Diagnostic & investigative approach of consultant neurologists in a real-world clinical setting

Chris Aitchison[1], Daniel Blackburn[2], Aijaz Khan[3], Richard Grünewald[3], and Tom Jenkins[2]

[1]The University of Sheffield
[2]The University of Sheffield Institute for Translational Neuroscience
[3]Sheffield Teaching Hospitals NHS Foundation Trust

July 8, 2020

Abstract

Background: Whilst core curricula in neurology are nationally standardised, in real-world clinical practice, different approaches may be taken by individual consultants. In this study, we investigated: (1) variance in diagnostic and investigative practice, using a case-based analysis of inter-rater agreement; (2) potential importance of any differences in terms of patient care; (3) relationships between clinical experience, diagnostic certainty, diagnostic peer-agreement and investigative approach; (4) development of novel individualised metrics to facilitate appraisal. Methods: Four neurologists with 0-23 years' experience at consultant level provided diagnosis, certainty (10-point Likert scale), and investigative approach for 200 consecutive general neurology outpatients seen by a newly qualified consultant. Diagnostic agreement was evaluated by percentage agreement. The potential importance of any diagnostic differences was assigned a score by the evaluating neurologist (6-point Likert scale). Associations between diagnostic agreement, certainty and investigative approach were assessed using Spearman correlation, logistic and ordinal regression, and reported as individualiszd metrics for each rater. Results: Diagnostic peer-agreement was 4/4, 3/4, 2/4 and 1/4 in 50%, 28%, 20% and 3% of cases, respectively. In 17%, differences in patient management were judged potentially important. Investigation rates were 42-73%. Mean diagnostic certainty ranged between 6.2/10 (SD 2.1) to 7.7/10 (SD 2.2) between least and most experienced consultants. Greater diagnostic certainty was associated with greater diagnostic peer-agreement (individual-rater regression coefficients 0.30-0.51, $p<0.01$) and lower odds of arranging investigations (individual-rater odds ratios 0.58-0.78, $p<0.01$). Conclusions: Variance in diagnostic and investigative practice between consultant neurologists exits and may result in differing management. Mean diagnostic certainty increased numerically with experience and was statistically associated with greater diagnostic peer-agreement and lower investigation rates. Metrics reflecting concordance with peers, and relationships to diagnostic confidence, could inform reflective practice.

Research paper statements

What is already known about this topic?

- There are relatively few research studies focusing on inter-rater variation in consultant practice in neurology, or any other specialty, especially in a real-world outpatient setting, in which uncertainty, balancing risks and benefits, and making clinical judgements, often in the face of incomplete information, are characteristic.
- Previous studies have tended to focus on approach to single symptoms or diseases.
- There is conflicting information on the impact of experience

What does this article add?

- This study broadens the assessment of inter-rater variation, focusing on a wider range of diseases as well as different aspects of practice.

- In addition, the study applies the use of clinical cases from a real-world setting, a novelty which better reflects the realities of practice.
- The article also adds to the existing literature regarding the impact of experience on clinical practice.

Introduction

A central goal of medical training is that qualified consultants should practice to the same gold standard, dictated by the speciality curriculum. In an ideal world, practice of consultants should be uniform. In practice, there will inevitably be differences stemming from multiple factors, including the complexity and nuance of medical disorders and presentations, differences in training and exposure to conditions during training, subspecialization interests, personal philosophy in terms of investigative approach (based on managing clinical risk and access to tests), and duration of clinical experience.

Perhaps surprisingly, there are relatively few research studies focusing on inter-rater variation in consultant practice in neurology, or any other specialty, especially in a real-world outpatient setting, in which uncertainty, balancing risks and benefits, and making clinical judgements, often in the face of incomplete information, are characteristic. Previous studies have tended to focus on approach to single symptoms or diseases, such as classification of dysarthria from speech recordings[1], management of Alzheimer's disease and vascular dementia[2], or skills in eliciting neurological signs[3]. Aspects such as clinical experience and diagnostic confidence have been studied; greater experience was associated with greater diagnostic confidence in patients with medically unexplained symptoms[4]. However, in real-world neurological practice, patients present with heterogeneous and undifferentiated problems, with varying quantity and quality accompanying data available. Consultants may rely, to a greater or lesser degree, on their clinical assessment, supported by investigations, the latter an expensive and finite resource. At the end of the clinical encounter, regardless of approach, the goal is that the patient receives an accurate diagnosis.

The primary aim of this study was to evaluate systematically similarities and differences in practice between consultant neurologists with differing levels of clinical experience, applied to a large series of consecutive patients seen in the general neurology outpatient setting by a newly qualified consultant neurologist, both in terms of diagnosis and investigative approach. The null hypothesis was that diagnosis, and investigations, would be uniform across participants. The alternative hypothesis was that differences in management approach would emerge, which could then be explored for practical utility as a potential tool for appraisal.

Methods

Ethical approval was obtained (University of Sheffield 022646). Four consultants working within the neurology department of Sheffield Teaching Hospitals NHS Foundation Trust with 0, 6, 16, and 23 years' experience participated in the study.

The first 200 consecutive cases seen by the first consultant immediately after qualification were anonymized to include only the history, examination findings, and any investigation results already available at the time of initial assessment. As this study was exploratory, a sample size calculation was not performed, and was based on previous studies[3]. To assess validity of the cases sampled, and place subsequent results into clinical context, cases were divided into disease groups mapped to the UK national neurology curriculum[5], based on primary presenting complaint. Each case was included in only one disease group. In cases where multiple symptoms were present, the symptom deemed most relevant to the clinical question was chosen. The number of major topics in the UK Neurology Curriculum[5] covered by each disease group was compared to the percentage of cases in that group observed in practice.

Consultants were asked to assess each of the real-world cases as if they had seen the patient in clinic themselves using a standardised questionnaire (Appendix 1). Briefly, they were asked to provide a single primary diagnosis, a differential diagnosis if relevant, and to rate their diagnostic certainty on a 1-10 Likert scale (1=Completely uncertain, 10=Completely certain). They were asked whether they would arrange any investigations and, if so, to specify them. Diagnostic peer-agreement was defined on an ordinal scale of 4 to 1 for each case as follows: 4= concordant diagnoses between all four consultants; 3= concordance between

three consultants; 2= concordance between two consultants; 1=complete discordance. Similarly, for each case, each individual consultant was assigned a diagnostic agreement score: 3=all three colleagues agree with their diagnosis; 2=two other consultants agree; 1=one other consultant agrees, 0= no other consultants agree.

Essentially identical diagnoses with minor differences in wording, judged clinically unimportant by the two principal authors (CA, TJ), were categorized as agreement. At the end of the study, the importance of any variance in diagnoses for each case was assessed by the consultant in charge of clinical care on a six-point Likert scale, taking into account subsequent investigation results and clinical follow-up data when available, and making an overall judgement as follows: 1= Discordant diagnoses could result in a very important difference in management with potentially severe consequences e.g. in patient with a headache, migraine versus glioblastoma; 2=A difference in management could result with potentially serious consequences e.g. epilepsy versus dissociative non-epileptic attacks; 3=A difference in management could result but with less serious potential consequence e.g. migraine versus tension headache; 4=A difference in management could result but with less immediate management consequences e.g. Parkinson's disease dementia versus Lewy body dementia; 5=A difference in wording but no meaningful difference in management e.g. tremor secondary to basal ganglia haemorrhage vs vascular parkinsonism; 6=Complete concordance. Percentages in each category were reported. Categories 1 and 2 were combined to define cases with important potential differences in clinical management. Data were reported for the cohort as a whole and by disease group.

Mean diagnostic certainty was calculated, with standard deviations, for each consultant for each disease group. Logistic regression models were applied using IBM SPSS Statistics for Windows, version 25 (IBM Corp, Ill., USA) to assess associations between diagnostic certainty, agreement and whether investigations were requested. The following metrics were reported for each individual rater: years' experience; number and percentage of cases investigations were arranged; mean investigations arranged per case; mean and median diagnostic agreement (with standard deviations); and Spearman's correlation coefficients between diagnostic agreement and diagnostic certainty (as a surrogate measure of accuracy of judgement of case difficulty). Ordinal logistic regression models were performed *post-hoc* to further explore associations between diagnostic agreement and diagnostic certainty, with diagnostic agreement entered as the dependent variable. In this model, a negative regression coefficient indicates that an increase in the independent variable (diagnostic certainty) is associated with an increase in the dependent variable (diagnostic agreement). Binary logistic regression models were performed to explore associations between the decision to arrange investigations and diagnostic certainty, with decision to investigate (yes/no) specified as the categorical dependent variable. In this model, a negative regression coefficient indicates that an increase in the independent variable (diagnostic certainty) is associated with lower odds of arranging investigations (i.e. dependent variable score=0 rather than 1). Odds ratios with associated 95% confidence intervals and p values were reported. $P<0.05$ was considered statistically significant.

Results

Case-mix

The percentage of cases within each disease group seen in practice compared to UK curriculum representation is reported in Table 1. Four curriculum areas were not represented in the cases seen (tumours, infections, cerebrospinal fluid disorders and toxic/metabolic states). The curriculum areas included for each disease group are reported in Appendix 2.

Diagnostic peer-agreement

Diagnostic peer-agreement by disease area is reported in Figure 1. All four consultants agreed on diagnosis in 99/200 cases. Three consultants agreed in 56/200 cases and two in 39/200 cases. There were 6 cases of complete disagreement: a case of ill-defined widespread sensory-motor disturbance (diagnoses of enhanced physiological tremor/cramp-fasciculation syndrome, uncertain, possible cervical spinal pathology and health anxiety were offered); an atypical blackout (diagnoses: uncertain, transient ischemic attack, focal seizure, fugue state); a case of non-specific dizziness (uncertain/exclude multiple sclerosis, possible cerebellar

ischemic inflammatory/neoplastic pathology, possible vestibular dysequilibrium, functional disorder); and three headache disorders: (1: atypical migraine, stress-related headache, atypical cluster headache, tension headache; 2: chronic pain/neurological cause unlikely, uncertain, mastoiditis, somatisation disorder; 3: idiopathic stabbing headache; primary headache; tension headache; migraine).

Importance of diagnostic differences

Diagnostic differences were judged potentially important (category 1 or 2) in 34 (17%) cases overall (4 (2%) category 1; (30 15%) category 2); 55 (27.5%) category 3, 21 (10.5%) category 4; 41 (20.5%) category 5 and 49 (24.5%) category 6 (complete concordance). Importance of differences by disease area are reported in Figure 2. No adverse events occurred; all potentially serious differential diagnoses provided were excluded by appropriate investigation and follow-up. The four cases in category 1 were: post-concussive headache with differential diagnosis of chronic subdural haemorrhage offered (imaging was subsequently normal); headaches with differential diagnoses of migraine, and posterior fossa structural pathology (imaging was normal); lower limb sensory symptoms with differential diagnoses of a functional disorder, intracranial structural pathology and lumbar radiculopathy (imaging and nerve conduction studies were normal); and sensory symptoms and bowel/bladder symptoms following a fall, with differential diagnoses of cervical myelopathy and subdural haemorrhage; imaging showed cervical spinal cord compression resulting in neurosurgical referral.

Associations between experience, diagnostic certainty, diagnostic agreement and investigative approach

These data are reported for each individual rater in Table 2. Mean diagnostic certainty increased numerically with years' experience but the number of raters was too small for statistical analysis. For each of the four raters, statistically significant associations were identified between increasing diagnostic certainty and greater diagnostic agreement. A one-unit increase in diagnostic certainty resulted in the odds of agreement with 0 colleagues rather than 1,2 or 3 reducing by 44%, 29%, 43% and 24%, respectively for the four raters.

For each of the four raters, significant associations were identified between increasing diagnostic certainty and decreased likelihood of arranging investigations. A one-unit increase in diagnostic certainty resulted in a reduction in the odds that a consultant would arrange any investigations of 22%, 29%, 44% and 42%, respectively for the four raters.

Discussion

The key results are: (1) there are differences in consultant neurologist diagnoses and investigative approach in the general neurology outpatient setting; (2) whilst differences in diagnosis are generally minor, a significant proportion have potentially important management implications; (3) greater diagnostic certainty is associated with higher diagnostic agreement and a lower likelihood of arranging investigations; and (4) individualized metrics can be produced representing different facets of clinical practice. Consideration could be given to developing this system of peer comparison as a tool to facilitate reflective practice at appraisal.

The case-mix encountered in this study by a newly qualified consultant neurologist, working in a district general hospital, appears representative of typical outpatient practice and current UK curriculum content, with headache, sensorimotor disorders, blackouts and tremor common problems. This differentiates our study from most others published to date, in which unselected cases have been studied infrequently. Previous literature on diagnostic consensus and investigative approach has focused on highly specific issues. One study[1] assessed classification of dysarthria by 72 neurologists from patients' speech recordings and found low agreement (Cohen's kappa 0.16-0.32). However, a different study[6]on clinical assessment of stroke found good inter-observer agreement in diagnosis (91/98 cases, Cohen's kappa=0.77). These conflicting data suggest that disease group may be an important factor, and this was supported by our study, in which balance disorders exhibited lowest diagnostic agreement, followed by headache, sensory/motor disorders, disorders of consciousness and tremor. Possible reasons for varying inter-rater agreement between disease groups include greater heterogeneity of disorders, fewer formalized diagnostic protocols, or more overlap between different disorders with similar phenotypes. For example, overlap in the clinical features of migraine and

tension headaches exists, with 58.4% of definite migraine patients having tension-type symptoms and 68.1% of tension-headache patients reporting migraine-type symptoms in one study[7]. Furthermore, in this context, the primary aim of a consultant neurologist may be to rule out secondary headaches. Despite not always agreeing on the primary headache disorder, there were only three cases in which one consultant suspected a secondary headache disorder in contradiction to the others and none were found when investigations were performed, implying agreement in decisions with implications for patient safety. Previous research has also highlighted the difficulty in diagnosing balance disorders; the sensation can be difficult to define, with numerous possible causes, overlapping phenotypes and frequently inconclusive investigations[8]. Previous studies[9]have found that a firm diagnosis may not be attainable in up to 20% of cases.

Relationships between diagnostic confidence and experience have been investigated in previous studies. In medically unexplained symptoms, greater clinician experience was associated with increased confidence in diagnosis amongst a mixed group of neurologists, cardiologists, gastroenterologists and rheumatologists[4]. A comparison of US and UK neurologists[10] using a similar case-based methodology to our study, found that more experienced neurologists ordered fewer investigations. This study had a far greater number of participants (705), but far fewer number of cases (3) than ours, so the two studies provide complementary data on this issue. In contrast, the GALATEA trial, an observational study into practice variability in Alzheimer's and cerebrovascular disease, found no relationship between experience and investigative decisions across 107 physicians[2]. Whilst we lacked statistical power to investigate relationships between experience and diagnostic confidence in our study, we were able to evaluate associations between diagnostic certainty, diagnostic peer-agreement and investigative approach, which are novel data, and we consider may have value as an appraisal tool. All UK clinicians undertake annual review and 5-yearly revalidation procedures to monitor and maintain competencies. Individualized metrics allow comparison of diagnostic consensus with peers, and could capture more complex aspects of clinical practice. For example, the strength of positive association between diagnostic certainty and diagnostic agreement may reflect aspects of diagnostic judgement, and associations between diagnostic certainty and investigative approach may indicate an individual's reliance on their clinical judgement, approach to resource usage and management of uncertainty. Our data are supported by a previous qualitative study[11] of general practitioners that found diagnostic uncertainty to be a positive determinant for ordering tests in 22 physicians surveyed. Neurologists in our study were sometimes uncertain. There is literature on teaching medical students the management of uncertainty[12] in an age when perhaps patients, and potentially doctors too, may trust in technology more than clinical judgement. Diagnostic uncertainty can result in over-investigation[13], which carries a risk of incidental findings that can be detrimental. Management of uncertainty may become especially important at transition to consultant level, with increased responsibility, and has not traditionally been part of medical school curricula. Further research into the influence of uncertainty on clinicians' decision-making may help develop tools to facilitate and support recognition, management, discussion and documentation of this normal aspect of clinical practice. This may be especially important in the current era of Covid-19, when face-to-face contacts and access to investigations are limited, potentially increasing diagnostic uncertainty.

A limitation of our study was that follow-up data, to confirm veracity of the diagnoses made, was often not available, and we recognize that diagnostic peer-agreement is not synonymous with correct diagnosis. However, this limitation accurately reflects the real-world general neurology clinic setting in which the study was conducted; many diagnoses in neurology are clinical, follow-up and investigations are not always necessary or conclusive and, conversely, a strength of our study is the external validity of investigating unselected consecutive cases. A further limitation is that, whilst our methodology was designed to simulate real-life practice, the raters not present in clinic with the patients did not have the benefit of taking the history and performing neurological examination themselves, and may have picked up additional signs, or non-verbal cues and clues, which can be important for diagnosis[14]. This potential bias between theoretical and actual practice could be investigated by two or more raters assessing the same patients and performing a reciprocal analysis, with the factor of interest management differences between consultants that met the patient or theoretically answered the questionnaire.

Despite these limitations, we propose that this methodology has potential as an appraisal tool, with the

advantages of producing quantitative data derived from a directly relevant clinical setting and enabling anonymized peer comparison of complex aspects of practice. Lack of standardized assessments[15], time burden[16] and subjectivity of current assessment[15] have been identified as problems with current appraisal tools. Only 43% of physicians reported changing their practice in response to their appraisal in one study[17], with another finding that 43% of physicians felt that patient safety had not been improved by revalidation[18]. New approaches based on real-world practice might help address some of these areas of need, although would have to be adapted to be feasible in a non-research setting. Collating larger datasets on diagnostic and investigative approaches may also help plan neurological services by helping direct resources appropriately and enabling feedback loops to be developed between resource providers and clinical practitioners.

In summary, there is variability in consultant neurologist practice, which may influence management and can be captured by metrics reflecting individual approaches. With refinement and consideration of the limitations listed above, the methodology applied in this study may be developed into new tools to facilitate reflection and appraisal, and maintain the current high standards of neurological practice.

References

1. Fonville S, Worp H, Maat P, Aldenhoven M, Algra A, Gijn J. Accuracy and inter-observer variation in the classification of dysarthria from speech recordings. Journal of Neurology. 2008;255:1545-1548.
2. Gil P, Ayuso J, Marey J, Antón M, Quilo C. Variability in the Diagnosis and Management of Patients with Alzheimer's Disease and Cerebrovascular Disease. Clinical Drug Investigation. 2008;28:429-437.
3. Hansen M, Sindrup S, Christensen P, Olsen N, Kristensen O, Friis M. Interobserver variation in the evaluation of neurological signs: observer-dependent factors. Acta Neurologica Scandinavica. 2009;90:145-149.
4. Warner A, Walters K, Lamahewa K, Buszewicz M. How do hospital doctors manage patients with medically unexplained symptoms: a qualitative study of physicians. Journal of the Royal Society of Medicine. 2017;110:65-72.
5. GMC 2019. SPECIALTY TRAINING CURRICULUM FOR NEUROLOGY [Internet]. Gmc-uk.org. 2010 [cited 22 June 2019]. Available from: https://www.gmc-uk.org/-/media/documents/2010-neurology–amendments-2013–admin-changes-170215_pdf-59818373.pdf.
6. Hand P, Haisma J, Kwan J, Lindley R, Lamont B, Dennis M et al. Interobserver Agreement for the Bedside Clinical Assessment of Suspected Stroke. Stroke. 2006;37:776-780.
7. Turkdogan D, Cagirici S, Soylemez D, Sur H, Bilge C, Turk U. Characteristic and Overlapping Features of Migraine and Tension-Type Headache. Headache: The Journal of Head and Face Pain. 2006;46:461-468.
8. Sloane P. Dizziness: State of the Science. Annals of Internal Medicine. 2001;134:823.
9. Post R., Dickerson L. Dizziness: A Diagnostic Approach. American Family Physician. 2010 Aug 15;82:361-368.
10. Vickrey B, Gifford D, Belin T, et al. Practice styles of US compared to UK neurologists. Neurology. 1998;50:1661-1668.
11. Weijden T, van Bokhoven M, Dinant G, van Hasselt C, Grol R. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. British Journal of General Practice. 2002;52:974-980.
12. Kim, K. and Lee, Y., 2018. Understanding uncertainty in medicine: concepts and implications in medical education. Korean Journal of Medical Education, 30, pp.181-188.
13. Kassirer, J. and Kopelman, R., 1987. Tolerating Uncertainty. Hospital Practice, 22, pp.21-28.
14. Plug L, Sharrack B, Reuber M. Conversation analysis can help to distinguish between epilepsy and non-epileptic seizure disorders: A case comparison. Seizure. 2009;18:43-50.
15. Griffin A, Furmedge D, Gill D et al. Quality and impact of appraisal for revalidation: the perceptions of London's responsible officers and their appraisers. BMC Medical Education. 2015;15.
16. Nath V, Seale B, Kaur M. Medical revalidation: From compliance to commitment. Mar 2014.
17. Archer J, Cameron N, Lewis M et al. UMbRELLA: Evaluating the regulatory impact of medical revalidation. Feb 2018.

18. Pearson K. Taking revalidation forward: Improving the process of relicensing for doctors. Jan 2017.

Tables

Table 1- Case-mix seen in the study (i.e. in the first 200 consecutive patients seen by a newly qualified neurology consultant in a general neurology clinic), categorized by disease group and the corresponding number of major neurology curriculum areas within disease group. Place of study: Sheffield Teaching Hospitals NHS Foundation Trust, UK

Year of study: 2019

Table 2- Data for four-way analysis between diagnostic decisions of 4 raters (clinical neurologists) assessing 200 real-world clinical case studies in a general neurology clinic. Association, and strength of the association, between diagnostic certainty and diagnostic agreement and the decision to investigate for each consultant.

Place of study: Sheffield Teaching Hospitals NHS Foundation Trust, UK

Year of study: 2019

Figures

Figure 1

Title**-** Inter-rater diagnostic agreement

Legend- Analysis of inter-rater diagnostic agreement for each consultant in primary diagnosis across 200 cases (the first 200 consecutive cases seen by a newly qualified neurology consultant in Sheffield Teaching Hospitals, NHS Foundation Trust, UK). Inter-rater diagnostic agreement has been divided into disease area. 4=Complete diagnostic concordance across all four consultants. 3=three consultants agree; 2=two consultants agree; 1-= complete discordance.

Place of study: Sheffield Teaching Hospitals NHS Foundation Trust, UK

Year of study: 2019

Figure 2

Title**-** The importance of diagnostic disagreements for patient management

Legend- Analysis of the importance of any diagnostic disagreements for patient management between four consultants. The consultants provided their diagnosis for the first 200 consecutive cases seen by a newly qualified consultant in Sheffield Teaching Hospitals, NHS foundation Trust, UK. The potential importance of any diagnostic differences was assigned a score by the evaluating neurologist (6-point Likert scale). (1)-Major disagreement with potentially life-threatening impact, to (6)-No disagreement; complete concordance. Data are shown overall and also by disease group.

Place of study: Sheffield Teaching Hospitals NHS Foundation Trust, UK

Year of study: 2019

Appendix

Appendix 1- Questionnaire sent out to all participants for completion, as if they had seen the patient in clinic themselves, for the first 200 consecutive patients seen by a newly qualified consultant in a General neurology clinic at Sheffield Teaching Hospitals.

Appendix 2- Major GMC Specialty Training for Neurology Curriculum Topics covered by each disease group

Hosted file

`Table 1.xlsx` available at https://authorea.com/users/340929/articles/467959-diagnostic-investigative-approach-of-consultant-neurologists-in-a-real-world-clinical-setting

Hosted file

`Table 2.xlsx` available at https://authorea.com/users/340929/articles/467959-diagnostic-investigative-approach-of-consultant-neurologists-in-a-real-world-clinical-setting