The classification of congenital talipes equinovarus

Andrew M. Wainwright, Tanya Auld, Michael K. Benson,
Tim N. Theologis

From the Nuffield Orthopaedic Centre, Oxford, England

We have assessed the reliability of four classification systems for club foot. Four observers evaluated nine children (18 feet) at different stages in the first six months of life, a total of 180 examinations. Each observer independently assessed all feet according to the classification systems described by Catterall, Diméglio et al, Harrold and Walker, and Ponseti and Smoley.

The variation between observers was assessed using the kappa test which for no more agreement than chance has a value of 0, and for complete agreement between observers a value of 1. The kappa values varied between 0.14 and 0.77 depending on which classification system was used. The system of Diméglio et al was found to have the greatest reliability.

Our findings suggest that current classification systems for the analysis of congenital talipes equinovarus are not entirely satisfactory.

J Bone Joint Surg [Br] 2002;84-B:1020-4.

Received 18 October 2001; Accepted after revision 14 February 2002

Congenital talipes equinovarus, or club foot, is one of the commonest congenital orthopaedic conditions. Its incidence in the UK is approximately 1:1000 live births and up to 50% of cases are bilateral. It should always be recognisable at birth but is now frequently diagnosed at 18 to 20 weeks of gestation by ultrasound.

It is important to be able to describe the treatment and probable outcome to the parents of a baby born with congenital talipes equinovarus. The condition is variable in its clinical course and severity. It may be difficult to assess the severity at initial presentation or to compare the results of treatment. Many classification systems have been proposed to address this problem.

The ideal classification system should be reliable and reproducible, practical enough for use in a clinical setting and should predict appropriate treatment at an early stage. The systems of Ponseti and Smoley,¹ Harrold and Walker,² Catterall³ and Diméglio et al⁴ are the most commonly used. Our aim was to assess the interobserver reliability of these systems.

Patients and Methods

Four observers independently assessed both feet of nine children with congenital talipes equinovarus on several visits during the first six months of life. In five children with unilateral involvement, the 'normal', uninvolved feet were also assessed. The first assessment was at a median of 11 days of age (1 to 89). The assessors were two consultant paediatric orthopaedic surgeons (MKB, TNT) a senior physiotherapist (TA) and an orthopaedic specialist registrar (AMW). It was not possible for every observer to be present at every outpatient visit to examine the child.

Classification. Each assessor had a copy of the original articles describing the classification systems which were being studied. A reference sheet summarising each classification system was also available.

Ponseti and Smoley¹ reported the results of treatment of congenital talipes equinovarus. Their classification system was based on ankle dorsiflexion, heel varus, forefoot supination and tibial torsion. Feet were classified on the basis of these measurements as either good, acceptable or poor (Table I).

Harrold and Walker² considered the ability to correct the deformity. The grade of deformity was determined by whether the foot could be held at or beyond the neutral position (grade 1), or whether there was fixed equinus or varus of <20° (grade 2) or >20° (grade 3) (Table II).

Catterall³ described four patterns depending on the evolution of the deformity which was classified as resolving, caused by tendon or joint contracture, or secondary to a false correction. Several clinical features are used for this classification (Table III).

The system of Diméglio et al⁴ is derived from a detailed scoring system based on the measurement of four parameters: 1) equinus in the sagittal plane; 2) varus deviation in the frontal plane; 3) 'derotation' around the talus of the

A. M. Wainwright, FRCS (Trauma & Orth), Specialist Registrar
T. Auld, MCSP, Senior Physiotherapist
M. K. Benson, FRCS, Consultant Orthopaedic Surgeon
T. N. Theologis, FRCS, Consultant Orthopaedic Surgeon
Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford OX3 7LD, UK.

Correspondence should be sent to Mr T. N. Theologis.

©2002 British Editorial Society of Bone and Joint Surgery
0301-620X/02/712909 \$2.00

Table I. The system of Ponseti and Smoley¹ for the classification of congenital talipes equinovarus

Ankle dorsi-flexion (degrees)	Heel varus (degrees)	Adduction of the fore part of the foot (degrees)	Tibial torsion (degrees)	Result
> 10	0	0 to 10	0	Good
0 to 10	0 to 10	10 to 20	Moderate	Acceptable
0	> 10	> 20	Severe	Poor

Table II. Summary of the system of Harrold and Walker² for the classification of congenital talipes equinovarus

Grade	Severity	Residual deformity with correction
1	Mild	Neutral or beyond
2	Moderate	< 20°
3	Severe	> 20°

Table III. Catterall's³ system for the classification of congenital talipes equinovarus

Foot	Resolving pattern	Tendon contracture	Joint contracture	False correction
Hindfoot				
Lateral malleolus	Mobile	Posterior	Posterior	Posterior
Equinus	No	Yes	Yes	Yes
Creases medial	No	No	Yes	No
Posterior	No	Yes	Yes	Yes
Anterior	Yes	No	No	Yes
Forefoot				
Lateral border	Straight	Straight	Curved	Straight
Mobile	Yes	Yes	No	Yes
Cavus	+/-	+/-	+/-	No
Supination	No	No	Yes	No

Table IV. The system of Diméglio et al⁴ for the classification of congenital talipes equinovarus

Classification grade	Type	Score	Reducibility
I	Benign	< 5	> 90% soft-soft, resolving
II	Moderate	5 to < 10	> 50% soft-stiff, reducible, partly resistant
III	Severe	10 to < 15	< 50% stiff-soft, resistant, partly reducible
IV	Very severe	15 to < 20	< 10% stiff-stiff, resistant

calcaneoforefoot block; and 4) adduction of the forefoot on the hindfoot in the horizontal plane. The scale includes four additional points for the presence of medial creases, a posterior crease, cavus and poor calf musculature. From the score, which has a maximum of 20 points, the deformity can be graded as benign, moderate, severe or very severe (Table IV). Diagrams and a video have been produced to aid assessment (Fig. 1).

Statistical analysis. The categories of classification were treated as nominal data since there was no absolute standard with which to compare them. The kappa statistic was used for evaluation of interobserver agreement. This is a chance-corrected measure of agreement for nominal data described by Cohen.⁵ It compares the observed agreement with the level of agreement expected by chance alone.⁶ The maximum value of 1.0 means that every assessor agrees with every other on every foot. A value of 0 indicates no more agreement than expected by chance alone. Inter-

Table V. Interpretation of the kappa statistic as proposed by Landis and Koch⁷

Kappa statistic	Strength of agreement
< 0.00	Poor
0.00 to 0.20	Slight
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Substantial
0.81 to 1.00	Almost perfect

pretation of the kappa value was based on the guidelines proposed by Landis and Koch⁷ (Table V). 'Normal' feet were defined as those assessed by all observers to be normal by all the classification systems.

Results

The interobserver agreement as described by the kappa statistic is shown for all observers assessing all feet in

Fig. 1

Diagrams of the evaluation of congenital talipes equinovarus used by Diméglio et al⁴ (Reprinted with permission from the *J Pediatr Orthop (B)* 1995;4:129-136).

Figure 2a. Most systems had moderate to substantial reliability when normal feet and affected feet were assessed together. The interobserver reliability for each system assessing only affected feet and excluding normal feet, is shown in Figure 2a. When only the affected feet were assessed, the Diméglio system alone showed moderate levels of reliability.

The level of agreement between the two consultants was also assessed (Fig. 2b). Compared with the interobserver reliability for all observers, they showed higher levels of agreement for most classification systems.

The system of Diméglio et al⁴ appears to be most reliable for consultants.

Discussion

It is important to be able to differentiate between various forms of congenital talipes equinovarus. Classification is difficult, however, because the deformity is complex, three-dimensional, and partly dynamic.

Classification systems are widely used in orthopaedic practice for the assessment of patients and comparison of

Fig. 2a

Fig. 2b

Kappa values obtained when congenital talipes equinovarus was assessed by a) all observers and b) the two consultant observers only using four classification systems.¹⁻⁴

treatments. Several studies have shown that many systems in current use do not have interobserver or intra-observer⁸⁻¹⁰ consistency.

There have been two previous studies which assessed the classification of club foot. An independent assessment of two classification systems¹¹ compared an orthopaedic specialist and a Fellow in paediatric orthopaedics. The authors assessed 55 feet using the systems of Diméglio et al⁴ and Pirani et al.¹¹ Statistical analysis was by recording the difference between the mean scores, the mean of differences between scores and by the use of Pearson correlation coefficients. Further analysis assessed the number of occasions that the examiners recorded scores which were within one or two points of each other. They concluded that there was very good interobserver reliability for both systems. The second assessment study examined the reproducibility of several measurements of congenital talipes equinovarus.¹² Both inter- and intraobserver agreement was assessed based on photographic and radiological measurements of the resting neonatal foot with congenital talipes equinovarus. This study showed that there was a mean measurement error of $>9^\circ$ between two photographs of the same foot. Postoperative clinical measurements of children between four and 16 years of age had a mean intraobserver difference of between 2° and 5° and a mean interobserver difference of between 5° and 14° . They showed that the measurement of deformity of a small foot is difficult at birth and remains difficult into late childhood.

Our analysis of the four classifications shows that each has specific problems. The system of Ponseti and Smoley¹

(Table I) was devised to assess the results of treatment of congenital talipes equinovarus. There are four clinical measurements which were evaluated to produce classification into three groups. The implication is that if one of the measurements of deformity is normal all measurements will be normal and the foot is therefore 'good'. This is not always the case, since one component of the deformity may be more severe than others. The authors do not give an explicit method of classifying a foot which has different degrees of deformity in any of the four components. In our study the classification was determined by the worst component of the deformity.

Harrold and Walker's system² (Table II) is based on the first examination of the foot and the estimated angle of fixed inversion and fixed equinus when firm pressure, insufficient to cause pain, was applied towards the everted and dorsiflexed position. This simple system partly quantifies the ability to correct the deformity. It may seem that this straightforward system would produce the most consistency, but reliability was not satisfactory.

Our study found that both of these systems produced moderate to substantial agreement when all feet were being assessed. When only affected feet were assessed, however, the interobserver reliability was only fair to moderate.

Catterall's³ system (Table III) is based on nine measurements found in the four patterns of congenital talipes equinovarus which he described. In our study, the individual measurements were recorded and the analysis undertaken on the basis of the best pattern match. If there were six, or fewer, matches of measurements in the four defined patterns, the

foot was deemed not to be classifiable by the Catterall system and 42% of the feet were not classifiable. There was poor to slight agreement when this system was used. The agreement was lowest between two experienced consultants assessing affected feet when the normal feet had been excluded. It may be necessary to have specific training in the application of this system before it can be used.

Diméglio et al⁴ have published a clear explanation of the assessment of the components of the deformity which produce the score in their classification system (Table IV). There are some discrepancies in the original paper regarding the score and the way in which it is converted into the four classes of deformity. First, the minimum score which is possible in this system is four since the minimum score for each of the four 'essential parameters' of deformity is 1. The abstract and text of the article describe the scale as 0 to 20. Secondly, the Table in the original article, and the text and abstract grade the foot differently based on the score. In the Table the grades I to IV are defined on the basis of scores of <5, 5 to <10, 10 to <15, and 15 to <20. In the abstract and text the grades are defined as being based on scores of 0 to 5 (or 1-5), 5 to 10, 10 to 15 and 15 to 20. For our analysis we have assumed that grade-I feet have a score of 4, grade-II feet a score of 5 to 9, grade-III feet a score of 10 to 14 and grade-IV feet a score of 15 to 20.

Although the system of Diméglio et al⁴ is complex, it gave the best agreement. In the original article the authors also assessed the consistency of their scoring system. With training, there was a reduction in the discrepancy of scoring from 40% to 6%. In our study the grading system of Diméglio et al⁴ was found to produce moderate to substantial agreement. When two consultants used this system to assess congenital talipes equinovarus excluding normal feet, there was substantial agreement.

There are several factors which may have affected the assessment in this study. Repeated examination by several observers may have led to greater flexibility of the foot towards the end of each session. Conversely, the child, and parents, may have tolerated earlier examinations better than later examinations. These factors may also be important in practice, making consistent classification difficult.

Further classification systems for club foot are necessary.

An ideal system should be reliable and reproducible. It should account for the three-dimensional characteristics of the deformity, yet be simple enough to apply in practice. It may need to include separate information for the hindfoot, midfoot and forefoot since the severity of the deformity may differ at each level. The system should also include information about the flexibility (or rigidity) of the deformity. It should be comprehensive enough to be usable before, during and after treatment, in children of all ages. Finally, treatment should be determined by the classification with reference to the different elements of the deformity, and it should predict the prognosis of the deformity at any stage and be used to compare the results of treatment.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References

1. Ponseti IV, Smoley EN. Congenital club foot: the results of treatment. *J Bone Joint Surg [Am]* 1963;45-A:261-344.
2. Harrold AJ, Walker CJ. Treatment and prognosis in congenital club foot. *J Bone Joint Surg [Br]* 1983;65-B:8-11.
3. Catterall A. A method of assessment of the clubfoot deformity. *Clin Orthop* 1991;264:48-53.
4. Diméglio A, Bensahel H, Souchet P, Mazeau P, Bonnet F. Classification of clubfoot. *J Pediatr Orthop B* 1995;4:129-36.
5. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46.
6. Seigal S, Castellan NJ. Measures of association and their tests of significance: nominally scaled data and kappa statistic. In: Siegel S, Castellan NJ Jr, eds. *Non parametric statistics for the behavioral sciences*. Second edition. New York, etc: McGraw Hill, 1988:284-91.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:363-74.
8. Brumback RJ, Jones AL. Interobserver agreement in the classification of open fractures of the tibia: the results of a survey of two hundred and forty-five orthopaedic surgeons. *J Bone Joint Surg [Am]* 1994;76-A:1162-6.
9. Neyt JG, Weinstein SL, Spratt K, et al. Stulberg classification system for evaluation of Legg-Calvé-Perthes disease: intra-rater and inter-rater reliability. *J Bone Joint Surg [Am]* 1999;81-A:1209-16.
10. Burstein AH. Editorial. Fracture classification systems: do they work and are they useful? *J Bone Joint Surg [Am]* 1993;75-A:1743-4.
11. Flynn JM, Donohoe M, Machenzie WG. An independent assessment of two clubfoot-classification systems. *J Pediatr Orthop* 1998;18:323-7.
12. Porter RW, Roy A, Rippstein J. Assessment in congenital talipes equinovarus. *Foot Ankle* 1990;11:16-21.