Reliability and utility analysis of the obstetric-gynecologic appropriateness of the SET Sistema Estructurado de Triaje® using clinical simulation

Análisis de fiabilidad y utilidad de la adecuación obstétrico-ginecológica del Sistema Estructurado de Triaje® (SET) mediante simulación clínica

M.ª Pilar Costa-Gil[1,2], Mireia Subirana-Casacuberta[2], Alexandra García-Muriana[1], Beatriz Fernández-Parralejo[3], Martina Pérez-Pérez[2], Ester Risco[2]

Obstetric triage is a key component of ensuring safe and timely care in maternal and child health within emergency departments.[1-3] The use of structured prioritization systems improves resource allocation and reduces delays in care,[4,5] although the available specific scales have limitations in terms of validity and reliability.[6-13] At the Consorci Corporació Sanitària Parc Taulí (CCSPT) in Sabadell (Barcelona), the Structured Triage System® (SET) has been used since 2009,[14] with triage performed by midwives. After analyzing its performance in more than 63,000 obstetric triage encounters, we observed that the assigned priority level did not determine when comprehensive clinical care was initiated and found that a high proportion of high-priority classifications did not correspond to indirect indicators of healthcare complexity. This suggests possible overestimation of case severity. Based on these findings, a modified version of the SET adapted to obstetric and gynecological emergencies was developed through a literature review and expert consensus. The objective of this study was to evaluate the reliability of this new version in a clinical simulation setting[15] and assess its perceived usability among health care professionals.

We conducted an observational, explanatory psychometric study in a computer-based simulation setting[15] between October and December 2024 at the CCSPT in Sabadell (Spain). The study included a total of 50 representative obstetric and gynecological emergency cases developed from the clinical experience of the research team. A total of 32 health care professionals (midwives, resident midwives, nurses, and obstetricians) with clinical experience and prior training in the new tool participated. They triaged the same cases at 2 different time points (T1 and T2). An expert consensus classification, established by a panel including the simulation instructors and the developer of the SET, was used as the reference standard (TS). A preproduction digital version incorporating the newly agreed decision

rules into the software was available, allowing triage assessments to be performed and recorded automatically. Interprofessional agreement or reliability was evaluated in three comparisons: T1 vs TS, T2 vs TS, and T1 vs T2. Intraprofessional reliability was assessed by analyzing the stability of measurements when the same professional encountered the same situation at a different time.

Data were collected on the characteristics of case reports and participating professionals, as well as the triage assessments performed (triage level). Statistical analyses were performed using R and RStudio software, versions 4.5.9 and 2025.05.0 Build 496, respectively (Vienna, Austria). Agreement percentages were used to quantify agreement between raters: exact modal agreement (EMA) and modal agreement with a tolerance of ±1 level (MA1). Cohen's and Fleiss' quadratically weighted kappa coefficients were used to assess interobserver and intraobserver agreement. The weighted kappa coefficient proposed by van der Wulp and van Stel was also considered, as it provides an alternative weighting scheme that accounts for the differences associated with misclassification by two or more triage levels. The 95% confidence interval (95% CI) was calculated to establish the precision limits of the estimates. Kendall's correlation coefficient was analyzed to measure the degree of agreement between ordinal classifications assigned by different raters. Participants also completed a questionnaire assessing the usability of the new software application (Likert scale). The study was approved by the Clinical Research Ethics Committee of the CCSPT (approval code: 2021/5026).

A total of 3,166 triage assessments were analyzed out of the 3,200 expected (Table 1). The mean age of the participating professionals was 35 years (SD, 11), and their mean professional experience was 12 years (SD, 10).

High EMA and MA1 values were obtained in all 3 comparisons (Tables 2 and 3). The study demonstrated high interprofessional reliability (overall kappa coefficients > 0.90), with low rates of over- and undertriage. The kappa coefficients calculated using the weighting scheme proposed by van der Wulp and van Stel were slightly lower (kappa > 0.89) but still indicated high reliability (Table 3).

The analysis demonstrated high reproducibility or intraprofessional reliability (Table 4), as indicated by the mean kappa coefficient (0.933; SD, 0.069) and Kendall's correlation coefficient (0.913; SD, 0.083). Most participants achieved agreement lev-

Table 1. Description of the variables used in the reliability tests (clinical cases and professionals)

Characteristics of clinical cases (N: 50)	Freq.	(%)
Distribution by ranges		
Obstetric	37	74.0
< 22 weeks	15	40.5
≥ 22 weeks and < 37 weeks	15	40.5
≥ 37 weeks	7	19.0
Puerperal: Puerperium (24 hours to 6 weeks after childbirth)	3	6.0
Gynecological	10	20.0
Distribution by maternal triage level (TS)		
Level I	3	6.0
Level II	18	36.0
Level III	14	28.0
Level IV	12	24.0
Level V	3	6.0
Distribution of reasons for consultation		
Shock	1	2.0
Other, unspecified problems and complications affecting the fetus	3	6.0
Loss or bleeding in early pregnancy, unspecified	4	8.0
Low mood and depression	2	4.0
Maternal hypertension, unspecified	3	6.0
Abdominal pain	2	4.0
General malaise, unspecified	1	2.0
Excessive vomiting during pregnancy	4	8.0
Convulsions	1	2.0
Loss or bleeding in advanced pregnancy, unspecified	3	6.0
Preeclampsia, unspecified	1	2.0
Urinary symptoms accompanied by fever and lower back or abdominal pain	1	2.0
Dyspnea in a previously healthy patient	1	2.0
Pruritus (itching)	1	2.0
Maternal care for insufficient fetal growth due to known or suspected placental insufficiency	1	2.0
Signs and symptoms of the onset of labor between 22 and 36 weeks of gestation	1	2.0
Preterm rupture of membranes	1	2.0
Preterm labor	1	2.0
Out-of-hospital delivery without placental expulsion	1	2.0
Decreased or absent fetal movements	2	4.0
Labor in progress	1	2.0
Signs and symptoms of the onset of labor at or after 37 weeks of gestation	1	2.0
Vaginal and/or uterine hemorrhage	2	4.0
Dyspareunia (painful intercourse)	1	2.0
Pelvic pain (acute or chronic) related to a neoplasm	1	2.0
Pelvic symptoms and pain in women	2	4.0
Amenorrhea, unspecified	1	2.0
Other types of vaginal and vulvar inflammation, unspecified		1.0
Unspecified breast lump	1	2.0
Postpartum hemorrhage	1	2.0
Headache related to anesthesia in the postpartum period	1	2.0
Mastitis associated with breastfeeding	1	2.0
Menstrual symptoms involving excess: excessive or frequent menstruation, menometrorrhagia, hypermenorrhea, polymenorrhea	1	2.0
Obstetric risk classification		
Normal/low	27	54.0
Medium	7	14.0
High	14	28.0
Very low	2	4.0
Location/Destination		
Waiting area	11	22.0
Emergency area – operating room	2	4.0
Examination area – medical care	13	26.0
Observation area – fetal monitoring	11	22.0
Observation area – maternal monitoring	9	18.0
Observation area – treatment	1	2.0
High-risk and very high-risk delivery area	2	4.0
Moderate- and low-risk/low-intervention delivery area	1	2.0

(Continued)

Table 1. Description of the variables used in the reliability tests (clinical cases and professionals) (Continuation)

Participant characteristics (n = 32)	Freq.	(%)
Professional category		
Midwife (M)	9	28.1
Resident midwife (RM)	8	25.0
Nurse (N)	12	37.5
Physician (MOb)	3	9.4
Age		
20–29 years	14	44.0
30–39 years	7	22.0
40–49 years	7	22.0
≥ =″ νεαρσ	4	12.0
Years of experience		
≤ = νεαρσ	12	38.0
6–9 years	7	22.0
10–20 years	4	12.0
> 20 years	9	28.0

Standard or theoretical triage: the clinical classification established and agreed upon by the expert team under the supervision of the SET developer (Dr. Gómez Jiménez).

els close to 100% when a tolerance of one triage level was allowed.

The professionals' overall assessment of the tool was favorable. Participants highlighted its ease of use, clarity in decision-making, and applicability to clinical practice, suggesting good acceptance of the system and appropriate integration into the triage process.

Analysis of the professionals' characteristics revealed no statistically significant differences (ANOVA) in agreement levels according to age, experience, or professional category.

The results of this study demonstrate a high degree of interprofessional agreement, a strong correlation with the reference standard (expert opinion), and high stability of intraprofessional responses. These findings are particularly relevant in a context in which the available specific triage tools have shown variable reliability.[1,2,12,13]

The algorithmic structure of the SET, based on predefined rules and interconnected decision trees, may help reduce variability in clinical interpretation and promote greater consistency in the assignment of urgency levels. Furthermore, participants' prior training supported by clinical simulation probably contributed to the high agreement observed. Both training and simulation are considered by the available evidence to be key elements in optimizing the performance of triage systems.[1,3,5,15]

The high intraprofessional reproducibility observed in this study indicates that the system maintains stability over time for the same rater, an essential property of structured triage tools.[7-10] This is particularly important for implementation in different healthcare settings, where decisions must be made rapidly and consistently.

Table 3. Statistical assessment of interobserver agreement or reliability: consistency of measurements performed by different observers in the 3 situations studied

Metrics	T1 vs TS	T2 vs TS	T1 vs T2
Percentage agreement	EMA: 91.62% MA1: 98.67%	EMA: 90.12% MA1: 96.75%	EMA: 97.56% MA1: 96.06%
Overall kappa (95% CI)[a]	0.947 (0.935-0.959)	0.926 (0.911-0.941)	0.924 (0.909-0.939)
Level I	0.954 (0.923-0.986)	0.952 (0.920-0.985)	0.933 (0.893-0.972)
Level II	0.907 (0.886-0.929)	0.903 (0.881-0.925)	0.874 (0.849-0.900)
Level III	0.866 (0.838-0.893)	0.840 (0.810-0.870)	0.797 (0.763-0.830)
Level IV	0.917 (0.894-0.942)	0.888 (0.861-0.914)	0.884 (0.857-0.910)
Level V	0.866 (0.814-0.918)	0.812 (0.725-0.873)	0.828 (0.770-0.885)
Alternative kappa (95%CI)[b]	0.927 (0.913-0.942)	0.909 (0.893-0.925)	0.892 (0.874-0.909)

T1: first triage assessment; T2: second triage assessment; TS: standard triage. EMA: exact modal agreement; MA1: modal agreement with a tolerance of ±1 level.
[a]Overall weighted kappa coefficient (Cohen's kappa) and weighted kappa coefficient for each level treated as a binary category.
[b]Alternative kappa coefficient developed by van der Wulp and van Stel.

Table 2. Comparison of triage level assignments

Triage level	I	II	III	IV	V	Total
T1 vs TS						
I	89	5	0	0	0	94
II	2	526	29	9	0	566
III	1	21	395	24	5	446
IV	0	1	0	369	11	381
V	0	0	5	4	87	96
Total	92	553	429	406	103	1.583
T2 vs TS						
I	85	5	2	0	0	92
II	0	521	35	15	1	572
III	1	14	390	31	6	442
IV	0	0	3	364	14	381
V	0	0	10	4	82	96
Total	86	540	440	414	103	1,583
T1 vs T2						
I	81	4	3	0	0	88
II	3	495	40	9	1	548
III	1	30	367	22	4	424
IV	0	2	18	372	11	403
V	0	0	9	8	86	103
Total	85	531	437	411	102	1,566

T1: first triage assessment; T2: second triage assessment; TS: standard triage.

Furthermore, the absence of relevant differences between professional profiles suggests that the system can be applied consistently by multidisciplinary teams,[2] reinforcing its potential usefulness in real-world clinical practice. Although this study focused primarily on the psychometric properties of the scale rather than the inherent capabilities of the professionals, the availability of midwives has been associated with shorter overall emergency department stays and higher perceived quality of care among women undergoing triage in hospital emergency departments.[1,3-5,14]

This study has several limitations. Evaluation in a controlled and predictable clinical simulation setting limits direct extrapolation to real-world practice, where additional factors such as health care demand, resource availability, and patient interaction influence triage decisions. The voluntary nature of participation and the unequal distribution of professional categories may limit the generalizability of the findings. Finally, in the absence of precise reference indicators, expert consensus was accepted as the reference triage classification, which may have influenced the agreement results observed. However, this approach is common in validation studies when no clearly defined reference standard is available. Nevertheless, the use of

clinical simulation allowed rigorous and consistent control of training and assessment conditions, as well as standardized comparison of classifications, representing an important methodological strength for the reliability analysis.[15]

In conclusion, the newly adapted obstetric and gynecological version of the SET demonstrated high inter- and intraobserver reliability and good acceptance among healthcare professionals in a controlled clinical simulation setting. These findings support its potential usefulness as a standardized triage tool in obstetric and gynecological emergencies, although studies conducted under real-world clinical conditions are needed to confirm its applicability and impact on health care delivery.

Table 4. Statistical assessment of intraobserver agreement or reliability: consistency of measurements performed by the same observer at different time points

Profesional	τ	EMA (%)	Kappa[a] (95% CI)	Kappa[b] (95% CI)
Ob1	0.898	85.417	0.935 (0.886-0.983)	0.867 (0.769-0.964)
Ob2	0.986	96.000	0.982 (0.959-1.000)	0.980 (0.955-1.000)
Ob3	0.983	98.000	0.991 (0.972-1.000)	0.977 (0.930-1.000)
Ob4	0.947	91.667	0.964 (0.928-1.000)	0.947 (0.898-0.997)
Ob5	0.611	67.391	0.674 (0.494-0.854)	0.586 (0.394-0.778)
Ob6	0.948	94.000	0.972 (0.939-1.000)	0.947 (0.886-1.000)
Ob7	0.768	78.723	0.810 (0.674-0.947)	0.757 (0.605-0.908)
Ob8	0.940	91.837	0.964 (0.926-1.000)	0.939 (0.879-0.998)
Ob9	1.000	100.000	1.000 (1.000-1.000)	1.000 (1.000-1.000)
Ob10	0.935	91.837	0.959 (0.916-1.000)	0.937 (0.875-0.998)
Ob11	0.801	79.545	0.845 (0.722-0.968)	0.743 (0.577-0.910)
Ob12	0.986	98.000	0.991 (0.973-1.000)	0.988 (0.965-1.000)
Ob13	0.814	79.592	0.850 (0.735-0.966)	0.805 (0.679-0.930)
Ob14	0.905	88.000	0.912 (0.830-0.993)	0.843 (0.719-0.967)
Ob15	0.968	95.918	0.980 (0.952-1.000)	0.965 (0.917-1.000)
Ob16	0.900	85.714	0.915 (0.841-0.989)	0.893 (0.818-0.968)
Ob17	0.953	98.000	0.963 (0.889-1.000)	0.960 (0.882-1.000)
Ob18	0.964	95.833	0.981 (0.953-1.000)	0.960 (0.904-1.000)
Ob19	0.970	96.000	0.981 (0.955-1.000)	0.973 (0.936-1.000)
Ob20	0.992	98.000	0.991 (0.974-1.000)	0.984 (0.953-1.000)
Ob21	0.934	95.918	0.952 (0.874-1.000)	0.944 (0.858-1.000)
Ob22	0.899	94.000	0.913 (0.806-1.000)	0.895 (0.776-1.000)
Ob23	0.947	93.878	0.972 (0.938-1.000)	0.937 (0.865-1.000)
Ob24	0.939	91.837	0.968 (0.934-1.000)	0.930 (0.862-0.999)
Ob25	0.927	90.000	0.954 (0.911-0.997)	0.928 (0.866-0.990)
Ob26	1.000	100.000	1.000 (1.000-1.000)	1.000 (1.000-1.000)
Ob27	0.797	74.000	0.860 (0.767-0.953)	0.758 (0.628-0.887)
Ob28	0.829	87.755	0.858 (0.731-0.985)	0.818 (0.670-0.967)
Ob29	0.867	87.500	0.902 (0.803-1.000)	0.857 (0.738-0.976)
Ob30	0.965	96.000	0.978 (0.946-1.000)	0.950 (0.881-1.000)
Ob31	0.927	94.000	0.945 (0.868-1.000)	0.930 (0.842-1.000)
Ob32	0.915	95.918	0.908 (0.750-1.000)	0.946 (0.865-1.000)

τ: Kendall's rank correlation coefficient; EMA: exact modal agreement.
[a]Weighted kappa coefficient (Cohen's kappa).
[b]Alternative kappa coefficient developed by Ineke van der Wulp and Henk F. van Stel.

References

1 Angelini D, Howard E. Obstetric triage: a systematic review of the past fifteen years: 1998-2013. MCN Am J Matern Child Nurs. 2014;39:284-97.
2 Committee on Obstetric Practice. Committee Opinion No. 667: Hospital-based triage of obstetric patients. Obstet Gynecol. 2016;128:228.
3 Royal College of Obstetricians and Gynaecologists. Maternity triage. Good Practice Paper No. 17. Londres: RCOG; 2023.
4 Paisley KS, Wallace R, DuRant PG. The development of an obstetric triage acuity tool. MCN Am J Matern Child Nurs. 2011;36:290-6.
5 Ruhl C, Garpiel SJ, Priddy P, Bozeman LL. Obstetric and fetal triage. Semin Perinatol. 2020;44:151240.
6 Smithson DS, Twohey R, Rice T, Watts N, Fernandes CM, Gratton RJ. Implementing an obstetric triage acuity scale: interrater reliability and patient flow analysis. Am J Obstet Gynecol. 2013;209:287-93.
7 Gratton RJ, Bazaracai N, Cameron I, Watts N, Brayman C, Hancock G, et al. Acuity assessment in obstetrical triage. J Obstet Gynaecol Can. 2016;38:125-33.
8 Ruhl C, Scheich B, Onokpise B, Bingham D. Interrater reliability testing of the maternal fetal triage index. J Obstet Gynecol Neonatal Nurs. 2015;44:710-6.
9 Kenyon S, Hewison A, Dann SA, Easterbrook J, Hamilton-Giachritsis C, Beckmann A, et al. The design and implementation of an obstetric triage system for unscheduled pregnancy-related attendances: a mixed methods evaluation. BMC Pregnancy Childbirth. 2017;17:309.
10 Veit-Rubin N, Brossard P, Gayet-Ageron A, Montandon CY, Simon J, Irion O, et al. Validation of an emergency triage scale for obstetrics and gynaecology: a prospective study. BJOG. 2017;124:1867-73.
11 Lindroos L, Korsoski R, Öhman MO, Elden H, Karlsson O, Sengpiel V. Improving assessment of acute obstetric patients: introducing a Swedish obstetric triage system. BMC Health Serv Res. 2021;21:1207.
12 Rashidi Fakari F, Simbar M, Zadeh Modares S, Alavi Majd H. Obstetric triage scales: a narrative review. Arch Acad Emerg Med. 2019;7:e13.
13 Moudi A, Iravani M, Najafian M, Zareiyan A, Forouzan A, Mirghafourvand M. Obstetric triage systems: a systematic review of measurement properties. BMC Pregnancy Childbirth. 2020;20:275.
14 Obregón-Gutiérrez N, Rebollo-Garriga G, Costa Gil MP, Puig-Calsina S, Calle-Del Fresno S, Garrido Domínguez M. El triaje obstétrico: nuevo reto para las matronas. Experiencia del Hospital Parc Taulí en la implementación de un sistema de triaje obstétrico. Matronas Prof. 2011;12:49-53.
15 Hernández Ruipérez T. Diseño y validación de un modelo docente en triaje de urgencias [tesis doctoral]. Murcia: Universidad Católica de San Antonio; 2016.

Author Affiliations: [1]Servicio de Ginecología y Obstetricia, Consorci Corporació Sanitària Parc Taulí, Sabadell, Barcelona, Spain. [2]Institut d´Investigació i Innovació Parc Taulí (I3PT-CERCA), Universitat Autònoma de Barcelona, Sabadell, Barcelona, Spain. [3]Servicio de Urgencias Generales, Consorci Corporació Sanitària Parc Taulí, Sabadell, Barcelona, Spain.
Email: Pcosta@tauli.cat / pcostagil48@gmail.com

Conflict of Interest Disclosures: To conduct validation studies at the Consorci Corporació Sanitària Parc Taulí (CCSPT) in Sabadell, the Institut d'Innovació i Investigació Parc Taulí (I3PT) entered into a contract—supplementing a prior memorandum of intent—with ESBARZER S.L., the company that holds the intellectual and industrial property rights to the SET (Structured Triage System®), and with Tree Technology S.A., the technology firm that developed and markets the software application for the SET (Structured Triage System®). This contract includes, among other provisions, compensation to cover the work performed by the research team. It is stipulated that this financial compensation is the sole payment arising from the aforementioned study, and that there is no other form of remuneration—whether in cash or in kind—that could be directly or indirectly attributed to this agreement. The agreement among I3PT, ESBARZER S.L., and Tree Technology S.A. included the consent and supervision of Dr. Josep Gómez Jiménez (author of the SET/Structured Triage System®), as well as the development and provision of a pre-production software version to be used for conducting reliability and validity studies of the obstetric-gynecological version of the system.
Author Contributions, Funding, and Ethical Responsibilities: The author has confirmed their authorship, the absence of external funding, and the maintenance of confidentiality and respect for patients' rights in the author's responsibilities document, the publication agreement, and the assignment of rights to EMERGENCIAS. This manuscript was completed with the support of the Department of Health of the Generalitat de Catalunya, under the 2024 call for grants from the Strategic Plan for Research and Innovation in Health (PERIS) 2022–2027—specifically within

the scope of "impact on the health and well-being of individuals"—under grant code SLT035/24/000013. The study was approved by the Clinical Research Ethics Committee of the CCSPT in Sabadell.

Acknowledgements: To the Catalan Association of Midwives (ACLL) and the Catalan Society of Obstetrics and Gynecology (SCOG) for their collaboration agreements with I3PT, which facilitated the formation of a group of experts in our field to conduct a Delphi process; the final report from this process established the clinical foundation for adapting the obstetric-gynecological scale of the Structured Triage System (SET®). To the technical team from the IT Department of CCSPT in Sabadell and to Tree Technology S.A., who made this research possible in its current form by developing a virtual simulation environment that facilitated access to the new software application, as well as the storage and retrieval of data. To the nurses, midwives, and obstetricians whose participation made this research possible. To Dr. J. Gómez Jiménez for reviewing the manuscript; his observations contributed to improving its readability and precision.

Use of Generative Artificial Intelligence Tools: During the preparation of this work, AI was used as a writing assistant; the resulting content was subsequently reviewed and edited as needed, with full responsibility for the publication's content being assumed by the authors.

Article not commissioned by the Editorial Board and with external peer review

Editor in Charge: Juan González del Castillo.

Corresponding Author: Pilar Costa Gil. Servicio de Ginecología y Obstetricia. Consorci Corporació Sanitària Parc Taulí (CCSPT). Parc Taulí s/n, 08208 Sabadell, Barcelona, Spain.