Clin Orthop Relat Res (2014) 472:1645–1651
DOI 10.1007/s11999-013-3142-0

Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

SURVEY (SYSTEMATIC REVIEW)

Another Look: Is There a Flaw to Current Hip Septic Arthritis Diagnostic Algorithms?

Chika Edward Uzoigwe MRCS

Received: 12 February 2013 / Accepted: 24 June 2013 / Published online: 28 August 2013

Abstract

Background Septic arthritis is an emergency. In 1999 Kocher et al. identified four clinical criteria to distinguish hip septic arthritis from transient synovitis in children (non-weightbearing, erythrocyte sedimentation rate $\geq$ 40 mm/L, white blood cell count $> 12 \times 10^9$/L, temperature > 38.5°C). Subsequent authors evaluating the same criteria produced conflicting results. This calls into question the use of such diagnostic algorithms. The reasons for the differences remain unclear.

Questions/purposes To what degree do studies, evaluating the predictive ability of diagnostic algorithms for septic arthritis, differ with regard to their results? Why do these differences exist? Is there a flaw in the statistical handling of the data?

Methods Using PubMed, original studies evaluating the clinical criteria for distinguishing hip septic arthritis and transient synovitis in children were identified. Clinical and statistical methods were examined.

Results Six studies evaluated the clinical criteria. Two found all four criteria able to distinguish septic arthritis from transient synovitis. There was significant variation between the studies in the risk engendered by the presence of each criteria. The differences were the result of the fact that in all cases, sample sizes were too small and in three cases, there were too few episodes of septic arthritis for a reliable predictive algorithm to be produced.

Conclusions Differing results between studies appear as a result of sample size and insufficient cases of septic arthritis in some cohorts. Transferable and reliable results can be achieved if sufficiently large samples with an adequate number of cases of septic arthritis are recruited.

Introduction

Septic arthritis is an orthopaedic emergency. The diagnosis is a challenge, especially in children, who may not be able to communicate their symptoms. Transient synovitis is a very common but benign phenomenon and mimics septic arthritis. Although septic arthritis must not be neglected, unnecessary surgery, in instances of transient synovitis, is undesirable. As a result of this clinical conundrum, diagnostic algorithms have been developed to help distinguish septic arthritis from transient synovitis. A reliable and transferable algorithm would be a useful clinical instrument.

Kocher et al. [6] published one such possible algorithm in 1999 on the ability of four variables to determine the probability of septic arthritis (nonweightbearing, erythrocyte sedimentation rate [ESR] $\geq$ 40 mm/L, white blood cell count [WCC] $> 12 \times 10^9$/L, temperature > 38.5°C). Five similar studies were performed by subsequent authors to validate these findings [3, 5, 8, 11, 12]. However, there was substantial disagreement among the studies on the probability of septic arthritis depending on the number of diagnostic features present but also on the predictive ability of each criterion [3, 8, 11, 12]. The notable differences

Each author certifies that he or she, or a member of his or her immediate family, has no funding or commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.
All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research* editors and board members are on file with the publication and can be viewed on request.

C. E. Uzoigwe (✉)
Department of Orthopaedics, Northampton General Hospital, 8 Harcourt Crescent, Sheffield S10 1DG, UK
e-mail: chika@doctors.org.uk

across these studies have limited the use of these diagnostic algorithms. The reasons for these differences have not been definitively explored and remain unclear.

To what degree do studies, evaluating the predictive ability of diagnostic algorithms for septic arthritis, differ? Why do these differences exist? Is there a flaw in the statistical handling of the data?

Methods

Search Criteria

Using PubMed, original papers evaluating the four diagnostic criteria for hip septic arthritis in children were identified (nonweightbearing, ESR $\geq$ 40 mm/L, WCC $>$ 12 $\times$ 10^9/L, temperature $>$ 38.5°C). The search term used was "septic arthritis transient synovitis". For each paper, the "related citations in PubMed" facility was explored to determine if there were any similar studies. The initial search term produced 80 results. Six satisfied the inclusion criteria [3, 5, 6, 8, 11, 12]. The "related citations in PubMed" facility did not yield any additional results.

For all the identified studies, the inclusion criteria and diagnostic criteria for septic arthritis and transient synovitis were noted. The statistical handling of data was also explored with particular reference to sample size, number of patients with septic arthritis, and analytical tools. It has previously been suggested the difference between the studies is the result of the ratio of cases of septic arthritis to cases of transient synovitis [12] in the studies. To explore this further, we determined this ratio for each study. We then see if there is a correlation between this and the probability of septic arthritis in the presence of all four predictors using Spearman rank correlation coefficient.

Results

Summary of Considered Studies: To What Degree Do They Differ?

Six studies, satisfying the search criteria, were all performed in the United Kingdom or the United States. The studies differed considerably with regard to parameters that were predictors and the risk engendered by each predictor (Tables 1, 2). The original work by Kocher et al. [6] looked only at those children who underwent joint aspiration for suspected septic arthritis. Those suspected of having septic arthritis were found to have either organisms isolated on aspirate culture or a hip aspirate WCC of more than 50,000/mL. Patients who did not satisfy these criteria were deemed to have transient synovitis. They identified four predictors that could discriminate between septic arthritis and transient synovitis: (1) inability to weightbear; (2) WCC of greater than 12 $\times$ 10^9/L; (3) ESR equal to or greater than 40 mm/hr; and (4) temperature of greater than 38.5°C. If all four predictors were present, Kocher et al. [6] calculated a probability of septic arthritis of 99.6%. Their validation study 5 years later produced similar results [5]. However three predictors resulted in a 93% probability of septic arthritis in the original 1999 [6] study but only 73% in his 2004 article [5] (Table 1). Luhmann

Table 1. Studies evaluating diagnostic algorithms for discriminating between septic arthritis and transient synovitis in children

Study	Predictor variables	Number of variables and probability of septic arthritis					
		0	1	2	3	4	5
Kocher et al., 1999 [6]	WB WCC $>$ 12 $\times$ 10^9, ESR $\geq$ 40 mm/hr, temperature $>$ 38.5°C	$<$ 0.2%	3%	40%	93%	99.6%	
Kocher et al., 2004 [5]	WB WCC $>$ 12 $\times$ 10^9, ESR $\geq$ 40 mm/hr, temperature $>$ 38.5°C	2%	9.5%	35%	73%	93%	
Luhmann et al., 2004 [8]	WB WCC $>$ 12 $\times$ 10^9, ESR $\geq$ 40 mm/hr, temperature $>$ 38.5°C					59%	
Caird et al., 2006 [3]	WB WCC $>$ 12 $\times$ 10^9, ESR $\geq$ 40 mm/hr, temperature $>$ 38.5°C, CRP $\geq$ 20 mg/L	17%	37%	62%	83%	93%	98%
Sultan and Hughes, 2010 [12]	WB WCC $>$ 12 $\times$ 10^9, ESR $\geq$ 40 mm/hr, temperature $>$ 38.5°C, CRP $\geq$ 20 mg/L	2.3%	5%	11%	22%	39%	60%
Singhal et al., 2011 [11]	WB WCC $>$ 12 $\times$ 10^9, Temperature $>$ 38.5°C, CRP $\geq$ 20 mg/L	1%				87%	

Shown are the diagnostic variables examined and the probability of septic arthritis given increasing numbers of predictor variables; WB = whole blood; WCC = white blood cell count; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein.

Table 2. The effect of each predictor variable on the odds of septic arthritis over transient synovitis

Variable	Study and odds ratio				
	Kocher et al. [6]	Kocher et al. [5]	Luhmann et al. [8]	Caird et al. [3]	Singhal et al. [11]
WB	**24.3**	**6**	0.5	3.2	**15**
	($p < 0.001$)	**($p < 0.001$)**	($p = 0.3$)	($p > 0.05$)	**($p < 0.001$)**
WCC $> 12 \times 10^9$	**14.4**	**4**	**3.5**	1.8	1.2
	($p < 0.001$)	**($p < 0.001$)**	**($p = 0.005$)**	($p > 0.05$)	($p < 0.79$)
Temperature > 38.5°C	**38.6**	**4**	**3.3**		3
	($p < 0.001$)	**($p < 0.001$)**	**($p = 0.01$)**		($p = 0.33$)
ESR ≥ 40 mm/hr	**25.9**	**5**	2.3	**7.0**	
	($p < 0.001$)	**($p < 0.001$)**	($p = 0.09$)	**($p < 0.05$)**	
CRP > 20 mg/L				**14.5**	**31**
				($p < 0.05$)	**($p < 0.001$)**

In parentheses are the p values obtained by the authors when they performed logistic regression using the predictors shown. Only Kocher et al. found all four of their predictors (criteria) statistically significant; statistically significant predictors are shown in bold; WB = whole blood; WCC = white cell count; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein.

et al. [8] found that only a temperature of greater than 38.5°C and WCC of greater than 12×10^9/L could discriminate septic arthritis from transient synovitis. If all four of nonweightbearing, WCC $>12 \times 10^9$/L, ESR ≥ 40 mm/hr, and temperature > 38.5°C were present, the probability of septic arthritis was only 59% in Luhmann et al.'s cohort (Tables 1, 2). Luhmann et al. could proffer no explanation for this difference but noted a greater proportion of patients with transient synovitis in her cohort compared with Kocher et al. [6]. Caird et al. [3] found ESR and C-reactive protein (CRP) as the only predictors of septic arthritis over transient synovitis (Table 2). There was significant discordance between the probability of septic arthritis engendered by fewer predictors. Importantly, the work in Sultan and Hughes [12] was methodologically different from the earlier studies [3, 5, 6, 8]. Preceding studies looked only at children who had arthrocentesis for suspected septic arthritis. Sultan and Hughes looked at all children presenting to the institution with an "irritable hip." In their study of 96 patients, only 10 underwent arthrocentesis. Sultan and Hughes' study hence lacked a gold standard that verified the diagnoses of septic arthritis or transient synovitis. They defined children with septic arthritis as those with either a positive culture on hip aspirate or with positive blood culture and "numerous white blood cells" on high-power microscopy of the hip aspirate and no other identified source of infection. What they meant by "numerous white blood cells" was not quantified. The diagnosis of transient synovitis was given to patients who had negative cultures, total resolution of symptoms, and no other identified pathology of the hip. However, all children who were classified with transient synovitis had complete resolution of symptoms without antibiotics. The final study in the series was performed by Singhal et al. [11]. This study was also methodologically different from the antecedent [3, 5, 6, 8]. Those with septic arthritis were those with a positive culture on hip aspirate or with a microscopic abundance of white cells (++ or more per high-power field). The remainder was given the diagnosis of transient synovitis. However, it is noteworthy that no child underwent diagnostic arthrocentesis. Ultrasound was performed in all children. Arthrotomy and lavage were reserved for those with an effusion and whom the treating clinician considered the risk of septic arthritis as likely. The basis for this suspicion is not expounded at all in the text. This nebulous uncertainty weakens the strength of this study. Three hundred eleven children were included. Only 42 underwent arthrotomy. Of these, 24 satisfied the criteria of septic arthritis. The remaining 18 were diagnosed with transient synovitis. All of the 269 who did not undergo arthrotomy had complete resolution of symptoms without antibiotics. The authors assumed that given there was complete recovery without washout and antibiotics, this cohort had transient synovitis. Singhal et al. identified only CRP and weightbearing status as discriminating between septic arthritis and transient synovitis.

Why Do These Differences Exist?

The studies do vary in the ratio of septic arthritis to transient synovitis (Tables 3, 4). Spearman rank correlation analysis, however, showed no relationship between this probability of septic arthritis in the presence of four predictors ($p = 0.10$). All the studies [3, 5, 6, 8, 11] with the exception of Sultan and Hughes [12] use logistic regression to determine the predictive ability of the clinical criteria. With regard to

Table 3. Sample size, number of predictor variables, and number of children with septic arthritis and transient synovitis

Study	Sample size	No variables	No episodes of septic arthritis	No episodes of transient synovitis	Minimum number of septic arthritis cases required
Kocher et al., 1999 [6]	168	4	82	86	40
Kocher et al., 2004 [5]	154	4	51	103	40
Luhmann et al., 2004 [8]	165	4	47	118	40
Caird et al., 2006 [3]	48	5	34	14	50
Sultan and Hughes 2010 [12]	96	5	5	91	50
Singhal et al., 2011 [11]	311	5	29	182	50

This table then goes on to determine the number of events required to support logistic regression.

Table 4. Ratio of transient synovitis (TS) to septic arthritis (SA) in the various studies

Study	TS:SA ratio	Probability of septic arthritis with 4 predictors present
Kocher et al., 1999 [6]	1:1	96.6%
Kocher et al., 2004 [5]	2:1	93%
Luhmann et al., 2004 [11]	2.5:1	59%
Caird et al., 2006 [3]	0.5:1	93%
Sultan and Hughes, 2010 [12]	18:1	39.4%
Singhal et al., 2011 [11]	10:1	87%

This is compared with the risk of septic arthritis when four predictors are present.

absolute number of cases, there is significant variability among the six studies. For example, in Sultan and Hughes, there are only four cases of septic arthritis. It is intuitive that with such few cases, it is difficult to clinically define the behavior of the disease. However, in Caird et al. [3], there are more cases of septic arthritis (34) than transient synovitis (14). Given that the logistic regression requires a specific minimum number of cases to produce reliable results, the difference may lie in any flaws in the statistical handing of the data and/or misuse of logistic regression.

Is There a Flaw in Statistical Handling of the Data?

Although five of the studies [3, 5, 6, 8, 11] use logistic regression, they vary in their sample sizes and in the number of cases of septic arthritis (Tables 1, 2). Hence, they do not appear to comply with any particular objective standard but merely recruit the number of patients available.

The identified differences, which we believe cannot be explained by regional variation (because two studies are from the same site) or differences in the underlying patient populations (because the pathogen profiles are similar), calls for an explanation. Absent such an explanation, the reliability and generalizability of such algorithms are called into question. An explanation may be identified by considering the statistical analyses used and that by doing so, we will be able to identify the kinds of studies that will be needed to settle this issue more definitively.

Discussion

Diagnostic algorithms do not obviate the need for independent thought but provide a framework to assist clinicians in decision-making with potentially life-threatening diagnoses. Transient synovitis mimics septic arthritis, yet the management of the two is very different. An algorithm discriminating between the two would be of some use. The algorithms that have been developed differ considerably in their predictive ability notwithstanding the fact that they use the same criteria. The studies are similar with regard to methodology. With the exception of Sultan and Hughes [12], they use multivariate analysis in the form of logistic regression to determine the effect of predictors. This requires a minimum number of patients and cases of septic arthritis. However, they recruit all the patients presenting within specific time periods rather than with a view to confirming minimum requirements of statistical tests. If we understand the reasons for the difference, we can produce a universally transferable and reliable diagnostic tool. The clinical methods of the studies are largely consonant; hence, an exposition of the authors' statistical methods may identify the source of discrepancy.

The potential limitations of the current study include the fact that I may have overlooked pertinent studies. This is unlikely given the search methodology. The ultimate ideal would be to pool the raw data from all the studies and perform statistically analysis from this much larger pool. One could then determine the nature of differences

between the individual studies and the pooled data. However, it is not possible to perform this analysis on the current data set.

To what degree do studies, evaluating the predictive ability of diagnostic algorithms for septic arthritis, differ with regard to their results?

The aim of all the studies we have considered was to determine the effect of the presence or absence of various predictor variables on the risk of septic arthritis. There are two means of doing this, both used by each group of authors. The first is univariate analysis and the second multivariate analysis. Univariate analysis involves examining each parameter individually. Hence, for each variable, the researcher would look at the proportion of children with septic arthritis who displayed a particular criterion. This then would be compared with the proportion of children with transient synovitis who displayed the same criterion. For example, the proportion of children diagnosed with septic arthritis with a temperature above 38.5°C is compared with the proportion of children with transient synovitis with a temperature of greater than 38.5°C. From this, a relative risk can be calculated. All the authors were cognizant of the limitations of univariate analysis; namely, that variations in one particular parameter may result in variations in another. Hence, the predictive effect of an increase in temperature may be the result of the fact that any increase in temperature is linked to an increase in ESR or CRP or both. For example, children with a higher temperature may be more likely to have septic arthritis than transient synovitis. However, some or all of this increased risk may be the result of the fact that children with higher temperatures also tend to have higher values for CRP and/or ESR. To avoid this issue, all the studies resort to multivariate analysis of potentially confounding variables. Logistic regression is a powerful statistical tool and is a form of multivariate analysis. It examines the effect of each predictor variable while maintaining each of the other predictor variables constant. Hence, it determines the real effect of change in each of the predictor parameters. For example, Kocher et al. in their 1999 [6] study found, with logistic regression, that a WCC of greater than 12×10^9/L increases the odds of septic arthritis by a factor of 14 compared with a WCC below that. This means that for any two children who have the same weightbearing status, ESR, and body temperature, the child with a WCC of greater than 12×10^9/L has a 14-fold increase in the odds of septic arthritis compared with a child with WCC of less than this value (Table 2). All the studies differ with regard to the results of the logistic regression. In so doing, they differ in the parameters that they consider as discriminators between septic arthritis and transient synovitis. They also differ with regard to the predictive effect of diagnostic algorithms.

Why Do These Differences Exist?

Although logistic regression is a powerful tool, it does have limitations. It is contingent on a number of assumptions. If these assumptions are not confirmed in the data under analysis, misleading results may ensue. The use of logistic regression has expanded in the medical literature in all disciplines, not just orthopaedics. In 2012 there were over 19,500 articles in PubMed with logistic regression in the title or abstract. In 2002 this figure was only 5600 [4]. It has been frequently noted that the limitations and assumptions on which logistic regression is dependent are not being respected [1, 10]. Given the widespread failure of appropriate application and hence potentially misleading results, there is published guidance on the appropriate use of logistic regression (and indeed all statistical tests) and the inherent assumptions that data must satisfy before it can be analyzed using this tool [7]. There are several critical assumptions that must be satisfied to perform logistic regression: sample size must be adequate, there must be a sufficient number of events per independent variable, and there must be an absence of collinearity (strong association) among variables. We believe that the wide variation in the predictive value of these diagnostic algorithms is the result of inconsistent and/or inappropriate application of logistic regression. In particular there may be a failure of the data entered to satisfy assumptions inherent to logistic regression.

Is There a Flaw in the Statistical Handling of the Data?

Assumption: Sample Size and Sufficient Events per Independent Variable

Logistic regression requires a relatively large sample sizes if reliable results are to be obtained. Minimum sample sizes of 400 are necessary to obtain accurate odds ratios [2, 9]. Nemes et al. [9] performed simulation studies using logistic regression and various sample sizes. They found that as the sample size increases, the odds ratio results tend to converge to the true value. At sample sizes above 400, the odds ratio had converged significantly to the true value. Similarly, with these large sample sizes, the 95% confidence interval of the odds ratio was very narrow. In contrast, when samples were less than 100, the odds ratio was far from the true value and the 95% confidence interval was very wide, compromising the use of the results. On this

basis, a number of sources advocate sample size of at least 400 to obtain accurate odds ratios [2, 9]. All the studies in my current series have comparatively small sample sizes. The largest is Kocher et al. in 1999 [6] with 168 children. The small sample sizes result in inaccurate odds ratios with very wide confidence intervals. The smaller sample size of subsequent authors may explain why they failed to find all of the statistically significant predictors of septic arthritis. In all of the studies, exact logistic regression should be used in preference to basic logistic regression. This is a logistic regression designed specifically for small sample sizes [15]. None of the studies analyzed here used exact logistic regression.

Now consider a situation in which a researcher is attempting explore the effect of CRP, ESR, WCC, weightbearing status, and body temperature on the risk of septic arthritis in a cohort of patients with septic arthritis and transient synovitis. Imagine, however, if within the study population, there are 298 patients with transient synovitis and two patients with septic arthritis. It is intuitively clear that with so few cases of the septic arthritis, the logistic regression computation program cannot determine the effect of so many variables on the risk of septic arthritis. Simulation studies have shown that for each variable examined, there must be a minimum of 10 of the least common event [1, 10]. If the events per variable are less than this, then odds ratios are inaccurate and confidence intervals too wide to be meaningful. Hence, if a study explores the effect of the weightbearing status, temperature, ESR, and WCC (ie, four variables), there must be at least 40 (four × 10) cases of septic arthritis or transient synovitis, depending on which is the least common of the two. Kocher et al. [5, 6] would require a minimum of 40 cases of septic arthritis to support a logistic regression analysis of their four variables on the probability of septic arthritis in a cohort of patients with septic arthritis and transient synovitis [5, 6] (Table 3). In all the studies with the exception of that of Caird et al., septic arthritis is diagnosed much less frequently than transient synovitis. It is apparent that only three [5, 6, 8] have sufficient events per variable to support logistic regression. Insufficient numbers of events per independent variable adversely affect the accuracy of the risk associated with this variable. It also tends to reduce the likelihood of obtaining statistically significant results.

Sultan and Hughes [12] in their article do not appear to use logistic regression or any form of multivariate analysis to determine probabilities. Rather, they calculate specificity and sensitivity [12]. They attribute the differences in the studies to the differing ratio of the septic arthritis to transient synovitis, which exists between the studies (Table 4). Spearman rank correlation test shows there is no correlation with septic arthritis, transient synovitis ratio, and the predicted probability of septic arthritis in the presence of four predictors ($p = 0.10$). However, direct comparison among all the studies is difficult because some include CRP, whereas others do not.

Assumption: No Collinearity

In logistic regression models, there often is an association between variables. Hence, an increase in body temperature may be associated with an increase in CRP. However, the logistic regression model cannot support two variables that are strongly correlated or have a strong linear relationship [1, 7, 10]. The most extreme example would be if the model included body temperature in Celsius and body temperature measured in Fahrenheit on the risk of septic arthritis. Obviously these are linearly correlated. The result of such a logistic regression would give an erroneous value as to the risk associated with all variables in the model.

Researchers are required to determine if there is a collinear relation between variables before including them in a logistic regression model. None of the authors determined whether there was a strong association between the potential predictors [5, 6, 8, 11, 12]. It is conceivable that there is a strong correlation among CRP, ESR, and WCC, which would compromise the reliability of the odds ratios and probabilities generated by logistic regression. Collinearity would occur, for example, if all or a very large proportion of those with a temperature of greater than 38.5°C temperature criterion) also had an ESR of greater than 40 mm/hr. Another example would be if all those with WCC of greater than 12×10^9/L also had an ESR of greater than 40 mm/hr. These are theoretical potential sources of collinearity. However, ESR and CRP have been found to have a strong correlation in children with monoarticular arthritis and in adults with rheumatoid arthritis [3, 14]. This might have affected the analysis of Caird et al. on this point, because they were the only group to use both ESR and CRP.

Conclusions

The common flaw of these six diagnostic algorithms appears to lie in the statistically handling of the data. All the articles use sample sizes too small to support logistic regression with sufficiently well-defined confidence intervals. Only three [5, 6, 8] of the six studies have sufficient events per variable to support logistic regression analysis. None the studies explores the data for the collinearity, whereas ESR and CRP have been found to be strongly

correlated in previous studies. Hence, Caird et al.'s study, which uses both as predictor variables, may not be able to justify the assumption that variables are not collinear, which is germane to the use of logistic regression. Other variables such as temperature and WCC may also be collinear, which calls into question the reliability of the results of the other studies. The flaws of the algorithms preclude a quantitative meta-analysis.

Using standard logistic regression, any future study should aim to have at least 400 patients. There should be at least 40 patients with a diagnosis of septic arthritis if a four-predictor model is used. In addition, preanalysis checks should be performed to ensure there is no strong correlation between predictor variables. In this way, a reliable and transferable diagnostic algorithm may be formulated, which would be of clinical use.

References

1. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol*. 2001;54:979–985.
2. Bewick V, Cheek L, Ball J. Statistics review 14: logistic regression. *Crit Care*. 2005;9:112–118.
3. Caird MS, Flynn JM, Leung YL, Millman JE, D'Italia JG, Dormans JP. Factors distinguishing septic arthritis from transient synovitis of the hip in children. A prospective study. *J Bone Joint Surg Am.* 2006;88:1251–1257.
4. Home—PubMed NCBI. Available at: http://www.ncbi.nlm.nih.gov/pubmed?term=logistic%20regression. Accessed January 5, 2013.
5. Kocher MS, Mandiga R, Zurakowski D, Barnewolt C, Kasser JR. Validation of a clinical prediction rule for the differentiation between septic arthritis and transient synovitis of the hip in children. *J Bone Joint Surg Am*. 2004;86:1629–1635.
6. Kocher MS, Zurakowski D, Kasser JR. Differentiating between septic arthritis and transient synovitis of the hip in children: an evidence-based clinical prediction algorithm. *J Bone Joint Surg Am.* 1999;81:1662–1670.
7. Lang TA, Secic M. *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers.* 2nd ed. Philadelphia, PA, USA: American College of Physicians; 2006.
8. Luhmann SJ, Jones A, Schootman M, Gordon JE, Schoenecker PL, Luhmann JD. Differentiation between septic arthritis and transient synovitis of the hip in children with clinical prediction algorithms. *J Bone Joint Surg Am*. 2004;86:956–956.
9. Nemes S, Jonasson JM, Genell A, Steineck G. Bias in odds ratios by logistic regression modelling and sample size. *BMC Med Res Methodol*. 2009;27:9:56.
10. Ottenbacher KJ, Ottenbacher HR, Tooth L, Ostir GV. A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions. *J Clin Epidemiol*. 2004;57:1147–1152.
11. Singhal R, Perry DC, Khan FN, Cohen D, Stevenson HL, James LA, Sampath JS, Bruce CE. The use of CRP within a clinical prediction algorithm for the differentiation of septic arthritis and transient synovitis in children. *J Bone Joint Surg Br*. 2011;93:1556–1561.
12. Sultan J, Hughes PJ. Septic arthritis or transient synovitis of the hip in children: the value of clinical prediction algorithms. *J Bone Joint Surg Br*. 2010;92:1289–1289.
13. Thompson A, Mannix R, Bachur R. Acute pediatric monoarticular arthritis: distinguishing Lyme arthritis from other etiologies. *Pediatrics*. 2009;123:959–965.
14. Vastesaeger N, Xu S, Aletaha D, St Clair EW, Smolen JS. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. *Rheumatology (Oxford).* 2009;48:1114–1121.
15. Venkataraman G, Ananthanarayanan V. Demystifying 'exact' logistic regression for pathologists. *J Clin Pathol*. 2008;61:237–238.