Clinical Prediction Rules for Identifying Children With Testicular Torsion

A Multicenter Prospective Study

Marcela Pía Valdivieso-Castro, MD, Lorena Vázquez-Gómez, MD,† Mikel Olabarri, MD,‡ Iván Presno-López, MD,§ Rocío Espinosa-Góngora, MD,‖ Anna Orejuela-Ribera, MD,¶ Amaia Cámara-Otegui, MD,# Sandra Rodríguez-Fernández, MD,** Myriam López-Rojo, MD,†† Laura Marfil-Godoy, MD,‡‡ Carmen Medina-Esquitino, MD,§§ Marta Soriano-Arola, MD,‖‖ Elena Aquino, MD,¶¶ Elena Faci, MD,## María Amalia Pérez-Sáez, MD,*** Alba Henares-Rodríguez, MD,††† Ignacio Javier Navarro- Lopez, MD,‡‡‡ Estefanía Romero-Castillo, MD,§§§ Irene Garralda-Torres, MD,‖‖‖ Zulema Lobato-Salinas, MD,¶¶¶ Amaia Lopez-Oreja, MD,### Carmen María Hinojosa-Mateo, MD,**** and Santiago Mintegi, MD, PhD‡*

Objectives: To validate clinical scores [Testicular Workup for Ischemia and Suspected Torsion (TWIST), testicular torsion (TT) score, Artificial Intelligence–based Score (AIS), Boettcher Alert Score (BALS)] when evaluating children under 18 with non-traumatic testicular pain in the emergency department. Our secondary objective was to create and compare a new TT score [Testicular Emergency Score for Torsion (TEST) score].

Methods: This was a multicenter prospective study in 21 Spanish pediatric emergency departments between 2020 and 2022, including 903 children 3 months to 18 years old with non-traumatic unilateral testicular pain, of them 93 TT (10.3%). To create a new score, the sample was randomly divided into derivation and validation set.

Results: The performance of the TWIST, TT score, AIS, and BALS was good, and the proportion of patients correctly classified as low risk was 37.9%, 52.7%, 30.3%, and 28%, respectively. The TEST score included the following predictors of TT identified by multivariable logistic regression analysis: age, duration of pain, nausea/vomiting, testicular volume increase, testicular elevation, induration, and absence of cremasteric reflex. TEST score had a higher area under the receiver operating curve (area under the curve) and correctly classified in the low-risk group of 63.6% of the patients.

Conclusions: Although TWIST, TT score, BALS, and AIS scores showed a good performance, the TEST score identifies a larger group of low-risk patients suitable for safe management without Doppler ultrasound.

Key Words: testicular torsion, ultrasound, testicular pain

(*Pediatr Emer Care* 2025;00:000–000)

Received for publication October 21, 2024; accepted March 13, 2025.

From the *Department of Pediatric Surgery, Cruces University Hospital, Barakaldo, Basque Country; †Department of Pediatrics, Ferrol University Hospital; ‡Department of Pediatric Emergency, Cruces University Hospital, University of the Basque Country, Barakaldo, Basque Country; §Teresa Herrera Hospital, A Coruña, Spain; ‖Department of Pediatric Surgery, Niño Jesús University Children's Hospital, Madrid; ¶Department of Pediatrics, Hospital Universitari Joan XXIII de Tarragona; #Department of Pediatric Emergency, Donostia University Hospital, San Sebastián, Spain; **Cabueñes University Hospital. Gijón, Spain; ††Department of Pediatrics, Hospital Universitario Rio Hortega, Valladolid; ‡‡Hospital Universitari Arnau Vilanova, Lleida; §§Hospital Consorci Sanitari Terrassa, Barcelona; ‖‖Pediatric Emergency, Department. Son Espases Hospital. Palma de Mallorca, Spain; ¶¶Hospital Universitario de Toledo; ##Department of Pediatrics, Hospital de Barbastro; ***Osakidetza Basque Health Service, Zumarraga Hospital, Pediatric Service, Zumarraga, España; †††Hospital Virgen de las Nieves, Granada; ‡‡‡Department of Pediatrics, Hospital Universitario Infanta Sofía, San Sebastián de los Reyes, Madrid; §§§Hospital Puerta del Mar; ‖‖‖Navarra University Hospital. Pamplona, Spain; ¶¶¶Althaia Foundation, Xarxa Assistencial i Universitària. Manresa, Spain; ###Osakidetza Basque Health Service, Mendaro Hospital, Pediatric Service, Mendaro, España; and ****Hospital del Tajo.

During the preparation of this work, the authors have only used generative AI and AI-assisted technologies for the translation of the manuscript.

Disclosure: The authors declare no conflict of interest

Reprints: Santiago Mintegi, MD, PhD, Department of Pediatric Emergency, Cruces University Hospital, University of the Basque Country, Plaza de Cruces s/n, Cruces-Barakado (Bizkaia), Barakaldo, Basque Country 48903, Spain (e-mail: santiago.mintegiraso@osakidetza.eus).

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.pec-online.com.

DOI: 10.1097/PEC.0000000000003394

Non-traumatic unilateral testicular pain is a common reason for pediatric consultations in emergency departments (EDs). While most cases are caused by benign conditions, it is crucial to rule out testicular torsion (TT) due to the risk of testicular loss. Testicular viability is highly dependent on the duration of symptoms.[1–3] Therefore, a prompt and accurate diagnosis of TT is essential.

Patients typically present with the sudden onset of severe testicular pain, often accompanied by nausea and vomiting.[4–6] On physical examination, the affected testis is usually slightly elevated, tender, and firm, often positioned horizontally. In addition, the cremasteric reflex is absent in nearly all cases. However, these findings can also occur in other acute scrotal conditions, such as torsion of the

appendix testis, epididymo-orchitis, and others.[6] In many children, the clinical presentation may be ambiguous, necessitating the use of color Doppler ultrasound for confirmation. However, the accuracy of Doppler ultrasound depends heavily on the operator's expertise, and its availability is not guaranteed in all EDs 24 hours a day, 7 days a week.

Several clinical scoring systems have been developed to guide initial decision-making in children presenting with testicular pain, aiming to minimize the misclassification of cases of TT. In 2014, Barbosa and colleagues introduced the Testicular Workup for Ischemia and Suspected Torsion (TWIST) score, a 7-point tool designed to assess acute scrotal pain. The score incorporates key clinical features such as testicular swelling, hard testis, absence of the cremasteric reflex, nausea or vomiting, and a high-riding testis. Based on these factors, patients are categorized into low, intermediate, and high-risk groups. The TWIST score demonstrated excellent performance when initially applied by urologists.[7]

Subsequent validations by non-urologist ED physicians confirmed its utility but revealed some cases of TT being misclassified in the low-risk category.[8–10] Despite this limitation, the TWIST score has proven to be a reliable tool, reducing reliance on scrotal ultrasound. Studies report a missed torsion rate of 1.6 per 100 acute scrotum presentations.[11] Following the TWIST score, other scoring systems have been developed, including the Boettcher Alert Score (BALS), the Artificial Intelligence–based Score (AIS), and the TT score.[12,13] However, prospective, independent, multicenter validations of these newer scores remain lacking.

The primary objective of this study was to evaluate the performance of various clinical scoring systems (TWIST, BALS, AIS, and TT score) in assessing children presenting to the ED with non-traumatic unilateral testicular pain.

The secondary objective was to identify independent risk factors for TT and to develop and validate a new scoring system—referred to as the Testicular Emergency Score for Torsion (TEST)—to improve the identification of TT in children presenting to the ED with non-traumatic unilateral testicular pain.

METHODS

This cohort study included children aged 3 months to 18 years who presented with non-traumatic unilateral testicular pain to 21 pediatric EDs that are part of the Spanish Pediatric Emergency Research Group (RISEUP-SPERG).

A diagnosis of TT was confirmed either through surgical intervention or by the presence of Doppler flow on ultrasound following successful manual detorsion. The decision regarding the type of intervention was at the discretion of the surgeon of each participating center.

Patients with a history of scrotal trauma or prior ipsilateral scrotal surgery (eg, orchidopexy, surgical removal of the appendix testis, and epididymal cyst) were excluded from the study.

Data Collection and Variable Selection

We included patients evaluated between January 2020 and June 2022.

Initial assessments for all children were conducted by either a resident (in pediatrics, surgery, or family and community medicine) or an attending physician (a pediatrician or pediatric surgeon) on call at the time of the ED visit. No child was initially evaluated by a urologist.

For all included patients, data were collected on variables previously associated with TT in the literature. These data were recorded on structured data sheets and included the following:

- Demographics and clinical history: Age, sexual activity, previous episodes of testicular pain.
- Pain characteristics: Duration and intensity of testicular pain, assessed using an age-adjusted pain scale:
- For children younger than 3 years, the "Face, Legs, Activity, Cry, Consolability" Scale.
- For children aged 3 to 7 years, the Faces Pain Scale–Revised.
- For children older than 7 years, the Visual Analog Scale.
- Associated symptoms: Nausea or vomiting, abdominal pain, fever, and dysuria.
- Physical examination findings: Location of pain, scrotal swelling, position of the testicle, testicular enlargement, firmness, cremasteric reflex, and presence of a blue dot sign.
- Diagnostics and management: Final diagnosis, treatment (manual detorsion or urgent surgery), patient disposition, and results of any additional investigations performed, including testicular Doppler ultrasound, urine dipstick tests, urine culture, and emergency surgical evaluations.

This systematic approach allowed for the standardized collection of relevant clinical information across all participating centers.

Data Collection Process

Electronic questionnaires were completed through Google Drive for all children included in the study. No identifiable data was recorded in Google Drive.

Before the study's initiation, the questionnaires, along with a study manual, were distributed to site investigators (ED physicians) to ensure the clarity and appropriateness of the content for data collection. This step allowed investigators to review and verify the comprehensibility of the text and ensure that the data collection process would be accurate and consistent.

Once completed, the questionnaires were electronically submitted to the principal investigator (M.P.V.C.). The scores for the various clinical assessment tools were calculated by the principal investigator based on the data provided (Supplemental Table 1, Supplemental Digital Content 1, http://links.lww.com/PEC/B421).

To identify any potential misdiagnoses, patients were followed up by telephone 2 weeks after their ED visit, with the family's prior consent (Supplemental Table 2, Supplemental Digital Content 2, http://links.lww.com/PEC/B422). Follow-up calls were conducted by the designated investigator at each participating center.

Statistical Analysis

Comparison of Existing Scores

To evaluate the diagnostic accuracy of the clinical scores, we calculated their sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In addition, we recorded the number of TT cases that were missed by each score.

For each score, the cutoff point for the low-risk group was determined as the value at which no cases of TT were

misclassified. Given that the primary goal of identifying a low-risk group is to minimize the need for ultrasound and ensure that no torsion cases are overlooked, an NPV of 100% was required. We also compared the proportion of patients categorized as low-risk by each scoring system.

Development of a New Score

To develop a TT predictor score, we divided the data into 2 sets: 70% was used to create the score (derivation set), and the remaining 30% was reserved for evaluation (validation set). Patients with any missing data were excluded from both the derivation and validation sets.

In the derivation set, we employed a multivariate logistic regression model, using TT as the outcome variable and related variables as predictors. Continuous variables were transformed into categorical ones by applying a cut-off point that maximized the Youden Index. Variables independently associated with TT were ranked according to the methodology proposed by Zhang et al.[14]

The performance of the score was assessed using the validation set. We utilized the area under the receiver operating characteristic curve (area under the curve) to evaluate its discriminatory ability. To determine the diagnostic accuracy, sensitivity, NPV, and negative likelihood ratio were calculated.

In previous studies, a Twist score of 5 points or higher was associated with a risk of TT exceeding 50%, serving as a high-risk cutoff point.[11] To ensure no cases of TT were missed, the new score's cutoff points were established as follows: low risk (0%), intermediate risk (1% to 50%), and high risk ($>50\%$).

Comparison of the New Score With Previous Scores

We compared the sensitivity, specificity, PPV, and NPV of the various scores. In addition, we assessed the missed TT rate (per 100 cases) in the low-risk group, as well as the negative exploration rate (per 100 cases) in the high-risk group for each score. For this analysis, we assumed that low-risk patients would be excluded from further intervention, high-risk patients would proceed to surgical exploration, and intermediate-risk patients would require additional evaluation with ultrasound.

The cost-saving analysis was conducted based on the billing rates for health care services for the year 2020 at our institution. Costs are expressed in euros, and the net savings were calculated based on avoiding additional diagnostic tests in low-risk patients.

Estimation of the Sample Size

At the hospital of the principal investigator, we recorded one child with non-traumatic testicular pain or inflammation for every 345 ED presentations in 2018 (unpublished data). Among these children, ~4.4% were diagnosed with TT and underwent surgery (equating to 1 case per 7800 ED presentations). We were not aware of the scores for all these patients. Nonetheless, considering that misidentifying children with TT is a significant error, we estimated that we should recruit at least 100 patients with TT. This estimation suggests a total of 780,000 ED presentations, which would result in ~2260 presentations for testicular pain or inflammation, ultimately leading to around 100 patients with TT.

All analyses were conducted using SPSS version 23 (IBM SPSS Statistics, IBM Corporation).

Ethics

The Clinical Research Ethics Committee of the Basque Country and the Institutional Review Boards of each participating institution approved the study. Consent was obtained from participants involved in the study. In addition, patients aged 6 to 12 years provided verbal consent, while patients over 12 years of age signed consent forms.

RESULTS

During the study period, we recorded 1299 episodes involving children with non-traumatic unilateral testicular pain evaluated at the 21 participating EDs. Of these, 903 patients (69.5%) were included in the study. The primary reason for non-recruitment was that the attending physician failed to offer participation (339 patients), while 13 patients refused to sign the informed consent, and 44 patients had other reasons for exclusion. We contacted 884 patients (97.8%) to inquire specifically about the need for surgical treatment after the ED visit; no child required surgical intervention. TT was diagnosed in 93 cases (of them, 82 were confirmed through surgical intervention and 11 through the presence of Doppler flow on ultrasound after manual detorsion).

Out of the total sample, 88 patients underwent urgent surgery. In 82 cases, TT was confirmed; 5 presented with alternative diagnoses (2 with torsion of the appendix testis, 2 with resolved intermittent TT, and 1 unspecified), while 1 case involved a normal testicle that underwent orchidopexy. The performance of ultrasound showed a sensitivity of 99% (95% CI: 97-100), specificity of 99% (95% CI: 98-99), PPV of 90% (95% CI: 84-96), and NPV of 100% (95% CI: 99-100). The median age of the patients was 11 years (interquartile range: 9 to 12), and the median duration of testicular pain was 14 hours (interquartile range: 4 to 48). A total of 93 patients (10.3%) were diagnosed with TT. The characteristics of the included patients are displayed in Table 1.

Performance of the Scores

We compared the accuracy of previous scores in Table 2, focusing on sensitivity and NPV using the cutoff points for each score where no TT cases were misclassified, thus achieving an NPV of 100%, as well as the area under the receiver operating characteristic curve. By applying these cutoff points, the TT score was able to identify a larger group of patients without TT. Detailed data are provided in Supplemental Table 3 (Supplemental Digital Content 3, http://links.lww.com/PEC/B423).

Development of a New Score

We created 2 subsamples from the overall database, including only those patients without missing values (765 patients). The most frequent missing item was pain intensity, which was recorded in 768 cases. Pain intensity is not included in previous scores, and in our data set, it did not correlate with TT. The first subsample (derivation set), used to generate the scale, consisted of 540 children (70%). We tested the scale in the second subsample (validation set, 225 patients, 30%).

We performed a multivariate analysis in the derivation set, identifying 7 independent risk factors for TT: age 1 year or younger or 10 years or older, duration of pain <6 hours, nausea and vomiting, increased testicular volume, testicular elevation, induration, and absence of the cremasteric reflex (Table 3). These 7 factors were incorporated into the TEST

TABLE 1. Characteristics of Patients Included in the Study

	Total (903); n/N (%)	No TT (810); n/N (%)	TT (93); n/N (%)	*P*
Age, median (IQR)	11 (9-12)	11 (8-12)	13 (12-14)	< 0.001
≤ 1 or ≥ 10 years old	624/903 (69.1)	534/810 (65.93)	90/93 (96.8)	< 0.001
Sexual activity	7/903 (0.8)	5/810 (0.6)	2/93 (2.2)	0.15
Previous episode	151/903 (16.7)	139/810 (17.2)	12/93 (12.9)	0.37
Duration of pain (h), median (IQR)	14 (4-48)	15 (6-48)	4 (2-13.5)	< 0.001
Short duration (< 6 h)	249/903 (27.6)	197/810 (24.3)	52/93 (55.9)	< 0.001
Pain intensity (median)	5.53	5 (4-7)	8 (6-8)	< 0.001
Severe pain (8-10)	146/768 (19)	104/686 (15.2)	42/82 (51.2)	< 0.001
Sudden pain	353/903 (39.1)	286/810 (35.3)	67/93 (72.1)	< 0.001
Nausea/vomiting	88/903 (9.8)	40/810 (4.9)	48/93 (51.6)	< 0.001
Abdominal pain	165/903 (18.3)	128/810 (15.8)	37/93 (39.8)	< 0.001
Fever	4/899 (0.4)	4/806 (0.5)	0/93	1
Dysuria	48/902 (5.3)	46/809 (5.7)	2/93 (2.2)	0.14
Testicular swelling	396/903 (43.9)	323/810 (39.9)	73/93 (78.5)	< 0.001
High riding testis	208/903 (23)	136/810 (16.8)	72/93 (77.4)	< 0.001
Blue dot	39/903 (4.3)	36/810 (4.4)	3/93 (3.2)	
Hard testis	177/903 (19.6)	99/810 (12.2)	78/93 (83.9)	< 0.001
Pain localization				< 0.001
Epididymis	85/903 (9.4)	82/810 (10.1)	3/93 (3.2)	—
Upper pole	279/903 (30.9)	270/810 (33.3)	9/93 (9.7)	—
Whole testis	539/903 (59.7)	458/810 (56.4)	81/93 (87.1)	—
Absent cremasteric reflex	193/903 (21.4)	127/810 (15.7)	66/93 (71)	< 0.001
Positive Prehn sign	139/813 (17.1)	121/727 (16.6)	18/86 (20.1)	0.88

IQR indicates interquartile range; TT, testicular torsion.

score, with values assigned based on the β-coefficient (Fig. 1). The resulting range of the TEST score was 0 to 11 points, with the following cut-off points established: ≤ 3 for low risk of TT (no cases of TT, 64.6% of the patients), 4 to 7 for intermediate risk (20 TT cases out of 147 patients, 27.2%), and ≥ 8 for high risk (35 TT cases out of 44 patients, 8.2%). The distribution of patients for each TEST score value is presented in Supplemental Table 3 (Supplemental Digital Content 3, http://links.lww.com/PEC/B423).

Comparison of Testicular Emergency Score for Torsion With Previous Scores in the Validation Set

All scores demonstrated strong performance, and the comparison of their abilities can be viewed in Figure 2. The stratification systems for the different scores are shown in Table 4. In our series, the cutoff points for identifying children with non-traumatic unilateral testicular pain were as follows: Twist < 1, BALS < 1, AIS < 1, TT score ≤ 2, and TEST score ≤ 3. With these cutoff points, the rates of patients identified as low risk were: Twist 37.9%, BALS 30.3%, AIS 28%, TT score 52.7%, and TEST score 63.3%.

We reviewed the telephone follow-up of the 200 patients in whom TT was ruled out without performing an ultrasound. Of them, we were unable to contact 5 of them. Among the remaining cases, none were finally diagnosed with TT.

Using this approach and not performing ultrasounds on TEST low-risk patients, a potential cost saving of €8636.92 (9065.76$) was achieved, equating to ~€92.62 (97.25$) per patient treated for testicular pain.

DISCUSSION

The TWIST, TT, AIS, and BALS scores demonstrate strong performance in evaluating children with non-traumatic unilateral testicular pain in the ED. To our knowledge, this is the largest multicenter prospective study assessing clinical scores for TT. Furthermore, we developed a new score (TEST) that incorporates age and duration of pain, effectively identifying a larger group of patients who do not have TT. Although this was not the primary

TABLE 2. Comparison of TWIST, TT, BALS, and AIS

Score	No. Cases (%)	Sensitivity, % (95% CI)	NPV, % (95% CI)	Area Under the Receiver Operating Curve (AUC), % (95% CI)
Twist = 0	342 (37.9)	100 (95.1-100)	100 (98.6-100)	0.94 (0.92-0.96)
TT score ≤ 2	476 (52.7)	100 (96.1-100)	100 (99-100)	0.96 (0.94-0.97)
BALS = 0	274 (30.3)	100 (95.1-100)	100 (98.3-100)	0.89 (0.86-0.93)
AIS = 0	253 (28)	100 (95.1-100)	100 (98.1-100)	0.92 (0.89-0.94)

AIS indicates Artificial Intelligence–based Score; AUC, area under the curve; BALS, Boettcher Alert Score; NPV, negative predictive value; TT, testicular torsion; TWIST, Testicular Workup for Ischemia and Suspected Torsion.

TABLE 3. Univarite and Multivariate Analysis in the Derivation Set of Predictive Factors for TT

	Univariate Analysis		Multivariate Analysis		
	OR (95% CI)	***P***	**OR (95% CI)**	***P***	**B-coefficient**
Age ≤1 or ≥10 years old	7.8 (2.8-33.3)	<0.001	6.5 (1.7-36.5)	0.015	1.88
Short duration of pain (<6 h)	4.6 (2.6-8.3)	<0.001	4.2 (1.6-11.7)	0.004	1.44
Severe pain (8-10)	6.1 (3.4-11)	<0.001	—	—	—
Sudden pain	6 (1.2-144)	<0.001	—	—	—
Nausea/vomiting	20.5 (10.5-40.9)	<0.001	5.4 (1.9-16.6)	0.002	1.7
Abdominal pain	3 (1.6-5.5)	<0.001	—	—	—
Fever	1 (0.05-18.1)	1	—	—	—
Dysuria	0.2 (0.01-2.7)	0.2	—	—	—
Testicular swelling	6.6 (3.4-14.3)	<0.001	3 (1.1-8.9)	0.039	1.1
High riding testis	21.1 (10.8-45)	<0.001	7.3 (3-18.7)	<0.01	1.98
Hard testis	29 (14.4-64)	<0.001	10.8 (4.3-29.6)	<0.001	2.38
Pain localization: whole testis	7.97 (1.7-190)	<0.001	—	—	—
Absent cremasteric reflex	11.3 (6.2-21.3)	<0.001	4.1 (1.7-10.5)	0.002	1.42

Hosmer-Lemeshow 3.89 (P = 0.8669)/area under the curve = 0.969 (0.954-0.985).
OR indicates odds ratio; TT, testicular torsion.

objective, a more selective use of ultrasound can lead to cost savings.

Ultrasound is commonly used to rule out TT, even in children who are at very low risk. While it is a safe procedure, overuse can extend ED stays unnecessarily and may result in transfers to facilities where ultrasound is available. Many of these patients initially visit their primary care centers, from which they are referred to hospital EDs for ultrasound. We anticipate that in centers lacking skilled ultrasonography, the data presented in Supplemental Table 3 (Supplemental Digital Content 3, http://links.lww.com/PEC/B423) can assist in decision-making. In addition, for patients with TT, imaging may cause delays of ∼1 hour,[15] which could affect the viability of testicle.

The classic symptoms and signs of TT are not always present and may overlap with other testicular pathologies.[16] Although we have identified items independently associated with TT (Table 2), none are exclusive to TT (Table 1), and the predictive value of each in isolation is low. Therefore, different combinations of signs and symptoms have been tested through clinical scoring methods to optimize the performance of medical history and physical examination.

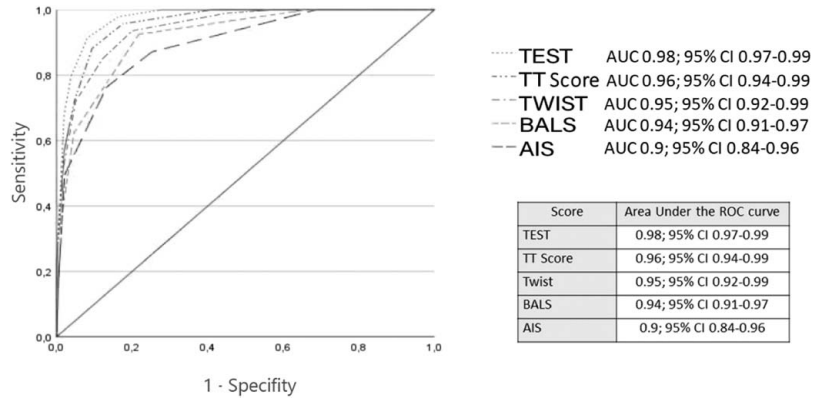
Despite the favorable performance of the scores, several critical details warrant attention. In our study, when applying the cutoff points proposed by Barbosa et al[7] for the TWIST score, some TT cases were misclassified as low risk. Specifically, 6 patients with TWIST scores ≤2 were subsequently diagnosed with TT (5 scoring 2 and one scoring 1), all of whom were older than 10 years and presented to the ED within the first 6 hours of pain onset. This misclassification aligns with prior findings.[11] Such occurrences may increase physicians' confidence in treating children in EDs and subsequently increase Doppler ultrasound utilization. Other researchers[8] have proposed lower cut-off points (TWIST <1) to avoid misclassification of TT cases; however, this approach decreases the number of children categorized as low risk. Adjusting cutoff points could lead to overfitting, where performance appears exceptional only in validation data sets. We firmly believe that for TT cases, achieving a NPV of 100% should be the ultimate aim. This strategy was initially employed by Barbosa et al in their preliminary studies, though subsequent research did not confirm a 100% NPV in the low-risk category. In this context, the TT score, which includes age, appears to perform better than the TWIST score. Further validation is necessary to either confirm these observations or suggest necessary adjustments to the cutoff points.

TEST (Teste's Emergency Score for Torsion) SCORE	
≤1 years old or ≥ 10 years old	2
Nausea/vomiting	2
High riding testis	2
Hard testicle	2
Testicular swelling	1
Absent cremasteric reflex	1
Duration of pain<6h	1

FIGURE 1. TEST score. TEST indicates Testicular Emergency Score for Torsion.

Incorporating new variables such as age and duration of pain—both significantly associated with TT—into a new score (the TEST score) enhances clinical score performance. The TEST score outperforms previous scores (TWIST, BALS, TT score, and AIS) and identifies a greater number of children at very low risk for TT who can be safely managed without Doppler ultrasound. Moreover, the TEST score is straightforward for emergency physicians to apply, leading to reduced ultrasound usage in low-risk patients—thereby decreasing costs and shortening ED stays—and minimizing testicular ischemia time in high-risk patients.

Nonetheless, Doppler ultrasound remains essential for evaluating selected children with non-traumatic unilateral testicular pain in the ED. Contrary to prior reports, we found no significant delays in entrance to the operating room among patients who underwent ultrasound compared with those who did not; patients with a strong suspicion of TT were likely prioritized. This suggests that, within our setting, performing an ultrasound did not impact testicular ischemia and effectively discriminated TT cases. Accordingly, all children classified in the intermediate-risk group should promptly receive a Doppler ultrasound due to the prevalence of TT among patients in this category.[7] Notably, 3 patients classified as intermediate risk by the TEST scale required orchiectomy, all presenting with pain lasting longer than 12 hours. Thus, if Doppler ultrasound is unavailable or there is any doubt, surgical exploration is the preferred option to prevent testicular loss. This consideration is vital



Score	Area Under the ROC curve
TEST	0.98; 95% CI 0.97-0.99
TT Score	0.96; 95% CI 0.94-0.99
Twist	0.95; 95% CI 0.92-0.99
BALS	0.94; 95% CI 0.91-0.97
AIS	0.9; 95% CI 0.84-0.96

FIGURE 2. Comparison of the performance of the different scores for the evaluation of children with unilateral non-traumatic testicular pain. AIS indicates Artificial Intelligence–based Score; BALS, Boettcher Alert Score; TEST, Testicular Emergency Score for Torsion; TT, testicular torsion; TWIST, Testicular Workup for Ischemia and Suspected Torsion.

when applying scores in settings without Doppler ultrasound.

Conversely, patients classified within the high-risk group should be evaluated in the operating room or manually detorsed, if Doppler ultrasound is not readily available. In our setting, Doppler ultrasounds are mainly done by radiologist physicians, who are on-site 24 hours a day in tertiary hospitals. However, some smaller hospitals have surgical on-site services 24 hours a day, but not a radiologist. In these cases, bedside ultrasound can be a useful tool, but in the absence of trained personnel, surgery or manual detorsion may be reasonable in high-risk patients. Compared with previous scores, the TEST score appears to be more selective. While TEST scores include fewer children in the high-risk category, they also exhibit the lowest rate of negative explorations, thus mitigating associated risks of general anesthesia and potential complications from scrotal surgery (such as infections or hematomas) in children without TT.[17–19]

Manual detorsion is a noninvasive technique that, when performed early by trained professionals, can improve outcomes.[20,21] In our series, 11 patients were treated with this technique, confirming the restoration of blood flow on ultrasound. It is worth noting that in one patient, a previous ultrasound was not performed; this patient presented with severe, recent-onset pain and scored the maximum on all described scales, with rapid clinical improvement following detorsion. The practice of manual detorsion is not consistent among the participating centers, and standardizing this technique is an improvement action we should consider.

An important aspect to consider in future studies is the effect of training on the performance of the scoring system by the physicians applying it. In our study, no pre-study training was conducted, as the variables in question are routinely assessed in these patients. However, a recent Australian study indicated improvements in testicular examination and documentation following a training period involving educational posters in the ED. Thus, it would be valuable to evaluate the impact of such training on score performance.[22]

This study has several limitations. During the study period, 394 eligible patients were not recruited, primarily because the attending physician did not offer recruitment, a common occurrence in many studies conducted in EDs. However, the duration of pain and the final diagnosis were similar to those observed in the overall sample (8.6% of patients diagnosed with TT). In addition, 138 patients were excluded from the development of the new score due to missing items required for multivariate analysis. Nonetheless, we believe that the final sample size of 765 is sufficient for our purposes, and no differences were observed in the TT rate.

A significant limitation of our study is the lack of a gold standard, as ultrasound was not performed on all patients. To avoid altering the usual practices of some participating centers and to limit patient recruitment, we opted to address this limitation by conducting telephone follow-ups. Despite this, there may be cases of misclassification where TT was incorrectly ruled out, particularly intermittent TTs that were not present at the time of ED evaluation. This situation continues to pose a challenge for emergency physicians, radiologists, and surgeons, as certain symptoms or signs may be absent during clinical assessment, sonography, or even surgical intervention. It is important to emphasize that no clinical scoring system can replace clinical experience. We believe that the high rate of telephone follow-ups, which did not reveal any cases of undetected torsion requiring surgical intervention, significantly mitigates this concern. The risk of misclassification is particularly elevated in patients who did not undergo ultrasonography. Among the 200 patients without ultrasonography, we were only unable to contact 5, and none of the remaining patients were ultimately diagnosed with TT. In the future, it would be valuable to validate our score with external samples, as we are confident that incorporating age and pain duration into the scoring model will enhance the performance of existing scores. Furthermore, the study was conducted in 21 Spanish EDs. It is possible that the characteristics of patients presenting with testicular pain may vary in other countries, potentially influencing the performance of clinical scores. Therefore, the TEST score should be applied with caution in other regions until external validation is achieved. However, we believe that the main elements included in the scores are likely to be similar in countries with comparable social and health care conditions. In contrast, we did not assess the performance of the score based on the experience of the evaluating physician. In our series, the score was utilized by various practitioners, including trainees (residents) and attending physicians from different medical specialties (pediatrics, pediatric emergency medicine, pediatric surgery, and

TABLE 4. Comparison of Different Scores' Stratification Systems in the Validation Set

	Low Risk			Intermediate Risk		High Risk			
	Sensitivity (95% CI) Specificity (95% CI)	Patients at Low Risk (%)	Missed Torsion Rate (Per 100 Presentations)	Patients at Intermediate Risk (%)	Ultrasound Rate (Per 100 Presentations)	Sensitivity (95% CI) Specificity (95% CI)	Patients at High Risk (%)	Negative Exploration Rate (Per 100 Presentations)	AUC (95% CI)
TWIST (0,1-4,5-7) *	100 (87.2-100) 41.9 (35-49.1)	36.9	0	49.3	49.3	77.8 (57.7-91.4) 95 (90.9-97.6)	13.7	4.4	0.95 (0.92-0.99)
TT (0-2, 3-5, 6-9) *	100 (87.2-100) 58.1 (50.8-65)	51.1	0	36	36	74.1 (53.7-88.9) 95.4 (91.6-97.9)	12.9	4	0.96 (0.94-0.99)
BALS (0,1, 2-4) *	100 (87.2-100) 33.3 (26.8-40.4)	29.3	0	38.7	38.7	66.7 (46-83.5) 77.3 (70.8-82.9)	28	20	0.94 (0.91-0.97)
AIS (0, 1-2, 3-5) *	100 (87.2-100) 30.3 (24-37.2)	26.7	0	51.6	51.6	74.1 (53.7-88.9) 85.4 (79.6-90)	21.8	12.8	0.9 (0.84-0.96)
TEST (0-3,4-7,8-11) *	100 (87.2-100) 72.2 (65.4-78.3)	63.6	0	25.8	25.8	77.8 (57.7-91.4) 98.5 (95.6-99.7)	10.7	1.3	0.98 (0.97-0.99)

Low-risk patients are ruled out with no further intervention. High-risk patients are ruled in and proceed to surgical exploration. Intermediate-risk patients require further workup with ultrasound. As all intermediate-risk patients receive ultrasound, "patients at intermediate risk (%)" and "ultrasound rate (per 100 presentations)" are equivalent.

*In brackets, values of the different scores for low, intermediate and high risk for TT.

AIS indicates Artificial Intelligence–based Score; AUC, area under the curve; BALS, Boettcher Alert Score; TEST, Testicular Emergency Score for Torsion; TT, testicular torsion; TWIST, Testicular Workup for Ischemia and Suspected Torsion.

community medicine). It would be beneficial to investigate whether differences exist based on the evaluator's specialty and years of experience, as well as the inter-rater agreement among different evaluators.[7] Finally, we did not do a concrete analysis of time savings per patient, which would have provided clearer evidence of the benefits for decision-makers.

CONCLUSIONS

The TWIST, TT score, BALS, and AIS scoring systems demonstrate good performance in identifying children with TT. The new TEST score, which incorporates age and duration of pain, can identify an even larger group of patients at low risk for TT, making them suitable for safe management without the need for Doppler ultrasound. Testing the impact of a short training program for proper application of the TEST score could enhance its clinical adoption and widespread use. Further multicenter, prospective validation studies are required to confirm our findings.

ACKNOWLEDGMENTS

The authors thank Borja Santos Zorrozua, PhD Bioinformatics, Biostatistics, And Information System Unit. Biocruces Bizkaia Health Research Institute, Hospital Universitario Cruces, Barakaldo, Bizkaia, Spain, for collaborating in the statistical design, carrying out the analyses and interpretation of results.

REFERENCES

1. Lian BS, Ong CC, Chiang LW, et al. Factors predicting testicular atrophy after testicular salvage following torsion. *Eur J Pediatr Surg*. 2016;26:17–21.
2. Bartsch G, Frank S, Marberger H, et al. Testicular torsion: late results with special regard to fertility and endocrine function. *J Urol*. 1980;124:375–378.
3. Mellick LB, Sinex JE, Gibson RW, et al. A systematic review of testicle survival time after a torsion event. *Pediatr Emerg Care*. 2019;35:821–825.
4. Kass EJ, Lundak B. The acute scrotum. *Pediatr Urol*. 1997;44:1251–1266.
5. Karmazyn B, Steinberg R, Kornreich L, et al. Clinical and sonographic criteria of acute scrotum in children: a retrospective study of 172 boys. *Pediatr Radiol*. 2005;35:302–310.
6. Kadish HA, Bolte RG. A retrospective review of pediatric patients with epididymitis, testicular torsion, and torsion of testicular appendages. *Pediatrics*. 1998;102:73–76.
7. Barbosa JA, Tiseo BC, Barayan GA, et al. Development and initial validation of a scoring system to diagnose testicular torsion in children. *J Urol*. 2013;189:1859–1864.
8. Sheth KR, Keays M, Grimsby GM, et al. Diagnosing testicular torsion before urological consultation and imaging: validation of the TWIST score. *J Urol*. 2016;195:1870–1876.
9. Frohlich LC, Paydar-Darian N, Cilento BG, et al. Prospective validation of clinical score for males presenting with an acute scrotum. *Acad Emerg Med*. 2017;24:1474–82.
10. Manohar CS, Gupta A, Keshavamurthy R, et al. Evaluation of testicular workup for ischemia and suspected torsion score in patients presenting with acute scrotum. *Urol Ann*. 2018;10:20–23.
11. Qin KR, Qu LG. Diagnosing with a TWIST: systematic review and meta-analysis of a testicular torsion risk score. *J Urol*. 2022;208:62–70.
12. Klinke M, Elrod J, Stiel C, et al. The BAL-score almost perfectly predicts testicular torsion in children: a two-center cohort study. *Front Pediatr*. 2020;8:601892.
13. Lim X, Angus MI, Panchalingam V, et al. Revisiting testicular torsion scores in an Asian healthcare system. *J Pediatr Urol*. 2020;16:821.e1–821.e7.
14. Zhang Z, Zhang H, Khanal MK. Development of scoring system for risk stratification in clinical medicine: a step-by-step tutorial. *Ann Transl Med*. 2017;5:436.
15. Pinar U, Duquesne I, Lannes F, et al. The use of doppler ultrasound for suspected testicular torsion: lessons learned from a 15-year multicentre retrospective study of 2922 patients. *Eur Urol Focus*. 2022;8:105–111.
16. Mellick LB. Torsion of the testicle: it is time to stop tossing the dice. *Pediatr Emerg Care*. 2012;28:80–86.
17. Swartz MA, Morgan TM, Krieger JN. Complications of scrotal surgery for benign conditions. *Urology*. 2007;69:616–619.
18. Smith C, Godbole P. Suture technique and complications following paediatric testicular fixation. *Int Arch Urol Complic*. 2018;4:039.
19. John M. Hutson, chapter 77 - undescended testis, torsion, and varicocele In: Arnold G, eds. *Coran, Pediatric Surgery*, 7th edn. Mosby; 2012: 1003–1019. ISBN 9780323072557
20. Cabral Dias Filho A, Rincon Cintra da Cruz P, Zanettini Riccetto CL. Testicular torsion patients should be manually detorsed at diagnosis: a propensity score matched analysis of the influence of interhospital transfer and surgical wait times on surgical organ salvage. *Pediatr Emerg Care*. 2022;38:e936–e942.
21. Dias Filho AC, Oliveira Rodrigues R, Riccetto CL, et al. Improving organ salvage in testicular torsion: comparative study of patients undergoing vs not undergoing preoperative manual detorsion. *J Urol*. 2017;197(pt 1):811–817.
22. Cheng EM, Chui JN, Crowe M, et al. Improving testicular examinations on paediatric patients in the emergency department: a quality improvement study to improve early diagnosis of testicular torsion. *Asian J Urol*. 2022;9:152–156.