



ORIGINAL ARTICLE

Diagnostic accuracy of the Spanish version of the 4AT scale (4AT-ES) for delirium screening in older inpatients

E. Delgado-Parada^{a,b,*}, D. Morillo-Cuadrado^{b,c}, J Saiz-Ruiz^{c,d,e}, A. Cebollada-Gracia^{d,e}, J.L. Ayuso-Mateos^{a,b,c}, A.J. Cruz-Jentoft^{e,f}

^a Department of Psychiatry, Hospital Universitario de La Princesa, Diego de León St, 62, Madrid 28006, Spain

^b Instituto de Investigación Sanitaria Hospital Universitario de la Princesa (IIS-Princesa), c/ Diego de León, 62, Madrid 28006, Spain

^c Network-based Biomedical Research Consortium, area of Mental Health (CIBERSAM), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0, Madrid 28029, Spain

^d Department of Psychiatry. Hospital Universitario Ramon y Cajal, Ctra. de Colmenar Viejo km. 9,100, Madrid 28034, Spain

^e Instituto Ramón y Cajal de Investigación Sanitaria (IRyCIS) Ctra. de Colmenar Viejo, Km. 9,100. Planta - 2 derecha Madrid 28034, Spain

^f Department of Geriatrics. Hospital Universitario Ramón y Cajal. Ctra. de Colmenar Viejo km. 9, 100,. Madrid 28034, Spain

Received 30 September 2021; accepted 12 January 2022

Available online 10 February 2022

KEYWORDS

Delirium screening;
Translation;
Diagnostic accuracy;
Validation

Abstract

Background and objectives: The 4AT scale is a sensitive tool for screening delirium, which can be applied rapidly in clinical settings without any specific training. It has not been translated, adapted, and validated to assess Spanish older adults. The aims of the study are: to translate and adapt to Spanish culture the 4AT scale, to present evidence of the diagnostic accuracy of this version (4AT-ES) when applied in non-specialized hospital wards, and to assess the loss of diagnostic accuracy in presence of risk factors.

Methods: A prospective sample was independently assessed on the 4AT-ES and the reference standard. One hundred and twenty-one inpatients (70+ years) for whom a psychiatric assessment was requested were included. Out of them, 50 were diagnosed with delirium. Nurses without specific training applied the 4AT-ES, and experienced psychiatrists cast the reference standard diagnosis (DSM-V criteria).

Results: Patients with delirium were older and had more risk factors (more previous delirium episodes, a higher likelihood of prior dementia/cognitive impairment) than controls. The 4AT-ES had excellent validity, sensitivity (96%) , and specificity (83.1%). The area under the curve was 0.918; in the subsample with any of those risk factors, its value did not decrease.

Conclusion: The 4AT-ES version of the 4AT scale was developed. When applied by non-specifically trained, nursing staff it showed excellent validity, sensitivity, and specificity, even in a

* Corresponding author at: Department of Psychiatry, Hospital Universitario de La Princesa, Diego de León St, 62, Madrid 28006, Spain.
E-mail address: eduardo.delgado@salud.madrid.org (E. Delgado-Parada).

subsample with previous risk factors. All indices were comparable to the original version. We recommend its use for efficient delirium screening in hospitalized older patients with suspected delirium.

© 2022 Asociación Universitaria de Zaragoza para el Progreso de la Psiquiatría y la Salud Mental. Published by Elsevier España, S.L.U. All rights reserved.

Introduction

Delirium is a neuropsychiatric disorder that presents with an acute onset and is fluctuating in nature. It is characterized by an overall cognitive impairment, being attention and awareness the most affected domains.^{1,2} Delirium is highly prevalent amongst older inpatients, especially within healthcare units treating severely ill patients (e.g., intensive care units, major surgery) or those treating frail individuals (e.g., palliative care units).³ Up to 50% of patients admitted within this age group suffer from delirium with different degrees of severity.⁴ The diagnosis is a clinical one and is based on the information gathered from the clinical assessment and from the collateral information obtained from relatives, and from health and care workers.^{2,5}

Delirium is linked with poor clinical outcomes —during admission, the risk of developing complications such as infections, falls, incontinence, or pressure ulcers is higher, which has a significant impact on admission length of stay and on its associated costs.^{4,5} Upon discharge, it increases the risk of functional and cognitive impairment, readmission and institutionalization, and ultimately, death.⁶ An under-diagnosis or a late diagnosis further worsens the prognosis.^{1,5} Therefore, delirium screening is essential to be considered as an earlier intervention. Having adequate screening tools, validated for their use by non-specialized health professionals on the wards, is thus necessary for effective detection.

The 4AT scale is a rapid and sensitive tool that requires no specific training⁷ and is freely available for download and use (www.the4at.com). The original English version has been validated since 2014 for the screening of delirium in older inpatients at different levels of the healthcare system.⁸ This scale is the screening instrument recommended in the latest clinical practice guidelines about delirium.⁹ It is recommended even in the presence of cognitive impairment when the differential diagnosis is rather difficult.¹⁰ Since its publication, the performance of the 4AT scale has been evaluated in several languages, showing that its use in routine clinical practice is well-supported by an excellent diagnosis accuracy.¹¹ Although there is a Spanish version,¹² the translation process carried out by the authors is unknown, and there is no validation study that we are aware of.

The aim of this paper is threefold: Firstly, to introduce a Spanish version of the 4AT scale, translated from the original version and adapted to the Spanish culture. Secondly, to show evidence of its validity and diagnostic accuracy for screening delirium in the older hospitalized population when applied by nursing staff. And lastly, to assess whether its diagnostic accuracy differs within the subpopulation of patients with cognitive impairment.

Material and methods

Translation and cultural adaptation to Spanish

A translation and cultural adaptation of the original 4AT instrument in English was carried out, following the guidelines for cross-cultural adaptation of self-report measurement instruments.¹³ In the first two phases, initial translation and synthesis, two bilingual (English/Spanish) translators, being Spanish their native language, participated independently. One of them had previous medical training, while the other did not. The third stage, inverse translation, was carried out by two bilingual translators without formal medical training, blind to the original English version. The fourth stage was carried out by an expert committee consisting of two experts in research methodology, one healthcare professional, one linguist, and the four translators that had participated in the previous stages. Finally, registered nurses, training to obtain a nursing specialty degree in geriatric or psychiatric nursing, proof-tested its application with 10 older inpatients. None of them reported any difficulties with the application.

Sample selection and data collection

A prospective, convenience sample of seventy-year-old or older inpatients admitted to a university hospital, for whom psychiatric assessment was requested, were considered for inclusion in the study. Patients that lacked verbal communication had a severe hearing impairment, were non-Spanish speakers, or refused to participate were excluded. The delirium diagnosis was based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) criteria and determined by an experienced psychiatrist, member of the Psychiatric Liaison team. Recruited patients who did not meet the criteria for delirium according to DSM-V were considered controls. Starting in July 2017, each patient that met the criteria (or a legal guardian) was asked to participate and sign the informed consent. A sample size of at least fifty participants in each relevant subgroup is recommended for health-related questionnaires.¹⁴ Thus, participants were recruited until a sample size of fifty participants was reached in both groups.

Measures

The 4AT scale consists of four items, addressing four specific domains suggestive of delirium.⁷ Item AT1 (level of alertness) scores alertness through direct observation, for a score of 0 or 4. The next two items, AT2 (AMT4 test) and AT3 (months backward) allow the screening of attention and cognition, respectively, through simple standardized questions; each of them is scored up to 2. Item AT4 (change or fluctuating course) requires collateral information to assess

the presence of fluctuations or a sudden change in mental state and is scored dichotomously as 0 or 4.

Each patient was assessed using the translated and adapted version by a registered nurse without specific training or experience in psychiatry/psychogeriatrics, and no prior training for the 4AT scale. In order to gather convergent evidence of the validity of the measures, patients were also assessed on the scale by the same psychiatrist who cast the DSM-V-based diagnosis, and in the same session. The time interval between this session and the assessment by the nurse was 24 h at most (in either order, at the convenience of the evaluators). The nurses were blind to the DSM-V-based diagnosis, and both evaluators were reciprocally blind to the 4AT assessment of one another. A total of five nurses and eleven psychiatrists participated in the assessments.

Other variables taken into account, drawn from the clinical records of each patient, were: gender, age (in years), time since admission (in days), number of previous delirium episodes, whether there was a previous diagnosis of dementia/cognitive impairment, and the medical ward where the patient was admitted. When necessary, the patient, legal guardian, or healthcare team were interviewed for the missing information. The following derived binary variables were also computed: whether they had had any previous delirium episode at all (delirium antecedents), and a variable of cognitive vulnerability, collapsing those cases that had a previous diagnosis of dementia or cognitive impairment with the ones that had at least one previous delirium episode. This latter variable separated the participants that were likely to suffer from some degree of cognitive impairment from those that were not suspected at all of having any prior cognitive impairment.

Statistical analyses

The following analyses were all performed using R version 4.0.4.¹⁵ Certain packages, indicated where appropriate, were used for some specific computations. Descriptive statistics were obtained for all the variables. Differences between the Delirium and the Control group were assessed using the t-test or the χ^2 -test (applying the correction for continuity) for the continuous and the categorical variables, respectively.

The 4AT score was computed from the assessment given by the nurse personnel (NUR score). The four 4AT items were assumed to be effect indicators of a single unidimensional latent variable standing for the Delirium construct. The internal evidence of validity was assessed by testing this hypothesis of unidimensionality with parallel analysis.¹⁶ Minimum rank parallel analysis with polychoric correlations failed to converge, so principal axes parallel analysis with Pearson correlations were used instead.¹⁷ Reliability was assessed by computing the Cronbach's α ¹⁸ and McDonald's Ω ¹⁹ internal consistency indices. Parallel analysis and Cronbach's α were computed with package psych v. 2.0.12,²⁰ and McDonald's Ω with packages lavaan version 0.6-7²¹ and semTools version 0.5-4.²²

The score given by the psychiatrist (PSY score) was used as evidence of convergent validity. Their internal validity and reliability were also assessed independently of the NUR measures, following the same procedure. Then, the NUR and PSY composite scores were correlated. This correlation was

corrected for attenuation using the Ω coefficients of both measures.

The NUR diagnostic accuracy was assessed by computing and plotting the ROC curve. The area under the curve (AUC) statistic was computed using R package pROC version 1.17.0.1.²³ Its 95%-coverage confidence interval was computed using DeLong's method with 2000 stratified replicates.²⁴ The Youden index²⁵ was also obtained for the total score, and the pre-specified cut-off point was compared to the optimal one (according to the Youden index), in order to assess its discriminant power. A predicted diagnostic was given based on the pre-specified cut-off point of 4 in the total score.⁸ Each individual item was also assessed by computing its AUC and the diagnostic accuracy statistics, using the non-null response categories as cut-off points. Sensitivity and specificity indices were computed for each of the cut-off points. Their 95%-coverage confidence intervals were computed using Wilson's score method,²⁶ which gives the closest estimation to nominal levels independently of sample size.²⁷ Positive and negative diagnostic likelihood ratios were also computed. The diagnostic prediction was computed as well for the PSY score, and the confusion matrix between both predictions was computed in order to assess their convergence. These analyses were also performed separately on the groups with and without suspected cognitive impairment (i.e., segmenting by cognitive vulnerability).

Results

The translation and adaptation process resulted in the Spanish version of the 4AT scale (4AT-ES). This version is translated to Spanish and adapted to the cultural context of Spanish residents. The instrument is shown in Table 1.

Twelve patients, who met at least one exclusion criteria, were excluded from the study. The sample was finally made up of a total of 121 participants aged 70 or older. Out of them, 50 were in the delirium and 71 in the control group. No data missingness was present in any of the variables of this sample. Descriptive statistics are shown in Table 2.

The groups differed significantly in all the quantitative variables ($\alpha = 0.05$), except for the time since admission ($t_{100.72} = -1.37$, $p\text{-value} = 0.175$). The mean was always higher in the delirium group in all the variables that were significantly different. They did not differ in the proportion of males and females, or the ward they were recruited from. In the remaining categorical variables, the proportion of patients with previous cognitive impairment or dementia ($\chi^2_2 = 18.57$, $p\text{-value} < 0.001$), with any previous episodes of delirium ($\chi^2_2 = 7.78$, $p\text{-value} = 0.020$), or with cognitive vulnerability ($\chi^2_2 = 9.77$, $p\text{-value} = 0.008$) was higher in the delirium group.

Fig. 1 shows the results of the parallel analyses with the responses given by the NUR and the PSY scorers. As the figures show, the 4AT scale was found to be unidimensional with both scores, and thus the measure was congeneric in both cases. Both measures were found to have excellent internal consistency: The NUR score had an α of 0.731 and an Ω of 0.809, while the values for the PSY score were 0.831 and 0.831, respectively. Regarding convergent

Table 1 Spanish version of the 4AT scale (4AT-ES).


Test 4AT: instrumento de cribado de deterioro cognitivo y delirium

(etiqueta)

Nombre del paciente:

Fecha de nacimiento:

Número:

Fecha:

Hora:

Evaluador:

Rodear con un círculo la respuesta correcta

[1] NIVEL DE CONCIENCIA

Este ítem se aplica en pacientes que pueden presentar marcada somnolencia (por ejemplo, dificultad para despertarse y/o con somnolencia evidente durante la evaluación) o agitados/hiperactivos. Observe al paciente. Si está dormido, intente despertarlo hablándole o tocándolo ligeramente el hombro. Pida al paciente que diga su nombre y dirección para ayudar a establecer la puntuación.

Normal (totalmente alerta, pero no agitado, durante la evaluación)	0
Somnolencia leve durante <10 segundos tras despertarse, luego normal	0
Claramente anormal	4

[2] AMT4

Edad, fecha de nacimiento, lugar (nombre del hospital o edificio), año actual.

Sin errores	0
1 error	1
2 o más errores/imposible de evaluar	2

[3] ATENCIÓN

Diríjase al paciente y pídale lo siguiente: “Por favor dígame los meses del año en orden inverso, empezando por diciembre.”

Para ayudar a la comprensión inicial está permitido preguntar “¿Cuál es el mes anterior a diciembre?”

Meses del año en orden inverso	Logra decir 7 o más meses correctamente	0
	Logra decir menos de 7 meses / se niega a comenzar	1
	Imposible de evaluar (debido a malestar, somnolencia, falta de atención)	2

[4] CAMBIO AGUDO O CURSO FLUCTUANTE

Evidencia de un cambio significativo o fluctuación en: nivel de conciencia, cognición, otra función mental (Por ejemplo: paranoia, alucinaciones) durante las últimas 2 semanas y que se mantiene en las últimas 24 horas

No	0
Sí	4

4 o más: posible delirium +/- deterioro cognitivo

1-3: posible deterioro cognitivo

0: delirium o deterioro cognitivo poco probables (pero la presencia de delirium es aún posible si la información [4] no es completa)

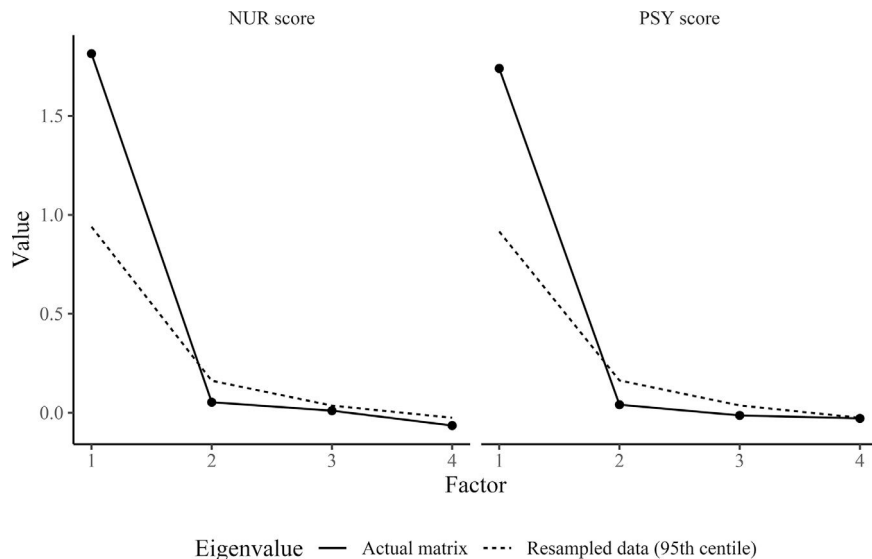
PUNTUACIÓN 4AT-ES

validity, the composite scores were found to be highly correlated; the Pearson correlation was 0.779 (0.697–0.841) between them. When corrected for attenuation by the Ω coefficient, its value ascended to 0.950. The confusion matrix is given in Table 3 (columns under the header *Diagnosis (PSY score)*) further shows that the scores given by the

two groups of evaluators are highly coincident; 88.43% of the cases were concordant, although the NUR scores tended to be more sensitive and less specific. The ROC curve of the NUR score is represented in Fig. 2. As the curve shows, the scale had a very high diagnostic accuracy, its estimated AUC statistic being 0.918. The optimal Youden index was 0.791,

Table 2 Descriptive statistics of the sample.

Variable	Delirium (N = 50)		No delirium (N = 71)		p-value
Age (years), Mean (Std. dev.)	82	(5.7)	78.3	(5.9)	0.001
Gender (Female), N (%)	25	(50%)	31	(43.7%)	0.789
Time since admission (days), Mean (Std. dev.)	9.2	(12.1)	12.1	(11.2)	0.175
Previous delirium episodes, Mean (Std. dev.)	0.7	(0.8)	0.3	(0.5)	0.003
Previous dementia or cognitive impairment, N (%)	23	(46%)	8	(11.3%)	< 0.001
Admission ward, N (%)					1.000
Medical	33	(66.0%)	47	(66.2%)	
Surgical	17	(34.0%)	24	(33.8%)	
Delirium antecedents, N (%)	25	(50.0%)	18	(25.4%)	0.020
Cognitive vulnerability, N (%)	29	(58.0%)	21	(29.6%)	0.008
NUR score, Mean (Std. dev.)	7.4	(2.8)	2.0	(2.6)	< 0.001
PSY score, Mean (Std. dev.)	7.0	(2.5)	1.4	(1.6)	< 0.001

**Fig. 1** Parallel analysis of the NUR and PSY scores of the 4AT-ES scale.

Abbreviations: NUR: nurse; PSY: psychiatrist.

corresponding to the score of 4, which corresponds to the cut-off point set by the original scale design.

Table 3 gives the confusion matrix of the NUR score-based diagnosis, under the header *Diagnosis (Reference standard)*. The indices related to the scale diagnostic accuracy, according to the cut-off point of 4, are given in Table 4, along with the diagnostic accuracy indices of the item response categories for the NUR scores. Both the total score and item indices are given for the complete sample, as well as for the groups with and without cognitive vulnerability.

Discussion

Our Spanish version of the 4AT scale (4AT-ES) showed evidence of high validity for screening delirium in hospitalized older people. Allegedly, strictly following the guidelines for translation and cross-cultural adaptation¹³ led to a rigorous version of the instrument in Spanish, suitable for validation in clinical settings in Spain. It should be noted though, when interpreting the results, that the prevalence of delirium in our sample was higher than in the validation of the original

Table 3 Confusion matrix, based on a NUR score equal to or higher than 4 with respect to the reference standard and the PSY-score based prediction.

Predicted	Diagnosis (Reference standard)		Diagnosis (PSY score)		Total
	No delirium	Delirium	No	Yes	
No	59	2	56	5	61
Yes	12	48	9	51	60
Total	71	50	65	56	121

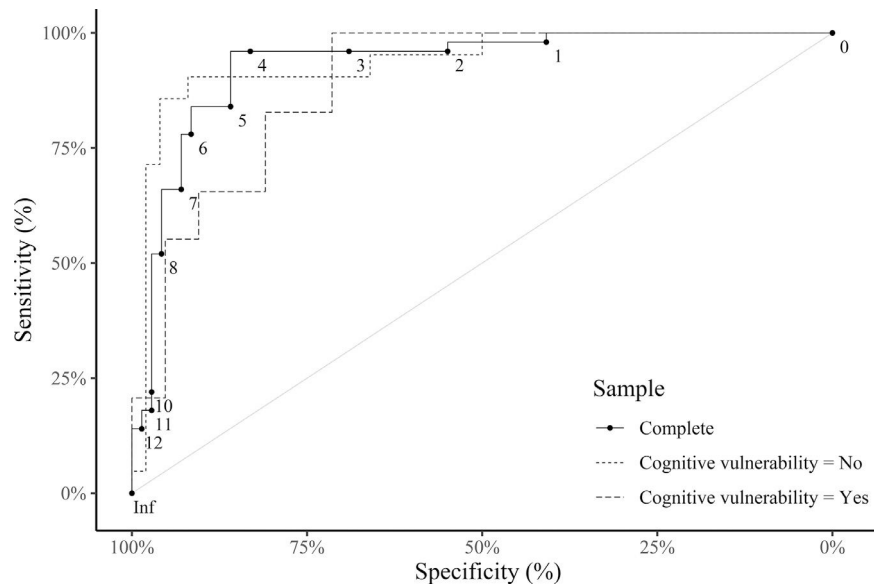


Fig. 2 ROC Curve of the 4AT-ES scale NUR scores.
Abbreviations: NUR: nurse.

version.⁸ The sample consisted of patients that were referred to the Psychiatry Liaison team, being delirium one of the most frequent conditions in the clinical practice of such teams,²⁸ thus the higher prevalence. Nevertheless, the sample was comparable to those used in the validation of the original sample⁸ and other validation procedures,^{29,30} the group with delirium having significantly more recognized predisposing factors: older age, previous cognitive impairment, and/or episodes of delirium.²

The scores obtained by the nursing staff and the psychiatrists were both unidimensional, as we would expect given the definition of the construct definition of delirium. The reliability of both scores was excellent, as evidenced by the estimated internal consistency indices. Furthermore, the high correlation between both measures yielded very solid evidence of the convergent validity between the two rater groups. Finally, the diagnostic accuracy of the nursing staff scores was also excellent: its estimated AUC was 0.918, very similar to the original version (0.927),⁸ and the cut-off point given by the Youden Index (0.791) confirmed the value expected from the design of the original version.

The sensitivity of the Spanish version was slightly higher than that of the original scale (96% vs. 89.7%),⁸ while the specificity was slightly lower (83.1% vs. 84.1%). We find the same pattern when comparing with the sensitivity (88%) and specificity (88%) pooled across studies with different adaptations of the scale in various settings.¹¹ The corresponding false-positive rate (16.9%) was higher than the omission rate (4%) but acceptable -given the higher cost of omitting a diagnosis of delirium, this imbalance is desirable in a screening tool.³¹ Nevertheless, under a positive diagnosis, having delirium was more than five times as likely as being a false positive, as given by the positive likelihood ratio. On the other hand, the negative likelihood ratio, close to 0, indicated that it was highly unlikely that a patient with delirium would go undiagnosed. Contrary to previous results,³² we

did not find worse diagnostic accuracy when the scale was applied by nurses without specific training.

Items AT2 and AT3, which measure cognition and attention respectively, contributed the most to sensitivity. On the other hand, items AT1 (level of awareness) and AT4 (change or fluctuating course) were the most specific. Consequently, the cognitive aspect does not contribute as much to the specific screening of delirium. This is consistent with the literature about screening delirium in a hospital setting when the level of consciousness is altered and/or information of a significant change or a fluctuating clinical course is available.³³

In the cognitive vulnerability subgroup, the AUC was similar (0.872) to the cognitive impairment subgroup in the original version (AUC = 0.891),⁸ and kept a good diagnostic accuracy. As in the original scale, sensitivity was higher in this subgroup, while specificity was higher in the subgroup without cognitive vulnerability. It should be taken into account though that the cognitive vulnerability definition included all cases with a history of cognitive impairment or dementia, as well as those who had previously presented an episode of delirium but without specifying the diagnosis of dementia. This differs considerably from Bellelli et al,⁸ who only considered participants with dementia in their cognitive impairment subgroup. Regarding the item properties, the most specific items in the whole sample (AT1 and AT4) had a lower specificity in the cognitive vulnerability group. Nevertheless, validity and diagnostic accuracy were similar to the original version; these conclusions are valid even in people with cognitive vulnerability and considering the diagnostic properties of each item.

In summary, the 4AT-ES has excellent diagnostic accuracy for screening delirium in hospitalized older patients. These properties do not depend on whether the evaluator is a psychiatrist or a nurse, both without specific training for its application. This allows its application in a non-specialized unit to assist in delirium screening. Moreover, the procedure

Table 4 Diagnostic accuracy of the 4AT-ES scale scored by the NUR personnel and its items, for the whole sample and for the subgroups with/without cognitive vulnerability.

Sample	Criterion	Cutoff point	Sensitivity (%)	Specificity (%)	LR +	LR -	AUC
Complete	4AT	4	96% (95.8% - 96.2%)	83.1% (82.8% - 83.4%)	5.680	0.048	0.918 (0.864 - 0.972)
	AT1	4	26% (25.6% - 26.4%)	95.8% (95.6% - 95.9%)	6.153	0.773	0.609 (0.0543 - 0.675)
	AT2	2	68% (67.6% - 68.4%)	78.9% (78.6% - 79.2%)	3.219	0.406	0.786 (0.709 - 0.864)
		1	88% (87.7% - 88.3%)	62% (61.6% - 62.3%)	2.314	0.194	
	AT3	2	52% (51.6% - 52.4%)	91.5% (91.3% - 91.8%)	6.153	0.524	0.775 (0.695 - 0.856)
		1	86% (85.7% - 86.3%)	50.7% (50.3% - 51.1%)	1.745	0.276	
	AT4	4	86% (85.7% - 86.3%)	84.5% (84.2% - 84.8%)	5.551	0.166	0.853 (0.788 - 0.917)
	Cognitive vulnerability						
	No						
Yes	4AT	4	90.5% (90.1% - 90.9%)	92% (91.8% - 92.2%)	11.310	0.104	0.918 (0.833 - 1.000)
	AT1	4	19% (18.5% - 19.6%)	98% (97.9% - 98.1%)	9.524	0.826	0.585 (0.497 - 0.673)
	AT2	2	66.7% (66% - 67.3%)	88% (87.7% - 88.3%)	5.556	0.379	0.824 (0.718 - 0.929)
		1	85.7% (85.2% - 86.2%)	68% (67.6% - 68.4%)	2.679	0.210	
	AT3	2	38.1% (37.4% - 38.8%)	92% (91.8% - 92.2%)	4.762	0.673	0.714 (0.585 - 0.843)
		1	76.2% (75.6% - 76.8%)	56% (55.6% - 56.4%)	1.732	0.425	
	AT4	4	85.7% (85.2% - 86.2%)	92% (91.8% - 92.2%)	10.714	0.155	0.889 (0.803 - 0.974)
	4AT	4	100% (100% - 100.0%)	61.9% (61.2% - 62.6%)	2.625	0.000	0.872 (0.767 - 0.977)
	AT1	4	31% (30.5% - 31.6%)	90.5% (90.1% - 90.9%)	3.259	0.762	0.608 (0.500 - 0.715)
	AT2	2	69% (68.4% - 69.5%)	57.1% (56.5% - 57.8%)	1.609	0.543	0.675 (0.531 - 0.818)
		1	89.7% (89.3% - 90.0%)	47.6% (46.9% - 48.3%)	1.712	0.217	
	AT3	2	62.1% (61.5% - 62.6%)	90.5% (90.1% - 90.9%)	6.517	0.419	0.804 (0.690 - 0.918)
		1	93.1% (92.8% - 93.4%)	38.1% (37.4% - 38.8%)	1.504	0.181	
	AT4	4	86.2% (85.8% - 86.6%)	66.7% (66% - 67.3%)	2.586	0.207	0.764 (0.643 - 0.886)

Abbreviations: LR +, positive likelihood ratio; LR -, negative likelihood ratio; AUC, area under the curve.

is simple, giving an accurate screening efficiently, and non-invasive, with minimal disturbance for the patient and no adverse effects. We hope that the 4AT-ES scale will facilitate the implementation of recommendations for the diagnosis and management of delirium in the clinical practice guidelines in Spain.

Limitations

One of the main weaknesses of this study is the absence of an instrument to use as an external criterion to test

convergent validity. As a result, we used the scores given by the psychiatrist as a criterion. However, these measures were given by the same specialist who cast the diagnosis, so they were not independent of the reference standard. This may make convergence evidence of validity redundant with diagnostic validation to some extent. However, since the original version of the scale was applied by expert physicians,⁸ using an equivalent criterion in the Spanish version allowed us to ensure validity under similar conditions.

There are other limitations that are worth mentioning. The characteristics of the patients excluded from the

sample could not be analyzed, as data was not available for them. Also, as each participant was assessed only once by a single member of the nurse team, we could neither obtain an index of interrater reliability for this measure nor assess its test-retest reliability. Nevertheless, it is worth noting that other authors^{8,29} have not assessed test-retest reliability either —when a construct is characterized precisely by its fluctuating course, as happens with delirium, test-retest reliability may be hazardous to assess.³⁴ The sample size was smaller than in the original study, since the estimation was carried out according to the recommendations for health-related questionnaires proposed by Terwee et al.¹⁴ In any case, it has allowed diagnostic accuracy, even in smaller sample sizes subgroups, those with and without cognitive vulnerability. Finally, the number and timing of administrations by each nursing professional who applied the scale was not collected in detail, nor was the potential impact of a learning curve considered. This information did not satisfy any of the objectives of the study but could be relevant in order to analyze the error variance components in the future, applying Generalizability Theory.

Conflict of interest

The authors report no conflicts of interest in this work.

Acknowledgments

All the listed authors have actively contributed to this work, meet the author criteria, and have approved the final version. First of all, the authors gratefully acknowledge the participation of the assessed participants, without whom this study would not have been possible. Second, the authors are also grateful for the willingness and commitment of the nursing staff, particularly Javier Álvarez Martín and María Verano de la Torre.

Ethical considerations

Participants were briefed about the study and signed informed consent before participating. When the patient was unable to provide informed consent, a legal guardian signed it instead. The study received ethical clearance from the Research Ethics Committee of the Ramón y Cajal Hospital (approved on April 24, 2017, agreement number 323).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet North Am Ed*. 2014;383:911–22.
- Wilson JE, Mart MF, Cunningham C, Shehabi Y, Girard TD, MacLulich AMJ. Delirium. *Nat Rev Dis Primers*. 2020;6:1–26.
- Marcantonio ER, Solomon CG. Delirium in hospitalized older adults. *N Engl J Med*. 2017;377:1456–66.
- Rieck KM, Pagali S, Miller DM. Delirium in hospitalized older adults. *Hosp Pract*. 2020;48:3–16.
- Oh ES, Fong TG, Hsieh TT, Inouye SK. Delirium in older persons: advances in diagnosis and treatment. *JAMA*. 2017;318:1161–74.
- Gibb K, Seeley A, Quinn T, Siddiqi N, Shenkin S, Rockwood. The consistent burden in published estimates of delirium occurrence in medical inpatients over four decades: a systematic review and meta-analysis study. *Age Ageing*. 2020;49:352–60.
- MacLulich AMJ, Anand A, Davis DHJ, Jackson T, Barugh AJ, Hall RJ. New horizons in the pathogenesis, assessment and management of delirium. *Age Ageing*. 2013;42:667–74.
- Bellelli G, Morandi A, Davis DHJ, Mazzola P, Turco R, Gentile S. Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age Ageing*. 2014;43:496–502.
- Network S.I.G.N. Risk reduction and management of delirium. 2019. <https://www.sign.ac.uk/sign-157-delirium>
- Morandi A, Bellelli G. Delirium superimposed on dementia. *Eur Geriatric Med*. 2020;11:53–62.
- Tieges Z, MacLulich AMJ, Anand A, Brookes C, Cassarino M, O'Connor M. Diagnostic accuracy of the 4AT for delirium detection in older adults: systematic review and meta-analysis. *Age Ageing*. 2021;50:733–43.
- Sander R., Corretge M., Dorrian M. 4AT Spanish version 1.0. 2016.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25:3186–91.
- Terwee CB, Bot SDM, de Boer MR, van der Windt D, Knol DL, Dekker J. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
- R. Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
- Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965;30:179–85.
- Timmerman ME, Lorenzo-Seva U. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol Methods*. 2011;16:209–20.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
- McDonald RP. Test Theory: A Unified Treatment. Mahwah, N.J.: Erlbaum Associates; 1999.
- Revelle, W., psych: Procedures for Personality and Psychological Research, 2020, Northwestern University, Evanston, Illinois, USA. <https://CRAN.R-project.org/package=psych> version = 2.0.12.
- Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012;48.
- Jorgensen T.D., Pornprasertmanit S., Schoemann A.M.; Rosseel Y., Miller P., Quick C., semTools: useful tools for structural equation modeling. 2020.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12:77.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837.
- Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–5.
- Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *J Am Statist Assoc*. 1927;22:209–12.
- Agresti A, Coull BA. Approximate is better than “Exact” for interval estimation of binomial proportions. *Am Stat*. 1998;52:119–26.
- Sánchez-González R, Baillès-Lázaro E, Bastidas-Salvado A, Ligoña A, Herranz-Villanueva S, Pérez-Solá V. Clinical

- profile of inpatients referred to a consultation-liaison psychiatry service: an observational study assessing changes over a 10-year period. *Actas Esp De Psiquiatr.* 2018;46:58–67.
29. Gagné A-J, Voyer P, Boucher V, Nadeau A, Carmichael PH, Pelletier M. Performance of the French version of the 4AT for screening the elderly for delirium in the emergency department. *Can J Emerg Med.* 2018;20:903–10.
 30. Kuladee S, Prachason T. Development and validation of the Thai version of the 4 'A's Test for delirium screening in hospitalized elderly patients with acute medical illnesses. *Neuropsychiatr Dis Treat.* 2016;12:437–43.
 31. Ascunce Elizaga N. Screening: why and how. *An Sist Sanit Navar.* 2015;38:5–7.
 32. Myrstad M, Watne LO, Johnsen NT, Bors-Lind E, Neerland BE. Delirium screening in an acute geriatric ward by nurses using 4AT: results from a quality improvement project. *Eur Geriatric Med.* 2019;10:667–71.
 33. European Delirium Association, American Delirium Society. The DSM-5 criteria, level of arousal and delirium diagnosis: inclusiveness is safer. *BMC Med.* 2014;12:141.
 34. Watson D. Stability versus change, dependability versus error: Issues in the assessment of personality over time. *J Res Personal.* 2004;38:319–50.