Using interactional and linguistic analysis to distinguish between epileptic and psychogenic nonepileptic seizures: A prospective, blinded multirater study

Markus Reuber a,*, Chiara Monzoni a, Basil Sharrack a, Leendert Plug b

a Academic Neurology Unit, University of Sheffield, Sheffield, UK
b Department of Linguistics and Phonetics, University of Leeds, Leeds, UK

A R T I C L E   I N F O

Article history:
Received 9 June 2009
Accepted 11 July 2009
Available online 11 August 2009

Keywords:
Conversation Analysis
Discourse Analysis
Epilepsy
Psychogenic nonepileptic seizures
Differential diagnosis
Communication
Interaction
Controlled study
Prospective

A B S T R A C T

This study was carried out to test the suggestion that close interactional and linguistic examination of the communication between neurologists and patients during a first encounter can contribute to the differential diagnosis of epilepsy or psychogenic nonepileptic seizures. Twenty unselected patients admitted for video/EEG telemetry because of diagnostic uncertainty were included. Two linguists blinded to all medical data independently studied video recordings and transcripts of 25- to 35-minute interactions. They attempted to predict the medical diagnosis on the basis of qualitative assessments addressing 17 separate observations. They also used a diagnostic scoring aid (DSA) to convert their qualitative assessments into a simple numeric score. Using qualitative assessment, both linguists predicted 17 of 20 (85%) diagnoses (κ = 0.59). With the DSA, diagnoses were predicted with a sensitivity of 85.7% (71.4%) and a specificity of 84.6% (92.3%). This blinded, prospective multirater study confirms the diagnostic value of linguistic and interactional observations in the seizure clinic setting.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

The three most common conditions causing patients to present to doctors with blackouts are epileptic seizures (57%), syncope (22%), and psychogenic nonepileptic seizures (PNES, 18%) [1]. The process of distinguishing between these and rarer causes of blacking out poses a difficult challenge. Misdiagnosis rates of 25 to 30% have been reported [2–4]. Intertictal tests including MRI and EEG are normal or show only nonspecific changes in more than two-thirds of patients presenting after an unprovoked epileptic seizure [1,5]. The same tests show (unexpected) abnormalities in more than one-fifth of patients with PNES [6]. Brief outpatient or prolonged inpatient video/EEG telemetry captures typical events in only one-half to two-thirds of patients referred [4,7–9]; and the recording of a single seizure does not necessarily produce a comprehensive diagnosis.

In view of this, history taking remains the key tool in this clinical setting [5]. Given how important it is to make a correct diagnosis before choosing treatment, and how crucial history taking is in this process, it is surprising how little research has focused on optimizing this procedure. Several studies suggest that clusters of factual items (such as the presence or absence of cyanosis, tongue biting, and the invariable occurrence of attacks from an upright position or at certain parts of the night) can help with relatively simple decisions, such as differentiating between tonic–clonic seizures and syncope [1,10,11]. However, there is no evidence so far that this approach works reliably for the differentiation of epilepsy and PNES. Given that PNES are defined by their resemblance to epileptic seizures, it is likely that this distinction is more difficult. This suspicion is supported by a study that demonstrated that it took a mean of more than 7 years to arrive at the diagnosis of PNES [12]. One study in which two experienced epileptologists, unaware of any other clinical information, were asked to rate detailed written seizure descriptions from patients with temporal lobe epilepsy or nonepileptic seizures found that the sensitivity of this approach for the detection of epileptic seizures was very good at 96%, but the specificity was only 50% [13].

Inspired by this challenge, a multidisciplinary research group in Germany started the Epiling project and directed their attention especially to how patients talk about their seizure experiences rather than what symptoms they mention (http://www.uni-bielefeld.de/lili/forschung/projekte/archiv/epiling). Using a method derived from Conversation Analysis, they examined more than 150 interactions between doctors and patients with seizures and identified a set of linguistically describable features, which appeared to have...
discriminating value [14,15]. The linguists involved in this project were aware of the medical diagnoses during most analyses. However, a total of five analyses were completed in which the linguist was "blind"ed to all medical information. Although these cases were medically complex, the linguist predicted the medical diagnosis in four of five patients [16].

Similar forms of interactional and linguistic analysis have been used successfully in other settings [17–21]. However, we are not aware of any previous attempts to use the analysis of interactional data for the differential diagnosis of medical conditions. We previously showed that the "diagnostic" features originally described in German patients can be replicated in English-speaking patients [22]. In this study, we pursued two additional aims:

First, we wanted to confirm that linguists "blind"ed to all medical information are able to predict the video/EEG-confirmed medical diagnosis in unselected patients with seizures admitted for video/EEG monitoring because of diagnostic uncertainty. To this end, two linguists were asked independently to analyze transcripts of 20 unselected doctor–patient encounters and predict the eventual medical diagnoses. The linguists demonstrated how they had reached their final judgment by providing brief written reports on 17 different aspects of linguistic and interactional analysis previously described as having differential diagnostic value [14,16,23,24].

Second, we wanted to explore whether this qualitative approach could be translated into a diagnostic score with acceptable sensitivity, specificity, and interrater reliability. For this part of the study, the linguists were asked to condense their written analysis into 17 separate numeric scores to allow further quantitative analyses.

2. Methods

2.1. Patients

Unselected adults with refractory seizure disorders were considered for inclusion if they had been referred for video/EEG to the Royal Hallamshire Hospital (Sheffield, UK) because the admitting neurologist was uncertain whether they had epilepsy or PNES. Clinical interviews with the patients were submitted to linguistic analysis only if a seizure was captured by video/EEG during the period of admission and if it was confirmed as typical by patients and witnesses of habitual attacks. Patients were included only if the ictal electroclinical appearances allowed an unequivocal diagnosis of epilepsy or PNES (all patients had seizures involving apparent loss of consciousness). "Gold standard" diagnoses were made by the patients' neurologists taking account of all available clinical data including the video/EEG reports issued by fully trained neurophysiologists with no other involvement in this research project. Patients were excluded from the project if they were thought to have epileptic seizures as well as PNES, had been admitted for epilepsy surgery evaluation, did not speak fluent English, or were not able to complete any of the self-report measures. All patients completed the Hospital Anxiety and Depression Scale (HADS) [25] and Trauma History Questionnaire (THQ) [26]. The Graded Naming Test [27] and the Test for Reception of Grammar (Version 2) [28] were used to assess linguistic competence.

2.2. Interviews

All participants were interviewed by a neurologist (M.R.) they were meeting for the first time. The neurologist was not aware of the outcome of the video/EEG recording. The interviews (lasting 25–35 minutes) were recorded using the pre-installed equipment in the video/EEG suite. To maximize patients' participation in the conversation, to limit the possible effect of the doctor's contribution, and to improve the linguists' ability to relate their findings to previous work in this area, the encounters followed the guidelines proposed by the German EpiLing project [14] (see Supplementary Table 2 for details). The interviews had a very open beginning which made no mention of seizures (What was your expectation when you came here?), allowing patients to determine the initial focus of the conversation. Even when patients were prompted to talk about specific seizures, (Could you tell me about your first/worst/last seizure?), the use of open questions left them free to choose what they considered most relevant. Direct questions about features such as ictal injuries, tongue biting, incontinence, seizures from sleep, past medical history, or previous treatments were deliberately avoided to ensure that linguists' diagnostic decisions could not be affected by information obtained in this way.

2.3. Linguistic analysis

The transcribed interviews and original video recordings were analyzed independently by two linguists (L.P., C.M.) blind to all other information about the patient. Their analysis was inspired by Conversation Analysis (CA). CA is a sociological research method that is thought to provide insights into the nature and quality of interactions independent of psychological or motivational properties of the speakers [29]. A range of previous publications describe the application of the approach to this clinical setting [14,23,30] and contain transcript extracts and examples of analysis [22,31–33].

For the first part of this study, the linguists were asked to develop a diagnostic hypothesis (epilepsy or PNES) based on their brief written qualitative assessments addressing 17 different aspects of analysis previously described as having differential diagnostic potential. To ensure that both linguists were using a similar approach to the data, they were discouraged from considering any new linguistic insights of potential diagnostic value when formulating their "diagnoses."

For the second part of the study, the linguists were encouraged to complete a DSA (see Supplementary Table 1) based on the condensation of their brief qualitative reports to 17 numeric statements for each patient ("1" = more in keeping with epilepsy; "0" = don't know or unable to rate; "-1" = more in keeping with PNES). The full version of the DSA (with more detailed instructions) that the linguists completed is available at http://www.personal.leeds.ac.uk/~lnplp/Seizures.html. The 17 points on the DSA were identical to the analytic headings the linguists were encouraged to address in their written reports. Given that it was uncertain that qualitative reports could be translated into numeric scores, the linguists were instructed to base their diagnostic hypothesis only on their written linguistic arguments.

The following examples are included to offer an impression of the interactional data and assessment procedure. Both extracts contribute to the rating of DSA item 8, focusing on whether speakers place more emphasis on subjective seizure symptoms or on the situation in which the seizure occurred. We should point out that the rating of each item should take account of the whole encounter and not be based on one excerpt. The notation used in the transcripts has been simplified to improve readability. The numbers in brackets indicate the duration of prolonged pauses >0.1 seconds, "I" is the interviewer, "P" the patient.
2.6. Statistical analysis

P: Well Wednesday afternoon when the huh young man was (.2)
that were the one when I sat the chip pan on fire.
I: Right.
P: That was the worst one.
I: Tell me what you can remember about that one, (.2).
P: All I can remember is putting the chip pan on, (.25) and I got the chips ready in the
basket to put in, I never let it get right up; cause when it
gets right up if often (.25) sometimes it either goes (.25)
bubbles up (.5) or it will overflow; (1.0) certain night (.2)
I got the chip pan on (.5) I got chips on the (.2) side in the basket (.16). I was doing summat else (.25) oh makin a
drink, (.15) I’d gone up at the kettle, and never got to my
drink, (.5) there were no drink made; (1.4) and all I can remember is, wha- comin’ round (.25) and just going like
that with the chip pan; (.5) put the chips into the basket, (.6)
that’s when the chip pan just went up in flames; (1.2)
as I put the chips into the chip (.25) basket (.25) it
overflowed it and the chip pan went up in flames; huh, I
screamed at boyfriend; can’t even remember what I’ve done;
(.8) cannot remember for all of me money what I’ve done, he
like I know you were makin’ a drink, (.6) he says I’ve been
shouting you for nearly ten minutes; (1.3) eh s- so you’ve
not answered me; (.25) so; I nev- I couldn’t even hear him
shouting them; (.25) that’s how bad it was; (.75)
I: You can’t remember what happened in that time at all?
P: No; (.25) cannot remember at all; (.4) I can’t remember
anything; (.25) I can never remember.

2.4. Example 1: Patient 5 (lines 595–637, worst seizure prompt)

I: Huh (.6) what about the WORST attack that you ever had
(1.3)?
P: The one where I (.2) I fell down (.2) when no that were the
one when I sat the chip pan on fire.

3. Results

3.1. Patients

Seven patients with an eventual diagnosis of epilepsy and 13
whose final diagnosis was PNES were included. The ratio of epi-
lepsy (35%) to PNES (65%) is typical of patients admitted to our unit
for monitoring because of uncertainty about the etiology of attacks.
Table 1 summarizes clinical and demographic features. The
portion of women was significantly greater in the PNES group.
Although patients in the epilepsy group were older, there were
no significant differences in terms of duration of the seizure disor-
der, seizure frequency, or proportion admitted to the hospital with
prolonged seizures. The two groups were matched for linguistic
abilities. Patients in the PNES group reported more traumatic expe-
riences, and had higher mean anxiety and depression levels than
the patients with epilepsy. Whereas no patients in the epilepsy
group achieved the "caseness" level on the HADS, in the PNES
group, six patients scored at the "caseness" level for anxiety and
four for depression. Eighteen of 20 patients in this study had been
treated with antiepileptic drugs; 15 were still receiving AEDs at the
time of admission. Clinical pre-admission diagnoses were proven
incorrect by video/EEG in 12 of 20 cases (60%) (see Table 2).

3.2. Diagnosis based on qualitative analysis

Based on their qualitative assessment, both linguists predicted
the correct diagnosis in 17 of 20 patients (85%) (see Table 2). Each
linguist would have predicted the correct diagnosis in one addi-
tional patient if they had been allowed to use linguistic insights
for the diagnosis which became apparent only after the DSA had
been constructed. The interrater reliability of this qualitative pro-
cEDURE was moderate to good (κ = 0.57).

3.3. Diagnosis based on the DSA

Table 2 summarizes the numeric translation of the two raters’
qualitative analyses of each of the 17 DSA items. Although the final
linguistic hypothesis was usually correct, in almost every case
some items were rated as favoring the incorrect diagnosis. Rating
of a patient’s conversational profile as entirely typical of epilepsy
or PNES was the exception. Both linguists agreed on 229 of 340
(67.4%) ratings. There was no agreement on 109 on 340 (32%) rat-
ings. Frank disagreement was recorded in only 12 of 340 (3.5%) rat-
ings.

Mean total DSA scores were significantly higher for patients
with epilepsy than for those with PNES (rater 1: 8.5 (range:
−11.0 to 16.0) versus −0.35 (range: −9.5 to 10.0); rater 2: 7.64
(range: −6 to 13.5) versus 1 (range: −10 to 11.5)). The differences
in the mean DSA scores were significant for both raters (rater 1:
P = 0.017, rater 2: P = 0.047). The area under the ROC curve for lin-
guistic rater 1 was 0.835 (SE = 0.127, asymptotic signifi-
cance = 0.016). Using the semiquantitative procedure of the DSA
and the optimal diagnostic cutoff score suggested by the ROC curve
(4.5), this rater would have categorized 80% of the cases correctly

2.4.1. Interpretation

Patient focuses on circumstantial details, does not report any
subjective symptoms or use a witness’ account to communicate
seizure manifestations visible to others.

Rating for item 8: −1/+1.

2.5. Example 2: Patient 14 (lines 438–454, last seizure prompt)

P: Well Wednesday afternoon when the huh young man was (.2)
actually attaching all this huh umm we were busy chatting
away about it when all of a sudden (.5) the sensation of the
déjà vu thing huh came into my head huh um... and then the next
thing I’m fr’ from being sat on a chair over there huh since I
was' (.25) the last thing I
hadn’t occurred to me how they’d got me in bed huh ummm.

2.5.1. Interpretation

The seizure description is from a first-person perspective and
emphasizes subjective symptoms. The episode is well contoured
(“the last thing I can remember”)“then the next thing”).

Rating for item 8: +1/+1.

2.6. Statistical analysis

Bivariate analyses were conducted with Fisher’s exact χ² test
for categorical variables and the Mann–Whitney U test for contin-
uous data. A total DSA score for each patient was derived by adding
the 17 subscores. A nonparametric receiver operating characteris-
tic (ROC) curve was based on the total DSA scores to assess the sen-
sitivity and specificity of the semiquantitative diagnostic
procedure. κ values were calculated to determine the interrater
reliability of qualitative diagnostic procedure and of individual
DSA items. κ values were considered to denote the following
levels of interrater reliability: <0.2 = poor, 0.21–0.4 = fair,
0.41–0.6 = moderate, 0.61–0.8 = good, and 0.81–1.00 = very good.
The predictive value of individual DSA items was assessed with χ²-tests. Two-sided P values <0.05 were considered significant.
3.4. Individual DSA item scores

Table 2 lists those items that usually attracted a diagnostic rating (+1 or −1) and those items that more commonly failed to contribute to the diagnosis (scored 0). Item 3 attracted sixteen 0/0 scores, suggesting that it rarely contributed to the linguists’ diagnoses; item 6 had ten 0/0 scores.

Table 3 summarizes the diagnostic value and interrater reliability of individual DSA items. Both raters’ assessments of items 8, 9, 10, 15, and 16 were significantly associated with the final medical diagnosis. The association was strongest for items 9 and 15. In addition, some items worked well for one but not the other rater. Items 3, 5, 11, 13, 14, and 17 were not significantly associated with the final medical diagnosis of either rater. Importantly, this list also

Table 1
Demographic and clinical details.

<table>
<thead>
<tr>
<th></th>
<th>Epilepsy group (n = 7)</th>
<th>PNES group (n = 13)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female gender</td>
<td>28.6%</td>
<td>84.6%</td>
<td>P = 0.022</td>
</tr>
<tr>
<td>Age (years)</td>
<td>46 (35–67)</td>
<td>32 (23–55)</td>
<td>P = 0.019</td>
</tr>
<tr>
<td>Duration (years)</td>
<td>17 (2–38)</td>
<td>8 (0.5–17)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Frequency per month</td>
<td>24 (1–300)</td>
<td>14 (0.5–120)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Emergency admissions with seizures</td>
<td>71.4%</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Current AED use</td>
<td>71.4%</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Graded Naming Test score</td>
<td>17 (14–21)</td>
<td>16 (8–23)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Test for Reception of Grammara</td>
<td>17 (15–20)</td>
<td>19 (12–20)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Anxiety</td>
<td>6 (3–9)</td>
<td>10 (1–16)</td>
<td>P = 0.043</td>
</tr>
<tr>
<td>Depression</td>
<td>3 (1–7)</td>
<td>9 (1–12)</td>
<td>P = 0.005</td>
</tr>
<tr>
<td>Trauma History Questionnaire score</td>
<td>28.6%</td>
<td>84.6%</td>
<td>n.s.</td>
</tr>
<tr>
<td>Physical and sexual events</td>
<td>0 (0–1)</td>
<td>2 (0–12)</td>
<td>P = 0.007</td>
</tr>
<tr>
<td>Total events</td>
<td>3 (0–5)</td>
<td>6 (1–25)</td>
<td>P = 0.028</td>
</tr>
</tbody>
</table>

a Median (range).

b Version 2.

Table 2
Rating of patients on individual items (rater 1/rater 2).a

<table>
<thead>
<tr>
<th></th>
<th>P1b</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
<th>P16</th>
<th>P17</th>
<th>P18</th>
<th>P19</th>
<th>P20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preadmission diagnosis</td>
<td>E</td>
<td>E</td>
<td>N</td>
<td>E</td>
<td>E</td>
<td>N</td>
<td>E</td>
<td>E</td>
<td>N</td>
<td>E</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>Final diagnosis after V-EEG</td>
<td>N</td>
<td>N</td>
<td>IGE</td>
<td>FCE</td>
<td>N</td>
<td>N</td>
<td>FCE</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>FCE</td>
<td>FCE</td>
<td>N</td>
<td>FCE</td>
<td>N</td>
<td>FCE</td>
<td>N</td>
<td>FCE</td>
<td>N</td>
<td>FCE</td>
</tr>
<tr>
<td>Summary rater diagnosis</td>
<td>N/N</td>
<td>E/E</td>
<td>E/E</td>
<td>E/N</td>
<td>E/E</td>
<td>N/N</td>
<td>E/N</td>
<td>N/N</td>
<td>E/N</td>
<td>E/E</td>
<td>E/N</td>
<td>N/N</td>
<td>E/N</td>
<td>N/N</td>
<td>E/N</td>
<td>N/N</td>
<td>E/N</td>
<td>N/N</td>
<td>E/N</td>
<td>N/N</td>
</tr>
<tr>
<td>11</td>
<td>N/N</td>
<td>E/E</td>
<td>E/E</td>
<td>E/E</td>
<td>N/O</td>
<td>0/O</td>
<td>N/N</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
</tr>
<tr>
<td>12</td>
<td>0/0</td>
<td>0/E</td>
<td>E/E</td>
<td>E/E</td>
<td>E/O</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
</tr>
<tr>
<td>13</td>
<td>0/0</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
</tr>
<tr>
<td>14</td>
<td>0/0</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
</tr>
<tr>
<td>15</td>
<td>0/E</td>
<td>0/E</td>
<td>E/E</td>
<td>E/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
</tr>
<tr>
<td>16</td>
<td>0/0</td>
<td>0/E</td>
<td>E/E</td>
<td>E/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
</tr>
<tr>
<td>17</td>
<td>0/E</td>
<td>0/E</td>
<td>E/E</td>
<td>E/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
</tr>
<tr>
<td>18</td>
<td>0/E</td>
<td>0/E</td>
<td>E/O</td>
<td>0/O</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>N/N</td>
<td>N/N</td>
<td>N/N</td>
<td>N/N</td>
<td>N/N</td>
</tr>
<tr>
<td>19</td>
<td>0/O</td>
<td>0/E</td>
<td>E/O</td>
<td>0/O</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>E/N</td>
</tr>
<tr>
<td>20</td>
<td>N/N</td>
<td>E/N</td>
<td>0/O</td>
<td>0/E</td>
<td>N/E</td>
<td>0/E</td>
<td>N/E</td>
<td>0/E</td>
<td>N/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
<td>0/E</td>
</tr>
<tr>
<td>21</td>
<td>0/0</td>
<td>0/E</td>
<td>E/O</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
</tr>
<tr>
<td>22</td>
<td>0/E</td>
<td>0/E</td>
<td>E/O</td>
<td>E/O</td>
<td>E/O</td>
<td>E/O</td>
<td>E/O</td>
<td>E/O</td>
<td>E/O</td>
<td>E/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>E/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
<td>0/O</td>
</tr>
<tr>
<td>23</td>
<td>0/E</td>
<td>0/E</td>
<td>E/O</td>
<td>0/O</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
</tr>
<tr>
<td>24</td>
<td>0/E</td>
<td>0/E</td>
<td>E/O</td>
<td>0/O</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
</tr>
<tr>
<td>25</td>
<td>0/E</td>
<td>0/E</td>
<td>E/O</td>
<td>0/O</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
</tr>
<tr>
<td>26</td>
<td>N/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>E/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
<td>0/N</td>
</tr>
</tbody>
</table>

a Items on which both raters agreed are in italic, those with disagreement are in boldface.

b P, patient; I, item; E, epilepsy/in favor of epilepsy; N, nonepileptic seizures (NES)/in favor of NES; IGE, idiopathic generalized epilepsy; FCE, focal or cryptogenic epilepsy; 0, not rated/uncertain.

(sensitivity = 85.7%, specificity = 84.6%). For rater 2, the area under the ROC curve was 0.824 (SE = 0.117, asymptotic significance = 0.019). The optimal diagnostic cutoff for this rater was 7.5. At this cutoff, rater 2 would have categorized 75% of the cases correctly (sensitivity = 71.4%, specificity = 92.3%).
includes the items with the lowest interrater agreement (items 5 and 14). Interrater agreement was highest for items 1, 2, 6, 15, and 16.

4. Discussion

Two linguists independently generated diagnostic hypotheses of epilepsy or PNES in an unselected group of patients admitted to a video/EEG unit. All patients had frequent seizures; 60% carried an incorrect working diagnosis at the time of admission although they had experienced seizures for a median of 8.5 years. In view of this, it is impressive that the linguists correctly predicted 85% of the eventual video/EEG-supported diagnoses. The process by which the linguists formulated their hypotheses ensured that their judgment was not simply based on a hunch: Written reports commented on 17 separate interactional, topical, and linguistic items identified previously identified as potentially “diagnostic” [14,16]. The lack of questions about many of the items of the neurologists traditionally used to come to a diagnosis (such as ictal features and past medical history), and the complete lack of epileptologist experience of the linguists means that it unlikely that they were strongly influenced by this kind of factual information. Their decisions reached an acceptable level of interrater reliability; achieving κ values similar to those from the diagnostic interview for the International Classification of Sleep Disorders Revised (ICSD-R) or the system of Operationalized Psychodynamic Diagnostics (OPD) [34,35]. In fact, the level of interrater agreement achieved by interactional and linguistic analysis also stands up to comparison with that achievable with more objective methods such as EEG and EEG–polysomnography [36,37].

Our groups were well matched in terms of linguistic competence. It is unlikely that the differences in the communication profiles are explained by demographic group differences, such as the lower proportion of women or the greater age of patients in the epilepsy group. However, this possibility cannot be dismissed completely because other studies have demonstrated that gender and age can affect various aspects of language processing and interaction, including doctor–patient interaction [38,39]. It is also possible that the interactional and linguistic differences observed between the two groups were related to differences in trauma history, psychopathology, and associated attachment styles [40]. Unfortunately, our study design does not allow us to differentiate between association and causality.

This study represents only a first attempt to develop a set of linguistic tools of proven differential diagnostic value. It is likely that the diagnostic yield could be improved by dropping items that helped neither of the two raters to predict the final diagnosis. Additional items first described since the inception of this study could enhance the diagnostic value of linguistic observations in this clinical setting: we now recognize that patients with PNES are significantly more likely to show resistance to use of the label seizure for their problem than patients with epilepsy (P = 0.004). Resistance can become apparent by complete avoidance of the word seizure, lack of self-initiation of this term, hesitation when using seizure, or self-repair (e.g., from seizure to fit) [32]. We have also improved our approach to analysis of the use of metaphors [31]. It is likely that the interrater reliability of the approach could be improved by a formalized training package for linguists, which was not available to the two raters involved in this study.

The fact that we were able to operationalize linguistically defined “diagnostic” observations in the form of a DSA, and to condense the qualitative findings into a numeric summary score, further supports the notion that the linguists were describing “real” and intersubjectively communicable phenomena. The match of gold standard final diagnoses to DSA diagnoses (75 and 80% accurate, respectively) was much better than the match to working diagnoses prior to admission (40% accurate). However, the observation that the diagnostic accuracy based on purely qualitative assessments was better than that based on the numeric score also indicates the limitations of the DSA approach. Whereas the qualitative procedure allows raters to place particular emphasis on one or several particular observations in their overall judgment, all items are given the same diagnostic value in the DSA. It may be possible to improve the diagnostic performance of the DSA in future studies by giving diagnostically more relevant items more weight.

Clearly, this study has a number of limitations. First, we were able to study only a limited number of cases. In view of the heterogeneity of both epileptic and nonepileptic seizure disorders, it would be desirable to replicate the findings in a larger patient group. Second, the patients in this study were not representative of patients with recent-onset blackouts. The sensitivity and specificity of the items described here could be different in an outpatient population with a smaller proportion of patients with PNES. We were keen to demonstrate that the linguistic procedure described here could make diagnostic contributions in a group of patients posing a particular diagnostic challenge. Given our own clinical experience and the findings of members of the German research group [16], we expect that the communication behavior exhibited by the chronic patients included in this study is very similar to that of patients with less refractory seizure disorders of more recent onset, but future studies are needed to confirm our findings in other patient groups. We are also aware that the interview procedure will need to be adapted for use in the outpatient clinic. In the United Kingdom, most neurologists have only 30 minutes for a whole visit. During this time they may need to examine the patient, explain the diagnosis, and initiate treatment. They may have to ask questions about driving, work, past medical history, and family planning. Although we suspect that most of the observations discussed here could be made in briefer encounters (as long as the consultations begin with an open question and the doctor avoids shaping the agenda for the first 5 to 10 minutes), future studies will need to prove the value of sociolinguistic observations in routine outpatient encounters. Further, although patient satisfaction can be increased by more open history taking [41–43], our study does not allow us to say whether investing outpatient time in the less directive way of history taking required to generate the observations discussed here works better diagnostically than more conventional question-and-answer sequences eliciting factual information. Finally, our study is limited by the fact that it focused exclusively on encounters between patients and a doctor. No seizure witnesses, family members, friends, or carers were involved in any of the conversations studied. It is likely that the analysis of third-party contributions would offer additional diagnostic insights.

Despite these limitations, this study demonstrates the differential diagnostic potential of the interactional and linguistic items discussed here. Our findings show that it is worthwhile to pay attention to whether information was volunteered (and at what point), and how much effort the doctor has to make before the patient will give a detailed account of a particularly memorable seizure (such as the first, last, or worst). Apparently redundant (and potentially) annoying conversational features such as pauses, hesitations, restarts, and formulation work can add a diagnostically relevant layer of meaning to the pure content of speech. We do not mean to suggest that listening for how patients talk should replace the focus on what they say, but suspect that history taking works best if more traditional fact-based inquiries are complemented by the less directive approach described here.

Ethical approval

This study was approved by the South Sheffield Research Ethics Committee.
Conflict of interest statement

None of the authors has any conflict of interest to disclose.

Acknowledgments

This study was an "own account" project financed by B.S. and M.R. We are grateful for comments on this article by Martin Schöndienst, Rod Duncan, and Jon Stone.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.yebeh.2009.07.018.

References