BMJ Open

Attribution of neuropsychiatric events in Systemic Lupus Erythematosus. Independent external validation of the Italian algorithm in an international multicenter cohort.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015546
Article Type:	Research
Date Submitted by the Author:	23-Dec-2016
Complete List of Authors:	Bortoluzzi, Alessandra; Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Department of Medical Sciences Fanouriakis, Antonis; Clinical Immunology and Allergy, University of Crete, Department of Rheumatology Appenzeller, Simone; Rheumatology Division, Department of Medicine, State University of Campinas Costallat, Lilian ; Rheumatology Division, Department of Medicine, State University of Campinas Scirè, Carlo; Italian Society for Rheumatology (SIR), Epidemiology Unit; Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Department of Medical Sciences Murphy, Elana; Dalhousie University and Queen Elizabeth II Health Sciences Center, Division of Rheumatology and Department of Pathology Bertsias, George; Clinical Immunology and Allergy, University of Crete, Department of Rheumatology Hanly, John; Dalhousie University and Queen Elizabeth II Health Sciences Center, Division of Rheumatology and Department of Pathology Govoni, Marcello; Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Department of Medical Sciences
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Diagnostics
Keywords:	RHEUMATOLOGY, systemic lupus erythematosus, neuropsychiatric disorders, peripheral nervous system, central nervous system, attribution algorithm

Original research article

Title page

Title: Attribution of neuropsychiatric events in Systemic Lupus Erythematosus. Independent external validation of the Italian algorithm in an international multicenter cohort.

Authors

A. Bortoluzzi[1], MD, PhD

A. Fanouriakis[2], MD,

S. Appenzeller[3], MD,

L.T.L. Costallat[3], MD,

C.A. Scirè[1,4], MD, PhD

E. Murphy[5], MD,

G. Bertsias[2], MD,

J.G. Hanly[5], MD,

M. Govoni[1], MD,

Affiliations

[1]Department of Medical Sciences, Section of Rheumatology, University of Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Cona, Ferrara, Italy

[2]Department of Rheumatology, Clinical Immunology and Allergy, University of Crete, Heraklion, Greece;

[3] Rheumatology Division, Department of Medicine, State University of Campinas, Campinas, Brazil;

[4]Epidemiology Unit, Italian Society of Rheumatology, Milan, Italy

[5]Dalhousie University and Queen Elizabeth II Health Sciences Center, Division of Rheumatology and Department of Pathology, Halifax, Nova Scotia, Canada

Correspondence to:

Alessandra Bortoluzzi,

Department of Medical Sciences, Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Cona, Ferrara

Cona (Ferrara) Via Aldo Moro 8

44124 Cona, Italy.

E-mail: brtlsn1@unife.it

Telephone number: +390532239651

FAX number +390532688109

Word count 3309

Title: Attribution of neuropsychiatric events in Systemic Lupus Erythematosus. Independent external validation of the Italian algorithm in an international multicenter cohort.

Keywords: neuropsychiatric disorders, attribution algorithm, systemic lupus erythematosus, focal manifestation, diffuse manifestation, peripheral nervous system, central nervous system, STARD.

Abstract

Objective: To validate the Italian algorithm of attribution of neuropsychiatric (NP) events to systemic lupus erythematosus (SLE) in an external international cohort of patients with SLE.

Methods: A retrospective cohort diagnostic accuracy design was followed. SLE patients attending three tertiary care lupus clinics, with one or more NP events, were included. The attribution algorithm, applied to the first NP event, considers four weighted items for each NP event: (i) time of onset of the event; (ii) type of NP event (major vs minor), (iii) concurrent non-SLE factors; (iv) favouring factors. To maintain blinding, two independent teams of assessors from each centre evaluated all NP events: the first provided an attribution diagnosis on the basis of their own clinical judgment, assumed as the "gold standard"; the second applied the algorithm, which provides a probability score ranging from 0 to 10. The performance of the algorithm was evaluated by calculating the area under the receiver operating characteristics curve (AUC).

Results: The study included 243 SLE patients with at least one NP event. The attribution score for the first NP event showed good accuracy with an AUC of 0.893 (95% CI, 0.849 - 0.937) using dichotomous outcomes for NPSLE (related vs uncertain/unrelated). The best single cut-off point to optimize classification of a first NP SLE-related event was ≥7 (sensitivity 87.9 %, specificity 82.6 %).

Conclusions: Validation exercise on an independent international cohort showed that the Italian attribution algorithm is a valid and reliable tool for the identification of NP events attributed to SLE.

Strengths and limitations of this study

- This study follows a retrospective cohort diagnostic accuracy design, nevertheless the collection of data from selected centers with medical expertise in neuropsychiatric systemic lupus erythematosus (NPSLE) may have favored a homogeneous diagnostic approach.
- The sample size is large and comprised of sufficient numbers of NP events observed in multiethnic patients.
- Some rare NP events are poorly represented in our cohort making our results not fully generalizable to all NP events included in the ACR glossary.

INTRODUCTION

Neuropsychiatric (NP) involvement is one of the most complex manifestations of systemic lupus erythematosus (SLE), characterized by a wide heterogeneity of clinical events affecting the central (CNS), peripheral (PNS) and autonomic nervous systems [1]. The phenomenology of NP involvement may include a variety of characteristics, such as NP events being focal or diffuse, acute or chronic, active or not active, single or multiple, synchronous or metachronous [2][3].

In 1999, the American College of Rheumatology (ACR) produced a standard nomenclature and set of case definitions for 19 NP syndromes (12 CNS and 7 PNS manifestations) known to occur in SLE. The ACR classification is considered a milestone in the field of NPSLE, providing definitions for each clinical NP syndrome; exclusion criteria, aimed to rule out NP events not directly related to SLE; associations, to consider potential concomitant or pre-existing confounding factors and a diagnostic work-up to assess each NP event[1]. In this respect, the ACR classification provided a useful tool for patient selection in clinical studies, offering standardized definitions that are primarily intended to create well-defined and homogenous cohorts of patients with NP involvement. However, up to date, the usefulness of ACR case definitions in clinical practice has proven to be of limited value; in fact, even if NP events (especially less specific ones, such as headaches, mood disorders, mild cognitive deficits or peripheral neuropathies not confirmed by electrophysiology [4]) has passed the ACR filter, it has been difficult to differentiate NPSLE patients from those with NP manifestations not related to SLE [5] and the final attribution still relies on the clinical judgment of experienced clinicians. Therefore, the optimal process to determine the attribution of NP events to SLE or other causes remains an unmet need.

In an attempt to address this issue, Monov and Monova proposed a model distinguishing major from minor or “common” NP events [6]. The latter were derived from a population-based study where the above mentioned less specific NP events have been considered as never being confidently attributed to SLE, since they are also frequently observed in the general population [4] [5]. In this model, it was proposed that a diagnosis of NPSLE can be reached, provided the exclusion of other causes, in the presence of at least one of the major NP events or, alternatively, in the presence of minor NP events combined with additional diagnostic data (i.e. neuroimaging, electrophysiology and laboratory abnormalities) [6]. Another attribution model, derived from the large SLE disease inception cohort recruited by the Systemic Lupus International Collaborating Clinics (SLICC) has been proposed by Hanly et al [7] [8]. This model - with two different levels of stringency (model A and B) - is based on three simple rules that take into account: i) the temporal relationship between the NP event and the diagnosis of SLE (6 months before to 15 months following SLE diagnosis, for a total period of 21 months - model A; within 10 years prior to SLE diagnosis and still present during the enrolment window - model B), ii) the type of NP event (major or minor) and iii) a comprehensive list of exclusions/associations derived from the ACR case definitions for 19 NP syndromes.

In a recent study of a large cohort of Italian SLE patients, we proposed and preliminarily validated a new algorithm, based on a probability score, to determine the attribution of NP events to SLE or to other causes [9]. The objective of the present study was to validate the Italian attribution algorithm in an international cohort of patients with SLE and at least one NP event, as per the 1999 ACR case definitions, with a first presenting NP manifestation.

METHODS

Study design

This study follows a retrospective cohort diagnostic accuracy design. Reporting complies with the "Standards for Reporting Diagnostic accuracy studies" (STARD) 2015 recommendations[10].

Participants

The study included a validating set of selected SLE patients attending 3 three tertiary care clinics dedicated to the management of patients with SLE from 1982 to 2015 (Department of Rheumatology, Clinical Immunology and Allergy, University of Crete, Heraklion, Greece; Medicine, State University of Campinas, Campinas, Brazil; Dalhousie University and Queen Elizabeth II Health Sciences Center, Halifax, Nova Scotia, Canada). Patients from each centre were selected if they satisfied the 1997 revised ACR classification criteria for SLE [11] and had one or more NP events, as defined in the ACR case definitions of 19 NP syndromes. The local ethics committees approved the study.

Attribution algorithm and case definition

A similar methodology to the one used in our original study was adopted (9). A dedicated electronic record was created, including demographic data and the core set of items for classification. Briefly, the algorithm included four items: (i) the timing of onset of the NP event (i.e. before, after or concurrent with SLE diagnosis); (ii) the type of NP event (major vs. minor or common, according to Ainiala et al[5]; (iil) the presence of confounding non-SLE factors (i.e. "associations" suggested in the glossary for the 1999 ACR case definitions); (iv) the presence of "favouring factors" (i.e. supporting attribution). The first two items applied to all NP events; for items (iii) and (iv) lists of variables specific for each NP event (derived from the glossary for the ACR case definitions for 19 NP syndromes and supplemented by systematic literature review and expert opinion) were generated (see Supplementary Tables S1 for the weight assigned to each item by the expert panel).

To maintain blinding, all first NP events, were evaluated by two independent teams of assessors from each centre, each of whom was assigned different tasks: the first provided an attribution diagnosis (related/uncertain/unrelated to SLE) on the basis of their own clinical judgment, utilizing all of the information available in the patient record; the second applied the attribution algorithm described above, using the same available information.

We chose to analyze the first NP events only, for two main reasons: a) to make results comparable to our original study and b) in order to validate rules for attribution of the first NP event before applying them also to subsequent NP events, since the attribution of subsequent events could be influenced by the classification of the first event.

Based on previously defined weights for each item [9], which sum up to a global score ranging from 0 to 10 points, two different attribution models were generated: an initial '*a priori*' model, based on the weights assigned by a Delphi round expert consensus, and an updated version, based on both '*a priori*' and 'data-driven' coefficients [9] (see below for more details).

Statistical analysis

The characteristics of the cohort are reported using descriptive statistics. Missing data were not imputed, and complete case analysis was performed. The international dataset has been evaluated separately and then compared and combined to the two previously published training and validating Italian cohorts (see Supplementary S2 for members of the Italian Study Group on Neuropsychiatric Systemic Lupus Erythematosus of the Italian Society of Rheumatology), in order to perform a pooled analysis [9].

The first analysis aimed to test discrimination of the previously reported algorithms ('a priori' and 'updated') on the international cohort. Discrimination was assessed with calculation of the area under curve (AUC) of the receiver operating characteristic (ROC) curve, using SLE related NP events (i.e. definite NPSLE) as positive and uncertain/unrelated as negative outcomes. The results from the international validating cohort were then compared to those of the training and validating Italian cohorts.

The second set of stratified analyses replicated the first, based on the type of NP event: major/minor, focal/diffuse and central/peripheral.

Further analyses replicated the process of adaptation of the *a priori* coefficient obtained by multivariate ordinal logistic models using importance weights to *a priori* and data-driven coefficients (3:1). These analyses were done in the new validating dataset and in the pooled data from all three cohorts. A final validated algorithm was defined based on robustness, discrimination and feasibility considerations.

Finally, based on the ROC tables using binary outcomes, the best threshold cut-off point for attribution, able to discriminate SLE-related (primary NPSLE) versus uncertain/not related NP events, was assessed in the international validating cohort and in the pooled dataset, based on the maximum proportion of correctly classified NPSLE cases. Other clinically relevant cut-off points with misclassification rates <10% were also defined. Results are presented as sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each cut-off point. All analyses were performed using Stata 11 (StataCorp, College Station, TX, USA).

RESULTS

International validation

The study included 243 patients with SLE and at least one NP event; patients were mainly women (219 female, 90.1%), with a mean (standard deviation, SD) age at first NP event of 39.0 (13.9) years. Mood disorder was the most frequent manifestation (n=43, 17.7%), followed by headache (n=36, 14.8%) and cerebrovascular disease (CVD, n=32, 13.2 %). 108 NP events were focal (44.4%) and 135 were diffuse (60.6%); 39 (16%) NP events involved the PNS and 204 (84%) events involved the CNS (Table 1).

Table 1. Demographic characteristics and distribution of NP events in the international cohort

	Cohort (243 pts) **N (%)**
Gender (M/F)	219/24 (90.1/9.9)
Age (years ± SD)	39 ± 13.9
Ethnicity	
Caucasian	197 (81.1)
African ancestry	24 (9.9)
Hispanic	22 (9)
CNS involvement (1st event)	
Mood disorder	43 (17.7)
Headache	36 (14.8)
CVD	32 (13.2)
Seizures	23 (9.5)
Anxiety	16 (6.6)
Cognitive dysfunction	12 (4.9)
MS-like syndrome	9 (3.7)
Myelopathy	8 (3.3)
Movement disorder	5 (2.1)
Acute confusional state	5 (2.1)
Aseptic meningitis	2 (0.8)
PNS involvement	
Cranial neuropathy	15 (6.2)
Polyneuropathy	9 (3.7)
Myasthenia gravis	9 (3.7)
Mononeuropathy	4 (1.6)
Guillain-Barré syndrome	2 (0.8)
Autonomic neuropathy	-
Plexopathy	-
Major/minor	145/98 (59.7/40.3)
Focal/diffuse	106/137 (43.6/56.4)
Central/peripheral	204/39 (83.9/16.1)

Applying the data driven and a priori coefficients (Supplementary S1), the ROC curve analysis related to the first NP event observed in the international cohort showed an AUC of 0.893 (95 % CI, 0.849 - 0.937) for the "*a priori*" model and 0.892 (95 % CI, 0.847 - 0.937) for the "data driven" model, using dichotomous outcomes (related vs. uncertain/unrelated, Figure 1), a performance comparable to the previously observed in the training and validating cohorts (Table 2).

Table 2. Comparison of the accuracy of the "a priori" and of the "updated" algorithms for attribution of the first NP events in the three cohorts. (AUC, area under the curve)

Cohort	N° of pts	A priori (original) algorithm		Updated algorithm		
		AUC	[95% Conf. Interval]	AUC	[95% Conf. Interval]	*P value**
Cohort 1 - Training (9)	225	0.845	0.797-0.892	0.857	0.811-0.904	0.03
Cohort 2 - Italian validating (9)	209	0.818	0.759-0.876	0.818	0.759-0.876	1.0
Cohort 3 – International	243	0.893	0.849-0.936	0.892	0.847-0.937	0.9
*p value^*		0.10		0.13		

**intra-cohorts comparison* *^inter-cohorts comparison*

The analysis of the "data-driven" coefficients, derived from the multivariate ordinal logistic model, and the "*a priori*" coefficients on the pooled data led to a final updated model where the weight assigned to each item was highly consistent with the assigned "*a priori*" coefficient (Supplementary S1).

Taking into account a global score ranging from 0 to 10, the best single cut-off score for correct classification of a first NP SLE-related event in the international cohort was 7 (Table 3) with a sensitivity of 87.9%, specificity of 82.64 %, a PPV of 77.68% and a NPV of 90.84. The best discriminating cut-off point was also assessed in the pooled cohorts, where the final score ≥ 7 was confirmed as the single best attribution threshold for a correct classification of the first SLE-related NP event (sensitivity 71.2 %, specificity 84.5 %, PPV 82.9 %, NPV 73.6 %); again, in the pooled cohort a score ≥ 8 was the cut-off point associated with a

misclassification probability <10% (sensitivity 47.5%, specificity 97.2 %, PPV 92.1 %, NPV 72.9 %), while a score ≤ 2 had a NPV of 90 % for a SLE-related event (Supplementary S1).

Table 3. Sensitivity, specificity, PPV and NPV for each defined cut-point derived from the application of the attribution algorithm (using "a priori" coefficients) to the first NP event observed in the international cohort

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
(>= 0)	100.0%	0.0%	40.7%	1		40.7%	
(>= 1)	100.0%	1.4%	41.6%	1.0141	0	41.1%	100.0%
(>= 2)	99.0%	6.2%	44.0%	1.0559	0.1616	42.1%	90.0%
(>= 3)	99.0%	16.7%	50.2%	1.1879	0.0606	45.0%	96.0%
(>= 4)	96.0%	31.9%	58.0%	1.41	0.1265	49.2%	92.0%
(>= 5)	92.9%	48.6%	66.7%	1.8084	0.1455	55.4%	90.9%
(>= 6)	91.9%	71.5%	79.8%	3.2284	0.113	68.9%	92.8%
(>= 7)	87.9%	82.6%	**84.8%**	5.0618	0.1467	77.7%	90.8%
(>= 8)	69.7%	91.0%	82.3%	7.7203	0.3331	84.1%	81.4%
(>= 9)	47.5%	97.2%	76.9%	17.0909	0.5403	92.1%	72.9%
(>= 10)	19.2%	99.3%	66.7%	27.6364	0.8137	95.0%	64.1%
(> 10)	0.0%	100.0%	59.3%	1			59.3%

LR, Likelihood ratio; PPV, Positive Predictive Value; NPV, Negative Predictive Value

Comparison of the performance of the algorithm in the three patient cohorts

The overall performance of the attribution algorithm applied to the three different cohorts showed some differences, being the results obtained in the international cohort even better, to the one of the original study (Table 2). To investigate the reasons for such a different performance we further analyzed the composition of the cohorts regarding the typology of the included NP events, since their heterogeneity could have impacted on the results.

As shown in table 4, the three cohorts have a different prevalence of individual NP events (Table 4): the international cohort had a higher prevalence of major, focal and peripheral NP events than the two previous cohorts.

Table 4. Prevalence rate of different NP events and performance of the algorithm in the international cohorts and comparison with the training and validating cohort.

Type of event	Training cohort (1)		Validating cohort (2)		International cohort (3)		p-values*
	%	AUC (95% CI)	%	AUC (95% CI)	%	AUC (95% CI)	
Minor	61	0.76 (0.68 - 0.84)	50.2	0.73 (0.62 - 0.83)	60.5	0.75 (0.60 - 0.90)	*0.88*
Major	39	0.93 (0.88 - 0.98)	49.8	0.81 (0.70 - 0.91)	39.5	0.89 (0.83 - 0.94)	*0.09*
p-values^		0.0006		0.124		0.055	
Focal	33.3	0.90 (0.83 - 0.97)	39.3	0.80 (0.69 - 0.92)	44.4	0.89 (0.84 - 0.96)	*0.31*
Diffuse	66.7	0.81 (0.76 - 0.88)	60.7	0.79 (0.70 - 0.87)	55.6	0.83 (0.74 - 0.92)	*0.78*
p-values^		0.101		0.542		0.110	
Central	89	0.85 (0.81 - 0.90)	91.9	0.81 (0.75 - 0.87)	83.9	0.89 (0.84 – 0.94)	*0.16*
Peripheral	11	Not applicable	8.1	0.89 (0.74 -1.00)	16.1	0.88 (0.76 – 0.98)	*0.83*
p-values^		-		0.270		0.871	

**p values inter-cohorts comparison between the AUC calculated for the different type of the event*

^p values intra-cohort comparison between the AUC calculated for the different type of the event

Stratified analyses based on the type of NP event: major/minor, focal/diffuse and central/peripheral.

The performance of the algorithm was evaluated separately by testing the events clustered by type of event. Comparing the accuracy of ROC curve in minor/major focal/diffuse and peripheral/central NP events there were no statistically significant differences in performance among the three cohorts, although, as expected, the accuracy of the model was better for major and focal events and similar for both central and peripheral manifestations (Table 4).

DISCUSSION

Recently, on behalf of the Study Group for NPSLE of the Italian Society of Rheumatology, an attribution model based on a simple numerical algorithm (ranging from 0 to 10) and derived from a robust statistical evaluation and large dataset was proposed. The original algorithm was tested on a single-center training cohort of SLE patients and then validated on an independent Italian cohort demonstrating good performance in terms of sensitivity, specificity, PPV and NPV when compared with expert clinical judgment (the current "gold" reference standard). To further validate this algorithm, taking also into account differences in ethnicity, we have tested its performance in a third independent international cohort including patients with one or more NP events, as per the 1999 ACR case definitions.

The first analysis, (based on '*a priori*' defined and 'updated' coefficients) aimed to test the discrimination power of the aforementioned algorithm on the external international cohort, demonstrated an overall performance of the algorithm highly comparable to our original study (Figure 1), confirming its high reliability. Further analyses replicated the process of adaptation of the *a priori* coefficients using data driven results of a) the new validating international cohort and b) the overall pooled dataset (all three cohorts) to validate the original model composed by pre-defined and weighted coefficients (9).

Based on the ROC tables and using binary outcomes, the best cut-off for discrimination (i.e. attribution threshold) was assessed in the international validating set and in the pooled data set. A total score ≥ 7 (range from 0 to 10) identified the maximum proportion of correctly classified NPSLE cases. Compared with the lower cut-off point we found in our original paper (≥ 6) (9) this result is worthy of comment. First, there were differences in the composition of the international and the original cohorts, with particular regard to the distribution of major NP events. Given the structure of the algorithm, higher scores are assigned to these types of NP events(12). In this way, 7 is the maximum score that can be reached by applying the model for a minor event. This implies a higher "attribution threshold" for minor NP events and, consequently, only a limited percentage of these events will be attributed to SLE using the algorithm in its current version. Accordingly, a greater prevalence of minor or diffuse events would influence the final performance of the attribution algorithm, which is derived from the cohort wherein it is applied. However, although the different composition of the individual cohorts (see Table 4) may have influenced the definition of the "attribution threshold", merging data of all three cohorts has balanced the proportion of major and minor events, thus making the newly identified cut point more reliable for attribution. Interestingly, in a recent study by Fanouriakis et al. [12] a similar result has been reached. In that study, different models of attribution, including our own, have been tested against "clinical judgment" in an independent and 'real life' cohort of SLE patients with NP involvement; applying our algorithm, the best performing cut-off point to ensure the discrimination between primary NPSLE from NP events not related to SLE was ≥7, i.e. the same as the one we found in the present validation study.

In our opinion, the "small window" of attribution for minor events is not a drawback; rather, it is in keeping with the evolution of the concept of NPSLE itself. In fact, inclusion of these minor events has substantially influenced the prevalence of NPSLE, especially in the past [13] [14] [15] [16], while in more recent years prospective studies derived from the SLICC inception cohort have challenged this concept of NPSLE, demonstrating that such events correlate poorly with conventional measures of SLE disease activity, autoantibodies, and lupus specific therapies. For this reason, these NP events require a more careful and rigorous clinical evaluation in order to determine the correct attribution [17] [18] [19]. For example, in the SLICC cohort, out of a total of 1732 patients, 17.8% had headache within the enrolment window, migraine in 60.7%, tension in 38.6%, intractable nonspecific in 7.1%, cluster in 2.6%, and intracranial hypertension in 1.0% [18]. Although the prevalence of headache rose to 58% by 10 years, only 26 patients (1.5% of the cohort) experienced "lupus headache" over the entire study, reported as a variable in the SLEDAI-2K [20] at annual assessments (19). Hanly et al also reported that mood disorders occurred in 12.7% of 1,827 patients in the SLICC cohort, and a little more than a third of the total (98 events, 38.3%) were attributed to SLE (18).

As a result of these and other studies, the frequency of NPSLE has been reevaluated[6] [21] [22] [9]. However, one must not forget that mood disorders, headache and mild cognitive deficits, all frequently observed in SLE patients, depend heavily on clinical assessment of mainly subjective symptoms; not surprisingly, it is in these cases that we observed the worst performance of the model, when compared with the current "gold standard", i.e. the judgment of experienced physicians. Nevertheless, given the intrinsic uncertainty of the diagnosis for some NP manifestations, especially the common minor NP events, to reach a confident diagnosis of primary NPSLE is sometimes only presumptive, despite the efforts to improve the tools available to the clinician. For this reason, the categorization of NPSLE events based upon a quantitative score could ensure a more standardized and consistent approach to the attribution of NP events in future studies of NPSLE [23]. Moreover, the model has characteristics of flexibility and versatility that could be adapted to the setting in which a clinician operates. It is possible to modulate the single cut-off in relation to clinical contingency, choosing from time to time sensitivity over specificity or vice versa, remembering that even more stringent cut-points (i.e. ≤ 2 and ≥ 8 meaning that the NP event has high chance to be unrelated or related to SLE, respectively) are also associated with a - relatively low -probability of misclassification (10%). It may be that more stringent cut points could be tested as a "therapeutic threshold" (i.e. to treat or not to treat). On this topic, a prospective study is already underway.

Our study has some limitations. First, the use of a retrospective design is a weakness that needs to be acknowledged, although the collection of data from selected centers with medical expertise in NPSLE may have favored a homogeneous diagnostic approach. A second limitation is the low number of some rare NP events, making our results not fully generalizable to all NP events included in the ACR glossary. Finally, this

model currently has to be considered as confidently tested and validated only for the evaluation and attribution of the first NP event. In fact, it is known that one of the most consistent risk factors for NP involvement in SLE includes the recurrence or the multiplicity of NP manifestations in the same patient [18] [17] [24] [25] [26][27] [28] and a proper attribution model, as a good clinical approach to NPSLE, has to be able to weigh a first event from a subsequent, depending on the typology and attribution of the antecedent event(s). These aspects are currently still uncovered, yet under fast development.

In summary, in this study we confirmed that the Italian attribution algorithm is a valid and robust tool for the correct identification of cases with NPSLE, with a validated score for attribution of the first NP event ≥ 7 (in a scale ranging from 0 to 10). The "*a priori* score" originally defined by the expert panel to weigh the single items included in the attribution model, was shown to be consistent and accurate and confirmed by the data driven analysis of both an external international cohort and of the pooled cohorts. In a medical setting as complex as NPSLE, we do not believe that our model should substitute the clinical judgment provided by experienced and multidisciplinary teams, but rather it could assist them in the attribution process. The categorization of NPSLE patients based upon a quantitative, reliable and validated probability score might provide a more standardized approach to the attribution of NP events, also to be used in future studies on NPSLE.

Acknowledgements: none.

Contributors:

A. Bortoluzzi: substantial contributions to the conception of the work and interpretation of data for the work; drafting the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

A. Fanouriakis : substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

S. Appenzeller: substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

L.T.L. Costallat: substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

C.A. Scirè: substantial contributions to the analysis of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

E. Murphy: substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

G. Bertsias: substantial contributions to the interpretation of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

J.G. Hanly: substantial contributions to the interpretation of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

M. Govoni: substantial contributions to the conception of the work and interpretation of data for the work; drafting the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published.

Competing interests: none declared.

Funding: S. Appenzeller has received grant from Conselho Nacional de Desenvolvimento Cientifico (CNPq 304255/2015-7).

Data sharing statement: the data set is available from corresponding author on request.

REFERENCES

1 The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. *Arthritis Rheum* 1999;42:599–608. doi:10.1002/1529-0131(199904)42:4<599::AID-ANR2>3.0.CO;2-F

2 Bertsias GK, Boumpas DT. Pathogenesis, diagnosis and management of neuropsychiatric SLE manifestations. *Nat Rev Rheumatol* 2010;6:358–67. doi:10.1038/nrrheum.2010.62

3 Govoni M, Bortoluzzi A, Padovan M, *et al.* The diagnosis and clinical management of the neuropsychiatric manifestations of lupus. *J Autoimmun* Published Online First: 11 July 2016. doi:10.1016/j.jaut.2016.06.013

4 Ainiala H, Loukkola J, Peltola J, *et al.* The prevalence of neuropsychiatric syndromes in systemic lupus erythematosus. *Neurology* 2001;57:496–500.

5 Ainiala H, Hietaharju A, Loukkola J, *et al.* Validity of the new American College of Rheumatology criteria for neuropsychiatric lupus syndromes: a population-based evaluation. *Arthritis Rheum* 2001;45:419–23.

6 Monov S, Monova D. Classification criteria for neuropsychiatric systemic lupus erythematosus: do they need a discussion? *Hippokratia* 2008;12:103–7.

7 Hanly JG. The neuropsychiatric SLE SLICC inception cohort study. *Lupus* 2008;17:1059–63. doi:10.1177/0961203308097568

8 Hanly JG, Urowitz MB, Su L, *et al.* Prospective analysis of neuropsychiatric events in an international disease inception cohort of patients with systemic lupus erythematosus. *Ann Rheum Dis* 2010;69:529–35. doi:10.1136/ard.2008.106351

9 Bortoluzzi A, Scirè CA, Bombardieri S, *et al.* Development and validation of a new algorithm for attribution of neuropsychiatric events in systemic lupus erythematosus. *Rheumatol Oxf Engl* 2015;54:891–8. doi:10.1093/rheumatology/keu384

10 Bossuyt PM, Cohen JF, Gatsonis CA, *et al.* STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. *Ann Transl Med* 2016;4. doi:10.3978/j.issn.2305-5839.2016.02.06

11 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725. doi:10.1002/1529-0131(199709)40:9<1725::AID-ART29>3.0.CO;2-Y

12 Fanouriakis A, Pamfil C, Rednic S, *et al.* Is it primary neuropsychiatric systemic lupus erythematosus? Performance of existing attribution models using physician judgment as the gold standard. *Clin Exp Rheumatol* 2016.

13 Brey RL, Holliday SL, Saklad AR, *et al.* Neuropsychiatric syndromes in lupus: prevalence using standardized definitions. *Neurology* 2002;58:1214–20.

14 Sanna G, Bertolaccini ML, Cuadrado MJ, *et al.* Neuropsychiatric manifestations in systemic lupus erythematosus: prevalence and association with antiphospholipid antibodies. *J Rheumatol* 2003;30:985–92.

15 Mok CC, Lau CS, Wong RW. Neuropsychiatric manifestations and their clinical associations in southern Chinese patients with systemic lupus erythematosus. *J Rheumatol* 2001;28:766–71.

16 Petri M, Naqibuddin M, Carson KA, *et al.* Cognitive function in a systemic lupus erythematosus inception cohort. *J Rheumatol* 2008;35:1776–81.

17 Hanly JG, Su L, Urowitz MB, *et al.* Mood Disorders in Systemic Lupus Erythematosus: Results From an International Inception Cohort Study. *Arthritis Rheumatol Hoboken NJ* 2015;67:1837–47. doi:10.1002/art.39111

18 Hanly JG, Urowitz MB, O'Keeffe AG, *et al.* Headache in Systemic Lupus Erythematosus: Results From a Prospective, International Inception Cohort Study. *Arthritis Rheum* 2013;65:2887–97. doi:10.1002/art.38106

19 Lockshin MD. Editorial: Splitting Headache (Off). *Arthritis Rheum* 2013;65:2759–61. doi:10.1002/art.38108

20 Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol* 2002;29:288–91.

21 Hanly JG, Urowitz MB, Su L, *et al.* Prospective analysis of neuropsychiatric events in an international disease inception cohort of patients with systemic lupus erythematosus. *Ann Rheum Dis* 2010;69:529–35. doi:10.1136/ard.2008.106351

22 Hanly JG, Urowitz MB, Sanchez-Guerrero J, *et al.* Neuropsychiatric events at the time of diagnosis of systemic lupus erythematosus: an international inception cohort study. *Arthritis Rheum* 2007;56:265–73. doi:10.1002/art.22305

23 Hanly JG. Attribution in the assessment of nervous system disease in SLE. *Rheumatology* 2015;54:755–6. doi:10.1093/rheumatology/keu458

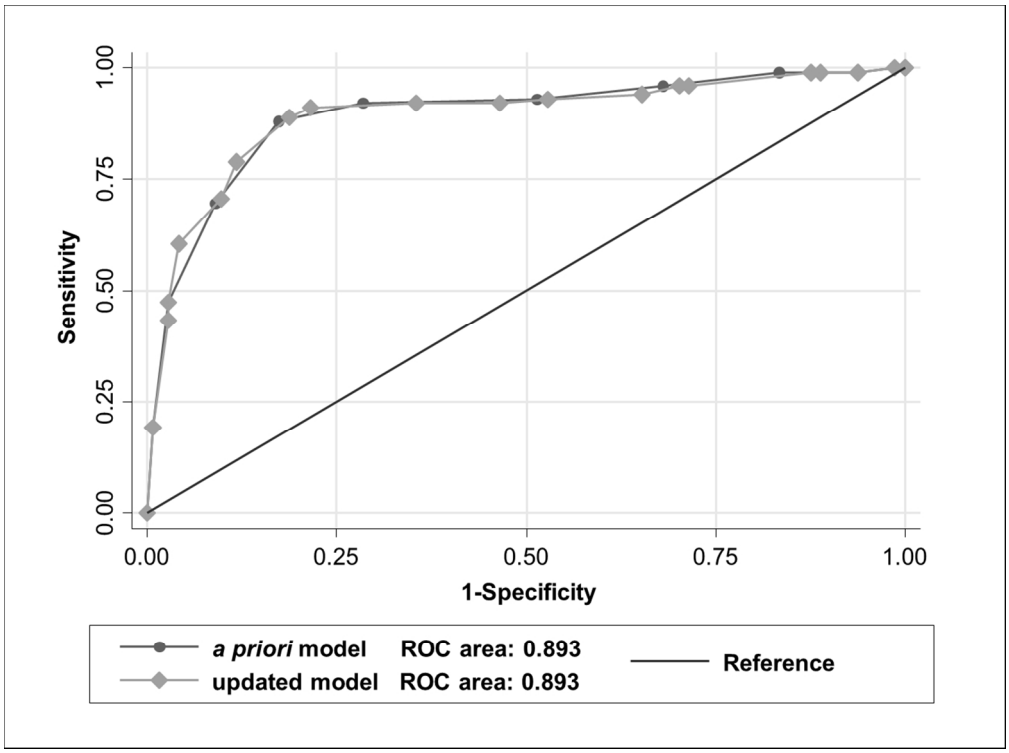
24 Bertsias G, Ioannidis JPA, Boletis J, *et al.* EULAR recommendations for the management of systemic lupus erythematosus. Report of a Task Force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics. *Ann Rheum Dis* 2008;67:195–205. doi:10.1136/ard.2007.070367

25 Hanly JG, Urowitz MB, Su L, *et al.* Seizure disorders in systemic lupus erythematosus results from an international, prospective, inception cohort study. *Ann Rheum Dis* 2012;71:1502–9. doi:10.1136/annrheumdis-2011-201089

26 Padovan M, Castellino G, Bortoluzzi A, *et al.* Factors and comorbidities associated with central nervous system involvement in systemic lupus erythematosus: a retrospective cross-sectional case–control study from a single center. *Rheumatol Int* 2010;32:129–35. doi:10.1007/s00296-010-1565-4

27 Appenzeller S, Cendes F, Costallat LTL. Epileptic seizures in systemic lupus erythematosus. *Neurology* 2004;63:1808–12.

28 Mikdashi J, Krumholz A, Handwerger B. Factors at diagnosis predict subsequent occurrence of seizures in systemic lupus erythematosus. *Neurology* 2005;64:2102–7. doi:10.1212/01.WNL.0000165959.98370.D5



Receiver operating characteristic (ROC) curve using dichotomous outcomes (related vs. uncertain/ not related), for attribution of the first NP event observed in the international cohort.

224x166mm (150 x 150 DPI)

Supplementary Table S1. Comparison between «a priori» vs «data driven» estimated scores (pooled analysis in 677 NP 1st events).

Item	CATEGORY	A PRIORI ORIGINAL COEFFICIENTS*	DATA DRIVEN (REFINED A POSTERIORI) COEFFICIENTS
Time onset of NP event	Before	**0**	0
	After	**2**	2.1
	Concurrent	**3**	3.1
Minor event (Ainiala list)	Yes	**0**	0
	No	**3**	2.9
Presence of Confounding Factors	≥ 1	**0**	0
	1	**1**	0.7
	No	**2**	1.9
Presence of favouring factors	No	**0**	0
	1	**1**	0.9
	≥ 1	**2**	2.1

(^) The resulting global score can range from 0 to 10; details for definition of each item category are reported elsewhere (9).

**A priori coefficients (original coefficients), identify the better scores to be used in the final version of the Italian attribution algorithm, the so-called "original algorithm" (^)*

Supplementary Table S1-bis. Detailed report of sensitivity, specificity, PPV and NPV for each defined cut-point derived from the application of the attribution algorithm to the first NP event observed in the training (cohort 1), validating (cohort 2) and pooled cohorts (all three cohorts, including the international cohort).

Cohort	Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
1	(>= 0)	100.00%	0.00%	62.67%	100.00%		62.67%	
1	(>= 1)	100.00%	0.00%	62.67%	1		62.67%	
1	(>= 2)	100.00%	1.19%	63.11%	1.012	0	62.95%	100.00%
1	(>= 3)	99.29%	7.14%	64.89%	1.0693	0.0993	64.22%	85.70%
1	(>= 4)	99.29%	16.67%	68.44%	1.1915	0.0426	66.67%	93.33%
1	(>= 5)	95.74%	35.71%	73.33%	1.4894	0.1191	71.43%	83.32%
1	**(>= 6)**	85.11%	61.90%	**76.44%**	2.234	0.2406	78.95%	71.24%
1	(>= 7)	58.87%	92.86%	71.56%	8.2411	0.443	93.26%	57.36%
1	(>= 8)	36.17%	100.00%	60.00%	0.6383		100.00%	48.28%
1	(>= 9)	21.99%	100.00%	51.11%	0.7801		100.00%	43.30%
1	(>= 10)	6.38%	100.00%	41.33%	0.9362		100.00%	
1	(> 10)	0.00%	100.00%	37.33%	1			

Cohort	Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
2	(>= 0)	100.00%	0.00%	51.20%	1		51.20%	
2	(>= 1)	100.00%	0.00%	51.20%	1		51.20%	
2	(>= 2)	100.00%	0.98%	51.67%	1.0099	0	51.44%	100.00%
2	(>= 3)	97.20%	5.88%	52.63%	1.0327	0.4766	52.00%	66.69%
2	(>= 4)	96.26%	17.65%	57.89%	1.1689	0.2118	55.08%	81.81%
2	(>= 5)	91.59%	36.27%	64.59%	1.4372	0.2319	60.12%	80.44%
2	(>= 6)	85.98%	61.76%	74.16%	2.2487	0.227	70.23%	80.77%
2	**(>= 7)**	71.96%	80.39%	**76.08%**	3.6701	0.3488	79.38%	73.21%
2	(>= 8)	58.88%	92.16%	75.12%	7.507	0.4462	88.74%	68.12%
2	(>= 9)	32.71%	95.10%	63.16%	6.6729	0.7076	87.50%	57.40%
2	(>= 10)	10.28%	98.04%	53.11%	5.243	0.9151	84.62%	51.02%
2	(> 10)	0.00%	100.00%	48.80%	1			48.80%

Cohort	Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
pooled	(>= 0)	100.00%	0.00%	51.26%	1		51.26%	
pooled	(>= 1)	100.00%	0.91%	51.70%	1.0092	0	51.48%	100.00%
pooled	**(>= 2*)**	99.71%	3.33%	52.73%	1.0315	0.0865	52.03%	**91.61%**
pooled	(>= 3)	98.56%	10.91%	55.83%	1.1063	0.1321	53.77%	87.81%
pooled	(>= 4)	97.41%	23.64%	61.45%	1.2756	0.1097	57.29%	89.67%
pooled	(>= 5)	93.66%	41.52%	68.24%	1.6014	0.1527	62.74%	86.17%
pooled	(>= 6)	87.32%	66.06%	76.96%	2.5728	0.1919	73.01%	83.21%
pooled	**(>= 7)**	71.18%	84.55%	**77.70%**	4.6059	0.3409	82.89%	73.61%
pooled	**(>= 8*)**	52.74%	93.64%	72.67%	8.2874	0.5047	**89.71%**	65.33%
pooled	(>= 9)	32.56%	97.27%	64.11%	11.9404	0.6933	92.62%	57.84%
pooled	(>= 10)	11.24%	99.09%	54.06%	12.3631	0.8958	92.85%	51.50%
pooled	(> 10)	0.00%	100.00%	48.74%	1			48.74%

***Cut points ensuring a misclassification probability less than 10%**

Supplementary S2

Members of the Italian Study Group on Neuropsychiatric Systemic Lupus Erythematosus of the Italian Society of Rheumatology:

Stefano Bombardieri[1],
Fabrizio Conti[2],
Salvatore De Vita[3],
Andrea Doria[4],
Gianfranco Ferraccioli[5],
Elisa Gremese[5],
Alessandro Mathieu[6],
Marta Mosca[1],
Melissa Padovan[7],
Matteo Piga[6],
Angela Tincani[8],
Paola Tomietto[9],
Guido Valesini[2],
Margherita Zen[4]

[1]Rheumatology Unit, Department of Clinical and Experimental Medicine, University of Pisa, Pisa

[2]Department of Internal Medicine and Medical Specialties, SapienzaUniversity of Rome, Rome

[3]Rheumatology Clinic, Azienda Ospedaliero Universitaria 'S. Maria della Misericordia' and DSMB, Department of Medical and Biological Sciences, University of Udine, Udine

[4]Department of Clinical and Experimental Medicine, Division of Rheumatology, University of Padova, Padova

[5]Division of Rheumatology and Internal Medicine, Institute of Rheumatology and Affine Sciences, CIC, Catholic University of the Sacred Heart, Rome,

[6]Rheumatology Unit, Department of Medical Sciences, University of Cagliari and AOU University Clinic, Cagliari

[7]Department of Medical Sciences, Section of Rheumatology, University of Ferrara and Azienda Ospedaliero Universitaria Sant'Anna di Cona, Ferrara

[8]Rheumatology and Clinical Immunology Unit, Spedali Civili and University of Brescia, Brescia

[9]Internal Medicine, AOU 'Ospedali Riuniti' of Trieste, Trieste, Italy

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	4
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	4
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	6-7
	4	Study objectives and hypotheses	6-7
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	8
Participants	**6**	Eligibility criteria	8
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	8
	8	Where and when potentially eligible participants were identified (setting, location and dates)	8
	9	Whether participants formed a consecutive, random or convenience series	8
Test methods	**10a**	Index test, in sufficient detail to allow replication	8
	10b	Reference standard, in sufficient detail to allow replication	8
	11	Rationale for choosing the reference standard (if alternatives exist)	8
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	8
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	8
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	8
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	8
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	9
	15	How indeterminate index test or reference standard results were handled	9
	16	How missing data on the index test and reference standard were handled	9
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	9
	18	Intended sample size and how it was determined	9
RESULTS			
Participants	**19**	Flow of participants, using a diagram	10
	20	Baseline demographic and clinical characteristics of participants	11
	21a	Distribution of severity of disease in those with the target condition	11
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	NA
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	12
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	12, 14
	25	Any adverse events from performing the index test or the reference standard	NA
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	17-18
	27	Implications for practice, including the intended use and clinical role of the index test	18
OTHER INFORMATION			
	28	Registration number and name of registry	NA
	29	Where the full study protocol can be accessed	NA
	30	Sources of funding and other support; role of funders	19

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Validity of the Italian algorithm for the attribution of neuropsychiatric events in systemic lupus erythematosus: a retrospective multicenter international diagnostic cohort study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015546.R1
Article Type:	Research
Date Submitted by the Author:	07-Mar-2017
Complete List of Authors:	Bortoluzzi, Alessandra; Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Department of Medical Sciences Fanouriakis, Antonis; Clinical Immunology and Allergy, University of Crete, Department of Rheumatology Appenzeller, Simone; Rheumatology Division, Department of Medicine, State University of Campinas Costallat, Lilian ; Rheumatology Division, Department of Medicine, State University of Campinas Scirè, Carlo; Italian Society for Rheumatology (SIR), Epidemiology Unit; Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Department of Medical Sciences Murphy, Elana; Dalhousie University and Queen Elizabeth II Health Sciences Center, Division of Rheumatology and Department of Pathology Bertsias, George; Clinical Immunology and Allergy, University of Crete, Department of Rheumatology Hanly, John; Dalhousie University and Queen Elizabeth II Health Sciences Center, Division of Rheumatology and Department of Pathology Govoni, Marcello; Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Department of Medical Sciences
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Diagnostics
Keywords:	RHEUMATOLOGY, systemic lupus erythematosus, neuropsychiatric disorders, peripheral nervous system, central nervous system, attribution algorithm

Original research article

Title page

Title: Validity of the Italian algorithm for the attribution of neuropsychiatric events in systemic lupus erythematosus: a retrospective multicenter international diagnostic cohort study.

Authors

A. Bortoluzzi[1], MD, PhD

A. Fanouriakis[2], MD,

S. Appenzeller[3], MD,

L.T.L. Costallat[3], MD,

C.A. Scirè[1,4], MD, PhD

E. Murphy[5], MD,

G. Bertsias[2], MD,

J.G. Hanly[5], MD,

M. Govoni[1], MD,

Affiliations

[1]Department of Medical Sciences, Section of Rheumatology, University of Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Cona, Ferrara, Italy

[2]Department of Rheumatology, Clinical Immunology and Allergy, University of Crete, Heraklion, Greece;

[3] Rheumatology Division, Department of Medicine, State University of Campinas, Campinas, Brazil;

[4]Epidemiology Unit, Italian Society of Rheumatology, Milan, Italy

[5]Dalhousie University and Queen Elizabeth II Health Sciences Center, Division of Rheumatology and Department of Pathology, Halifax, Nova Scotia, Canada

Correspondence to:

Alessandra Bortoluzzi,

Department of Medical Sciences, Section of Rheumatology, University of Ferrara, Ferrara and Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara, Cona, Ferrara

Cona (Ferrara) Via Aldo Moro 8

44124 Cona, Italy.

E-mail: brtlsn1@unife.it

Telephone number: +390532239651

FAX number +390532688109

Word count 3309

Title: Validity of the Italian algorithm for the attribution of neuropsychiatric events in SLE: a retrospective multicenter international diagnostic cohort study.

Keywords: neuropsychiatric disorders, attribution algorithm, systemic lupus erythematosus, focal manifestation, diffuse manifestation, peripheral nervous system, central nervous system, STARD.

Abstract

Objective: To validate the Italian algorithm of attribution of neuropsychiatric (NP) events to systemic lupus erythematosus (SLE) in an external international cohort of patients with SLE.

Methods: A retrospective cohort diagnostic accuracy design was followed. SLE patients attending three tertiary care lupus clinics, with one or more NP events, were included. The attribution algorithm, applied to the NP manifestations, considers four weighted items for each NP event: (i) time of onset of the event; (ii) type of NP event (major vs minor), (iii) concurrent non-SLE factors; (iv) favouring factors. To maintain blinding, two independent teams of assessors from each centre evaluated all NP events: the first provided an attribution diagnosis on the basis of their own clinical judgment, assumed as the "gold standard"; the second applied the algorithm, which provides a probability score ranging from 0 to 10. The performance of the algorithm was evaluated by calculating the area under the receiver operating characteristics curve (AUC).

Results: The study included 243 SLE patients with at least one NP manifestation, for a total of 336 events. 285 (84.8 %) NP events involved the CNS and 51 (15.2%) the PNS. The attribution score for the first NP event showed good accuracy with an AUC of 0.893 (95% CI, 0.849 - 0.937) using dichotomous outcomes for NPSLE (related vs uncertain/unrelated). The best single cut-off point to optimize classification of a first NP SLE-related event was ≥7 (sensitivity 87.9 %, specificity 82.6 %). Satisfactory accuracy was observed also for subsequent NP events

Conclusions: Validation exercise on an independent international cohort showed that the Italian attribution algorithm is a valid and reliable tool for the identification of NP events attributed to SLE.

Strengths and limitations of this study

- This study follows a retrospective cohort design that could have influenced the proper attribution of neuropsychiatric (NP) events; nevertheless the collection of data from selected centers with medical expertise in neuropsychiatric systemic lupus erythematosus (NPSLE) may have favored a homogeneous diagnostic approach.
- The sample size is large and comprised of sufficient numbers of NP events observed in multiethnic patients.
- Some rare NP events are poorly represented in our cohort making our results not fully generalizable to all NP events included in the ACR glossary.

INTRODUCTION

Neuropsychiatric (NP) involvement is one of the most complex manifestations of systemic lupus erythematosus (SLE), characterized by a wide heterogeneity of clinical events affecting the central (CNS), peripheral (PNS) and autonomic nervous systems [1]. The phenomenology of NP involvement may include a variety of characteristics, such as NP events being focal or diffuse, acute or chronic, active or not active, single or multiple, synchronous or metachronous [2][3].

In 1999, the American College of Rheumatology (ACR) produced a standard nomenclature and set of case definitions for 19 NP syndromes (12 CNS and 7 PNS manifestations) known to occur in SLE. The ACR classification is considered a milestone in the field of NPSLE, providing definitions for each clinical NP syndrome; exclusion criteria, aimed to rule out NP events not directly related to SLE; associations, to consider potential concomitant or pre-existing confounding factors and a diagnostic work-up to assess each NP event[1]. In this respect, the ACR classification provided a useful tool for patient selection in clinical studies, offering standardized definitions that are primarily intended to create well-defined and homogenous cohorts of patients with NP involvement. However, up to date, the usefulness of ACR case definitions in clinical practice has proven to be of limited value; in fact, even if NP events (especially less specific ones, such as headaches, mood disorders, mild cognitive deficits or peripheral neuropathies not confirmed by electrophysiology [4]) has passed the ACR filter, it has been difficult to differentiate NPSLE patients from those with NP manifestations not related to SLE [5] and the final attribution still relies on the clinical judgment of experienced clinicians. Therefore, the optimal process to determine the attribution of NP events to SLE or other causes remains an unmet need.

In an attempt to address this issue, Monov and Monova proposed a model distinguishing major from minor or “common” NP events [6]. The latter were derived from a population-based study where the above mentioned less specific NP events have been considered as never being confidently attributed to SLE, since they are also frequently observed in the general population [4] [5]. In this model, it was proposed that a diagnosis of NPSLE can be reached, provided the exclusion of other causes, in the presence of at least one of the major NP events or, alternatively, in the presence of minor NP events combined with additional diagnostic data (i.e. neuroimaging, electrophysiology and laboratory abnormalities) [6]. Another attribution model, derived from the large SLE disease inception cohort recruited by the Systemic Lupus International Collaborating Clinics (SLICC) has been proposed by Hanly et al [7] [8]. This model - with two different levels of stringency (model A and B) - is based on three simple rules that take into account: i) the temporal relationship between the NP event and the diagnosis of SLE (6 months before to 15 months following SLE diagnosis, for a total period of 21 months - model A; within 10 years prior to SLE diagnosis and still present during the enrolment window - model B), ii) the type of NP event (major or minor) and iii) a comprehensive list of exclusions/associations derived from the ACR case definitions for 19 NP syndromes.

In a recent study of a large cohort of Italian SLE patients, we proposed and preliminarily validated a new algorithm, based on a probability score, to determine the attribution of NP events to SLE or to other causes [9]. The objective of the present study was to validate the Italian attribution algorithm in an international cohort of patients with SLE and at least one NP event, as per the 1999 ACR case definitions.

METHODS

Study design

This study follows a retrospective cohort diagnostic accuracy design. Reporting complies with the "Standards for Reporting Diagnostic accuracy studies" (STARD) 2015 recommendations [10].

Participants

The study included a validating set of selected SLE patients attending 3 three tertiary care clinics dedicated to the management of patients with SLE from 1982 to 2015 (Department of Rheumatology, Clinical Immunology and Allergy, University of Crete, Heraklion, Greece; Medicine, State University of Campinas, Campinas, Brazil; Dalhousie University and Queen Elizabeth II Health Sciences Center, Halifax, Nova Scotia, Canada). Patients from each centre were selected if they satisfied the 1997 revised ACR classification criteria for SLE [11] and had one or more NP events, as defined in the ACR case definitions of 19 NP syndromes. The local ethics committees approved the study.

Attribution algorithm and case definition

A similar methodology to the one used in our original study was adopted [9]. A dedicated electronic record was created, including demographic data and the core set of items for classification. Briefly, the algorithm included four items: (i) the timing of onset of the NP event (i.e. before, >6 months; concurrent, within 6 months or after SLE diagnosis); (ii) the type of NP event (major vs. minor or common, according to Ainiala et al[5]; (iiI) the presence of confounding non-SLE factors (i.e. "associations" suggested in the glossary for the 1999 ACR case definitions); (iv) the presence of "favouring factors" (i.e. supporting attribution). The first two items applied to all NP events; for items (iii) and (iv) lists of variables specific for each NP event (derived from the glossary for the ACR case definitions for 19 NP syndromes and supplemented by systematic literature review and expert opinion) were generated (see Supplementary S1 and S2 for the complete lists and Supplementary S3, Tables S1 for the weight assigned to each item by the expert panel).

To maintain blinding, all first NP events, were evaluated by two independent teams of assessors from each centre, each of whom was assigned different tasks: the first provided an attribution diagnosis (related/uncertain/unrelated to SLE) on the basis of their own clinical judgment, utilizing all of the information available in the patient record; the second applied the attribution algorithm described above, using the same available information.

We chose to analyze primarily the first NP events, for two main reasons: a) to make results comparable to our original study and b) in order to validate rules for attribution of the first NP event before applying them also to subsequent NP events, since the attribution of subsequent events could be influenced by the classification of the first event. To verify this point, we evaluated separately subsequent NP events.

Based on previously defined weights for each item [9], which sum up to a global score ranging from 0 to 10 points, two different attribution models were generated: an initial '*a priori*' model, based on the weights assigned by a Delphi round expert consensus, and an updated version, based on both '*a priori*' and 'data-driven' coefficients [9] (see below for more details).

Statistical analysis

The characteristics of the cohort are reported using descriptive statistics. Missing data were not imputed, and complete case analysis was performed. The international dataset has been evaluated separately and then compared and combined to the two previously published training and validating Italian cohorts (see Supplementary S4 for members of the Italian Study Group on Neuropsychiatric Systemic Lupus Erythematosus of the Italian Society of Rheumatology), in order to perform a pooled analysis [9].

The first analysis aimed to test discrimination of the previously reported algorithms ('a priori' and 'updated') on the international cohort. Discrimination was assessed with calculation of the area under curve (AUC) of the receiver operating characteristic (ROC) curve, using SLE related NP events (i.e. definite NPSLE) as positive and uncertain/unrelated as negative outcomes. The results from the international validating cohort were then compared to those of the training and validating Italian cohorts.

The second set of stratified analyses replicated the first, based on the type of NP event: major/minor, focal/diffuse, ischemic/non-ischemic and central/peripheral.

Further analyses replicated the process of adaptation of the *a priori* coefficient obtained by multivariate ordinal logistic models using importance weights to *a priori* and data-driven coefficients (3:1). These analyses were done in the new validating dataset and in the pooled data from all three cohorts. A final validated algorithm was defined based on robustness, discrimination and feasibility considerations.

Finally, based on the ROC tables using binary outcomes, the best threshold cut-off point for attribution, able to discriminate SLE-related (primary NPSLE) versus uncertain/not related NP events, was assessed in the international validating cohort and in the pooled dataset, based on the maximum proportion of correctly classified NPSLE cases. Other clinically relevant cut-off points with misclassification rates <10% were also defined. Results are presented as sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each cut-off point. All analyses were performed using Stata 11 (StataCorp, College Station, TX, USA).

RESULTS

International validation

The study included 243 patients with SLE (178 from Heraklion, Greece; 53 from Campinas, Brazil and 12 Nova Scotia, Canada) and at least one NP event for a total of 336 events. 197 patients (81.1 %) were of European ancestry, 24 (9.9 %) of African ancestry and 22 (9 %) Hispanic; they were mainly women (219 female, 90.1%; 24 males, 9.9%), with a mean (standard deviation, SD) age at first NP event of 39.0 (13.9) years. 285 (84.8 %) NP events involved the CNS and 51 (15.2%) events involved the PNS (Table 1). Mood disorder was the most frequent manifestation (n=55, 16.4%), followed by headache (n=50, 14.9%) and cerebrovascular disease (CVD, n=38, 11.3 %).

Table 1. Distribution of NP events in the international cohort.

	First, N (%)	Following, N (%)	All, N (%)
PNS and CNS involvement	243 (100)	93 (100)	336 (100)
CNS involvement	202 (82.3)	83 (89.2)	285 (84.8)
Mood disorder	43 (17.7)	12 (12.9)	55 (16.4)
Headache	35 (14.4)	15 (16.1)	50 (14.9)
CVD	33 (13.6)	5 (5.4)	38 (11.3)
Seizures	22 (9.1)	12 (12.9)	34 (10.1)
Anxiety	16 (6.6)	2 (2.1)	18 (5.4)
Cognitive dysfunction	13 (5.3)	20 (21.5)	33 (9.8)
Psychosis	11 (4.5)	6 (6.5)	17 (5.1)
MS-like syndrome	9 (3.7)	1 (1.08)	10 (3)
Myelopathy	8 (3.3)	4 (4.3)	12 (3.6)
Movement disorder	5 (2.1)	1 (1.1)	6 (1.8)
Acute confusional state	5 (2.1)	3 (3.2)	8 (2.4)
Aseptic meningitis	2 (0.8)	2 (2.1)	4 (1.2)
PNS involvement	41 (17.7)	10 (10.8)	51 (15.2)
Cranial neuropathy	15 (6.2)	3 (3.2)	18 (5.4)
Polyneuropathy	10 (4.1)	4 (4.3)	14 (4.2)
Myasthenia gravis	9 (3.7)	-	9 (2.7)
Mononeuropathy	4 (1.6)	2 (2.1)	6 (1.8)
Guillain-Barré syndrome	3 (1.2)	-	3 (0.9)
Autonomic neuropathy	-	-	-
Plexopathy	-	1 (1.1)	1 (0.3)
Major/minor	148/95 (60.9/39.1)	64/29 (68.8/31.2)	212/124 (63.1/36.9)
Focal/diffuse	109/134 (44.9/55.1)	61/32 (65.6/34.4)	170/166 (50.6/49.4)
Central/peripheral	202/41 (83.1/16.9)	83/10 (89.2/10.8)	287/49 (85.4/14.6)

Applying the data driven and a priori coefficients (Supplementary S3, Table S1), the ROC curve analysis related to the first NP event observed in the international cohort showed an AUC of 0.893 (95 % CI, 0.849 - 0.937) for the "*a priori*" model and 0.892 (95 % CI, 0.847 - 0.937) for the "data driven" model, using

dichotomous outcomes (related vs. uncertain/unrelated, Figure 1), a performance comparable to the previously observed in the training and validating cohorts (Table 2).

Table 2. Comparison of the accuracy of the "a priori" and of the "updated" algorithms for attribution of the first NP events in the three cohorts. (AUC, area under the curve)

Cohort	N° of pts	A priori (original) algorithm		Updated algorithm		
		AUC	[95% Conf. Interval]	AUC	[95% Conf. Interval]	*P value**
Cohort 1 - Training (9)	225	0.845	0.797-0.892	0.857	0.811-0.904	0.03
Cohort 2 - Italian validating [9]	209	0.818	0.759-0.876	0.818	0.759-0.876	1.0
Cohort 3 – International [9]	243	0.893	0.849-0.936	0.892	0.847-0.937	0.9
*p value^*		0.10		0.13		

**intra-cohorts comparison* *^inter-cohorts comparison*

The analysis of the "data-driven" coefficients, derived from the multivariate ordinal logistic model, and the "*a priori*" coefficients on the pooled data led to a final updated model where the weight assigned to each item was highly consistent with the assigned "*a priori*" coefficient (Supplementary S3, Table S1).

The ROC curve analysis stratified for the timing of onset of the fist NP events, before, concomitant or after the diagnosis of SLE, showed an AUC of 0.68 (95 % CI, 0.12 – 1.00), 0.78 (95 % CI, 0.64 – 0.92) and 0.85 (95 % CI, 0.64 – 0.92), respectively. A similar analysis applied to subsequent NP events, showed again a good performance with an AUC of 0.80 (95 % CI, 0.71 – 0.88) (see details in Supplementary S3, Table S2).

Taking into account a global score ranging from 0 to 10, the best single cut-off score for correct classification of a first NP SLE-related event in the international cohort was 7 (Table 3) with a sensitivity of 87.9%, specificity of 82.64 %, a PPV of 77.68% and a NPV of 90.84. The best discriminating cut-off point was also assessed in the pooled cohorts, where the final score ≥ 7 was confirmed as the single best attribution threshold for a correct classification of the first SLE-related NP event (sensitivity 71.2 %, specificity 84.5 %, PPV 82.9 %, NPV 73.6 %); again, in the pooled cohort a score ≥ 8 was the cut-off point associated with a misclassification probability <10% (sensitivity 47.5%, specificity 97.2 %, PPV 92.1 %, NPV 72.9 %), while a

score ≤ 2 had a NPV of 90 % for a SLE-related event (Supplementary S3, Table S3). Including subsequent NP events, the same cut-off points have been deemed applicable as best discrimination threshold.

Table 3. Sensitivity, specificity, PPV and NPV for each defined cut-point derived from the application of the attribution algorithm (using "a priori" coefficients) to the first NP event observed in the international cohort.

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
(>= 0)	100.0%	0.0%	40.7%	1		40.7%	
(>= 1)	100.0%	1.4%	41.6%	1.0141	0	41.1%	100.0%
(>= 2)	99.0%	6.2%	44.0%	1.0559	0.1616	42.1%	90.0%
(>= 3)	99.0%	16.7%	50.2%	1.1879	0.0606	45.0%	96.0%
(>= 4)	96.0%	31.9%	58.0%	1.41	0.1265	49.2%	92.0%
(>= 5)	92.9%	48.6%	66.7%	1.8084	0.1455	55.4%	90.9%
(>= 6)	91.9%	71.5%	79.8%	3.2284	0.113	68.9%	92.8%
(>= 7)	87.9%	82.6%	**84.8%**	5.0618	0.1467	77.7%	90.8%
(>= 8)	69.7%	91.0%	82.3%	7.7203	0.3331	84.1%	81.4%
(>= 9)	47.5%	97.2%	76.9%	17.0909	0.5403	92.1%	72.9%
(>= 10)	19.2%	99.3%	66.7%	27.6364	0.8137	95.0%	64.1%
(> 10)	0.0%	100.0%	59.3%	1			59.3%

LR, Likelihood ratio; PPV, Positive Predictive Value; NPV, Negative Predictive Value

Comparison of the performance of the algorithm in the three patient cohorts

The overall performance of the attribution algorithm applied to the three different cohorts showed some differences, being the results obtained in the international cohort even better, to the one of the original study (Table 2). To investigate the reasons for such a different performance we further analyzed the composition of the cohorts regarding the typology of the included NP events, since their heterogeneity could have impacted on the results.

As shown in table 4, the three cohorts have a different prevalence of individual NP events (Table 4): the

International cohort had a higher prevalence of major, focal and peripheral NP events than the two previous cohorts.

Table 4. Prevalence rate of different NP events and performance of the algorithm in the international cohorts and comparison with the training and validating cohort.

Type of event	Training cohort (1)		Validating cohort (2)		International cohort (3)		p-values*
	N (%)	AUC (95% CI)	N (%)	AUC (95% CI)	N (%)	AUC (95% CI)	
Minor	136 (60.4)	0.76 (0.68 - 0.84)	104 (50.2)	0.73 (0.62 - 0.83)	95 (60.9)	0.75 (0.60 - 0.90)	*0.88*
Major	89 (39.6)	0.93 (0.88 - 0.98)	105 (49.8)	0.81 (0.70 - 0.91)	148 (39.1)	0.89 (0.83 - 0.94)	*0.09*
p-values^		0.0006		0.12		0.06	
Focal	76 (33.8)	0.90 (0.83 - 0.97)	83 (39.7)	0.80 (0.69 - 0.92)	109 (44.9)	0.89 (0.84 - 0.96)	*0.31*
Diffuse	149 (66.2)	0.81 (0.76 - 0.88)	122 (60.3)	0.79 (0.70 - 0.87)	134 (55.1)	0.83 (0.74 - 0.92)	*0.78*
p-values^		0.10		0.54		0.11	
Ischemic	38 (16.7)	0.87 (0.75 -0.98)	28 (13.4)	0.82 (0.62 -1.00)	33 (13.6)	0.85 (0.69 -1.00)	*0.92*
Non-Ischemic /inflammatory	187 (83.3)	0.84 (0.78-0.89)	181 (86.6)	0.82 (0.76 - 0.88)	210 (86.4)	0.89 (0.84-0.94)	*0.14*
p-values^		0.65		0.99		0.59	
Central	200 (88.9)	0.85 (0.81 - 0.90)	192 (91.9)	0.81 (0.75 - 0.87)	202 (83.1)	0.89 (0.84 – 0.94)	*0.16*
Peripheral	25 (11.1)	Not applicable	17 (8.1)	0.89 (0.74 -1.00)	41 (16.9)	0.88 (0.76 – 0.98)	*0.83*
p-values^		-		0.27		0.87	

**p values inter-cohorts comparison between the AUC calculated for the different type of the event*

^p values intra-cohort comparison between the AUC calculated for the different type of the event

Stratified analyses based on the type of NP event: major/minor, focal/diffuse and central/peripheral.

The performance of the algorithm was evaluated separately by testing the events clustered by type of event. Comparing the accuracy of ROC curve in minor/major, focal/diffuse, ischemic/non-ischemic and peripheral/central NP events there were no statistically significant differences in performance among the three cohorts, although, as expected, the accuracy of the model was better for major and focal events and similar for ischemic, non-ischemic, central and peripheral manifestations (Table 4).

DISCUSSION

Recently, on behalf of the Study Group for NPSLE of the Italian Society of Rheumatology, an attribution model based on a simple numerical algorithm (ranging from 0 to 10) and derived from a robust statistical evaluation and large dataset was proposed. The original algorithm was tested on a single-center training cohort of SLE patients and then validated on an independent Italian cohort demonstrating good performance in terms of sensitivity, specificity, PPV and NPV when compared with expert clinical judgment (the current "gold" reference standard). To further validate this algorithm, taking also into account differences in ethnicity, we have tested its performance in a third independent international cohort including patients with one or more NP events, as per the 1999 ACR case definitions.

The first analysis, (based on '*a priori*' defined and 'updated' coefficients) aimed to test the discrimination power of the aforementioned algorithm on the external international cohort, demonstrated an overall performance of the algorithm highly comparable to our original study (Figure 1), confirming its high reliability. Further analyses replicated the process of adaptation of the *a priori* coefficients using data driven results of a) the new validating international cohort and b) the overall pooled dataset (all three cohorts) to validate the original model composed by pre-defined and weighted coefficients [9].

Based on the ROC tables and using binary outcomes, the best cut-off for discrimination (i.e. attribution threshold) was assessed in the international validating set and in the pooled data set. A total score ≥ 7 (range from 0 to 10) identified the maximum proportion of correctly classified NPSLE cases (both first and subsequent NP events). Compared with the lower cut-off point we found in our original paper (≥ 6) [9] this result is worthy of comment. First, there were differences in the composition of the international and the original cohorts, with particular regard to the distribution of major NP events. Given the structure of the algorithm, higher scores are assigned to these types of NP events [12]. In this way, 7 is the maximum score that can be reached by applying the model for a minor event. This implies a higher "attribution threshold" for minor NP events and, consequently, only a limited percentage of these events will be attributed to SLE using the algorithm in its current version. Accordingly, a greater prevalence of minor or diffuse events would influence the final performance of the attribution algorithm, which is derived from the cohort wherein it is applied. However, although the different composition of the individual cohorts (see Table 4) may have influenced the definition of the "attribution threshold", merging data of all three cohorts has balanced the proportion of major and minor events, thus making the newly identified cut point more reliable for attribution. Interestingly, in a recent study by Fanouriakis et al. [12] a similar result has been reached. In that study, different models of attribution, including our own, have been tested against "clinical judgment" in an independent and 'real life' cohort of SLE patients with NP involvement; applying our algorithm, the best performing cut-off point to ensure the discrimination between primary NPSLE from NP events not related to SLE was ≥7, i.e. the same as the one we found in the present validation study.

In our opinion, the "small window" of attribution for minor events is not a drawback; rather, it is in keeping with the evolution of the concept of NPSLE itself. In fact, inclusion of these minor events has substantially influenced the prevalence of NPSLE, especially in the past [13] [14] [15] [16], while in more recent years prospective studies derived from the SLICC inception cohort have challenged this concept of NPSLE, demonstrating that such events correlate poorly with conventional measures of SLE disease activity, autoantibodies, and lupus specific therapies. For this reason, these NP events require a more careful and rigorous clinical evaluation in order to determine the correct attribution [17] [18] [19]. For example, in the SLICC cohort, out of a total of 1732 patients, 17.8% had headache within the enrolment window, migraine in 60.7%, tension in 38.6%, intractable nonspecific in 7.1%, cluster in 2.6%, and intracranial hypertension in 1.0% [18]. Although the prevalence of headache rose to 58% by 10 years, only 26 patients (1.5% of the cohort) experienced "lupus headache" over the entire study, reported as a variable in the SLEDAI-2K [20] at annual assessments [19]. Hanly et al also reported that mood disorders occurred in 12.7% of 1,827 patients in the SLICC cohort, and a little more than a third of the total (98 events, 38.3%) were attributed to SLE [18].

As a result of these and other studies, the frequency of NPSLE has been reevaluated[6] [9] [21] [22]. However, one must not forget that mood disorders, headache and mild cognitive deficits, all frequently observed in SLE patients, depend heavily on clinical assessment of mainly subjective symptoms; not surprisingly, it is in these cases that we observed the worst performance of the model, when compared with the current "gold standard", i.e. the judgment of experienced physicians. Nevertheless, given the intrinsic uncertainty of the diagnosis for some NP manifestations, especially the common minor NP events, to reach a confident diagnosis of primary NPSLE is sometimes only presumptive, despite the efforts to improve the tools available to the clinician. For this reason, the categorization of NPSLE events based upon a quantitative score could ensure a more standardized and consistent approach to the attribution of NP events in future studies of NPSLE [23]. Moreover, the model has characteristics of flexibility and versatility that could be adapted to the setting in which a clinician operates. It is possible to modulate the single cut-off in relation to clinical contingency, choosing from time to time sensitivity over specificity or vice versa, remembering that even more stringent cut-points (i.e. ≤ 2 and ≥ 8 meaning that the NP event has high chance to be unrelated or related to SLE, respectively) are also associated with a - relatively low -probability of misclassification (10%). It may be that more stringent cut points could be tested as a "therapeutic threshold" (i.e. to treat or not to treat). On this topic, a prospective study is already underway.

There are several study limitations that should be mentioned. First, the use of a retrospective design is a weakness that could have influenced the proper attribution of some NP events, thus at risk of bias, due to incomplete data collection, especially for NP events observed before the publication of the ACR nomenclature. However a supplementary analysis restricted to the subset of events observed after 1999 gave similar results to those obtained using all first NP events (data not shown). A second limitation is the

low number of some rare NP events, making our results not fully generalizable to all NP events included in the ACR glossary. Finally, this model currently has to be considered as confidently tested and validated for the evaluation and attribution of the first NP event since the attribution of subsequent events could be influenced by history or recurrence of NP manifestations in the same patient, recognized as a risk factor for primary NPSLE involvement [17][18][24][25][26][27][28] However, when the algorithm was applied to subsequent NP events, it demonstrated a similar and satisfactory performance as for the first one, especially for antecedent events unrelated to SLE.

In summary, in this study we confirmed that the Italian attribution algorithm is a valid and robust tool for the correct identification of cases with NPSLE, with a validated score for attribution of NP events ≥ 7 (in a scale ranging from 0 to 10). The "*a priori* score" originally defined by the expert panel to weigh the single items included in the attribution model, was shown to be consistent and accurate and confirmed by the data driven analysis of both an external international cohort and of the pooled cohorts. In a medical setting as complex as NPSLE, we do not believe that our model should substitute the clinical judgment provided by experienced and multidisciplinary teams, but rather it could assist them in the attribution process. The categorization of NPSLE patients based upon a quantitative, reliable and validated probability score might provide a more standardized approach to the attribution of NP events, also to be used in future studies on NPSLE.

Acknowledgements: none.

Contributors:

A. Bortoluzzi: substantial contributions to the conception of the work and interpretation of data for the work; drafting the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

A. Fanouriakis : substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

S. Appenzeller: substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

L.T.L. Costallat: substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

C.A. Scirè: substantial contributions to the analysis of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

E. Murphy: substantial contributions to the acquisition of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

G. Bertsias: substantial contributions to the interpretation of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

J.G. Hanly: substantial contributions to the interpretation of the data for the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published;

M. Govoni: substantial contributions to the conception of the work and interpretation of data for the work; drafting the work and revising it critically for important intellectual content; agreement to be accountable for all aspects of the work related to accuracy and integrity; final approval of the version to be published.

Competing interests: none declared.

Funding: S. Appenzeller has received grant from Conselho Nacional de Desenvolvimento Cientifico (CNPq 304255/2015-7).

Data sharing statement: the data set is available from corresponding author on request.

REFERENCES

1 The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. *Arthritis Rheum* 1999;**42**:599–608. doi:10.1002/1529-0131(199904)42:4<599::AID-ANR2>3.0.CO;2-F

2 Bertsias GK, Boumpas DT. Pathogenesis, diagnosis and management of neuropsychiatric SLE manifestations. *Nat Rev Rheumatol* 2010;**6**:358–67. doi:10.1038/nrrheum.2010.62

3 Govoni M, Bortoluzzi A, Padovan M, *et al.* The diagnosis and clinical management of the neuropsychiatric manifestations of lupus. *J Autoimmun* Published Online First: 11 July 2016. doi:10.1016/j.jaut.2016.06.013

4 Ainiala H, Loukkola J, Peltola J, *et al.* The prevalence of neuropsychiatric syndromes in systemic lupus erythematosus. *Neurology* 2001;**57**:496–500.

5 Ainiala H, Hietaharju A, Loukkola J, *et al.* Validity of the new American College of Rheumatology criteria for neuropsychiatric lupus syndromes: a population-based evaluation. *Arthritis Rheum* 2001;**45**:419–23.

6 Monov S, Monova D. Classification criteria for neuropsychiatric systemic lupus erythematosus: do they need a discussion? *Hippokratia* 2008;**12**:103–7.

7 Hanly JG. The neuropsychiatric SLE SLICC inception cohort study. *Lupus* 2008;**17**:1059–63. doi:10.1177/0961203308097568

8 Hanly JG, Urowitz MB, Su L, *et al.* Prospective analysis of neuropsychiatric events in an international disease inception cohort of patients with systemic lupus erythematosus. *Ann Rheum Dis* 2010;**69**:529–35. doi:10.1136/ard.2008.106351

9 Bortoluzzi A, Scirè CA, Bombardieri S, *et al.* Development and validation of a new algorithm for attribution of neuropsychiatric events in systemic lupus erythematosus. *Rheumatol Oxf Engl* 2015;**54**:891–8. doi:10.1093/rheumatology/keu384

10 Bossuyt PM, Cohen JF, Gatsonis CA, *et al.* STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. *Ann Transl Med* 2016;**4**. doi:10.3978/j.issn.2305-5839.2016.02.06

11 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;**40**:1725. doi:10.1002/1529-0131(199709)40:9<1725::AID-ART29>3.0.CO;2-Y

12 Fanouriakis A, Pamfil C, Rednic S, *et al.* Is it primary neuropsychiatric systemic lupus erythematosus? Performance of existing attribution models using physician judgment as the gold standard. *Clin Exp Rheumatol* 2016.

13 Brey RL, Holliday SL, Saklad AR, *et al.* Neuropsychiatric syndromes in lupus: prevalence using standardized definitions. *Neurology* 2002;**58**:1214–20.

14 Sanna G, Bertolaccini ML, Cuadrado MJ, *et al.* Neuropsychiatric manifestations in systemic lupus erythematosus: prevalence and association with antiphospholipid antibodies. *J Rheumatol* 2003;**30**:985–92.

15 Mok CC, Lau CS, Wong RW. Neuropsychiatric manifestations and their clinical associations in southern Chinese patients with systemic lupus erythematosus. *J Rheumatol* 2001;**28**:766–71.

16 Petri M, Naqibuddin M, Carson KA, *et al.* Cognitive function in a systemic lupus erythematosus inception cohort. *J Rheumatol* 2008;**35**:1776–81.

17 Hanly JG, Su L, Urowitz MB, *et al.* Mood Disorders in Systemic Lupus Erythematosus: Results From an International Inception Cohort Study. *Arthritis Rheumatol Hoboken NJ* 2015;**67**:1837–47. doi:10.1002/art.39111

18 Hanly JG, Urowitz MB, O'Keeffe AG, *et al.* Headache in Systemic Lupus Erythematosus: Results From a Prospective, International Inception Cohort Study. *Arthritis Rheum* 2013;**65**:2887–97. doi:10.1002/art.38106

19 Lockshin MD. Editorial: Splitting Headache (Off). *Arthritis Rheum* 2013;**65**:2759–61. doi:10.1002/art.38108

20 Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol* 2002;**29**:288–91.

21 Hanly JG, Urowitz MB, Su L, *et al.* Prospective analysis of neuropsychiatric events in an international disease inception cohort of patients with systemic lupus erythematosus. *Ann Rheum Dis* 2010;**69**:529–35. doi:10.1136/ard.2008.106351

22 Hanly JG, Urowitz MB, Sanchez-Guerrero J, *et al.* Neuropsychiatric events at the time of diagnosis of systemic lupus erythematosus: an international inception cohort study. *Arthritis Rheum* 2007;**56**:265–73. doi:10.1002/art.22305

23 Hanly JG. Attribution in the assessment of nervous system disease in SLE. *Rheumatology* 2015;**54**:755–6. doi:10.1093/rheumatology/keu458

24 Bertsias G, Ioannidis JPA, Boletis J, *et al.* EULAR recommendations for the management of systemic lupus erythematosus. Report of a Task Force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics. *Ann Rheum Dis* 2008;**67**:195–205. doi:10.1136/ard.2007.070367

25 Hanly JG, Urowitz MB, Su L, *et al.* Seizure disorders in systemic lupus erythematosus results from an international, prospective, inception cohort study. *Ann Rheum Dis* 2012;**71**:1502–9. doi:10.1136/annrheumdis-2011-201089

26 Padovan M, Castellino G, Bortoluzzi A, *et al.* Factors and comorbidities associated with central nervous system involvement in systemic lupus erythematosus: a retrospective cross-sectional case–control study from a single center. *Rheumatol Int* 2010;**32**:129–35. doi:10.1007/s00296-010-1565-4

27 Appenzeller S, Cendes F, Costallat LTL. Epileptic seizures in systemic lupus erythematosus. *Neurology* 2004;**63**:1808–12.

28 Mikdashi J, Krumholz A, Handwerger B. Factors at diagnosis predict subsequent occurrence of seizures in systemic lupus erythematosus. *Neurology* 2005;**64**:2102–7. doi:10.1212/01.WNL.0000165959.98370.D5

Figure Legends

Figure 1. Receiver operating characteristic (ROC) curve using dichotomous outcomes (related vs. uncertain/ not related), for attribution of the first NP event observed in the international cohort.

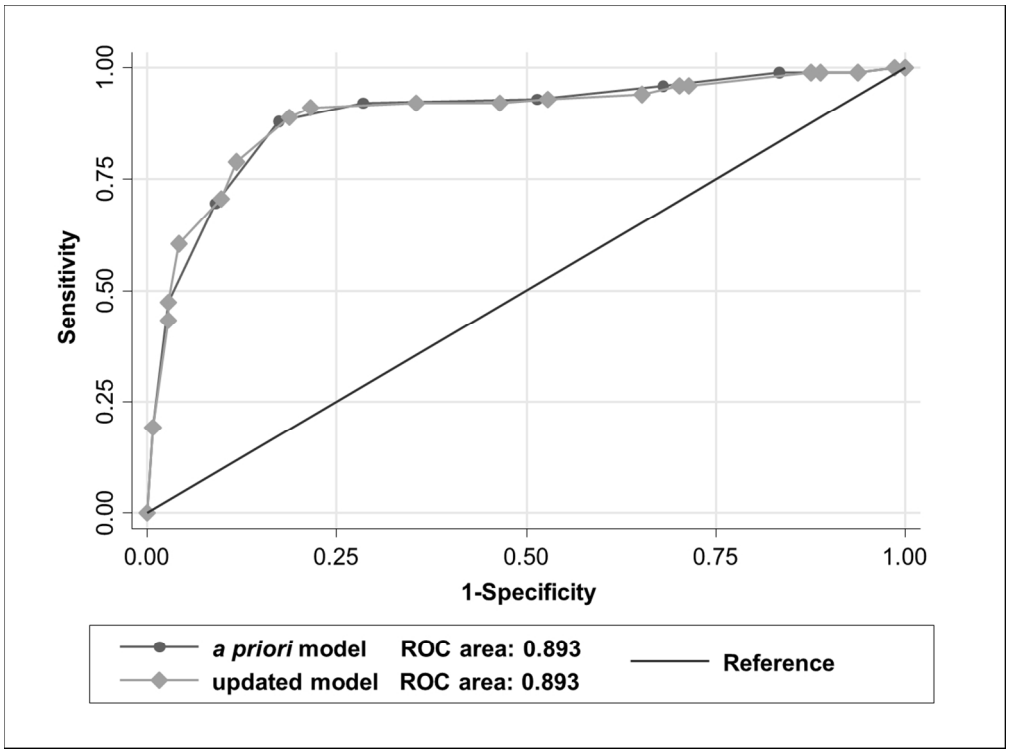


Figure 1. Receiver operating characteristic (ROC) curve using dichotomous outcomes (related vs. uncertain/ not related), for attribution of the first NP event observed in the international cohort.

112x83mm (300 x 300 DPI)

Supplementary_S1.

Check-list of favouring factors (including SLE-specific risk factors recognized by EULAR recommendations and those arbitrarily defined by the expert panel), deemed as potentially related to each NP event.

Central Nervous System pictures (according to 1999 ACR classification)	**SLE-specific risk factors** (as listed by 2010 EULAR recommendations (citazione)	**Reference**	**Favouring Factors** (deemed as relevant by the expert panel)	**Evidence**	**Notes**
CVD	• SLE disease activity • aPL antibodies[1] • Heart valve disease[2] • Previous CVD • Age	(1) (2)(3)(4)(5)(6)(7)(8)(9)(10) (11)(5)(12) (13)(3) (14)(15)(16)	 • Age< 50 years	 (EP)	SLEDAI > 6 Libman-Sack endocarditis (aPL+)
Seizure disorder	• SLE disease activity • SLE damage • aPL antibodies[1] • Past or concurrent Major NPSLE[3] • Anti-Sm antibodies	(17)(13)(18)(19)(20)(21) (21) (22)(17)(13)(23)(19)(21) (2)(17)(13)(3)(20)(21) (21)(24)	 • No familiar history • Abnormal neuroimaging	(EP) (EP) (EP)	SLEDAI > 6 SDI > 1.5 Seizure disorder, CVD, psychosis MRI or SPECT
Cognitive dysfunction	• SLE disease activity • SLE damage • Past or concurrent Major NPSLE[3] • aPL antibodies[1] • Heart valve disease[2] • Education level • Age	(2)(25)(26)(27)(28) (29)(30) (22)(31)(32)(33)(34)(35) (22)(30)(36)(37)(38) (11)(39) (40)(37)(41) (40)(2)(42)(30)(41)(43)	 • Age< 50 years • Response to IS or GC Rx • Abnormal neuroimaging	 (EP) (EP) (EP) (EP)	SLEDAI ≥16 SDI ≥1.0 Libman sacks endocarditis (aPL+) At least secondary school MRI or SPECT
Movement disorders	• aPL antibodies[1]	(22) (44)(45)(46)(47)'	 • Response to IS or GC Rx	 (EP)	

			• Abnormal neuroimaging • High SLE disease activity	(EP) (EP)	MRI or SPECT SLEDAI > 6
Acute confusional state	• SLE disease activity • Past or concurrent Major NPSLE[3]	(11)(3) (48)	• Abnormal neuroimaging • Response to IS or GC Rx	(EP) (EP)	MRI or SPECT
Psychiatric disorders	• SLE disease activity • Past or concurrent Major NPSLE[3] • Anti-ENA	(2)(49)(50)(51)(52)(53) (32)(54)(55) (24)(56)(57) (58)(59)(60)(61)(62)(63)	• Anti-ribosomal-p antibodies • No familiar history • Abnormal neuroimaging	(EP) (EP) (EP)	SLEDAI ≥16 Anti-Sm, anti-RNP, anti-Ro MRI or SPECT
Myelopathy	• aPL antibodies[1]	(64)(65)(66)(67)	• Response to IS or GC Rx • High SLE disease activity	(EP) (EP)	SLEDAI > 6
MS-like syndrome			• aPL antibodies • CSF < 4 OCB • High SLE disease activity	(EP) (EP)	Persistently medium-high titres SLEDAI > 6
Aseptic Meningitis			• Response to IS or GC Rx • High SLE disease activity	(EP) (EP)	SLEDAI > 6
Mood abnormalities			• Anti-rib-P antibodies • Abnormal neuroimaging • High SLE disease activity	(EP) (EP) (EP)	MRI or SPECT SLEDAI > 6
Anxiety			• Abnormal neuroimaging • High SLE disease activity	(EP) (EP)	MRI or SPECT SLEDAI > 6
Headache			• No familiar history • Abnormal neuroimaging • aPL antibodies (^) • Response to IS or GC Rx • High SLE disease activity • High SLE disease activity	(EP) (EP) (EP) (EP) (EP) (EP)	MRI or SPECT Persistently medium-high titres SLEDAI > 6

Peripheral Nervous System pictures (according to 1999 ACR classification)	SLE-specific Risk factors (as listed by 2010 EULAR recommendations) cit	Reference	Favouring Factors (deemed as relevant by the expert panel)	Evidence	Notes
Cranial neuropathy	• aPL antibodies[1]	(22)(2)	• High SLE disease activity	(EP)	SLEDAI > 6
Peripheral neuropathy	• Anti-ENA	(2)(68)(24)	• High SLE disease activity	(EP)	anti-dsDNA, anti-Sm, anti-RNP
	• Anti-DNA • Past or concurrent Major NPSLE[3]	(2) (17)(69)(3)	• High SLE disease activity	(EP)	SLE-related Seizure SLEDAI > 6
Mononeuritis			• Vasculitis • High SLE disease activity	(EP) (EP)	SLEDAI > 6
Myasthenia Gravis			• Response to IS or GC Rx • High SLE disease activity	(EP) (EP)	SLEDAI > 6
Autonomic Neuropathy			• Response to IS or GC Rx • High SLE disease activity	(EP) (EP)	SLEDAI > 6
Acute Demyelinating Poliradiculopathy			• High SLE disease activity	(EP)	SLEDAI > 6
Plexopathy			• High SLE disease activity	(EP)	SLEDAI > 6

*CVD, cerebrovascular disease, SLEDAI, SLE disease activity index, aPL, antiphospholipid antibodies (including lupus anticoagulant), SDI, SLE International Collaborating Clinics Damage Index, MRI: Magnetic resonance imaging, SPECT: Single-photon emission computed tomography, CSF: Cerebrospinal fluid, anti rib-P antibodies, anti ribosomal-P antibodies.*0CB, oligo-clonal bands, IS (immunosuppressant), GC (glucocorticosteroids),

^ relevant in case of migraine.

(EP) correspond to the shared opinion of the Expert Panel.

The Expert Panel unanimously assumed that high SLE disease activity should be considered as a generic additional potential favoring factor also for those NP events where evidence is not established yet.

[1] Persistently positive, moderate-to-high titers of aPL antibodies

[2] Not septic Libman Sacks endocarditis

[3] Refers to past or concurrent major NPSLE syndrome

References

1. Mikdashi J, Handwerger B, Langenberg P, Miller M, Kittner S. Baseline disease activity, hyperlipidemia, and hypertension are predictive factors for ischemic stroke and stroke severity in systemic lupus erythematosus. Stroke J Cereb Circ. 2007 Feb;38(2):281–5.

2. Mikdashi J, Handwerger B. Predictors of neuropsychiatric damage in systemic lupus erythematosus: data from the Maryland lupus cohort. Rheumatol Oxf Engl. 2004 Dec;43(12):1555–60.

3. Buján S, Ordi-Ros J, Paredes J, Mauri M, Matas L, Cortés J, et al. Contribution of the initial features of systemic lupus erythematosus to the clinical evolution and survival of a cohort of Mediterranean patients. Ann Rheum Dis. 2003 Sep;62(9):859–65.

4. Picillo U, Migliaresi S, Marcialis MR, Longobardo A, La Palombara F, Tirri G. Longitudinal survey of anticardiolipin antibodies in systemic lupus erythematosus. Relationships with clinical manifestations and disease activity in an Italian series. Scand J Rheumatol. 1992;21(6):271–6.

5. Roldan CA, Gelgand EA, Qualls CR, Sibbitt WL Jr. Valvular heart disease as a cause of cerebrovascular disease in patients with systemic lupus erythematosus. Am J Cardiol. 2005 Jun 15;95(12):1441–7.

6. Donders RC, Kappelle LJ, Derksen RH, Algra A, Horbach DA, de Groot PG, et al. Transient monocular blindness and antiphospholipid antibodies in systemic lupus erythematosus. Neurology. 1998 Aug;51(2):535–40.

7. Ishii Y, Nagasawa K, Mayumi T, Niho Y. Clinical importance of persistence of anticardiolipin antibodies in systemic lupus erythematosus. Ann Rheum Dis. 1990 Jun;49(6):387–90.

8. Kushner MJ. Prospective study of anticardiolipin antibodies in stroke. Stroke J Cereb Circ. 1990 Feb;21(2):295–8.

9. Sturfelt G, Eskilsson J, Nived O, Truedsson L, Valind S. Cardiovascular disease in systemic lupus erythematosus. A study of 75 patients form a defined population. Medicine (Baltimore). 1992 Jul;71(4):216–23.

10. Locht H, Wiik A. IgG and IgM isotypes of anti-cardiolipin and anti-beta2-glycoprotein i antibodies reflect different forms of recent thrombo-embolic events. Clin Rheumatol. 2006 Mar;25(2):246–50.

11. Roldan CA, Gelgand EA, Qualls CR, Sibbitt WL Jr. Valvular heart disease by transthoracic echocardiography is associated with focal brain injury and central neuropsychiatric systemic lupus erythematosus. Cardiology. 2007;108(4):331–7.

12. Molad Y, Levin-Iaina N, Vaturi M, Sulkes J, Sagie A. Heart valve calcification in young patients with systemic lupus erythematosus: a window to premature atherosclerotic vascular morbidity and a risk factor for all-cause mortality. Atherosclerosis. 2006 Apr;185(2):406–12.

13. Appenzeller S, Cendes F, Costallat LTL. Epileptic seizures in systemic lupus erythematosus. Neurology. 2004 Nov 23;63(10):1808–12.

14. Toloza SMA, Uribe AG, McGwin G Jr, Alarcón GS, Fessler BJ, Bastian HM, et al. Systemic lupus erythematosus in a multiethnic US cohort (LUMINA). XXIII. Baseline predictors of vascular events. Arthritis Rheum. 2004 Dec;50(12):3947–57.

15. Csépány T, Bereczki D, Kollár J, Sikula J, Kiss E, Csiba L. MRI findings in central nervous system systemic lupus erythematosus are associated with immunoserological parameters and hypertension. J Neurol. 2003 Nov;250(11):1348–54.

16. Morelli S, Bernardo ML, Viganego F, Sgreccia A, De Marzio P, Conti F, et al. Left-sided heart valve abnormalities and risk of ischemic cerebrovascular accidents in patients with systemic lupus erythematosus. Lupus. 2003;12(11):805–12.

17. Andrade RM, Alarcón GS, González LA, Fernández M, Apte M, Vilá LM, et al. Seizures in patients with systemic lupus erythematosus: data from LUMINA, a multiethnic cohort (LUMINA LIV). Ann Rheum Dis. 2008 Jun;67(6):829–34.

18. Jeffries M, Hamadeh F, Aberle T, Glenn S, Kamen DL, Kelly JA, et al. Haemolytic anaemia in a multi-ethnic cohort of lupus patients: a clinical and serological perspective. Lupus. 2008 Aug;17(8):739–43.

19. Shoenfeld Y, Lev S, Blatt I, Blank M, Font J, von Landenberg P, et al. Features associated with epilepsy in the antiphospholipid syndrome. J Rheumatol. 2004 Jul;31(7):1344–8.

20. Ramsey-Goldman R, Alarcón GS, McGwin G, Petri M, Vilá LM, Edberg JC, et al. Time to seizure occurrence and damage in PROFILE, a multi-ethnic systemic lupus erythematosus cohort. Lupus. 2008 Mar;17(3):177–84.

21. Mikdashi J, Krumholz A, Handwerger B. Factors at diagnosis predict subsequent occurrence of seizures in systemic lupus erythematosus. Neurology. 2005 Jun 28;64(12):2102–7.

22. Sanna G, Bertolaccini ML, Cuadrado MJ, Laing H, Khamashta MA, Mathieu A, et al. Neuropsychiatric manifestations in systemic lupus erythematosus: prevalence and association with antiphospholipid antibodies. J Rheumatol. 2003 May;30(5):985–92.

23. Herranz MT, Rivier G, Khamashta MA, Blaser KU, Hughes GR. Association between antiphospholipid antibodies and epilepsy in patients with systemic lupus erythematosus. Arthritis Rheum. 1994 Apr;37(4):568–71.

24. Hitchon CA, Peschken CA. Sm antibodies increase risk of death in systemic lupus erythematosus. Lupus. 2007;16(3):186–94.

25. Gladman DD, Urowitz MB, Slonim D, Glanz B, Carlen P, Noldy N, et al. Evaluation of predictive factors for neurocognitive dysfunction in patients with inactive systemic lupus erythematosus. J Rheumatol. 2000 Oct;27(10):2367–71.

26. Panopalis P, Julian L, Yazdany J, Gillis JZ, Trupin L, Hersh A, et al. Impact of memory impairment on employment status in persons with systemic lupus erythematosus. Arthritis Rheum. 2007 Dec 15;57(8):1453–60.

27. Hanly JG, Fisk JD, Sherwood G, Jones E, Jones JV, Eastwood B. Cognitive impairment in patients with systemic lupus erythematosus. J Rheumatol. 1992 Apr;19(4):562–7.

28. Kozora E, Ellison MC, West S. Depression, fatigue, and pain in systemic lupus erythematosus (SLE): relationship to the American College of Rheumatology SLE neuropsychological battery. Arthritis Rheum. 2006 Aug 15;55(4):628–35.

29. Petri M, Naqibuddin M, Carson KA, Sampedro M, Wallace DJ, Weisman MH, et al. Cognitive function in a systemic lupus erythematosus inception cohort. J Rheumatol. 2008 Sep;35(9):1776–81.

30. Tomietto P, Annese V, D'agostini S, Venturini P, La Torre G, De Vita S, et al. General and specific factors associated with severity of cognitive impairment in systemic lupus erythematosus. Arthritis Rheum. 2007 Dec 15;57(8):1461–72.

31. Hay EM, Black D, Huddy A, Creed F, Tomenson B, Bernstein RM, et al. Psychiatric disorder and cognitive impairment in systemic lupus erythematosus. Arthritis Rheum. 1992 Apr;35(4):411–6.

32. Monastero R, Bettini P, Del Zotto E, Cottini E, Tincani A, Balestrieri G, et al. Prevalence and pattern of cognitive impairment in systemic lupus erythematosus patients with and without overt neuropsychiatric manifestations. J Neurol Sci. 2001 Feb 15;184(1):33–9.

33. Emori A, Matsushima E, Aihara O, Ohta K, Koike R, Miyasaka N, et al. Cognitive dysfunction in systemic lupus erythematosus. Psychiatry Clin Neurosci. 2005 Oct;59(5):584–9.

34. Loukkola J, Laine M, Ainiala H, Peltola J, Metsänoja R, Auvinen A, et al. Cognitive impairment in systemic lupus erythematosus and neuropsychiatric systemic lupus erythematosus: a population-based neuropsychological study. J Clin Exp Neuropsychol. 2003 Feb;25(1):145–51.

35. Sailer M, Burchert W, Ehrenheim C, Smid HG, Haas J, Wildhagen K, et al. Positron emission tomography and magnetic resonance imaging for cerebral involvement in patients with systemic lupus erythematosus. J Neurol. 1997 Mar;244(3):186–93.

36. Denburg SD, Carbotte RM, Ginsberg JS, Denburg JA. The relationship of antiphospholipid antibodies to cognitive function in patients with systemic lupus erythematosus. J Int Neuropsychol Soc JINS. 1997 Jul;3(4):377–86.

37. McLaurin EY, Holliday SL, Williams P, Brey RL. Predictors of cognitive dysfunction in patients with systemic lupus erythematosus. Neurology. 2005 Jan 25;64(2):297–303.

38. Leritz E, Brandt J, Minor M, Reis-Jensen F, Petri M. Neuropsychological functioning and its relationship to antiphospholipid antibodies in patients with systemic lupus erythematosus. J Clin Exp Neuropsychol. 2002 Jun;24(4):527–33.

39. Roldan CA, Gelgand EA, Qualls CR, Sibbitt WL Jr. Valvular heart disease is associated with nonfocal neuropsychiatric systemic lupus erythematosus. J Clin Rheumatol Pract Rep Rheum Musculoskelet Dis. 2006 Feb;12(1):3–10.

40. Brey RL, Holliday SL, Saklad AR, Navarrete MG, Hermosillo-Romo D, Stallworth CL, et al. Neuropsychiatric syndromes in lupus: prevalence using standardized definitions. Neurology. 2002 Apr 23;58(8):1214–20.

41. Ginsburg KS, Wright EA, Larson MG, Fossel AH, Albert M, Schur PH, et al. A controlled study of the prevalence of cognitive dysfunction in randomly selected patients with systemic lupus erythematosus. Arthritis Rheum. 1992 Jul;35(7):776–82.

42. Carlomagno S, Migliaresi S, Ambrosone L, Sannino M, Sanges G, Di Iorio G. Cognitive impairment in systemic lupus erythematosus: a follow-up study. J Neurol. 2000 Apr;247(4):273–9.

43. Glanz BI, Slonim D, Urowitz MB, Gladman DD, Gough J, MacKinnon A. Pattern of neuropsychologic dysfunction in inactive systemic lupus erythematosus. Neuropsychiatry Neuropsychol Behav Neurol. 1997 Oct;10(4):232–8.

44. Cervera R, Asherson RA, Font J, Tikly M, Pallarés L, Chamorro A, et al. Chorea in the antiphospholipid syndrome. Clinical, radiologic, and immunologic characteristics of 50 patients from our clinics and the recent literature. Medicine (Baltimore). 1997 May;76(3):203–12.

45. Asherson RA, Derksen RH, Harris EN, Bouma BN, Gharavi AE, Kater L, et al. Chorea in systemic lupus erythematosus and “lupus-like” disease: association with antiphospholipid antibodies. Semin Arthritis Rheum. 1987 May;16(4):253–9.

46. Asherson RA, Mayou SC, Merry P, Black MM, Hughes GR. The spectrum of livedo reticularis and anticardiolipin antibodies. Br J Dermatol. 1989 Feb;120(2):215–21.

47. Cervera R, Piette J-C, Font J, Khamashta MA, Shoenfeld Y, Camps MT, et al. Antiphospholipid syndrome: clinical and immunologic manifestations and patterns of disease expression in a cohort of 1,000 patients. Arthritis Rheum. 2002 Apr;46(4):1019–27.

48. Futrell N, Schultz LR, Millikan C. Central nervous system disease in patients with systemic lupus erythematosus. Neurology. 1992 Sep;42(9):1649–57.

49. Ward MM, Studenski S. The time course of acute psychiatric episodes in systemic lupus erythematosus. J Rheumatol. 1991 Apr;18(4):535–9.

50. Purandare KN, Wagle AC, Parker SR. Psychiatric morbidity in patients with systemic lupus erythematosus. QJM Mon J Assoc Physicians. 1999 May;92(5):283–6.

51. Lim LC, Lee TE, Boey ML. Psychiatric manifestation of systemic lupus erythematosus in Singapore. A cross-cultural comparison. Br J Psychiatry J Ment Sci. 1991 Oct;159:520–3.

52. Nery FG, Borba EF, Hatch JP, Soares JC, Bonfá E, Neto FL. Major depressive disorder and disease activity in systemic lupus erythematosus. Compr Psychiatry. 2007 Feb;48(1):14–9.

53. Seguí J, Ramos-Casals M, García-Carrasco M, de Flores T, Cervera R, Valdés M, et al. Psychiatric and psychosocial disorders in patients with systemic lupus erythematosus: a longitudinal study of active and inactive stages of the disease. Lupus. 2000;9(8):584–8.

54. Utset TO, Golden M, Siberry G, Kiri N, Crum RM, Petri M. Depressive symptoms in patients with systemic lupus erythematosus: association with central nervous system lupus and Sjögren's syndrome. J Rheumatol. 1994 Nov;21(11):2039–45.

55. Hanly JG, Fisk JD, McCurdy G, Fougere L, Douglas J-A. Neuropsychiatric syndromes in patients with systemic lupus erythematosus and rheumatoid arthritis. J Rheumatol. 2005 Aug;32(8):1459–1456.

56. Tikly M, Gould T, Wadee AA, van der Westhuizen E, Mokgethwa BBN. Clinical and serological correlates of antinucleosome antibodies in South Africans with systemic lupus erythematosus. Clin Rheumatol. 2007 Dec;26(12):2121–5.

57. Shortall E, Isenberg D, Newman SP. Factors associated with mood and mood disorders in SLE. Lupus. 1995 Aug;4(4):272–9.

58. Bonfa E, Golombek SJ, Kaufman LD, Skelly S, Weissbach H, Brot N, et al. Association between lupus psychosis and anti-ribosomal P protein antibodies. N Engl J Med. 1987 Jul 30;317(5):265–71.

59. Schneebaum AB, Singleton JD, West SG, Blodgett JK, Allen LG, Cheronis JC, et al. Association of psychiatric manifestations with antibodies to ribosomal P proteins in systemic lupus erythematosus. Am J Med. 1991 Jan;90(1):54–62.

60. Teh LS, Hay EM, Amos N, Black D, Huddy A, Creed F, et al. Anti-P antibodies are associated with psychiatric and focal cerebral disorders in patients with systemic lupus erythematosus. Br J Rheumatol. 1993 Apr;32(4):287–90.

61. Yoshio T, Masuyama J, Ikeda M, Tamai K, Hachiya T, Emori T, et al. Quantification of antiribosomal P0 protein antibodies by ELISA with recombinant P0 fusion protein and their association with central nervous system disease in systemic lupus erythematosus. J Rheumatol. 1995 Sep;22(9):1681–7.

62. Briani C, Lucchetta M, Ghirardello A, Toffanin E, Zampieri S, Ruggero S, et al. Neurolupus is associated with anti-ribosomal P protein antibodies: an inception cohort study. J Autoimmun. 2009 Mar;32(2):79–84.

63. Karassa FB, Afeltra A, Ambrozic A, Chang D-M, De Keyser F, Doria A, et al. Accuracy of anti-ribosomal P protein antibody testing for the diagnosis of neuropsychiatric systemic lupus erythematosus: an international meta-analysis. Arthritis Rheum. 2006 Jan;54(1):312–24.

64. Lavalle C, Pizarro S, Drenkard C, Sánchez-Guerrero J, Alarcón-Segovia D. Transverse myelitis: a manifestation of systemic lupus erythematosus strongly associated with antiphospholipid antibodies. J Rheumatol. 1990 Jan;17(1):34–7.

65. Mok MY, Chan EYT, Fong DYT, Leung KFS, Wong WS, Lau CS. Antiphospholipid antibody profiles and their clinical associations in Chinese patients with systemic lupus erythematosus. J Rheumatol. 2005 Apr;32(4):622–8.

66. Alarcón-Segovia D, Delezé M, Oria CV, Sánchez-Guerrero J, Gómez-Pacheco L, Cabiedes J, et al. Antiphospholipid antibodies and the antiphospholipid syndrome in systemic lupus erythematosus. A prospective analysis of 500 consecutive patients. Medicine (Baltimore). 1989 Nov;68(6):353–65.

67. D'Cruz DP, Mellor-Pita S, Joven B, Sanna G, Allanson J, Taylor J, et al. Transverse myelitis as the first manifestation of systemic lupus erythematosus or lupus-like disease: good functional outcome and relevance of antiphospholipid antibodies. J Rheumatol. 2004 Feb;31(2):280–5.

68. Huynh C, Ho SL, Fong KY, Cheung RT, Mok CC, Lau CS. Peripheral neuropathy in systemic lupus erythematosus. J Clin Neurophysiol Off Publ Am Electroencephalogr Soc. 1999 Mar;16(2):164–8.

69. Feinglass EJ, Arnett FC, Dorsch CA, Zizic TM, Stevens MB. Neuropsychiatric manifestations of systemic lupus erythematosus: diagnosis, clinical spectrum, and relationship to other features of the disease. Medicine (Baltimore). 1976 Jul;55(4):323–39.

Supplementary_S2.

Check-list, related to each NP event for the presence of concurrent or confounding non-SLE factors (as suggested by the 1999 ACR criteria).

CNS EVENTS	
CVA	Diabetes mellitus
	Dyslipidemia
	Atherosclerotic vascular disease
	Atrial fibrillation
	Valvular heart disease
	Atrial septal defect
	Hypercoagulability state
	Antiphospholipid antibody syndrome
	Hypertension
	Smoking
	Cocaine or amphetamine abuse
Seizures	Thrombotic-thrombocytopenic purpura/microangiopathy
	Stroke or transient ischemic attack
	Migraine
	Metabolic: hypoglycemia, hypoxemia, uremia
	Tumor
	Infection
Movement disorders (*)	Stroke, vascular malformation, hypoxic damage
	Tumor
	Pregnancy (chorea gravidarum)
	Rheumatic fever (Sydenham's chorea)
Myelopathy (*)	Preexisting demyelinating syndrome
	Infections: herpes zoster, HIV
MS-like syndrome	Structural lesions, e.g., tumor, arteriovenous malformation
	Familial disorders, e.g., hereditary spastic paraplegia, ataxia, and leukodystrophies
	Sarcoid, Behçet's disease, other vasculitis
	Multiple sclerosis
Headache (*)	Cranial neuropathies
	Headache associated with eye, ear, sinus, teeth, TMJ, cervical spine disease
Aseptic Meningitis	None
Cognitive Dysfunction	Substance abuse
	Medication (steroids, sedatives)
	History of learning disabilities
	History of head injury
	Other primary neurologic and psychiatric disorders
	Metabolic disturbances, particularly uremia and diabetes
	Antiphospholipid antibody syndrome
	Coexisting emotional distress, fatigue, and pain.
Acute Confusional State	Marked psychosocial stress
	Corticosteroid use (#)
	Thrombotic thrombocytopenic purpura/hemolytic uremic syndrome
Psychosis	Marked psychosocial stress
	Corticosteroids (#)

PNS EVENTS	
Cranial nerve neuropathy	Nutritional: thiamine deficiency
	Metabolic: diabetes mellitus
	Inflammatory: multiple sclerosis
	Ischemic: giant cell arteritis, brainstem stroke
	Infiltrative: sarcoid
Mononeuritis	Diabetic neuropathy
	Local damage from mechanical injury, radiation, malignancy, sarcoid
	Infection: Lyme, HIV, herpes
	Vasculitis; polyarteritis nodosa, Wegener's granulomatosis, cryoglobulinemia, rheumatoid arthritis, Sjögren's syndrome, etc.
Myasthenia Gravis	Pure red cell aplasia
	Thyroid abnormalities
	Thymoma
Autonomic Neuropathy	Diabetic neuropathy and peripheral neuropathy of other causes
	Autonomic failure in elderly
Acute Demyelinating Poliradiculopathy	None
Polineuropathy	Heavy metal and solvent exposures: arsenic, lead, mercury, n-hexane, etc.
	Drug toxicity: isoniazid, vincristine, phenytoin, colchicine, etc.
	Leprosy, HIV, diphtheria, Lyme disease
	Diabetes, uremia, amyloid, alcoholism, porphyria, etc.
	Paraproteinemia, cryoglobulinemia
	Sjögren's syndrome
	Inherited forms: Charcot-Marie-Tooth, Fabry's disease, Tangier's disease, familial amyloid polyneuropathy, etc.
Plexopathy	Diabetes mellitus
	Polyarteritis nodosa or other vasculitis
	Sarcoid

(*) Antiphosphlipid antibodies: item reviewed by the study group and not included
(#) Corticosteroid use > 10 g

(*) (#) Govoni M, Bombardieri S, Bortoluzzi A, Caniatti L, Casu C, Conti F, et al. Factors and comorbidities associated with first neuropsychiatric event in systemic lupus erythematosus: does a risk profile exist? A large multicentre retrospective cross-sectional study on 959 Italian patients. *Rheumatol Oxf Engl* 2012;51:157–168.

Supplementary_S3 Table S1. Comparison between «a priori» vs «data driven» estimated scores (pooled analysis in 677 NP 1st events).

Item	CATEGORY	A PRIORI ORIGINAL COEFFICIENTS*	DATA DRIVEN (REFINED A POSTERIORI) COEFFICIENTS
Time onset of NP event	Before	**0**	0
	After	**2**	2.1
	Concurrent	**3**	3.1
Minor event (Ainiala list)	Yes	**0**	0
	No	**3**	2.9
Presence of Confounding Factors	≥ 1	**0**	0
	1	**1**	0.7
	No	**2**	1.9
Presence of favouring factors	No	**0**	0
	1	**1**	0.9
	≥ 1	**2**	2.1

(^) The resulting global score can range from 0 to 10; details for definition of each item category are reported elsewhere (9).
*A priori coefficients (original coefficients), identify the better scores to be used in the final version of the Italian attribution algorithm, the so-called “original algorithm” (^)

Supplementary_S3 Table S2. ROC curve analysis related to the following NP event observed in the international cohort, stratified by all subsequent events and in relation to the first event (related or unrelated).

Subsequent event	N° of events	AUC	[95% Conf. Interval]
All	93	0.80	(0.71 – 0.88)
After a 1st unrelated NP event	33	0.83	(0.68 – 0.97)
After a 1st related NP event	60	0.79	(0.68 – 0.91)

Supplementary_S3 Table S3. Detailed report of sensitivity, specificity, PPV and NPV for each defined cut-point derived from the application of the attribution algorithm to the first NP event observed in the training (cohort 1), validating (cohort 2) and pooled cohorts (all three cohorts, including the international cohort).

Cohort	Cut-point	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
1	(>= 0)	100.00%	0.00%	62.67%	100.00%		62.67%	
1	(>= 1)	100.00%	0.00%	62.67%	1		62.67%	
1	(>= 2)	100.00%	1.19%	63.11%	1.012	0	62.95%	100.00%
1	(>= 3)	99.29%	7.14%	64.89%	1.0693	0.0993	64.22%	85.70%
1	(>= 4)	99.29%	16.67%	68.44%	1.1915	0.0426	66.67%	93.33%
1	(>= 5)	95.74%	35.71%	73.33%	1.4894	0.1191	71.43%	83.32%
1	**(>= 6)**	85.11%	61.90%	**76.44%**	2.234	0.2406	78.95%	71.24%
1	(>= 7)	58.87%	92.86%	71.56%	8.2411	0.443	93.26%	57.36%
1	(>= 8)	36.17%	100.00%	60.00%	0.6383		100.00%	48.28%
1	(>= 9)	21.99%	100.00%	51.11%	0.7801		100.00%	43.30%
1	(>= 10)	6.38%	100.00%	41.33%	0.9362		100.00%	
1	(> 10)	0.00%	100.00%	37.33%	1			

Cohort	Cut-point	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
2	(>= 0)	100.00%	0.00%	51.20%	1		51.20%	
2	(>= 1)	100.00%	0.00%	51.20%	1		51.20%	
2	(>= 2)	100.00%	0.98%	51.67%	1.0099	0	51.44%	100.00%
2	(>= 3)	97.20%	5.88%	52.63%	1.0327	0.4766	52.00%	66.69%
2	(>= 4)	96.26%	17.65%	57.89%	1.1689	0.2118	55.08%	81.81%
2	(>= 5)	91.59%	36.27%	64.59%	1.4372	0.2319	60.12%	80.44%
2	(>= 6)	85.98%	61.76%	74.16%	2.2487	0.227	70.23%	80.77%
2	**(>= 7)**	71.96%	80.39%	**76.08%**	3.6701	0.3488	79.38%	73.21%
2	(>= 8)	58.88%	92.16%	75.12%	7.507	0.4462	88.74%	68.12%
2	(>= 9)	32.71%	95.10%	63.16%	6.6729	0.7076	87.50%	57.40%
2	(>= 10)	10.28%	98.04%	53.11%	5.243	0.9151	84.62%	51.02%
2	(> 10)	0.00%	100.00%	48.80%	1			48.80%

Cohort	Cut-point	Sensitivity	Specificity	Correctly Classified	LR+	LR-	PPV	NPV
pooled	(>= 0)	100.00%	0.00%	51.26%	1		51.26%	
pooled	(>= 1)	100.00%	0.91%	51.70%	1.0092	0	51.48%	100.00%
pooled	(>= 2*)	99.71%	3.33%	52.73%	1.0315	0.0865	52.03%	91.61%
pooled	(>= 3)	98.56%	10.91%	55.83%	1.1063	0.1321	53.77%	87.81%
pooled	(>= 4)	97.41%	23.64%	61.45%	1.2756	0.1097	57.29%	89.67%
pooled	(>= 5)	93.66%	41.52%	68.24%	1.6014	0.1527	62.74%	86.17%
pooled	(>= 6)	87.32%	66.06%	76.96%	2.5728	0.1919	73.01%	83.21%
pooled	(>= 7)	71.18%	84.55%	77.70%	4.6059	0.3409	82.89%	73.61%
pooled	(>= 8*)	52.74%	93.64%	72.67%	8.2874	0.5047	89.71%	65.33%
pooled	(>= 9)	32.56%	97.27%	64.11%	11.9404	0.6933	92.62%	57.84%
pooled	(>= 10)	11.24%	99.09%	54.06%	12.3631	0.8958	92.85%	51.50%
pooled	(> 10)	0.00%	100.00%	48.74%	1			48.74%

*Cut points ensuring a misclassification probability less than 10%

Supplementary S4

Members of the Italian Study Group on Neuropsychiatric Systemic Lupus Erythematosus of the Italian Society of Rheumatology:

Stefano Bombardieri[1],
Fabrizio Conti[2],
Salvatore De Vita[3],
Andrea Doria[4],
Gianfranco Ferraccioli[5],
Elisa Gremese[5],
Alessandro Mathieu[6],
Marta Mosca[1],
Melissa Padovan[7],
Matteo Piga[6],
Angela Tincani[8],
Paola Tomietto[9],
Guido Valesini[2],
Margherita Zen[4]

[1]Rheumatology Unit, Department of Clinical and Experimental Medicine, University of Pisa, Pisa

[2]Department of Internal Medicine and Medical Specialties, SapienzaUniversity of Rome, Rome

[3]Rheumatology Clinic, Azienda Ospedaliero Universitaria 'S. Maria della Misericordia' and DSMB, Department of Medical and Biological Sciences, University of Udine, Udine

[4]Department of Clinical and Experimental Medicine, Division of Rheumatology, University of Padova, Padova

[5]Division of Rheumatology and Internal Medicine, Institute of Rheumatology and Affine Sciences, CIC, Catholic University of the Sacred Heart, Rome,

[6]Rheumatology Unit, Department of Medical Sciences, University of Cagliari and AOU University Clinic, Cagliari

[7]Department of Medical Sciences, Section of Rheumatology, University of Ferrara and Azienda Ospedaliero Universitaria Sant'Anna di Cona, Ferrara

[8]Rheumatology and Clinical Immunology Unit, Spedali Civili and University of Brescia, Brescia

[9]Internal Medicine, AOU 'Ospedali Riuniti' of Trieste, Trieste, Italy

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Pag. 1;3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Pag.4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Pag. 6-7
Objectives	3	State specific objectives, including any prespecified hypotheses	Pag. 7
Methods			
Study design	4	Present key elements of study design early in the paper	Pag. 8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Pag.8-9
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Pag.8-9
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	N.A.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Pag.8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Pag.8-9
Bias	9	Describe any efforts to address potential sources of bias	Pag.8-9
Study size	10	Explain how the study size was arrived at	NA
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Pag. 9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Pag. 9
		(*b*) Describe any methods used to examine subgroups and interactions	Pag. 9
		(*c*) Explain how missing data were addressed	Pag. 9
		(*d*) If applicable, explain how loss to follow-up was addressed	NA
		(*e*) Describe any sensitivity analyses	Pag. 9
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Pag.10-11
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Pag.10-11
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) Summarise follow-up time (eg, average and total amount)	Pag.10-

			11
Outcome data	15*	Report numbers of outcome events or summary measures over time	Pag.10-11
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Pag 11-14
		(*b*) Report category boundaries when continuous variables were categorized	Pag 11-14
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Pag 15
Discussion			
Key results	18	Summarise key results with reference to study objectives	Pag16
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Pag17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pag17
Generalisability	21	Discuss the generalisability (external validity) of the study results	Pag 17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Pag 20

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	4
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	4
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	6-7
	4	Study objectives and hypotheses	6-7
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	8
Participants	**6**	Eligibility criteria	8
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	8
	8	Where and when potentially eligible participants were identified (setting, location and dates)	8
	9	Whether participants formed a consecutive, random or convenience series	8
Test methods	**10a**	Index test, in sufficient detail to allow replication	8
	10b	Reference standard, in sufficient detail to allow replication	8
	11	Rationale for choosing the reference standard (if alternatives exist)	8
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	8
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	8
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	8
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	8
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	9
	15	How indeterminate index test or reference standard results were handled	9
	16	How missing data on the index test and reference standard were handled	9
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	9
	18	Intended sample size and how it was determined	9
RESULTS			
Participants	**19**	Flow of participants, using a diagram	10
	20	Baseline demographic and clinical characteristics of participants	11
	21a	Distribution of severity of disease in those with the target condition	11
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	NA
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	12
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	12, 14
	25	Any adverse events from performing the index test or the reference standard	NA
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	17-18
	27	Implications for practice, including the intended use and clinical role of the index test	18
OTHER INFORMATION			
	28	Registration number and name of registry	NA
	29	Where the full study protocol can be accessed	NA
	30	Sources of funding and other support; role of funders	19

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.