9			
E	**58.0%**	5.9%	0.1%		**98.1%**	0.1%	0.1%
M	41.0%	**76.2%**	0.2%		1.1%	**99.8%**	2.1%
D	1.1%	17.9%	**99.8%**		0.9%	0.2%	**97.8%**
Rater 4				Rater 4[b]			
E	**80.9%**	2.1%	0.1%		**91.5%**	0.1%	0.1%
M	18.0%	**82.0%**	1.8%		7.6%	**99.8%**	0.3%
D	1.0%	15.9%	**98.2%**		0.9%	0.2%	**99.6%**
Rater 5				Rater 10			
E	**75.3%**	0.1%	0.1%		**98.1%**	0.1%	0.1%
M	23.7%	**95.7%**	0.2%		1.1%	**99.7%**	0.3%
D	1.0%	4.2%	**99.7%**		0.9%	0.2%	**99.6%**
Rater 6							
E	**98.3%**	0.1%	0.1%				
M	0.8%	**99.6%**	0.2%				
D	0.9%	0.3%	**99.7%**				

[a] Although we used the same sample of fractures, the estimated distribution changed due to modification of the category definitions between the two classifications sessions
[b] Only one rater participated in both classification sessions. We present results for that rater on the same line for easier comparison between sessions

the last two pilot studies (classification sessions 3 and 4) and illustrate the benefit of such an approach. In our experience, developers of a classification system should conduct such studies early, as they still have great flexibility to make relevant changes either in the category definitions or in the diagnostic method used in the process of classifying fractures. The classification of pediatric fractures as E, M or D is clinically relevant [22], as it distinguishes different fracture entities requiring different management approaches.

The sample of cases used was prospectively collected and thus was representative of common pediatric fractures seen in a large children's hospital. There were at least 15 cases per fracture type in our study, thus allowing detailed analysis and data modeling. When evaluating a classification, it is critical that all important fracture categories are represented in the sample. The expert panel reviewed results from each classification session to make a decision on appropriate changes. Although results from session 3 were very promising, a number of limitations were highlighted. Regarding the coding of fracture type, it was found that (1) one rater applied the rule provided by the Müller AO classification of fractures–long bones for adults [21] regarding the position of the square, (2) the square side was not large enough to include all clinically recognized M fractures, and (3) the strict application of the classification definition was impractical without a measuring device. Some raters did not perform measurements systematically, while others used rulers, either physical or computer-generated. Following these observations, the definition to distinguish between E, M and D fractures was changed (Figure 2), and a transparency with a series of colored predrawn squares was developed for use on the computer screen. Consequently, 13 cases changed categories in a consensus rating from D to M between the two sessions; clinicians agreed that they would handle all these cases as being M fractures, hence justifying the definition change. In session 3, rater 2 applied clinical judgment in classifying twelve of these 13 fractures as M fracture, while the other raters classified them more objectively as D fractures according to the definition. This problem was easily identified through a pilot agreement study.

Classification reliability and accuracy were quantified using the Kappa coefficient as well as an alternative technique called Latent Class Modeling. In most reliability studies of fracture classification published in the orthopedic literature, a statistic reflecting some level of agreement (e.g., the Kappa statistic) was used to conclude on the overall reliability of classification systems [2]. But as mentioned by Kraemer [26], there are problems associated with this approach, in particular that "there is no consensus on the most appropriate measure to use, the interpretation of their magnitude and the type of inferences that can be made from the results". We calculated Kappa coefficients, however, to explore variation in level of agreement between classification categories, and between raters. In session 3, Kappa coefficients differed significantly from one, and varied between pairs of raters. We suspected that rater 2 differed from the other raters, and that raters 1, 5 and 6 were more similar in their classification results compared with the other raters. This warranted more exploration of the reasons for this variation. Results following session 4 showed almost perfect agreement between raters, which likely illustrates a clearer understanding and better use of the definitions. However, Kappa coefficients remained difficult to interpret [2] and an alternative analysis was needed to conclude on the accuracy of the classification proposal.

Latent Class Modeling has proven useful and successful in other medical disciplines for the assessment of the accuracy of diagnostic classifications [17]. This technique has been developed over the last 30 years [27] but to our knowledge has not been applied in the evaluation of fracture classification. The major strengths of this technique are (1) it allows studying the profile of agreement or disagreement between raters; it identifies raters having particular problems (e.g., in session 3, rater 2 showed poorer accuracy in the classification of D radius fractures) and fractures that are particularly difficult to classify; (2) it is a flexible approach that can be adapted to a wide range of classification systems, and can take large numbers of cases and raters into consideration; (3) it provides a quantification of the degree of accuracy, under the assumption that the different latent classes represent the true fracture types; and (4) the interpretation of the output probabilities is straightforward and clinically relevant. Benchmark values can be defined for each classification category in order to judge a classification process as "accurate". This method is based on important assumptions, i.e., that the ratings are obtained independently (raters did not influence each other) and that within each latent class, any observed patterns of agreement (disagreement) are due to chance (statistician refer to this assumption as "local independence"). This last assumption is verified and considered in the analyses.

In this study, there was a perfect agreement between raters' consensus and the fracture allocation obtained via the modeling process. There is, however, a conceptual difference between the two approaches. The raters' consensus describes the composition of the sample used. By exploring reasons for disagreement, raters can propose appropriate modifications in the classification. However, a consensus may not be obtained or raters may not be members of the expert panel. Latent Class Modeling uses the agreement data to describe the population of fractures and allocates sampled fractures to identified classes depending on rating profiles. Fractures with the lowest risk of correct allocation are easily identified for discussion by the expert panel. This is a more practical and flexible approach, as raters' meetings may be spared without invalidating results.

In some circumstances, latent classes may not reflect true categories, but artificial categories occurring because of shared bias among raters. Interpretation of the results depends, therefore, on the assumptions that can be reasonably made about the clinical significance of latent classes. If the category definition or diagnostic images are inappropriate, the identified classes may not reflect "true" fracture types. This probably occurred in session 3 for the distinction between M and D fractures. When in doubt, investigators should consider revising the category definitions as in our example. Alternatively, additional relevant clinical information should be included in the classification process, such as observation during operations and/or postoperative follow-up examinations.

Pilot agreement studies should be systematically conducted in the evaluation process of fracture classifications to quantify classification accuracy. Latent Class Modeling provides results of greater relevance and interpretability in this context. Our example showed that standard radiographic anteroposterior and lateral views provide surgeons with sufficient information to accurately classify fracture type (E, M or D). Classification accuracy estimates of at least 95% can be reached providing surgeons fully understand and correctly apply the definition.

Acknowledgments

The authors are grateful to Dr. Mike John (Martin Luther University Halle/Wittenberg, Germany) and Dr. Beate Hanson (AOCID, AO Foundation, Switzerland) for their fruitful comments regarding the edition of this manuscript.

References

1. Martin JS, Marsh JL. Current classification of fractures. Rationale and utility. Radiol Clin North Am 1997;35:491–506.
2. Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. Acta Orthop Scand 2004;75:184–94.
3. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. New York: Oxford University Press Inc, 1995.

4. Burstein AH. Fracture classification systems: do they work and are they useful? J Bone Joint Surg Am 1993;75:1743–4.
5. Bland JM, Altman DG. Statistics notes: Validating scales and indexes. BMJ 2002;324:606–7.
6. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993;46:423–9.
7. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. J Clin Epidemiol 1996;49:431–4.
8. Cook RJ. Kappa and its dependence on marginal rates. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: Wiley, 1998:2166–8.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
10. Altman DG. Inter-rater agreement. Practical statistics for medical research. London: Chapman & Hall/CRC, 1990:403–9.
11. Fleiss JL. Statistical methods for rates and proportions, 2nd edn. New York: John Wile Sons, 1981:217–8.
12. Svanholm H, Starklint H, Gundersen HJ, et al. Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. APMIS 1989;97:689–98.
13. Garrett ES, Eaton WW, Zeger S. Methods for evaluating the performance of diagnostic tests in the absence of a gold standard: a latent class model approach. Stat Med 2002;21:1289–307.
14. Valenstein PN. Evaluating diagnostic tests with imperfect standards. Am J Clin Pathol 1990;93:252–8.
15. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol 1995;141:263–72.
16. Uebersax JS, Grove WM. Latent class analysis of diagnostic agreement. Stat Med 1990;9:559–72.
17. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. Stat Med 1986;5:21–7.
18. Hagenaars J, McCutcheon A, eds. Applied latent class analysis. Cambridge: Cembridge University Press, 2002.
19. Laer L von, Gruber R, Dallek M, et al. Classification and documentation of children's fractures. Eur J Trauma 2000;26:2–14.
20. Slongo T, Schaerli AF, Koch P, et al. Klassifikation und Dokumentation der Frakturen im Kindesalter – Pilotstudie der internationalen Arbeitsgemeinschaft für Kindertraumatologie. Zentralbl Kinderchir 1995:157–63.
21. Müller M, Narzarian S. The comprehensive classification for fractures of long bones. Berlin–Heidelberg–New York: Springer, 1990.
22. Von Kap-Herr H. Lecture notes. University of Mainz: Clinic of Paediatric Surgery, 2000.
23. Vermunt J, Magidson J. Latent Gold 2.0 computer manual. Belmont: Statistical Innovations, 2000.
24. Vermunt J, Magidson J. Latent Gold 3.0 upgrade manual. Belmont: Statistical Innovations, 2002.
25. Audigé L, Kellam J, Hanson B. Validation concept for fracture classifications. AO/ASIF Dialogue 2002;15:1–5.
26. Kraemer HC. Measurement of reliability for categorical data in medical research. Stat Methods Med Res 1992;1:183–99.
27. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 1974;61: 215–31.

Address for Correspondence
Laurent Audigé, PhD
AO Clinical Investigation and Documentation
AO Foundation
Stettbachstraße 10
8600 Dübendorf
Switzerland
Phone (+41/44) 200-2462, Fax -2460
e-mail: laurent.audige@aofoundation.org