Discovering themes in medical records of patients with psychogenic non-epileptic seizures

Joshua Lay,1 Udaya Seneviratne,1,2 Anthony Fok,2 Helene Roberts,2 Thanh Phan1,2

ABSTRACT

Introduction Epileptic and psychogenic non-epileptic seizures (PNES) are common diagnostic problems encountered in hospital practice. This study explores the use of unsupervised machine learning in discovering themes in medical records of patients presenting with PNES. We hypothesised that themes generated by machine learning are comparable with the classification by human experts.

Methods This is a retrospective analysis of the medical records in the emergency department of patients (age >18 years) with PNES who underwent inpatient video-electroencephalography monitoring from May 2009 to June 2014 and received a final diagnosis of PNES. Prior to machine learning of written text, we applied a standardised approach in natural language processing to create a document-term matrix (removal of numbers, stop-words and punctuations, transforming fonts to lower case). The words were separated into tokens and treated as if existing within a bag-of-words. A probability of each word existing within a topic (theme) was modelled on multivariate Dirichlet distribution (R Foundation, V.3.5.0). Next, we asked four experts to independently provide a clinical interpretation of the generated topics. When the majority of (≥3) experts agreed, it was regarded as highly congruent. Interactive data are available on the web at (https://gntem2.github.io/PNES/#topic=1&lambda=0.6&term=).

Results There were 39 patients (74.4% women, median age 35 years with range 20–82). A total of 121 documents were converted to text files for text mining. There were 15 generated topics with 12/15 topics rated as highly congruent. The main themes were about descriptors of seizures and medication use.

Conclusions The findings from machine learning on PNES-related documentation provides evidence for the feasibility of applying machine-learning methodology to analyse large volumes of medical records. The topics generated by machine learning were congruent with interpretations by clinicians indicating this method can be used for screening of medical conditions among large volumes of medical records.

INTRODUCTION

Psychogenic non-epileptic seizures (PNES) are involuntary experiential and behavioural responses to a psychological process which are not associated with abnormal electrical discharges in the brain.1 PNES has some features similar to epileptic seizures (ES) and can be mistaken for ES by those who are not familiar with this condition.2 The frequency of PNES misdiagnosed as epilepsy is as high as 30%,3 and in some cases diagnosis can be delayed by 7–9 years.4 5 As a result, patients with PNES are exposed to iatrogenic harms which might cause significant morbidity and mortality,6 as well as contribute to excessive utilisation of scarce medical resources.7

Many researchers have been evaluating different approaches to improve the diagnosis and classification of ES versus PNES.8–10 These approaches can improve the accuracy of visual discrimination of seizures short term and medium term.10 In hypothetical treatment scenarios, this newly learnt knowledge did not necessarily translate to logical decision-making in the use of medication for the management of PNES.11 The lack of congruence between knowledge and clinical practice has led us to consider different approaches for understanding the thinking process at the time of the management of patients with PNES in the emergency department. We have chosen the emergency department as it is the first point of contact between the patients and the doctors in the hospital setting. One way of understanding the thinking at the time of the initial management is to employ a natural language-processing approach to analyse the written material by doctors working in the emergency department. Probabilistic topic modelling is a machine-learning method that generates topics or discovers themes among a collection of documents. This task is performed without the aid of an observer hence, the method is listed as unsupervised.12 13 We hypothesised that thematic analysis (topic modelling) permits an ‘unbiased’ approach to discover the themes documented in medical records.
This is a retrospective analysis of the medical records in the Emergency Department of adult patients (age >18) with psychogenic nonepileptic seizures (PNES) who underwent inpatient video-EEG monitoring from May 2009 to June 2014 and received a final diagnosis of PNES by epileptologists.

Prior to machine learning of written text, we applied a standardised approach in natural language processing to create a document-term matrix. This process includes removal of numbers, stop-words, and punctuations, transforming fonts to lower case.

Next, the words were separated into tokens and treated as existing within a bag-of-words. A probability of each word existing within a topic (theme) was generated in R (R Foundation, version 3.5.0). The distributions of words in each topic were modelled on multivariate Dirichlet distribution.

The topic model was visualised using a package in R called LDAvis.

Finally, we asked four experts to independently provide clinical interpretation of the generated topics. When the majority of experts agreed, it was regarded as highly congruent.

Figure 1 The flowchart illustrates the key steps in data collection and analysis. EEG, electroencephalograph.

Figure 2 Wordcloud of frequent words in documents. The size of the words represents the frequency of word appearance in documents. eeg, electroencephalograph; gcs, Glasgow Coma Scale; gtc, generalised tonic–clonic seizure; hopc, history of presenting conditions; loc, loss of consciousness; mas, Melbourne ambulance system; nad, no abnormality detected.

METHODS

Subjects

We reviewed all medical records of patients who underwent inpatient video electroencephalography monitoring (VEM) from May 2009 to June 2014. This study included adult patients (age ≥18 years) with a final diagnosis of PNES. The consensus opinion of at least two epileptologists based on history, examination and investigations including the video-electroencephalograph was required for the final diagnosis. Those patients with a mix of both ES and PNES were excluded. The methodology was detailed in our previous publications.7 14

Data retrieval, pre-processing and analysis

The medical records pertaining to clinical notes and observations during the emergency department visits were manually transcribed into plain text format. These documents were then compiled into a collection of documents (corpus). We used tm package (in R Foundation for Statistical Analysis V.3.5.0) for the pre-processing of the data.15 These steps included tokenisation where words in a sentence were separated so that they existed within a bag-of-words (unigram analysis). There was no importance placed with regards to the order of words in the documents. The words in the document-term matrix were further processed to enhance the accuracy of subsequent analysis. Phrases considered to be important were paired together. Additionally, a ‘stop word’ filter was used to remove common English words that provide little information (eg, ‘I, he, am, was, don’t and so on’). Next, we used the term frequency–inverse document frequency to prioritise important terms. This enhances the detection of meaningful terms and downweights common but unimportant terms that were not previously removed. Topic models were then generated from the corpus of these documents to extract ‘hidden themes’. This step was performed using the Latent Dirichlet Allocation algorithm via the topicmodels package in R.16 Finally, the topics and words related to each topic were visualised using the LDAvis package for R (interactive data available on the web https://gntem2.github.io/PNES/#topic=1&lambda=0.6&term=).17 The clinical interpretations of the topics were provided by four experienced neurologists (US, TP, AF, HR). Each rater was asked to independently provide three possible interpretations of each topic. These interpretations had to be in the form of a single word; they were not asked to provide a description of the topic. If three or four experts provided a similar interpretation of a topic, it was considered to be highly congruent. When the agreement was found between two
experts only, the results were rated equivocal. Figure 1 highlights the key steps in data collection and analysis.

**RESULTS**

A total of 39 patients fulfilled the inclusion criteria. Women constituted 74.4% of the cohort (29 cases). The median age of the cohort was 35 years (range 20–82) and the median duration of the condition was 3 years (range 1–432 months). A total of 121 documents were analysed and used in this analysis. Fifteen topics (themes) were generated based on analysis of the harmonic mean. Frequent words appearing in the document are displayed as wordcloud in figure 2. Words used in the description of ES were frequently observed among these patients with PNES and included ‘tonic’, ‘clonic’, ‘gtc’ (abbreviation for generalised tonic–clonic seizure), ‘phenytoin’, ‘midazolam’ and ‘clonazapam’. Words suggestive of PNES were used rarely (‘pseudoseizure’).

The results of the topic modelling are displayed on the web to allow user interaction (https://gntem2.github.io/PNES/#topic=1&lambda=0.6&term=). The user can query the topic by clicking on the topic number or number under the tab ‘Selected Topic’. By hovering over words, the topics in which the terms exist will appear. Screenshots of the web display are available as figures 3 and 4. Table 1 summarises the interpretation of each topic independently by the four experts. Agreement among the experts was observed in 12/15 topics. An example of agreement on a topic is illustrated in figure 3. The grouping of words in this topic suggests a description of the clinical observation in the emergency department. The terms used to describe this topic by the raters in table 1 were: phenomenology (rater 1), semiology (rater 2), phenomenology (rater 3), semiology (rater 4). The raters were congruent in their interpretation of the theme of this topic.

Table 2 provides the frequent terms observed in the 15 topics and the percentage of tokens (words). The data for topics 1 and 5 are the same as those in figures 3 and 4. It is recommended that the reader visit the interactive web page to explore the relationship among words as well as their appearances in the topic. For example, the term ‘status’ appears in topics 3, 5 and 12, and ‘ambulance’...
appears in topics 2, 7, 8 and 11, but ‘arching’ only appears in topic 1.

**DISCUSSION**

In this proof-of-concept study, we have demonstrated the use of unsupervised machine-learning methods to identify different themes in medical records of patients with PNES. The majority of these themes were interpreted as congruent by four experts. This method is efficient in gaining quick comprehension of the medical records. Furthermore, topic models provide an insight into the thinking process of health professionals at the time of initial management. While the method is applied to a small dataset from the scanned medical records here as proof of concept, it is scalable to large volumes of medical data available in electronic medical records and for any disease process. The topics generated by machine learning were congruent with interpretations by clinicians indicating this method can be used for screening of medical conditions among large volumes of medical records in clinical practice.

Natural language processing (NLP), also described as text mining, is a way of converting unstructured text into a structured format paving the way for automated analysis. With the increasing use of electronic medical records, the need for NLP to handle large volumes of medical information has also risen. There are three main approaches adopted in NLP. The machine-learning approach involves fully automated text processing, whereas the rule-based approach depends on predefined rules by experts. The hybrid approach is a combination of the two methods.

The topic modelling approach has been previously adopted by medical researchers to identify patterns of comorbid medical conditions, detect medication non-compliance using posts in patient forums, describe public health information from the social media, cluster analysis of large biomedical datasets and predict inpatient clinical order patterns. The method used in this study assigns words to topics (themes) based on the probability of membership of the topic but the clinical interpretation of the topic is still needed. We approached this by asking four experts to assign meaning to the collection of words in each topic. This aspect of the work can be prone to bias as different experts can interpret the collection of words in different ways or use various words to convey a similar meaning. For example, the terms ‘phenomenology’ and ‘semiology’ were used to described observations on seizures. A drawback of topic modelling...
Table 1  Clinical interpretation of the 15 topics by four raters

| Topic | Rater 1 | Rater 2 | Rater 3 | Rater 4 | Rater 1 | Rater 2 | Rater 3 | Rater 4 |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|
|       | Interpreted theme 1 | Interpreted theme 1 | Interpreted theme 1 | Interpreted theme 1 | Interpreted theme 2 | Interpreted theme 2 | Interpreted theme 2 | Interpreted theme 2 |
| 1     | Phenomenology of seizure | Semiology | Seizure phenomenology | Semiology | Phenomenology of seizure | Semiology | Seizure phenomenology | Semiology | Phenomenology of seizure | Semiology | Seizure phenomenology | Conscious state |
| 2     | Post-ictal phenomenology | Findings on examination | Examination | Semiology | History of event | Findings on examination | Left-sided weakness | Semiology | PNES event | Findings on examination | Anxiety | Severity |
| 3     | Phenomenology | Type of movement | Tonic-clonic seizure | Generalised phenomenology | Phenomenology | Type of movement | Tonic-clonic seizure | Semiology | Consciousness | Awareness | Tonic-clonic seizure | Semiology | Severity |
| 4     | Phenomenology | Treatment | Seizure-phenytoin | Medication | Phenomenology | Awareness | Seizure-phenytoin | Semiology | Phenomenology | Awareness | Seizure-phenytoin | Semiology | Severity |
| 5     | Background history | Type of movement | Seizure-valproate | Conscious state | Background history | Awareness | Seizure-valproate | Medication | Phenomenology | Treatment | Seizure-valproate | Recurrence |
| 6     | Treatment | Treatment | Anti-epileptic medications | Medication | Post-ictal signs | Treatment | Anti-epileptic medications | Generalised | Phenomenology | Clinical features | Anti-epileptic medications | Severity |
| 7     | Phenomenology | Post-ictal phase | Post-ictal | Recurrence | Treatment | Treatment | Post-ictal midazolam | Semiology | Treatment | Examination findings | Post-ictal behaviour | Conscious state |
| 8     | Consciousness | Comorbidity | Vasovagal | Semiology | Consciousness | Treatment | Preceding seizure history | Differential diagnosis | Confusion | Semiology | Consciousness | Treatment | Fever | Severity |
| 9     | Preceding seizure history | Responsiveness | Blackout | Conscious state | Responsiveness | History prior to Seizure | Responsiveness | Tremor | Semiology | History prior to seizure | Responsiveness | Headache | Recovery |
| 10    | History prior to seizure | Ambulance assessment | Ambulance assessment | Semiology | History prior to seizure | Awareness | Fall | Conscious state | History prior to seizure | Awareness | Conversion | Post-ictal |
| 11    | Phenomenology of seizure | Clinical features | Observations in resuscitation bed | Semiology | Phenomenology of seizure | Type of movements | Observations in resuscitation bed | Generalised | Phenomenology of seizure | Awareness | Observations in resuscitation bed | Severity |
| 12    | Consciousness | Diagnosis | Airway assessment | Semiology | Consciousness | Duration | Agitation | Recovery | Post-ictal psychosis | Investigations | Assault | Provoked |
| 13    | Treatment | Post-ictal phase | Post-ictal | Semiology | Phenomenology | Treatment | Post-ictal | Medication | Phenomenology | Responsiveness | Post-ictal | Post-ictal |
| 14    | Consciousness | Airway management | History | Conscious state | Pre-ictal history | Investigations | History | Conscious state | Medication | Awareness | History | Recovery |
| 15    |                   |       |       |       |       |       |       |       |       |       |       |       |

The raters provided three possible interpretations for each topic. Shading represents congruent themes. PNES, psychogenic non-epileptic seizures.
### Table 2  Top 30 most salient words for 15 topics

| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 | Topic 11 | Topic 12 | Topic 13 | Topic 14 | Topic 15 |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Left    | Lower   | Clonic  | Phenytoin | Arrival | Clonazepam | Episode | Small | Post | Episode | Ambulance | Twitching | Activity | Postictal | States   |
| Weakness | Arm     | Tonic   | Fit      | Valproate | Diazepam | Postictal | Limb   | Right | Blackouts | Ceased    | Neck     | Seizure type | Sodium valproate | Period |
| Unable  | Examination | Selfresolve | Limb | Gcs | Midazolam | Past | Brain | Body | Period | Cal | Prior | Stopped Medication | Trolley |
| Previous | Without | Epilepsy | Blood | Headache | Gtc | Jerking | Fevers | Said | Epileptic | Drowsy | Discharge | Mas | Injuries | Airway |
| Face    | Memory | Place | Arms | Similar | Describes | Injury | Admission | Levetiracetam | Floor | Head strike | Response | Reports | Contact | Frequent |
| Speak   | Right | Etoh   | Months | Loading | Hour | Developed | Came | See | Similar | Loc | Spontaneously | Stunned | Urine |
| Typical  | Agitated | Level | Movements | Onset | Arrived | Examination | Supported | Found | Used | Confused | Insulin | Pseudoseizure | Stimuli | Disorder |
| Grunting | Falls   | Room   | Settled | Felt | Hour | Loc | Poor | Side | Triage | Regularly | Gtc | Intermittent | Generalised | Happened |
| Noises  | Work    | Anxious | Unconscious | Consciousness | Sits | Sitting | Description | Short | Bears | Better | Responsive | Guedel | Consciousness | Pins |
| Unresponsive | Associated | Aware | Reports | Responsive | Frothing | Shakes | Started | Woke | Become | Reassurance | Decision | Fingers | Unknown | Conversation |
| Multiple | Previous | Regained | Strilled | Strailed | Variably | Fit | Duration | Mas | Sob | Tumor | Can | Tolerating | Call | Movements | Still |
| See     | Ambulance | Times | Opening | Struck | Whole | Resus | Meds | Treated | Looked | Way | Blue | Xsy | Red | Investigation |
| Resolved | Hold | Status | Holding | Saying | Phenytoin | Midazolam | Taking | Diagnosed | Two | Complaints | Slowly | Collapsed | Furry | Regain |
| Make    | Earlier | State | Altered | Changes | Ground | Calling | Saw | Headache | Legs | Med | Commoened | Flat | Requiring | Fell |
| Drooping | Event   | Total | Feeling | Kind | Lying | Behaviour | Vasovagal | Repeated | Turns | Conscious | Applied | Assault | Increase | Arrived |
| Instance | Ear    | Ground | Fell | Like | Jerking | Displays | Drowsy | Syncope | Awake | Come | Lasts | Prolonged | Like | Conscious |
| Gcs     | Worse | Meds | State | Mal | Fall | Legs | Self | Complain | Responding | Asked | Focal | Taken | Unresponsive | Gags |
| Hands   | Claims | Pupils | Sitting | Test | Legs | Bit | Rolling | Looking | Became | Impression | Gastro | Fitted | Hands | Opens |
| Guite   | Anxiety | Resists | Shakes | Didn’t | Arms | Current | Depression | Maintained | Activity | Became | Drowsiness | State | Prior | Attempted |
| Antiepileptic | Combined | Handed | Leg | Brief | Cyanosis | Gcs | Alert | Associated | Arriving | Focal | Fitted | Meds | Arm | Agitation |
| Arching | Episode | Notify | Past | Status | Husband | Feel | Mouth | Move | Awoke | Reactive | Hours | Squeezes | Spasm | Converse |
| Talking | Left | Disorientated | Shook | Hands | Violently | Recovery | Became | Loaded | Man | Dusky | Observed | Discs | Close | Never |
| Appeared | Move | Quiet | Injuries | Collapse | Returned | Collapse | Ground | Hours | Security | Equivocal | Resus | Aggressive | Investigate | Develops |
| Consciousness | Weakness | Difference | Hips | Sounds | Sodium | valproate | Sounds | Focal | Confused | Leg | Interact | Daughter | Aprooa | Definitely | Note | Coryzal |
| Full    | Fall    | Rapid | Lots | Staring | Speaking | Nasal | Gcs | Total | Ordination | Labels | Gym | Verbal | Frustrated | Seat |
| Feel    | Recovery | Support | Padded | Btwn | Sol | Slumped | Work | Duration | Headache | Stayed | Cuff | Phx | Occurred | Visible |
| Feeling  | Unknown | Enolute | Lifting | Phase | Constrictions | Surgery | Ambulance | Still | Mildy | Violent | Fists | Oblivious | Gazing | Drop |
| Falls   | Epilepsy | Aura | Big | Tell | Pregnancy | Vitals | Light | Speak | Nature | Causing | Inflating | Sure | Breath | Pillow |
| Opening | Floor | Stop | Remembering | Lowered | Rash | Vomiting | Conus | Consciousness | Far | Admissions | Promuding | Witness | Surrounding | Crushed |
| Levetiracetam | Fit | Immediately | Requested | Mobilise | Sats | Top | Circulation | Increased | During | Conversion | Saliva | Expression | Sphincter | Ct brain |

The words are ordered from top to bottom in terms of probability of belonging to the topic.
is that it is not a classification tool and cannot separate PNES from ES. Machine-learning methods for classification include generalised linear model, naive Bayes classification, tree-based approaches, support vector machine and neural network. These methods were not used here as they are not adept at evaluating the thematic structures of the documents we used in the study.

This study provides useful insights into how doctors in the emergency department view or conceptualise seizures among patients presenting with PNES. This approach allows one to postulate that the clinicians were considering the diagnosis of ES rather than PNES. This was inferred from the frequent use of descriptive semiological terms of ES such as ‘tonic’ and ‘clonic’ in the documents. Furthermore, ‘epilepsy’, ‘epileptic’ and ‘GTC’ (implying generalised tonic–clonic seizure) appeared more often in topics (2, 3, 6, 10, 12), whereas ‘pseudoseizure’ featured only once in topic 13 indicating the focus of doctors was ES. Topic 1 is a collection of terms describing the seizure semiology, but descriptions of terms relating to PNES such as ‘arching’ appeared only once. Other typical terms used in the description of PNES such as pelvic thrusting, ‘eye closure’, ‘head-shake’, ‘asymmetry’ and ‘asynchrony’ were not observed at all. This view is consistent with the frequent mention of antiepileptic medications used in the treatment of seizure and status epilepticus; these medications include phenytoin, clonazepam, diazepam and midazolam (topic 6 on acute treatment). The frequent use of the term ‘loading’ is a likely reference to intravenous loading of phenytoin in status epilepticus and appeared in topics 6 and 14. Aligned with that, ‘status’ appeared in topics 3 and 5 suggesting that the doctors were considering the diagnosis of status epilepticus. Misdiagnosis of non-epileptic psychogenic status as true status epilepticus leads to inappropriate interventions resulting in considerable morbidity, healthcare utilisation cost and even mortality. These observations raise the need to improve education on seizure diagnosis among medical professionals.

Related to this matter is the need to document observations of these events in free-text form avoiding the use of jargon. This is a potential trap with electronic medical record whereby commonly used phrases are saved for repeated use. This situation may lead to homogenising of neurological descriptions. What we have illustrated here is only one use of topic modelling with medical records in the setting of the emergency department. This approach does not have to be restricted to this location or neurological disorders. It can be used for other medical conditions and in any location. Furthermore, the method can also be applied to qualitative data from surveys or suggestions from team-building meetings.

Limitations
There are several limitations to this study. The unigram (bag-of-words) approach we adopted does not arrange words according to their meanings. In this approach, each word is treated equally and the relationships among words are not explored. In order to overcome this limitation, we used word combinations when we considered the sequence of words to be important (example: ‘status epilepticus’). An alternative method is bigram approach. However, the package ‘topicmodels’ does not handle the bigram analysis. Additionally, the use of stop-word filter denotes that a negative meaning of the sentence may not be discovered in the themes (example: ‘no incontinence’ vs ‘incontinence’). At the time of the data acquisition, electronic medical records were not operational in our institution. The data were transcribed from scanned medical records. This task introduced a potential source of error with manual copying of texts. It is hoped that the introduction of electronic medical records will render the analysis of written text easier. Another limitation of this study is that we have only studied patients with PNES who had VEM. This approach was undertaken to ensure the analysis was related to patients with confirmed PNES. The drawback of this approach is that patients with PNES who did not have VEM could not be captured. Hence, it is possible that more severe cases of PNES were included in the study introducing potential bias. Furthermore, our study did not include patients presenting with ES. As such the results cannot be directly extrapolated to all patients presenting with seizures.

CONCLUSIONS
Our study shows that unsupervised machine learning can be used to objectively and efficiently evaluate large chunks of unsorted medical data. This provides a good starting point for subsequent deeper analysis by highlighting key themes. Additionally, our study also has implications for the understanding of the thinking process of healthcare workers at the time of evaluating patients presenting with seizures. The analysis of a larger sample of patients presenting with ES and PNES would be useful to detect different topics between the two seizure types in order to design a way of improving the diagnosis based on the topic modelling approach.

Contributors JL—literature search, data collection, data analysis and interpretation, drafting and revision of the manuscript. US—literature search, data interpretation and critical revision of the manuscript. AF—data interpretation and critical revision of the manuscript. HR—data interpretation and critical revision of the manuscript. TP—study concept and design, data analysis and interpretation, and critical revision of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests There are no competing interests.

Patient consent for publication Not required.

Ethics approval This retrospective study was conducted at Monash Medical Centre, Melbourne, Australia with approval from the Human Research Ethics Committee of Monash Health, Australia.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request by researchers subject to approval of the institution.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which provides a good starting point for subsequent deeper analysis by highlighting key themes. Additionally, our study also has implications for the understanding of the thinking process of healthcare workers at the time of evaluating patients presenting with seizures. The analysis of a larger sample of patients presenting with ES and PNES would be useful to detect different topics between the two seizure types in order to design a way of improving the diagnosis based on the topic modelling approach.

Contributors JL—literature search, data collection, data analysis and interpretation, drafting and revision of the manuscript. US—literature search, data interpretation and critical revision of the manuscript. AF—data interpretation and critical revision of the manuscript. HR—data interpretation and critical revision of the manuscript. TP—study concept and design, data analysis and interpretation, and critical revision of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests There are no competing interests.

Patient consent for publication Not required.

Ethics approval This retrospective study was conducted at Monash Medical Centre, Melbourne, Australia with approval from the Human Research Ethics Committee of Monash Health, Australia.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request by researchers subject to approval of the institution.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which provides a good starting point for subsequent deeper analysis by highlighting key themes. Additionally, our study also has implications for the understanding of the thinking process of healthcare workers at the time of evaluating patients presenting with seizures. The analysis of a larger sample of patients presenting with ES and PNES would be useful to detect different topics between the two seizure types in order to design a way of improving the diagnosis based on the topic modelling approach.

Contributors JL—literature search, data collection, data analysis and interpretation, drafting and revision of the manuscript. US—literature search, data interpretation and critical revision of the manuscript. AF—data interpretation and critical revision of the manuscript. HR—data interpretation and critical revision of the manuscript. TP—study concept and design, data analysis and interpretation, and critical revision of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests There are no competing interests.

Patient consent for publication Not required.

Ethics approval This retrospective study was conducted at Monash Medical Centre, Melbourne, Australia with approval from the Human Research Ethics Committee of Monash Health, Australia.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request by researchers subject to approval of the institution.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which provides a good starting point for subsequent deeper analysis by highlighting key themes. Additionally, our study also has implications for the understanding of the thinking process of healthcare workers at the time of evaluating patients presenting with seizures. The analysis of a larger sample of patients presenting with ES and PNES would be useful to detect different topics between the two seizure types in order to design a way of improving the diagnosis based on the topic modelling approach.

Contributors JL—literature search, data collection, data analysis and interpretation, drafting and revision of the manuscript. US—literature search, data interpretation and critical revision of the manuscript. AF—data interpretation and critical revision of the manuscript. HR—data interpretation and critical revision of the manuscript. TP—study concept and design, data analysis and interpretation, and critical revision of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests There are no competing interests.

Patient consent for publication Not required.

Ethics approval This retrospective study was conducted at Monash Medical Centre, Melbourne, Australia with approval from the Human Research Ethics Committee of Monash Health, Australia.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request by researchers subject to approval of the institution.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which provides a good starting point for subsequent deeper analysis by highlighting key themes. Additionally, our study also has implications for the understanding of the thinking process of healthcare workers at the time of evaluating patients presenting with seizures. The analysis of a larger sample of patients presenting with ES and PNES would be useful to detect different topics between the two seizure types in order to design a way of improving the diagnosis based on the topic modelling approach.

Contributors JL—literature search, data collection, data analysis and interpretation, drafting and revision of the manuscript. US—literature search, data interpretation and critical revision of the manuscript. AF—data interpretation and critical revision of the manuscript. HR—data interpretation and critical revision of the manuscript. TP—study concept and design, data analysis and interpretation, and critical revision of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests There are no competing interests.

Patient consent for publication Not required.

Ethics approval This retrospective study was conducted at Monash Medical Centre, Melbourne, Australia with approval from the Human Research Ethics Committee of Monash Health, Australia.
permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD
Udaya Seneviratne http://orcid.org/0000-0003-0182-9702

REFERENCES
1. Reuber M, Elger CE. Psychogenic nonepileptic seizures: review and update. Epilepsy Behav 2003;4:205–16.
2. Oto MM. The misdiagnosis of epilepsy: appraising risks and managing uncertainty. Seizure 2017;44:143–6.
3. Chowdhury FA, Nashef L, Elwes RDC. Misdiagnosis in epilepsy: a review and recognition of diagnostic uncertainty. Eur J Neurol 2008;15:1034–42.
4. Reuber M, Fernández G, Bauer J, et al. Diagnostic delay in psychogenic nonepileptic seizures. Neurology 2002;58:493–5.
5. Seneviratne U, Briggs B, Lowenster et al. The spectrum of psychogenic non-epileptic seizures and comorbidities seen in an epilepsy monitoring unit. J Clin Neurosci 2011;18:361–5.
6. Wilner AN, Bream PR. Status epilepticus and pseudostatus epilepticus. Seizure 1993;2:257–60.
7. Seneviratne U, Low ZM, Low ZX, et al. Medical health care utilization cost of patients presenting with psychogenic nonepileptic seizures. Epilepsia 2019;60:349–57.
8. Ochoa JG, Wludyka P. Randomized comparison between traditional plus interactive web-based methods for teaching seizure disorders. Teach Learn Med 2008;20:114–7.
9. Seneviratne U, Minato E, Paul E. How reliable is ictal duration to differentiate psychogenic nonepileptic seizures from epileptic seizures? Epilepsy Behav 2017;66:127–31.
10. Seneviratne U, Ding C, Bower S, et al. Video-based training improves the accuracy of seizure diagnosis. J Neurol Neurosurg Psychiatry 2014;85:466–70.
11. Seneviratne U, Ma H, Phan TG. How do doctors in training react to seizures? Epilepsy Behav 2016;54:104–9.
12. Blei DM. Probabilistic topic models. Commun ACM 2012;55:77–84.
13. Blei DM, AY N, Jordan MI. Latent Dirichlet allocation. J Machine Learn Res 2003;3:993–1022.
14. Seneviratne U, Rajendran D, Brusco M, et al. How good are we at diagnosing seizures based on semiology? Epilepsia 2012;53:e63–6.
15. Feinerer I, Hornik K, Meyer D. Text Mining Infrastructure in R. J Stat Softw 2008:25.
16. Grün B, Hornik K. topicmodels: An R Package for Fitting Topic Models. J Stat Softw 2011;40.
17. Sievert C, Shirley K. LDavis: a method for visualizing and interpreting topics. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, Maryland, USA, 2014.
18. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. J Biomed Inform 2017;73:14–29.
19. Pons E, Braun LMM, Hunink MGM, et al. Natural language processing in radiology: a systematic review. Radiology 2016;279:329–43.
20. Bhattacharya M, Jurkovicz C, Shatkay H. Identifying patterns of associated-conditions through topic models of electronic medical records. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 15-18 Dec 2016, 2016.
21. Abdellaoui R, Foulquié P, Texier N, et al. Detection of cases of Noncompliance to drug treatment in patient forum posts: topic model approach. J Med Internet Res 2018;20:e85.
22. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9:e103408.
23. Zhao W, Zou W, Chen JJ. Topic modeling for cluster analysis of large biological and medical datasets. BMC Bioinformatics 2014;15(Suppl 11):S11.
24. Chen JH, Goldstein MK, Asch SM, et al. Predicting inpatient clinical order patterns with probabilistic topic models vs conventional order sets. J Am Med Inform Assoc 2016;77:ocw136–80.