

# Comparison of the predictive ability of two scoring systems for the diagnosis of tuberculosis in children

Marina L. Delgado, M.D.<sup>a</sup> and Norma E. González, M.D.<sup>a</sup>

## ABSTRACT

Diagnosing childhood tuberculosis requires a strong diagnostic suspicion due to its paucibacillary manifestation. Different scoring systems and algorithms have been developed to deal with such challenges skillfully. These include the Keith Edwards diagnostic score and Stegen's criteria modified by Toledo (Stegen-Toledo).

**Objectives.** To compare the predictive ability of the Keith Edwards diagnostic score to that of the Stegen-Toledo scoring system. To estimate the predictive diagnostic ability of chest X-rays individually.

**Population, Material and Methods.** Patients seen at the Department of Pediatric Pulmonology of the City of Buenos Aires between 2009 and 2012. A  $\chi^2$  test was used to analyze the predictive diagnostic ability, and performance was compared using the extended McNemar test.

**Results.** Keith Edwards: 19.62% sensitivity, 97.62% specificity ( $p < 0.0001$ ). Stegen-Toledo: 43.54% sensitivity, 97.82% specificity ( $p < 0.0001$ ). Sensitivity was significantly different between both scoring systems ( $p < 0.05$ ), but no differences were observed in terms of specificity. Chest X-ray: 91.15% sensitivity, 87.72% specificity ( $p < 0.0001$ ).

**Conclusion.** The Keith Edwards scoring system showed a lower diagnostic sensitivity than the Stegen-Toledo score in this group of patients, with no differences observed in terms of specificity. The chest X-ray showed a 91.15% sensitivity to diagnose tuberculosis in this group of patients.

**Key words:** tuberculosis, children, score, diagnostic test, chest X-ray.

851 in adolescents were reported to the Tuberculosis Control Program.<sup>2</sup>

Diagnosing childhood tuberculosis requires a strong diagnostic suspicion due to its paucibacillary presentation. Diagnosis is made based on a series of components, such as a positive tuberculin test (PPD), an abnormal chest X-ray, and a history of contact with a tuberculosis source.<sup>3</sup> The problem with obtaining a bacteriological confirmation is that in primary pulmonary tuberculosis the culture of sputum or gastric lavage has a low yield ranging from 30% to 40%.<sup>4</sup> Different scoring systems and diagnostic algorithms have been developed to overcome such challenge and provide physicians with a reasonable tool to identify sick children requiring antituberculosis therapy.<sup>5,6</sup>

One of the most widely-known scales is the Keith Edwards scoring system for the diagnosis of childhood tuberculosis,<sup>7</sup> recommended by the WHO in 1996 in the guidelines for the management of tuberculosis and human immunodeficiency virus (HIV) co-infection (Table 1).<sup>8</sup> Among published diagnostic scores is Stegen's criteria modified by Toledo (Stegen-Toledo) (Table 2).<sup>9</sup> In addition to clinical and epidemiological criteria, this scoring system includes radiological criteria. At present, the WHO does not recommend using the Keith Edwards score<sup>10</sup> or any other diagnostic scores due to their variability and low performance.<sup>11</sup> However, we believe that assessing and comparing their performance may be of great help to identify what elements or variables are most commonly associated with the diagnosis of tuberculosis and

a. Department of Pulmonary Medicine, Hospital General de Niños Pedro de Elizalde.

E-mail Address:  
Marina L. Delgado,  
M.D.: marinaldelgado@hotmail.com

Funding:  
Research fellowship granted by the Sociedad Argentina de Pediatría, 2013.

Conflict of Interest:  
None.

Received: 10-22-2014  
Accepted: 5-11-2015

## INTRODUCTION

Childhood tuberculosis is a major health problem worldwide. In 2012, the World Health Organization (WHO) estimated that more than 500,000 children get sick with tuberculosis every year.<sup>1</sup>

In Argentina, although the total number of people sick with tuberculosis has decreased, the tendency in infectious cases, mainly in young individuals, has not reduced accordingly. In 2012, 901 new cases in children younger than 15 years old and

should therefore be of greater importance when attempting to develop new diagnostic scores.

## OBJECTIVES

To compare the predictive ability of the Keith Edwards diagnostic score to that of the Stegen-Toledo score. To estimate the predictive diagnostic ability of chest X-rays individually.

## MATERIAL AND METHODS

**Design:** observational, analytical, cross-sectional, retrospective study. Population: patients aged between 1 month old and 19 years old diagnosed and treated for tuberculosis, and non-sick patients in contact with tuberculosis receiving chemoprophylaxis and who had started being follow-up at the Department of Pulmonary Medicine of Hospital Pedro de Elizalde between January 1<sup>st</sup>, 2009 and December 31<sup>st</sup>, 2012. The reference standard was the diagnosis of tuberculosis by a medical team of pediatric pulmonologists based on the presence of two or more of the following criteria: epidemiological (close contact with a tuberculosis patient in the past two years), immunological (positive PPD test:  $\geq 10$  mm in immunocompetent patients or  $\geq 5$  mm in immunocompromised patients), clinical (some of the following factors: 15-day history of cough, fever, fatigue and/or weight loss, hemoptysis, malnutrition, meningitis with

compatible cerebrospinal fluid [CSF], lymph node enlargement, abdominal mass, ascites, pericarditis, bone tumor and/or central nervous system [CNS] tumor with no other apparent cause, lack of response to broad spectrum antibiotic therapy), radiological (some of the following factors: condensation, atelectasis,

TABLE 2. Stegen's criteria modified by Toledo

Criteria	Score
Finding of Koch's bacilli	7
Specific granuloma (histological)	4
Positive tuberculin test*	3
Epidemiological history <sup>†</sup>	2
Suggestive X-ray <sup>‡</sup>	2
Suggestive clinical picture <sup>§</sup>	2

\* Induration larger than 10 mm in diameter in immunocompetent patients, or larger than 5 mm in immunocompromised patients.

<sup>†</sup> Close contact or living with a tuberculosis patient in the past two years.

<sup>‡</sup> Presence of pneumonia-like alveolar infiltrate, atelectasis, pleural effusion, hilar adenopathy or miliary pattern.

<sup>§</sup> More than 15 days of history of one or more of the following symptoms: cough, fever, malaise, weight loss, hemoptysis, or organ involvement (e.g., seizure).

$\geq 7$ : definitive diagnosis. Treatment should be initiated.

5-6: feasible diagnosis, treatment may be initiated.

4-3: possible diagnosis, additional tests may be done.

$\leq 2$ : no tuberculosis.

TABLE 1. Keith Edwards scoring system for the diagnosis of childhood tuberculosis

General characteristics		Score
Duration of illness	<2 weeks	0
	2-4 weeks	1
	>4 weeks	3
Nutrition (% of weight for age)	>80%	0
	60%-80%	1
	<60%	3
Family history of tuberculosis	None	0
	Reported by a family member	1
	Confirmed positive sputum bacilloscopy	3
Local findings	Unexplained fever or night sweats	2
	Positive tuberculin test	3
	Malnutrition that does not improve after four weeks of treatment	3
	Lack of response to broad spectrum antibiotic therapy	2
	Painless lymph node enlargement	3
	Joint or bone inflammation	3
	Abdominal mass or ascites	3
	Abnormal cerebrospinal fluid	3
Angle deformity of the spine	4	

$\geq 7$ : highly probable tuberculosis.

5-6: probable tuberculosis; may warrant treatment.

3-4: monitoring and test repetition.

0-2: no tuberculosis.

effusion, hilar adenopathy, miliary pattern), bacteriological (positive sputum bacilloscopy and/or culture for *Mycobacterium tuberculosis* done on the study specimen) and/or anatomical-pathological (granuloma with caseous necrosis). Information was collected based on data recorded in these patients' office medical records.

**Variables:** age, sex, diagnosis and components of both scores were recorded (*Table 1* and *Table 2*). This information corresponded to data present at the time of diagnosis and treatment or chemoprophylaxis initiation. For both scoring systems, a score  $\geq 7$  (positive) indicated a highly probable diagnosis of tuberculosis.

Medical records with insufficient information to apply any of the scores were excluded.

### Statistical analysis

The sample size estimated was 613 patients based on data from a reference study conducted in South Africa,<sup>12</sup> for an estimated ratio of 53% of cases detected by the Stegen-Toledo scoring system (this is higher than the ratio obtained in said study by the Keith Edwards scoring system), for a 3% accuracy level, and a 95% confidence level.

The  $\chi^2$  test was used to analyze the sensitivity, specificity, positive predictive value and negative predictive value of both scoring systems.

A multiple logistic regression analysis was done using the SPSS software to identify independent predictors of the Stegen-Toledo score.

The statistical extended McNemar test, as described by Hawass, was used to assess differences in diagnostic precision (sensitivity and specificity) between both scoring systems and between Stegen-Toledo and chest X-ray (as single diagnostic element).<sup>13</sup>

### Ethical considerations

This study was authorized by the Head of the Outpatient Clinic, the Bioethics Committee, and the Teaching and Research Committee of Hospital General de Niños Pedro de Elizalde. We hereby state that all information included in this study is true, and that the study was conducted in accordance with the Declaration of Helsinki, the International Good Clinical Practice Standards, and applicable national and local regulations regarding clinical research.

### RESULTS

Medical records corresponding to the 2009-2012 period were reviewed. Out of a total of 954 medical records of tuberculosis, 32 were excluded because data were insufficient for the implementation of a diagnostic score.

Out of 922 patients, epidemiological criteria were observed in 726 (78.7%); immunological criteria, in 265 (28.7%); clinical criteria, in 247 (26.8%); radiological criteria, in 443 (48%); bacteriological criteria, in 97 (10.5%); and anatomical-pathological criteria, in 22 (2.4%). *Table 3* shows the characteristics of the study population.

*Table 4* details the predictive diagnostic ability of the Keith Edwards score, the Stegen-Toledo score, and chest X-ray considered as a single diagnostic element.

*Figure 1* shows the percentage of tuberculosis patients diagnosed by the different scoring systems.

The extended McNemar test evidenced a general significant difference in terms of sensitivity and specificity between the Keith Edwards and Stegen-Toledo scores,  $p < 0.05$ . Sensitivity of both scoring systems was compared separately using the McNemar test or correlated

TABLE 3. Characteristics of the study population

	Tuberculosis patients	Contact with tuberculosis	Total
Included patients	418	504	922
Hospitalized	157	21	178
Male	211	251	462
Median age	7.5 $\pm$ 5.34	7 $\pm$ 5.13	7.08 $\pm$ 5.24
<5 years old	150	190	340
5 to 14 years old	206	265	471
15 to 19 years old	62	49	111
HIV+	11	6	17
Extrapulmonary tuberculosis	68	-	-
Severe tuberculosis	141	-	-

HIV: human immunodeficiency virus.

proportions, which showed a significant difference between both scales,  $p < 0.05$  (95% CI: 0.24-0.29). However, no significant differences were observed in both tests' specificity,  $p = 0.7410$ .

A multiple logistic regression analysis that included the elements that make up the Stegen-Toledo score (Table 2) allowed to observe that both clinical and radiological aspects behaved as independent predictors for the rest of the score,  $p < 0.05$  in both cases (Table 5).

The extended McNemar test evidenced a general significant difference in terms of sensitivity and specificity between the Keith

Edwards and Stegen-Toledo scores,  $p < 0.05$ . Sensitivity of both scoring systems was compared separately using the McNemar test or correlated proportions, which showed a significant difference between both scores,  $p < 0.05$  (95% CI: 0.45-0.50). However, no significant differences were observed for both tests' specificities,  $p = 0.7431$ .

## DISCUSSION

Given the diagnostic difficulties observed in childhood tuberculosis, it is necessary to study the usefulness of available tools, such as

TABLE 4. Results of predictive diagnostic ability

	p	OR	95% CI	Se	Sp	PPV	NPV	LR+
Keith Edwards	<0.0001	10.01	5.37-18.63	19.62	97.62	87.23	59.42	8.24
Stegen-Toledo	<0.0001	34.56	18.44-64.77	43.54	97.82	94.30	67.63	19.95
X-ray	<0.0001	73.58	47.88-113.06	91.15	87.72	86	92.29	7.42

Se: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; OR: odds ratio; p: significance level of  $\chi^2$  test.

FIGURE 1. Percentage of tuberculosis patients diagnosed based on the different scoring systems

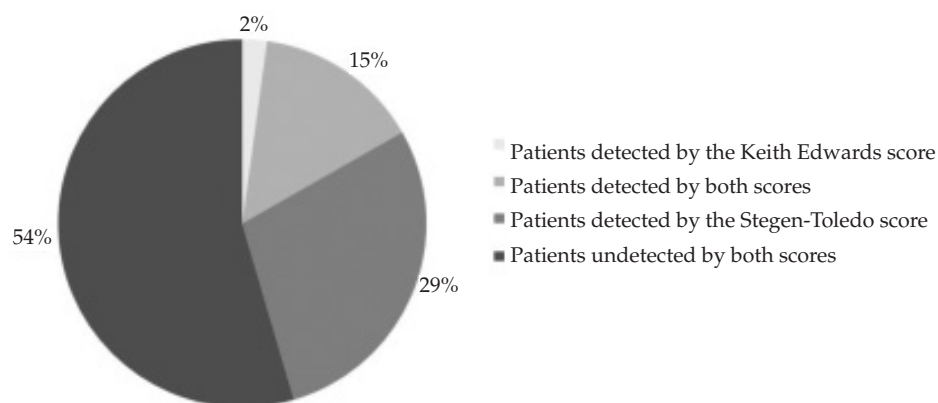


Table 5. Multivariate analysis including all components of the Stegen-Toledo scoring system

	p	OR	95% CI	
			Lower	Upper
Epidemiological criteria	0.366	1.43	0.66	3.10
X-ray	0.000	83.19	50.42	137.24
Clinical criteria	0.000	5.00	2.40	10.44
Microbiological confirmation	0.997	985221272.94	0.00	-
Specific granuloma	0.997	5863494468.32	0.00	-
PPD	0.055	1.69	0.99	2.88

PPD: tuberculin test.

the diagnostic scores analyzed in this study. According to our results, the Stegen-Toledo score was more sensitive than the Keith Edwards score in this group of patients and, although the specificity of the latter is higher, this information may not be statistically significant.

In a prospective study conducted in South Africa<sup>12</sup> in a population of children younger than two years old who received the BCG vaccine between 2001 and 2006, 1445 children with a history of contact with tuberculosis or clinical characteristics compatible with tuberculosis were included. Different diagnostic approaches were applied on these patients (in addition to supplementary diagnostic methods, such as X-ray, microbiological research, PPD, HIV serology) to assess agreement among them. Among these patients, 611 (42%) had a tuberculosis diagnosis based on clinical data and supplementary studies. The Keith Edwards score diagnosed 175 (10%) cases, while the Stegen-Toledo score, 772 (53.4%), with statistically significant differences in terms of frequency ( $p < 0.0001$ ) among the three groups. Consistency between the Stegen-Toledo score and clinical diagnosis was higher than with the Keith Edwards score (85.3% and 64.4%, respectively, and this was a significant difference,  $p < 0.005$ ). This lower consistency for the Keith Edwards score probably is due to the fact that this scale assigns a high diagnostic score to clinical characteristics present in chronic diseases, severe malnutrition and extrapulmonary tuberculosis, which may be more commonly observed in hospitalized patients, not in children with suspected disease included in the context of an immunization program. In this case, the score would have a very low sensitivity.

In the original Keith Edwards score study, sensitivity was much higher, approximately 80%-90%, because the group of patients analyzed in that study had clinical characteristics with a high risk for tuberculosis, especially extrapulmonary. Inclusion criteria were unexplained fever for  $\geq 7$  days plus one of the following characteristics: malnutrition and fatigue, abdominal mass or ascites, coma, generalized/partial seizures or focal neurological deficit, lymph node enlargement. In addition, the study was conducted in areas of India and Africa with a high prevalence of tuberculosis and HIV.<sup>7,14,15</sup>

In the original study and those that continued with the Stegen-Toledo criteria, the priority was to improve the diagnostic sensitivity, so data

that may be present in completely asymptomatic patients were included in assessed variables.<sup>9,16</sup>

The group of patients included in our study was more heterogeneous than those of the above mentioned studies. We included both asymptomatic or oligosymptomatic patients who attended an outpatient clinic (as in the case of the South African study), severe patients hospitalized with extrapulmonary presentations and greatly affected general status (as in the original study for the Keith Edwards score). This may explain why, when analyzing consistency between clinical diagnosis and both scores, consistency is 73.13% for Stegen-Toledo and 61.21% for Keith Edwards, with no significant differences between both values ( $p = 0.072$ ). In addition, the fact that we collected data from outpatient medical records probably enabled us to obtain a better specificity for both scores, especially for the Keith Edwards scoring system, because patients with characteristics of severe disease hospitalized with other diagnoses were not included in this study because they had not been followed-up by the tuberculosis service.

In order to assess what components of the Stegen-Toledo score had a stronger association with the diagnosis of tuberculosis, a multiple logistic regression analysis was done using the variables included in this scoring system; findings evidenced that X-ray and clinical criteria behaved as independent predictors. Considering that the latter variable is present in both diagnostic scores, it could be inferred that differences in sensitivity may be due to data collected in relation to X-ray criteria, with an OR of 83 for X-rays and of 5 for clinical criteria (Table 5).

Given that sensitivity results for the Stegen-Toledo score were similarly low, the next step to diagnose tuberculosis in this group of patients was to assess whether using X-ray alone as a diagnostic procedure would achieve greater sensitivity and specificity than the full score itself. This was confirmed by the sensitivity shown by chest X-ray: 91.15%, with a significant difference when compared to Stegen-Toledo's sensitivity ( $p < 0.05$ ).

In a retrospective study conducted in Lima (Peru),<sup>16</sup> chest X-ray had a 90.4% sensitivity. This was also clearly relevant in the diagnosis of tuberculosis, although the study demonstrated that an X-ray with positive findings did not account for a significantly relevant variable to make a definitive diagnosis using the Stegen-Toledo score, probably due to the low value

assigned to such variable. As in this study, X-ray behaved as an independent statistically significant variable of the scale.

In a prospective study that assessed tuberculosis prevalence and screening methods conducted in Kenya,<sup>17</sup> 20,566 individuals aged 15 to 60 years old were assessed using a survey on tuberculosis symptoms, chest X-ray, and sputum bacilloscopy. A sputum culture was requested in suspected cases, and the reference standard was the presence of bacteriological confirmation. In these patients, the sensitivity of the chest X-ray as a single diagnostic factor was higher than the symptom survey (100% and 82%, respectively, in non HIV-positive patients). The multiple logistic regression analysis in this study also showed that the X-ray behaved as an independent predictor (OR: 32, 95% CI: 16.3–64.2,  $p < 0.0001$ ).

In our study, in relation to the population co-infected with HIV, a positive serology was not statistically and significantly associated with the diagnosis of tuberculosis, with both systems with a  $\geq 7$  score, or with a chest X-ray with positive findings. The number of cases included probably does not allow to draw conclusions in this regard.

It is worth noting that the results of this study may not be extrapolated because they were obtained in a population attending a pulmonology clinic, therefore, in general they were patients who had been in contact with tuberculosis, and most had compatible clinical criteria or a positive PPD test as a result of assessments made by pediatricians or other specialists. This may be the reason why a chest X-ray with positive findings had a significant diagnostic predictive value and power.

Lastly, it should be noted that a weakness of this kind of assessments is that studies comparing different diagnostic scores published to date have evidenced great disagreement among one another and their predictive value may not be established in the absence of a single reference standard.

## CONCLUSION

The Keith Edwards scoring system for the diagnosis of tuberculosis evidenced a low diagnostic sensitivity in this group of patients. The Stegen-Toledo score showed a higher sensitivity, but no statistically significant differences were observed in terms of specificity.

The chest X-ray behaved as an independent variable within the Stegen-Toledo scoring system.

X-ray sensitivity as a single diagnostic element was higher than that of the full Stegen-Toledo score. ■

## REFERENCES

1. World Health Organization. Global Tuberculosis Report 2013. Geneva, 2013. [Accessed on: September 15, 2014]. Available at: [http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf).
2. Fernández HR, Arias SJ, Garcilazo DA. Situación de la Tuberculosis. Argentina 2011-2012. Santa Fe: Administración Nacional de Laboratorios e Institutos de Salud; 2012. [Accessed on: September 15, 2014]. Available at: [http://www.anlis.gov.ar/iner/wpcontent/uploads/2014/05/SituaciónTb\\_2012\\_Web.pdf](http://www.anlis.gov.ar/iner/wpcontent/uploads/2014/05/SituaciónTb_2012_Web.pdf).
3. Sociedad Argentina de Pediatría. Comité Nacional de Neumonología, Comité Nacional de Infectología. Criterios de diagnóstico y tratamiento de la tuberculosis infantil. *Arch Argent Pediatr* 2002;100(2):159-78.
4. Sant'Anna CC, Orfalais CT, March Mde F. A retrospective evaluation of a score system adopted by the Ministry of Health, Brazil in the diagnosis of pulmonary tuberculosis in childhood: a case control study. *Rev Inst Med Trop Sao Paulo* 2003;45(2):103-5.
5. Stegen G, Jones K, Kaplan P. Criteria for guidance in the diagnosis of tuberculosis. *Pediatrics* 1969;43(2):260-3.
6. Nair PH, Philip E. A scoring system for the diagnosis of tuberculosis in children. *Indian Pediatr* 1981;18(5):299-303.
7. Edwards K. The diagnosis of childhood tuberculosis. *P N G Med J* 1987;30(2):169-78.
8. Harries AD, Maher D. TB/HIV: a clinical manual. Geneva: World Health Organization; 1996.
9. Toledo A, Katz F, Montiel J, Rico FG. Criterios de diagnóstico en tuberculosis infantil. *Rev Mex Pediatr* 1979;46:239-43.
10. Rieder HL, Chen-Yuan C, Gie RP, Enarson DA. Crofton's clinical tuberculosis. 3<sup>rd</sup> ed. Oxford: Macmillan; 2009.
11. World Health Organization. Guidance for national tuberculosis programmes on the management of tuberculosis in children. 2<sup>nd</sup> ed. Geneva, 2014. [Accessed on: May 12, 2015]. Available at: [http://www.who.int/tb/publications/childtb\\_guidelines/en/](http://www.who.int/tb/publications/childtb_guidelines/en/).
12. Hatherill M, Hanslo M, Hawkrigde T, Little F, et al. Structured approaches for the screening and diagnosis of childhood tuberculosis in a high prevalence region of South Africa. *Bull World Health Organ* 2010;88(4):312-20.
13. Hawass NE. Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients. *Br J Radiol* 1997;70(832):360-6.
14. Van Beekhuizen HJ. Tuberculosis score chart in children in Aitape, Papua New Guinea. *Trop Doct* 1998;28(3):155-60.
15. Van Rheeën P. The use of paediatric tuberculosis score chart in an HIV-endemic area. *Trop Med Int Health* 2002;7(5):435-41.
16. Vega Aguilar MD. Evaluación de los criterios de Stegen y Toledo en niños con tuberculosis pulmonar en el Hospital de Pichanaki 2000-2011. Lima: Universidad Nacional Mayor de San Marcos; 2012. [Accessed on: May 12, 2015]. Available at: [http://ateneo.unmsm.edu.pe/ateneo/bitstream/123456789/2608/1/Vega\\_Aguilar\\_Manuel\\_Delfin\\_2012.pdf](http://ateneo.unmsm.edu.pe/ateneo/bitstream/123456789/2608/1/Vega_Aguilar_Manuel_Delfin_2012.pdf).
17. Van't Hoog AH, Meme HK, Laserson KF, Agaya JA, et al. Screening strategies for tuberculosis prevalence surveys: the value of chest radiography and symptoms. *PLoS ONE* 2012;7(7):e38691.