Interobserver variability in Pirani clubfoot severity scoring system between the orthopedic surgeons

Saurabh Jain, Anand Ajmera, Mahendra Solanki, Alok Verma

ABSTRACT

Background: Congenital talipes equinovarus (clubfoot) is one of the most common congenital pediatric orthopedic foot deformity, which varies in severity and clinical course. Assessment of severity of the club foot deformity is essential to assess the initial severity of deformity, to monitor the progress of treatment, to prognosticate, and to identify early relapse. Pirani's scoring system is most acceptable and popular for club foot deformity assessment because it is simple, quick, cost effective, and easy. Since the scoring system is subjective in nature it has inter- and intra-observer variability, it is widely used. Hence, the interobserver variability between orthopedic surgeons in assessing the club foot severity by Pirani scoring system.
Materials and Methods: We assessed the interobserver variability between five orthopedic surgeons of comparable skills, in assessing the club foot severity by Pirani scoring system in 80 feet of 60 children (20 bilateral and 40 unilateral) with club foot deformity. All the five different orthopedic surgeons were familiar with Pirani clubfoot severity scoring and Ponseti cast manipulation, as they had already worked in CTEV clinics for at least 2 months. Each of them independently scored, each foot as per the Pirani clubfoot scoring system and recorded total score (TS), Midfoot score (MFS), Hind foot score (HFS), posterior crease (PC), emptiness of heel (EH), rigidity of equnius (RE), medial crease (MC), curvature of lateral border (CLB), and lateral head of talus (LHT). Interobserver variability was calculated using kappa statistic for each of these signs and was judged as poor (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.00).
Results: The mean age was 137 days (range 21-335) days. The mean Pirani score was 3.86. We found the overall consistency to be substantial for overall score (total score kappa - 0.71) and also for midfoot (0.68) and hindfoot (0.66) separately. The consistency was least for the emptiness of heel (kappa - 0.39), and best for rigidity of equnius (kappa - 0.68) and rest of the parameters were moderate (kappa between 0.40 and 0.60).
Conclusion: The Pirani scoring system had got substantial reliability in assessing the clubfoot deformity even when the reliability test was extended to five different orthopedic surgeons simultaneously.

Key words: Clubfoot, interobserver variability, Pirani score
MeSH terms: Club foot, reliability and validity, congenital abnormalities, foot

INTRODUCTION

Congenital talipes equinovarus (clubfoot, CTEV) is one of the most common congenital pediatric orthopedic foot deformity which requires correction.[1,2] Assessment of severity of the club foot deformity is essential to assess the initial severity of deformity, to monitor the progress of treatment, to prognosticate, and to identify early relapse.[1,3] There are various clinical assessment scoring systems such as Ponseti and Smoley[4], Catterall[5], Dimeglio[6], Harrold and Walker.[7] Pirani score is reliable, quick, and easy to use, hence it is used both for the initial assessment and for followup of the treatment.[8,9] Being subjective nature of the scoring system makes it prone to interobserver variability. Different studies

Department of Orthopaedics, MGMMC, Indore, Madhya Pradesh, India

Address for correspondence: Dr. Saurabh Jain,
Department of Orthopaedics, Mahavir Hospital, A-2, Sec. C, Sch. 71,
Footi Khothi Sq., Indore, Madhya Pradesh, India.
E-mail: jaindrsaurabh@yahoo.com

have compared the interobserver variability of the Pirani score among orthopedic surgeon and physiotherapist or allied health worker. But to the best of our knowledge none of the study has been done to compare the interobserver variability of Pirani score, among orthopedic surgeons themselves, who are the most frequent users of the scoring system.[10,11] Hence, purpose of this study was to assess the interobserver variability between orthopedic surgeons in assessing the club foot severity by Pirani scoring system.

MATERIALS AND METHODS

A foot deformity correction camp was organized at our institute in September 2015. All patients coming to the camp with foot deformity were examined by a senior orthopedic surgeon to screen patients of club foot. All patients of idiopathic club foot coming to the camp with age <1 year were included in the study. Secondary club foot, previously operated patients, atypical club foot, and children more than 1 year age were excluded from the study. All the clubfoot children included in the study were independently examined and assessed by five different orthopedic surgeons of comparable clinical experience and skill, who were familiar with Pirani clubfoot severity scoring and the Ponseti cast manipulation. All the five orthopedic surgeons were senior resident, who had atleast 2 years experience after completion of their postgraduation in orthopedics and had worked in CTEV clinics for at least 2 months which is being run in the department at our institute weekly.

All these clubfoot children were then started on treatment by Ponseti cast manipulation and thereafter, were asked to review weekly and regularly in CTEV clinics.

All five orthopedic surgeons independently scored, each foot as per the Pirani clubfoot scoring system (total score [TS]), which is the sum total of midfoot score (MFS) and hind foot score (HFS). The HFS is the sum total of three signs – posterior crease (PC), emptiness of heel (EH), and rigidity of equinus (RE). The MFS is sum total of three signs – medial crease (MC), curvature of lateral border (CLB), and lateral head of talus (LHT). Each of these six signs was graded as either 0 (no abnormality), 0.5 (moderate abnormality), or 1 (severe abnormality) as per the deformity. Thus, TS, i.e., sum of MFS and HFS of all six signs of the club foot, can range from 0 to 6, with 6 being the most deformed foot and 0 being the normal [Table 1].[8]

The data was analyzed for interobserver variability using kappa statistic for each of the six signs (PC, EH, RE, MC, CLB, LHT) and also for MFS, HFS, and TS between all five orthopedic surgeons. The kappa statistic interobserver reliability (strength of agreement) was judged as poor (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.00).

RESULTS

A total of 112 children were enrolled for the foot deformity camp, out of which 78 children were of clubfoot deformity. After fulfilling the inclusion criteria, 60 children were enrolled for the study. Twenty had bilateral involvement and 40 were unilateral (24 right and 16 left). Thus, a total of 80 feet in 60 children were included in the study. Out of these 60 children, 46 were male and 14 were female. Age of the patients ranged from 21 days to 335 days (mean age 137 days). None of these children were ambulatory at the time of assessment.

The mean of the six Pirani score parameters for all the five observers was as 0.61, 0.55, 0.76, 0.62, 0.67, and 0.67 for PC, EH, RE, MC, CLB, and LHT, respectively. The mean for all the five observers for HFS, MFS, and TS was 1.92, 1.98, and 3.86, respectively [Table 2]. Mean overall Pirani score was 3.86, whereas mean Pirani score in unilateral cases was 3.28, and mean in bilateral cases was 4.44.

The overall kappa value, i.e., interobserver reliability for total Pirani score (TS) was 0.71, with substantial degree of agreement present between the observers. There was substantial reliability in HFS and MFS also as the kappa value in both the groups was more than 0.6, i.e., 0.66 and 0.68, respectively.

Table 1: Pirani scoring system

Parameters	Score 0	Score 0.5	Score 1
Hind foot score			
Severity of PC	Multiple fine creases	One or two deep creases	Deep creases change contour of arch
EH	Tuberosity of calcaneus easily palpable	Tuberosity of calcaneus more difficult to palpate	Tuberosity of calcaneus not palpable
RE	Normal ankle dorsiflexion	Ankle dorsiflexes beyond neutral, but not fully	Cannot dorsiflex ankle to neutral
Mid foot score			
Severity of MC	Multiple fine creases	One or two deep creases	Deep creases change contour of arch
CLB of foot	Straight	Mild distal curve	Curve at calcaneocuboid joint
Palpation of LHT	Navicular completely "reduces;" lateral talar head cannot be felt	Navicular partially "reduces;" lateral head less palpable	Navicular does not "reduce;" lateral talar head easily felt

LHT=Lateral part of head of talus, CLB=Curvature of lateral border, MC=Medial crease, RE=Rigidity of equinus, EH=Emptiness of heel, PC=Posterior crease

The interobserver reliability, i.e., kappa value for hind foot signs were as 0.46 (moderate) for PC, 0.39 (fair) for EH, and 0.68 (substantial) for RE and of the mid foot signs were as 0.43 (moderate) for MC, 0.56 (moderate) for CLB, and 0.53 (moderate) for LHT, respectively [Table 3].

DISCUSSION

The incidence of clubfoot is 1:1000 live birth and 50% are bilateral.[1,2] The condition is variable in its clinical course, severity, and expected response to the treatment, leading to the unpredictability in the duration and type of the treatment required.[3] Hence, while treating clubfoot, it is important to classify and grade between various forms and severity of CTEV. These classification systems help to assess the initial degree and severity of the composite deformity before treatment, to monitor and guide the progress of treatment and to predict and compare outcome as well as to identify the early relapse and plan the treatment accordingly.[1,3]

An ideal classification should describe the deformity, correlate, and compare the outcomes, determine the treatment and predict prognosis without having intra- and inter-observer variability.[1,3] It should be simple, easy, user friendly, objective, uniformly accepted, cost effective, reliable, reproducible, and retrievable from retrospective analysis. It should be comprehensive accounting for the three-dimensional characteristics of deformity and include separate information for forefoot, midfoot, and hindfoot deformities and applicable to all forms of deformity, at all ages and at all stages of treatment.[1,3]

Assessment by radiography and magnetic resonance imaging is not recommended in the child due to various reasons such as nonvisualization of unossified cartilige, projection errors, difficult positioning, radiation exposure, noncost-effectiveness, and lack of uniform interpretation universally.[12,13] Authors also have emphasized on clinical evaluation as the yard stick for the assessment of deformity.[6,14,15]

Even after improved understanding of the pathoanatomy of clubfoot, a reliable classification system based on the clinical evaluation still remains elusive and there is no agreed ideal grading system. Many authors such as Maceven *et al.*,[16] Wynne-Davies,[17] Chacko and Mathew,[18] McKay,[19] Ponseti and Smoley,[4] Harrold and Walker,[7] Catterall,[5] Diméglio *et al.*,[6] and Pirani *et al.*[8,9] have developed the classification systems. None of them had proved superior over the other and gold standard is yet to be established. But Pirani's classification has gained wide clinical acceptability and popularity because it is simple, reliable, quick, cost effective, easy to learn, use, and apply.[3,20] It can predict the number of casts required to correct the deformity and the probability of achilies tendon tentomy.[20,21] Scher *et al.* found that significantly higher Pirani score requires significantly more number of cast and HFS rather than the MFS of the Pirani score predicts the need for tenotomy, as it is the hindfoot equinius that the tenotomy is correcting.[22]

Table 2: Mean of the Pirani score parameters by different observers

Pirani score parameter	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Total mean
PC	0.625	0.675	0.55	0.5125	0.725	0.6175
EH	0.625	0.7125	0.3625	0.375	0.6875	0.5525
RE	0.775	0.8	0.6875	0.75	0.7875	0.76
HFS	2.025	2.187	1.6	1.63	2.2	1.92
MC	0.7125	0.6375	0.6	0.55	0.6375	0.6275
CLB	0.7125	0.575	0.725	0.7875	0.5875	0.6775
LHT	0.7125	0.75	0.8125	0.625	0.4875	0.6775
MFS	2.1375	1.962	2.1375	1.962	1.712	1.9825
TS	4.1625	4.1496	3.737	3.6	3.912	3.864

LHT=Lateral part of head of talus, CLB=Curvature of lateral border, MC=Medial crease, RE=Rigidity of equinus, EH=Emptiness of heel, PC=Posterior crease

Table 3: Inter-observer reliability of Pirani score parameters

Pirani score parameter	Intraclass correlation coefficient	Kappa value	Strength of agreement
PC	0.4022	0.46	Moderate
EH	0.3255	0.39	Poor
RE	0.6546	0.68	Substantial
HFS	0.6221	0.66	Substantial
MC	0.4294	0.43	Moderate
CLB	0.5322	0.56	Moderate
LHT	0.534	0.53	Moderate
MFS	0.6407	0.68	Substantial
TS	0.7004	0.71	Substantial

LHT=Lateral part of head of talus, CLB=Curvature of lateral border, MC=Medial crease, RE=Rigidity of equinus, EH=Emptiness of heel, PC=Posterior crease

Several studies such as Catterall[5] and Cummings *et al.*[23] commented on problem of, lack in intra- and inter-observer consistency in classification systems owning to subjective nature of these classifications and despite the lack of reliable data, surgeons have been using them regularly as a dependent measure.[23,24]

Since Pirani score is among one of the most commonly used score, we thought it was worthwhile to find its interobserver consistency among five different orthopedic surgeons using kappa value.

Flynn concluded that there is good interobserver reliability of 89% for both Demeglio and Pirani classification systems between orthopedic specialist and a fellow in pediatric orthopedics, but only after a short initial training phase.[25] Porter assessed the inter- and intra-observer agreement of photographic and radiological measurements of the resting neonatal foot with club foot and showed mean measurement of error of more than 9°.[26]

Wainwright compared four club foot assessment systems and found that Ponseti and Smoley classification, which is based on worst component of the deformity and Harrold and Walker's system, which is based on the ability to correct the deformity, both of these systems produced moderate to substantial agreement when all feet were being assessed, whereas Catterall's system had only poor to slight agreement. For all the three systems, the agreement was lowest and was only fair to moderate when the normal feet had been excluded and only affected feet were assessed. Diméglio-system although gave the best agreement with moderate to substantial agreement, but it is complex and needs training for reduction in the discrepancy from 40% to 6%. They finally concluded that all current classifications are still not entirely satisfactory as they are subjective in nature and have inter- and intra-observer variation. Jillani *et al.*[10] in a two staged study, i.e., before training and after training, compared orthopedic surgeon and a lower level allied health worker, i.e., a plaster technician who had 2-year operation theater technician diploma and showed the overall kappa values for the parameters as 0.716, 0.625, 0.696, 0.675, 0.391, 0.543, 0.457, and 0.362, respectively, for CLB, MC, LHT, PC, RE, EH, HFS, and TS with conclusion that prior training and supervision in the early phase improves the reliability. They found interobserver reliability to be fair to substantial (fair for TS and equines rigidity, other parameters substantial to moderate) with point-to-point interobserver agreement for all components of deformity to be 82%.[10] Another study showed moderate to substantial interobserver reliability between a pediatric orthopedic surgeon and a physiotherapy assistant, with point-to-point interobserver agreement for all components of deformity to be 83%,[11] with κ statistic was 0.61 for PC, 0.72 for EH, 0.51

for RE, 0.54 for HFS, 0.57 for MC, 0.54 for CLB, 0.56 for LHT, 0.50 for MFS, and 0.50 for TS. Flynn found higher agreement of 89% when comparison done between two physicians of comparable skills, i.e., orthopedic specialist and a fellow in pediatric orthopedics with correlation coefficients of 0.90 for the Pirani classification, and 0.83 for the Dimeglio classification. Correlation coefficients were much lower for the first 15 feet scored and were also lower when the therapist's scores were included.[25] In similar study, Pirani *et al.* found the interobserver strength of agreement in clubfoot scoring to be substantial or almost perfect among three independent observers, with kappa score of TS, MFS, and hindscore to be 0.92, 0.91, and 0.86, respectively.[9] However, in their study, the second observer was an orthopedic resident, not a paramedic.

Although all the studies had done comparison between two persons alone, hence we thought it would be interesting to extend the comparison between five orthopedic surgeons of comparable skill and experience. We found the overall consistency to be substantial for overall score (TS kappa - 0.66) and also for midfoot and hindfoot separately. But when the components were visualized separated, the consistency was least for the EH (kappa - 0.39), and best for RE (kappa - 0.68) and rest of the parameters were moderate (kappa between 0.40 and 0.60). Thus the assessment of EH was the parameter which was least and the rigidity of equinus was most reliable as per our study. Since both the parameters are part of HFS, the HFS agreement remained marginally on the substantial side.

Our study is limited by factors such as repeated examination by several observers may have led to greater flexibility of the foot and the child and parents may have tolerated earlier examinations better than later examinations. Further collecting static measurements from infants is challenging because of the size of the foot, the less evident anatomical landmarks and the degree of cooperation.

These interobserver variations can be also attributed to differences in the training and background of observers, which we tried to remove by taking orthopedic surgeon of comparable skill and experience in our study. Further our agreement was substantial in only two of the Pirani's parameter and rest of the parameters had poor or moderate agreement because Pirani system is also not so sensitive and it tends to give a diagnosis of moderate abnormality as there are only three levels of scoring 0, 0.5, and 1, but the overall Pirani score had substantial agreement. Another limitation of the study is low number of feet, but even with this number of feet the power of the study is more than 0.80 with alpha error of 0.05. Further the study includes only the children coming in the camp on that single day on which camp was done, hence study is limited to 80 feet only.

Conclusion

The Pirani scoring system has got substantial reliability in assessing the clubfoot deformity even when the reliability test was extended to five different orthopedic surgeons simultaneously. This consistency was seen in the various parameters of Pirani score also when assessed separately, except for the EH, which is the least reliable among all the parameters. We recommend to do further studies including the many persons simultaneously, such as surgeons, physiotherapist, or assistants for the assessment of the reliability of these classification systems.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

References

1. Wainwright AM, Auld T, Benson MK, Theologis TN. The classification of congenital talipes equinovarus. J Bone Joint Surg Br 2002;84:1020-4.
2. Kelly DM. Congential anomalies of lower extremity. In: Canale ST, Beaty JH, editors. Campbell's Operative Orthop. 12th ed. St. Louis: Mosby Elsevier; 2008. p. 1078-100.
3. Jain P, Mehtani A, Goel M, Jain S, Sood A, Kumar Jain A. Correlation of foot bimalleolar angle with Pirani score to assess the severity of congenital talipes equinovarus deformity. J Pediatr Orthop B 2012;21:68-72.
4. Ponseti IV, Smoley EN. Congenital club foot: The results of treatment. J Bone Joint Surg Am 1963;45-A: 261-344.
5. Catterall A. A method of assessment of the clubfoot deformity. Clin Orthop Relat Res 1991;264:48-53.
6. Diméglio A, Bensahel H, Souchet P, Mazeau P, Bonnet F. Classification of clubfoot. J Pediatr Orthop B 1995;4:129-36.
7. Harrold AJ, Walker CJ. Treatment and prognosis in congenital club foot. J Bone Joint Surg Br 1983;65:8-11.
8. Pirani S, Outerbridge H, Moran M, Sawatsky BJ. A Method of Evaluating the Virgin Clubfoot with Substantial Interobserver Reliability. Miami, Florida: POSNA; 1995;71;99.
9. Pirani S, Hodges D, Sekeramyi F. A reliable and valid method of assessing the amount of deformity in the congenital clubfoot deformity (The Canadian Orthopaedic Research Society and the Canadian Orthopaedic Association conference proceeding). J Bone Joint Surg Br 2008;90-B Suppl I: 53.
10. Jillani SA, Aslam MZ, Chinoy MA, Khan MA, Saleem A, Ahmed SK. A comparison between orthopedic surgeon and allied health worker in Pirani score. J Pak Med Assoc 2014;64 12 Suppl 2:S127-30.
11. Shaheen S, Jaiballa H, Pirani S. Interobserver reliability in Pirani clubfoot severity scoring between a paediatric orthopaedic surgeon and a physiotherapy assistant. J Pediatr Orthop B 2012;21:366-8.
12. Meena S, Sharma P, Gangary SK, Lohia LK. Congenital clubfoot. J Orthop Allied Sci 2014;2:34-9.
13. Pirani S, Zeznik L, Hodges D. Magnetic resonance imaging study of the congenital clubfoot treated with the Ponseti method. J Pediatr Orthop 2001;21:719-26.
14. Herbsthofer B, Eckardt A, Rompe JD, Küllmer K. Significance of radiographic angle measurements in evaluation of congenital clubfoot. Arch Orthop Trauma Surg 1998;117:324-9.
15. Bensahela H, Kuo K, Duhaime M; International Clubfoot Study Group. Outcome evaluation of the treatment of clubfoot: The international language of clubfoot. J Pediatr Orthop B 2003;12:269-71.
16. Macewen GD, Scott DJ Jr., Shands AR Jr. Followup survey of clubfoot treated at the Alfred I. du Pont Institute with special reference to the value of plaster therapy, instituted during earliest signs of recurrence, and the use of night splints to prevent or minimize the manifestations. JAMA 1961;175:427-30.
17. Wynne-Davies R. Talipes equinovarus. A review of eighty-four cases after completion of treatment. J Bone Joint Surg Br 1964;46:464-76.
18. Chacko V, Mathew T. Some observations in the treatment of congenital clubfoot. Indian J Orthopaedics 1976;10:127-31.
19. McKay DW. New concept of and approach to clubfoot treatment: Section II – Correction of the clubfoot. J Pediatr Orthop 1983;3:10-21.
20. Dyer PJ, Davis N. The role of the Pirani scoring system in the management of club foot by the Ponseti method. J Bone Joint Surg Br 2006;88:1082-4.
21. Goriainov V, Judd J, Uglow M. Does the Pirani score predict relapse in clubfoot? J Child Orthop 2010;4:439-44.
22. Scher DM, Feldman DS, van Bosse HJ, Sala DA, Lehman WB. Predicting the need for tenotomy in the Ponseti method for correction of clubfeet. J Pediatr Orthop 2004;24:349-52.
23. Cummings RJ, Davidson RS, Armstrong PF, Lehman WB. Congenital clubfoot. J Bone Joint Surg Am 2002;84-A: 290-308.
24. Gelfer Y, Durham S, Daly K, Ewins D. Intraobserver reliability of static measures in the normally developing infant foot and clubfoot. J Pediatr Orthop B 2009;18:214-9.
25. Flynn JM, Donohoe M, Mackenzie WG. An independent assessment of two clubfoot-classification systems. J Pediatr Orthop 1998;18:323-7.
26. Porter RW, Roy A, Rippstein J. Assessment in congenital talipes equinovarus. Foot Ankle 1990;11:16-21.