



Inter- and Intra-rater Reliability of the Checketts' Grading System for Pin-site Infections across All Skin Colours

Rosalind Groenewoud¹, Harpreet Chhina², Jeffrey Bone³, Sanjeev Sabharwal⁴, Chris Iobst⁵, Anirejuoritse Bafor⁶, Anthony Cooper⁷

Received on: 21 July 2022; Accepted on: 01 May 2023; Published on: 31 May 2023

ABSTRACT

The Checketts' grading system (CGS) is the only classification that provides both a description of how to visually grade the infection and the appropriate course of treatment. There are no studies on the reliability of this system nor on whether skin colour can influence applicability. This study aims to determine the inter-rater and intra-rater reliability of the CGS to assess whether this scale could be used as a universal grading system across all skin colours.

A survey consisting of 134 anonymised photographs of pin-site infections was sent out to orthopaedic surgeons specialising in limb lengthening and reconstruction and to patients or carers of individuals who had external fixators. For each photograph, the participants were asked to grade the infection using the CGS, rate their confidence in their chosen grade on a Likert scale and assign a treatment option. The participants were supplied with the CGS at the beginning of the survey, after the 45th and 90th photographs.

The inter-rater reliability of the CGS between the surgeons, expressed as an intraclass correlation coefficient (ICC), was poor-to-moderate at both time points (ICC = 0.56 for baseline survey and ICC = 0.48 for follow-up). This was similar for the patient or caretaker group. There was a lower inter-rater reliability for grading of dark skin as opposed to light skin by surgeons but not for patients or caretakers. The inter-rater reliability of treatment decisions between the surgeons was poor at both time points (kappa = 0.30 and 0.22) with similar inter-rater reliability for dark (kappa = 0.26 and 0.23) compared with light skin (kappa = 0.29 and 2.6). This was similar for the patient or caretaker group. The surgeons' confidence (Table 4) in grading was low (median = 1). The patient or caretaker group's confidence in their grading was modest (median = 2).

The reliability of the CGS as assessed here demonstrates poor-to-moderate inter-rater reliability which makes interpretation of published pin site infection rates using this scale difficult. The design of new grading systems will need to consider skin colour to reduce inequities in medical decision-making.

Keywords: Checketts grading system, Inter-rater reliability, Paediatric, Pin site infections, Skin.

Strategies in Trauma and Limb Reconstruction (2023): 10.5005/jp-journals-10080-1581

INTRODUCTION

Pin site infections are the most common complication resulting from external fixation devices (EFDs). There is no accepted definition for a pin site infection nor an accepted grading criteria to diagnose the severity of an infection¹ universally. Pin site infections can lead to severe consequences: there can be a failure of the bone-pin interface leading to pin loosening, non-union of fractures and chronic osteomyelitis, as well as rare reports of life- and limb-threatening conditions of toxic shock syndrome or necrotising fasciitis.² The reported incidence of pin track infection associated with EFD usage ranges from 11 to 89%.³⁻⁶ This wide range suggests a disparity in the grading and reporting of these infections.⁷ Much research exists on treatment and prevention of pin site infections but a comparison is difficult without a common terminology. Future studies on pin site infections need prior establishment of a universally accepted definition for pin site infection and grading system for both severity and the counting of pin site infections.

A grading of pin site infections is needed to evaluate the severity of infection and subsequent treatment. A review of literature found eight existing classifications for pin site infection. Four are based on the degree of infection.⁸⁻¹¹ The Saleh and Scott classification¹² is based on the response to treatment. Simpler systems also exist, such as good, bad and ugly;¹³ calm, irritated and infected¹⁴ and a major or a minor system.¹⁵

¹Faculty of Medicine, University of British Columbia, Vancouver, British Columbia

²Department of Orthopaedics, BC Children's Hospital, Vancouver, British Columbia

³BC Children's Hospital Research Institute, Vancouver, British Columbia

⁴UCSF Benioff Children's Hospital, Oakland, California

^{5,6}Nationwide Children's Hospital, Columbus, Ohio

⁷Department of Orthopaedics, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia

Corresponding Author: Anthony Cooper, Department of Orthopaedics, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Phone: +604-875-2642, e-mail: externalfixators@cw.bc.ca

How to cite this article: Groenewoud R, Chhina H, Bone J, *et al.* Inter- and Intra-rater Reliability of the Checketts' Grading System for Pin Site Infections across All Skin Colours. *Strategies Trauma Limb Reconstr* 2023;18(1):2-6.

Source of support: No direct funding for the study. The Limb Reconstruction Research Program is supported by BC Children's Hospital Foundation.

Conflict of interest: None

Checketts' grading system (CGS; Table 1) is the only classification that provides both a description of how to visually grade the

Table 1: Checketts' pin site infection grading system

Grade	Characteristics	Treatment
Minor infection		
1	Slight redness, little drainage	Improved pin site care
2	Redness of the skin, discharge, pain and tenderness of the soft tissue	Improved pin site care, oral antibiotics
3	Grade 2 but no improvement with oral antibiotics	Affected pins re-sited and external fixation can be continued
Major infection		
4	Severe soft tissue infection involving several pins, sometimes associated with loosening of the pin	External fixation must be abandoned
5	Grade 4 but with radiographic changes	External fixation must be abandoned
6	Infection after fixator removal. Pin track heals initially but will subsequently breakdown and discharge in intervals	Curettage of the pin track

infection and the appropriate course of treatment. Therefore, this system may have the greatest utility and potential for a universal grading of pin site infections. The ability to grade a pin site infection visually has important implications in an era adjacent to the recent COVID-19 pandemic which has brought about keeping non-emergency hospital visits to a minimum. The need for remote reporting of pin site infections has increased. In CGS, Grades 0, 1 and 2 can be determined based on visual assessment. Higher grades require knowledge of previous treatment, clinical assessment of pins and radiographic findings.

There are no studies that have evaluated the reliability of the CGS. Dermatological conditions often have a different presentation on dark skin tones, such as decreased redness or increased hyperpigmentation.¹⁶ As the CGS partly relies on redness of the pin site as a descriptive factor, the difference between the grades may be less observable in darker skin tones. The criteria for many skin conditions were created with a focus on light skin tones, individuals with darker skin are routinely under-diagnosed and experience higher rates of morbidity and mortality in many dermatologic conditions.¹⁷ The CGS does not provide any clarification on how to assess different skin colours, nor did the original study distinguish the findings of pin site infections based on skin colour, thereby providing no evidence that the visual grading system is applicable to all skin colours.

This study aims to determine the inter-rater and intra-rater reliability of the CGS to assess whether this scale could be used as a universal grading system across all skin colours.

MATERIALS AND METHODS

A survey consisting of 134 anonymised photographs of pin site infections was sent out to orthopaedic surgeons specialising in limb lengthening and reconstruction as well as to a group of patients who have had EFDs in the past or are currently undergoing treatment with EFDs. For patients under the age of 14 years, only parents or caretakers were asked to grade the infections. Patients of ages 14–18 years with EFD experience were asked to complete their own surveys.

For each photograph, the participants were asked to grade the infection using the CGS (Table 1), rate their confidence in their chosen grade on a Likert scale (0—Not at all confident, 1—Slightly confident, 2—Somewhat confident, 3—Mostly confident, 4—Very confident), and assign a treatment option (0—no treatment, 1—increased cleaning, 2—oral antibiotics, 3—intravenous [IV] antibiotics, 4—pin removal). The participants were supplied with the CGS at the beginning of the survey, after the 45th photograph and after the 90th photograph.

For the two groups of raters (surgeons and caretakers or patients), reliability was calculated using the intraclass correlation coefficient (ICC) for CGS, Fleiss' kappa coefficients for inter-reliability for treatment, and Cohen's kappa coefficients for intra-reliability for treatment. A sub-group analysis was also performed for the reliability of using this grading system in pin site infections in dark skin compared with light skin. All statistics are provided with 95% confidence intervals and were calculated using R statistical software version 4.0.3.^{18,19}

Skin colour was assessed using the Fitzpatrick Phototype Scale Type, which is a common numerical scale used to visually classify skin colours and has shown good reliability for digital images of skin.¹⁸ Types I–III were assigned to the 'light skin' category and types IV–VI were assigned to the 'dark skin' category for analysis. The survey consisted of 97 'light skin' pin sites and 37 'dark skin' pin sites. Photographs were chosen to ensure at least 25 photographs of each minor infection grade—Grades 0, 1 and 2 infections (as graded by the research team)—with at least 10 photographs of each infection grade in dark skin. Assuming a kappa statistic of good agreement (i.e., 0.63) with at least 6 raters, 22 photographs per grade were required to ensure the lower boundary of the confidence interval for the kappa statistic is at least 0.5.

At the end of the survey, participants were asked to provide a short answer to the question 'when assessing a pin site, what do you think are the most important aspects for deciding the infection grade and which treatment is needed?' This was to help understand the thought-process for grading decisions in pin site infections. An open coding thematic analysis of the responses determined common themes within both groups. A qualitative frequency distribution was determined of the common themes to compare the responses from the two groups.

After a four-week period, the same survey was sent to the same participants again for grading. Reminders for survey completion were sent out every week.

RESULTS

The survey was sent to 14 surgeons, with a 65% response rate ($n = 9$), and 13 patients and caretakers, with a 62% response rate ($n = 8$).

The inter-rater reliability of the CGS between the surgeons, expressed as an ICC, was poor-to-moderate at both time points (ICC = 0.56 [0.50–0.62] for baseline survey and ICC = 0.48 [0.38–0.57] for follow-up) though slightly lower in the follow-up survey (Table 2). The sub-analysis for skin colour showed a reduction in inter-rater reliability at both time points between surgeons when grading dark skin (ICC = 0.46 [0.34–0.61], ICC = 0.4 [0.27–0.56]) compared with light skin (ICC = 0.56 [0.5–0.63], ICC = 0.48 [0.38–0.58]). The CGS showed

Table 2: Inter-rater reliability: ICC for grading and Fleiss kappa for treatment

Time point	Measure	Surgeons (N = 9)			5 Parents and 3 patients (N = 8)		
		Overall	Light	Dark	Overall	Light	Dark
Baseline	CGS grading	0.56 (0.50, 0.62)	0.56 (0.50, 0.63)	0.46 (0.34, 0.61)	0.51 (0.43, 0.58)	0.51 (0.42, 0.59)	0.49 (0.37, 0.64)
	Treatment	0.30	0.29	0.26	0.19	0.20	0.12
One month follow-up	CGS grading	0.48 (0.38, 0.57)	0.48 (0.38, 0.58)	0.40 (0.27, 0.56)	0.41 (0.28, 0.53)	0.41 (0.27, 0.53)	0.41 (0.26, 0.57)
	Treatment	0.22	0.26	0.23	0.16	0.15	0.13

CGS, Checketts' grading system

Table 3: Intra-rater reliability: ICC for grading and Cohen's kappa for treatment

	Surgeons (N = 7)			4 Parents and 2 patients (N = 6)		
	Overall	Light skin	Dark skin	Overall	Light skin	Dark skin
CGS grading (ICC)	0.85	0.85	0.91	0.90	0.92	0.97
Treatment (Cohen's kappa)	0.45	0.46	0.43	0.36	0.37	0.32

CGS, Checketts' grading system; ICC, intraclass correlation coefficient

Table 4: Summary of confidence by group

	N	Median (min-max)
Baseline	2,295	1 (0-4)
1 month	2,025	2 (0-4)
Parent and patients	1,890	2 (0-4)
Surgeons	2,430	1 (0-4)
Surgeons: dark skin	486	1 (0-4)
Surgeons: light skin	1,944	1 (0-4)
Parent and patients: dark skin	378	2 (0-4)
Parent and patients: light skin	1,512	2 (0-4)

good intra-rater reliability (Table 3) for surgeons (ICC = 0.85), with a slight increase in reliability for dark skin (ICC = 0.91) compared with light skin (0.85).

The patient or caretaker group showed poor-to-moderate inter-rater reliability at both time points (ICC = 0.51 [0.43-0.58] for baseline survey and ICC = 0.41 [0.28-0.53] for follow-up) and, similar to the surgeon group, showed lower coefficients in the follow-up survey (Table 2). The patient or caretaker group had no modification for skin colour at either time point. This group also had a high intra-rater reliability (ICC = 0.90) using the CGS (Table 3), with a slight increase in reliability for dark skin (ICC = 0.97) compared with light skin (0.92).

The inter-rater reliability of treatment decisions (Table 2) between the surgeons was poor at both time points (kappa = 0.30 and 0.22) with similar inter-rater reliability for dark (kappa = 0.26 and 0.23) compared with light skin (kappa = 0.29 and 0.26). The intra-rater reliability for treatment decisions was poor or moderate (kappa = 0.45), with no discrepancy based on skin colour. The results from the patient or caretaker group were similar; there was poor inter-reliability at both time points (kappa = 0.19 and 0.13) with no modification based on skin colour, and poor intra-reliability (kappa = 0.36).

Overall, the surgeons' confidence (Table 4) in grading was low (median = 1). The patient or caretaker group's confidence in their grading was modest (median = 2). For both groups, there was no change in confidence based on skin colour. The study participants felt more confident in the follow-up survey (median = 2), compared to the baseline survey (median = 1).

Qualitative analysis of the short answer question at the end of survey showed several themes in the thought-processes of participants (Table 5). In the surgeon group, four themes emerged,

Table 5: Qualitative analysis of short answer question: 'When assessing a pin site, what do you think are the most important aspects for deciding the infection grade and which treatment is needed?'

	Surgeons (n = 8)	Parents/patients (n = 7)
Redness	6 (75%)	7 (100%)
Pain/tenderness	6 (75%)	2 (29%)
Discharge/drainage	6 (65%)	7 (100%)
Swelling	1 (13%)	0
Warmth	0	2 (29%)
Response to previous treatment	4 (50%)	0
Evidence of loosening	1 (13%)	0
Frustration/dissatisfaction with grading solely on pictures	4 (50%)	2 (29%)

which corresponded directly to the Checketts' scale: 'redness', 'pain and tenderness', 'discharge or drainage' and 'response to previous treatment'. Other important clinical signs that were mentioned were 'swelling' and 'loosening of the pin sites'. The surgeons also expressed frustration with grading just based on photographs as clinical symptoms and history could not be elicited.

In the caretaker or patient group, 'redness' and 'discharge or drainage' were the most common themes. Interestingly, 'pain' was not a common theme in this group. This group also mentioned 'warmth' around the pin site as an important factor for distinguishing the infection grade which was not a theme brought up by the surgeon group.

DISCUSSION

The CGS demonstrated good intra-rater reliability and poor-to-moderate inter-rater reliability for surgeons, and this was reduced further when evaluating darker skin. The patient or caretaker group had similar results; good intra-rater reliability and poor-to-moderate inter-rater reliability but without any reduction for skin colour. Both groups had poor inter-rater and intra-rater reliability for treatment decisions. Overall, surgeons and patient or caretakers did not feel confident in using the scale to grade the photographs.

There are several limitations to the interpretation of the results. The study has shown suboptimal inter- and intra-reliability for treatment decisions by the surgeons. However, the lower

coefficients for treatment decisions compared with grading may be explained by a limitation mentioned by one surgeon in the short answer question: the survey forced participants to choose one treatment decision, even though the CGS suggests a multi-modal approach to treatment. The survey did not provide guidance on this decision and therefore participants may have alternated between choosing the 'bare minimum' treatment option (such as improved pin site care for a grade 2) or the most efficacious treatment (such as oral antibiotics for a grade 2) for each pin site.

Other limitations of the study include analysing patients and caretakers as one group due to the low survey response numbers and the arbitrary division of 'light' and 'dark' skin groups. Patients and caretakers have different experiences; grouping them together may have obscured separate and important findings from both groups. The creation of just two skin colour groups fails to consider that redness is progressively less evident as skin colour gets darker. Therefore, the GSC may perform poorer in Fitzpatrick type VI skin (Black skin) than Fitzpatrick Type IV and Type V skin (medium and dark brown), but this cannot be confirmed by our study.

The results of this study suggest the need for a revision to the CGS to reduce subjectivity and increase confidence and consistency with grading. The current system does not incorporate all clinical symptoms and signs used by clinicians, patients and their families for diagnosing infection. One possibility for a revision could include all four classic symptoms of infection as described by first century AD Roman scholar Celsus: 'calor (heat), dolor (pain), tumour (swelling), rubor (redness)'²⁰ as all four qualities were mentioned in the short answer portion of the survey. Alternatively, inspiration from the recently validated scale for assessing the need for IV antibiotics in paediatric patients with cellulitis, the Melbourne ASSET score,²¹ may be timely. The ASSET score uses five features—area, systemic features of sepsis, severity of swelling, eye involvement and severity of tenderness—to determine the need for IV antibiotics. Redness and previous oral antibiotic treatment were originally considered for the ASSET scale but did not improve performance of the scale and were thus eliminated. A similar direction may be taken for the CGS by eliminating redness and previous treatment, and to focus on the severity of pain, swelling and discharge to delineate minor infections.

The lack of standardisation of grading in pin site infections¹ has been lamented. Future studies on pin site infections need prior establishment of a universally accepted definition for pin site infection and grading system for both severity and the counting of pin site infections.

An accurate and reproducible scale for grading pin site infections has many relevant applications. The expected increase in the accuracy of diagnosing the severity of pin site infections should lead to a decrease in the need for in-person visits as well as reduce the inappropriate use of antibiotics used to treat 'infections' that are simply inflamed pin sites. Furthermore, a revision to the Checketts' criteria that increases the utility of the scale for all skin colours will aid in proper diagnosis and treatment and potentially reduce an area of racial inequity within the medical system. A reliable scale is also essential for future research on pin site infections, such as comparing the effects of different coating of the pins and wires used in EFDs or investigating alternative pin site care measures. Pre-existing research is difficult to summarise without a consistent definition.

The COVID-19 pandemic has led to more virtual patient visits, with the use of photographs and video becoming more commonplace for patient follow-ups with EFDs.²² A grading system which is less subjective and easier for a patient to utilise would benefit from reducing the need for hospital visits and inappropriate

antibiotic prescriptions. This study shows the inadequate reliability of grading pin site infections and making treatment decisions using photographs solely. Therefore, the incorporating of clinical symptoms during telehealth visits that can be assessed by a patient and their families, such as pain or warmth, is needed for adequate grading and assessment.

Clinical Significance

This is the first study to assess the reliability of the CGS. The results demonstrate poor-to-moderate inter-rater reliability of the CGS which makes interpretation of published pin site infection rates using this scale challenging. Furthermore, treatment decisions showed poor inter-rater and intra-rater reliability, indicating that photographs alone are not sufficient for determining optimal treatment for a patient. When designing new grading systems, considering skin colour is essential to reduce racial inequity in medical treatment.

ORCID

Harpreet Chhina  <https://orcid.org/0000-0002-7656-457X>

Anthony Cooper  <https://orcid.org/0000-0002-7864-2981>

REFERENCES

1. Iobst CA. Pin-track infections: Past, present and future. *J Limb Lengthen Reconstr* 2017;3(2):78–84. DOI: 10.4103/jllr.jllr_17_17.
2. Jauregui JJ, Bor N, Thakral R, et al. Life- and limb-threatening infections following the use of an external fixator. *Bone Joint J* 2015;97-B(9):1296–1300. DOI: 10.1302/0301-620X.97B9.35626.
3. Patterson MM. Multicenter pin care study. *Orthop Nurs* 2000;24(5):349–360. DOI: 10.1097/00006416-200509000-00011.
4. Parameswaran AD, Roberts CS, Seligson D, et al. Pin tract infection with contemporary external fixation: How much of a problem? *J Orthop Trauma* 2003;17(7):503–507. DOI: 10.1097/00005131-200308000-00005.
5. Schalamon J, Petnehazy T, Ainoedhofer H, et al. Pin tract infection with external fixation of pediatric fractures. *J Pediatr Surg* 2007;42(9):1584–1587. DOI: 10.1016/j.jpedsurg.2007.04.022.
6. Davies R, Holt N, Nayagam S. The care of pin sites with external fixation. *J Bone Joint Surg* 2005;87(5):716–719. DOI: 10.1302/0301-620X.87B5.15623.
7. Lethaby A, Temple J, Santy-Tomlinson J. Pin site care for preventing infections associated with external bone fixators and pins. *Cochrane Database of Systematic Reviews* 2013. Available at: <https://doi.org/10.1002/14651858.CD004551.pub3>.
8. Paley D. Problems, obstacles, and complications of limb lengthening by the Ilizarov technique. *Clin Orthop Relat Res* 1990;250:81–104.
9. Dahl MT, Gulli B, Berg T. Complications of limb lengthening: A learning curve. *Clin Orthop Relat Res* 1994;301:10–18. PMID: 8156659.
10. Checketts' RG, Otterburn M, MacEachern G. Checketts' Pin track infection definition, incidence and prevention. *Int J Orthop Trauma Suppl* 1993;3(3):16–18.
11. Saw A, Chan C, Penafort R, Sengupta S. A Simple Practical protocol for care of metal-skin interface of external fixation. *Med J Malaysia*. 2006;61(A):62–65.
12. Saleh M, Scott BW. Pitfalls and complications in leg lengthening: The Sheffield experience. *Semin Orthop* 1992;7:207–222.
13. Clint SA, Eastwood DM, Chasseaud M, et al. The "Good, Bad and Ugly" pin site grading system. A reliable and memorable method for documenting and monitoring ring fixator pin sites. *Injury* 2010;41(2):147–150. DOI: 10.1016/j.injury.2009.07.001.
14. Santy J, Vincent M, Duffield B. The principles of caring for patients with Ilizarov external fixation. *Nurs Stand* 2009;23(26):50–55. DOI: 10.7748/ns2009.03.23.26.50.c6835.
15. Ward P. Care of skeletal pins: a literature review. *Nurs Stand* 1998;12(39):34–38. DOI: 10.7748/ns1998.06.12.39.34.c2514.

16. Taylor SC. Skin of colour: Biology, structure, function, and implication for dermatologic disease. *J Am Acad Dermatol* 2002;46(2):S41–S46. DOI: 10.1067/mjd.2002.120790.
17. Buster KJ, Stevens EI, Elmets CA. Dermatologic health disparities. *Dermatol Clin* 2012;30(1):53. DOI: 10.1016/j.det.2011.08.002.
18. Altieri L, Hu J, Nguyen A, et al. Interobserver reliability of teledermatology across all Fitzpatrick skin types. *J Telemed Telecare* 2017;23(1):68–73. DOI: 10.1177/1357633X15621226.
19. R Core Team. R. A language and environment for statistical computing. R Foundation for Statistical Computing, 2021. Available at: <https://www.R-project.org/>.
20. Spencer WG. *Celsus De Medicina*. Cambridge, Massachusetts. Harvard University Press. 1971 (Republication of the 1935 edition).
21. Ibrahim LF, Hopper SM, Donath S, et al. Development and validation of a cellulitis risk score: The Melbourne ASSET score. *Pediatrics* 2019;143(2):e20181420. DOI: 10.1542/peds.2018-1420.
22. Rizzi AM, Polachek WS, Dulas M, et al. The new 'normal': Rapid adoption of telemedicine in orthopaedics during the COVID-19 pandemic. *Injury* 2020;51(12):2816–2821. DOI: 10.1016/j.injury.2020.09.009.