

Original Research Article

Number Needed to Diagnose, Predict, or Misdiagnose: Useful Metrics for Non-Canonical Signs of Cognitive Status?

A.J. Larner

Cognitive Function Clinic, Walton Centre for Neurology and Neurosurgery, Liverpool, UK

Keywords

Dementia · Diagnosis · Mild cognitive impairment · Neurological signs · Number needed

Abstract

Background/Aims: “Number needed to” metrics may hold more intuitive appeal for clinicians than standard diagnostic accuracy measures. The aim of this study was to calculate “number needed to diagnose” (NND), “number needed to predict” (NNP), and “number needed to misdiagnose” (NNM) for neurological signs of possible value in assessing cognitive status. **Methods:** Data sets from pragmatic diagnostic accuracy studies examining easily observed and dichotomised neurological signs (“attended alone” sign, “attended with” sign, head turning sign, applause sign, *la maladie du petit papier*) were analysed to calculate the NND, NNP, and NNM. **Results:** All measures of discrimination showed broad ranges. The range of NND and NNP suggested that these signs were, with a single exception, of value for correctly diagnosing or predicting cognitive status (presence or absence of cognitive impairment) when between 2 and 4 patients were examined. However, NNM showed similar values (range 1–5 patients) suggesting risk of misdiagnosis. **Conclusion:** NND, NNP, and NNM may be useful, intuitive, metrics in assessing the utility of diagnostic tests in day-to-day clinical practice. A ratio of NNM to either NND or NNP, termed the likelihood to diagnose or misdiagnose, may clarify the utility or inutility of diagnostic tests.

© 2018 The Author(s)

Published by S. Karger AG, Basel

A.J. Larner
Cognitive Function Clinic
Walton Centre for Neurology and Neurosurgery
Lower Lane, Fazakerley, Liverpool L9 7LJ (UK)
E-Mail a.larner@thewaltoncentre.nhs.uk

Introduction

Many measures of discrimination have been used to describe the utility of diagnostic tests [1, 2]. Most usually, diagnostic test accuracy studies report paired values of test sensitivity and specificity and positive and negative predictive values (PPV, NPV). Other, single, global or unitary, indicators of test diagnostic performance have also been described, including:

- correct classification accuracy, the total number of true positives and true negatives divided by the total number of patients assessed, and inaccuracy, the total number of false positives and false negatives divided by the total number of patients assessed ($= 1 - \text{accuracy}$);
- Youden index (Y), a combination of sensitivity and specificity, given by $(\text{sensitivity} + \text{specificity} - 1)$ [3];
- predictive summary index (PSI, or Ψ), a combination of positive and negative predictive values given by $(\text{PPV} + \text{NPV} - 1)$ [4].

All these parameters have values ranging between 0 and 1, sometimes expressed as percentages. It may be difficult for clinicians to relate these numeric outcomes to individual patients in day-to-day clinical practice.

Cook and Sackett [5] introduced the “number needed to treat” (NNT) metric as a way to represent the “impact” of treatments. This measure is arguably more intuitive to clinicians and patients than more traditional measures of discrimination. Adaptations of NNT have been described (e.g., “number needed to harm” (NNH) [6]; “number needed to see” (NNS) [7]). Analogous adaptations may be relevant to diagnostic test accuracy studies.

The inverse of the Youden index ($1/Y$) has been defined as the “number needed to diagnose” (NND), that is, the number of patients who need to be examined in order to correctly detect one person with the disease of interest in a study population of persons with and without the known disease [4]. For diagnostic tests, low values of NND will be desirable.

Linn and Grunau [4] also suggested a new statistic, the inverse of PSI ($1/\Psi$ or $1/\Psi$), which they termed the “number needed to predict” (NNP), interpreted as the number of patients who need to be examined in the patient population in order to correctly predict the diagnosis of one person. Whilst NND is insensitive to variation in disease prevalence, since it depends entirely on sensitivity and specificity, NNP is dependent on prevalence and may therefore be deemed a better descriptor of diagnostic tests in patient populations with different prevalence of disease [4]. For diagnostic tests, low values of NNP will be desirable.

Habibzadeh and Yadollahie [8] have proposed another index, the “number needed to misdiagnose” (NNM), as a measure of diagnostic test effectiveness, defined as the inverse of $(1 - \text{accuracy}) = 1/\text{inaccuracy}$. NNM is the number of patients who need to be tested in order for one to be misdiagnosed by the test. For diagnostic tests, high values of NNM will be desirable.

A number of simple, non-canonical, neurological signs of possible value in the diagnosis of cognitive status have been described, whose utility is based in part on their being easily observed and categorised as present or absent: the “attended alone” sign and its converse the “attended with” sign, the head turning sign, the applause sign, and *la maladie du petit papier* [9]. The aim of the current study was to reanalyse data sets from diagnostic test accuracy studies of these signs in order to calculate and compare the parameters NND, NNP, and NNM.

Methods

Data from pragmatic prospective diagnostic accuracy studies undertaken in a dedicated cognitive disorders clinic, located in a secondary care setting (regional neuroscience centre) and using a standardised methodology [1, 9], were analysed.

These studies examined the following non-canonical neurological signs:

- the attended alone sign [10]: defined as the patient attending the clinic appointment without a knowledgeable informant, despite prior provision of written instructions to do so;
- the attended with sign [11]: the converse of the attended alone sign, the patient attending the clinic appointment with an informant in accordance with prior provision of written instructions to do so;
- the head turning sign [11–13]: the patient turning her/his head towards an accompanying informant when asked open questions about memory symptoms during the history taking phase of the clinical assessment;
- the applause sign [14]: in the clinical examination phase of the assessment the patient is asked to clap hands three times, and responds with more than three claps;
- *la maladie du petit papier* [15, 16]: the patient presents a self-written list of symptoms (on paper or iPad) during the clinical assessment.

All these signs are easily observed and dichotomised as present/absent. The attended alone sign and *la maladie du petit papier* have been suggested to indicate absence of cognitive impairment, whereas the attended with, head turning, and applause signs have been suggested to indicate the presence of cognitive impairment [9].

Data from these studies, pooled where appropriate, were used to calculate the following parameters: sensitivity and specificity, Youden index [3], and NND [4]; PPV and NPV, PSI [4], and NNP [4]; accuracy, inaccuracy, and NNM [8].

Reference standard diagnoses were dementia, mild cognitive impairment, or subjective memory complaint, by judgment of an experienced clinician based on standard diagnostic criteria for dementia (DSM-IV) and mild cognitive impairment (Petersen) [9].

A further, novel, metric was also derived, the “likelihood to be diagnosed or misdiagnosed” (LDM). Analogous to the previously described “likelihood to be helped or harmed” (LHH) metric, calculated as the ratio of NNH to NNT [6], LDM is given by the ratio of NNM to either NND or NNP. Since for diagnostic tests low values of NND and NNP and high values of NNM are desirable, higher values of LDM (>1) would suggest a test more likely to diagnose than misdiagnose.

Prevalence (P) of cognitive impairment for each study was calculated as the number of patients receiving a criterion diagnosis of dementia or mild cognitive impairment (true positives and false negatives) divided by the total number of patients assessed. Level of the test (Q) was calculated as the number of patients with a positive test in the population studied (true positives and false positives) divided by the total number of patients assessed.

All studies followed either STARD [17] or STARDdem [18] guidelines, depending on the exact time at which they were undertaken. In all studies subjects gave informed consent and study protocols were approved by the institute’s committee on human research.

Results

A summary of the different studies (Table 1) showed a broadly similar prevalence of patients with cognitive impairment (range 0.32–0.63), the outlier being the study of the head turning sign which logically required exclusion of those who attended alone. Level of the test showed a broad range, from low frequency (*la maladie du petit papier* = 0.05) to high frequency (attended with = 0.66).

The sensitivity and specificity of the different signs varied (Table 2), from very sensitive (attended alone for diagnosis of no cognitive impairment = 0.93, or no dementia = 1.00; attended with for diagnosis of any cognitive impairment = 0.93) to insensitive (*la maladie du*

Table 1. Study demographics

Sign	N	P	Q	Ref.
Attended alone	726	0.32	0.34	10
Attended with	726	0.32	0.66	11
Head turning	246	0.63	0.43	11–13
Applause	275	0.45	0.22	14
<i>La maladie du petit papier</i>	258	0.41	0.05	15, 16

P, prevalence of any cognitive impairment = (TP + FN)/N; Q, level of the test = (TP + FP)/N; TP, true positive; FN, false negative; FP, false positive.

Table 2. Number needed to diagnose by sign

Sign	Diagnosis	Sensitivity	Specificity	Y	NND = 1/Y
Attended alone	No cognitive impairment	0.93	0.45	0.38	2.63
Attended alone	No dementia	1.00	0.45	0.45	2.22
Attended with	Any cognitive impairment	0.93	0.47	0.40	2.50
Head turning	Any cognitive impairment	0.65	0.95	0.60	1.67
Applause	Any cognitive impairment	0.36	0.89	0.25	4.00
Applause	Dementia	0.54	0.85	0.39	2.56
<i>La maladie du petit papier</i>	No cognitive impairment	0.07	0.98	0.05	20.0

Y, Youden index; NND, number needed to diagnose.

Table 3. Number needed to predict by sign

Sign	Diagnosis	PPV	NPV	PSI	NNP = 1/PSI
Attended alone	No cognitive impairment	0.47	0.93	0.40	2.50
Attended alone	No dementia	0.48	1.00	0.48	2.08
Attended with	Any cognitive impairment	0.45	0.93	0.38	2.63
Head turning	Any cognitive impairment	0.95	0.61	0.56	1.79
Applause	Any cognitive impairment	0.72	0.63	0.35	2.86
Applause	Dementia	0.46	0.89	0.35	2.86
<i>La maladie du petit papier</i>	No cognitive impairment	0.85	0.43	0.28	3.57

PPV, positive predictive value; NPV, negative predictive value; PSI, predictive summary index; NNP, number needed to predict.

petit papier for diagnosis of no cognitive impairment = 0.05). The expected trade-off between sensitivity and specificity was observed, with less sensitive signs being more specific. A range of values for the Youden index was observed (0.05–0.60), and hence also for NND (1/Y), ranging from 1.67 (head turning sign for any cognitive impairment) to 20 (*la maladie du petit papier* for no cognitive impairment).

The PPV and NPV of the different signs varied (Table 3), with a PPV range of 0.45–0.95 and NPV range of 0.43–1.00. A range of values for PSI was observed (0.28–0.56) and hence for NNP (1/PSI), ranging from 1.79 (head turning sign for any cognitive impairment) to 3.57 (*la maladie du petit papier* for no cognitive impairment).

Table 4. Number needed to misdiagnose by sign

Sign	Diagnosis	Acc	Inacc	NNM = 1/Inacc
Attended alone	No cognitive impairment	0.61	0.39	2.56
Attended alone	No dementia	0.64	0.36	2.78
Attended with	Any cognitive impairment	0.61	0.39	2.56
Head turning	Any cognitive impairment	0.76	0.24	4.17
Applause	Any cognitive impairment	0.65	0.35	2.86
Applause	Dementia	0.79	0.21	4.76
<i>La maladie du petit papier</i>	No cognitive impairment	0.45	0.55	1.82

Acc, correct classification accuracy; Inacc, inaccuracy; NNM, number needed to misdiagnose.

Table 5. Likelihood to be diagnosed or misdiagnosed by sign

Sign	Diagnosis	LDM = NNM/NND	LDM = NNM/NNP
Attended alone	No cognitive impairment	0.97	1.02
Attended alone	No dementia	1.25	1.37
Attended with	Any cognitive impairment	1.02	0.97
Head turning	Any cognitive impairment	2.50	2.33
Applause	Any cognitive impairment	0.72	1.00
Applause	Dementia	1.86	1.66
<i>La maladie du petit papier</i>	No cognitive impairment	0.09	0.51

LDM, likelihood to be misdiagnosed; NNM, number needed to misdiagnose; NND, number needed to diagnose; NNP, number needed to predict.

Despite the spread of values for sensitivity and specificity, PPV and NPV, Y, and PSI, the values for NND and NNP were, with a single exception, ≤ 4 (Tables 2, 3, right-hand column).

The accuracy (range 0.45–0.79) and inaccuracy (range 0.21–0.55) of the different signs varied (Table 4), with NNM ranging from 1.82 (*la maladie du petit papier* for no cognitive impairment) to 4.76 (applause sign for dementia).

Values for the LDM (Table 5) were high for some signs (>1), suggesting balance in favour of diagnosis over misdiagnosis (e.g., head turning sign for any cognitive impairment), and low (<1) for others, suggesting balance in favour of misdiagnosis over diagnosis (e.g., *la maladie du petit papier* for no cognitive impairment).

Discussion

Clinicians generally think in terms of patients, rather than probabilities. Thus, “number needed to” parameters may hold particular intuitive appeal for clinicians. To the author’s knowledge, this study represents a first attempt to characterise neurological signs in terms of the number needed to diagnose, predict, and misdiagnose metrics suggested by Linn and Grunau [4] and Habibzadeh and Yadollahie [8].

Values for NNP for all the signs examined suggested that between 2 and 4 patients need to be examined in the patient population for correct prediction of either the diagnosis of cognitive impairment in someone with a positive test result or absence of cognitive impairment

in someone with a negative test result. These numbers suggest that these signs may be of clinical use in day-to-day practice, an observation which might influence clinician uptake.

Conversely, values for NNM suggested that similar numbers, between 2 and 5 patients, need to be examined in order for one to be misdiagnosed by the test. Generally, tests with low NND or NNP had higher NNM (e.g., head turning sign for diagnosis of any cognitive impairment) whilst those with high NND or NNP had low NNM (e.g., *la maladie du petit papier* for diagnosis of no cognitive impairment). Other signs had similar values for NND, NNP, and NNM (e.g., attended lone, attended with, applause).

The study has a number of limitations. All index studies were undertaken in the same clinic, with the risks of patient-based (selection, spectrum) and test performance biases [1], and all were cross-sectional studies with risk of diagnostic error. Studies of these signs in settings with different disease prevalence (e.g., primary care, community) would be of interest, and with follow-up for delayed verification of diagnosis.

The neurological signs examined are non-canonical, and currently not widely used (with the possible exception of the applause sign, particularly in the context of movement disorder clinics), although potentially widely applicable, since they are quick to perform, cost free, and easily interpreted and categorised. Some validation studies in independent patient cohorts have been reported for some of these signs [19, 20], but studies of possible relationships to disease biomarkers are in their infancy [21]. The signs examined are easily dichotomised, thus facilitating calculation of NND, NNP, and NNM, which may not be the case for cognitive screening instruments which require the application of test cut-offs [22]. Nevertheless, calculation of these metrics may help clinicians to decide on the possible value of specific signs and tests in the clinical setting.

The utility or inutility of these “numbers needed to” parameters will, as for measures of discrimination, depend on the clinician’s purpose in doing the test. If the clinician wishes to identify all cases (no false negatives), a highly sensitive test with low NND or NNP, with consequent risk of false positives, may be acceptable despite low NNM. If the clinician’s purpose is to exclude all non-cases (false positives), for example in a treatment trial, a low NNM may outweigh low NND or NNP. LDM may give a more global measure of diagnostic gain.

Disclosure Statement

The author declares no conflicts of interest.

References

- 1 Larner AJ. [Diagnostic test accuracy studies in dementia. A pragmatic approach](#). London: Springer; 2015. <https://doi.org/10.1007/978-3-31916697-1>.
- 2 Quinn TJ, Takwoingi Y. Assessment of the utility of cognitive screening instruments. In: Larner AJ, editor. [Cognitive screening instruments. A practical approach](#). 2nd ed. London: Springer; 2017. p. 15–34.
- 3 Youden WJ. Index for rating diagnostic tests. [Cancer](#). 1950 Jan;3(1):32–5.
- 4 Linn S, Grunau PD. New patient-oriented summary measure of net total gain in certainty for dichotomous diagnostic tests. [Epidemiol Perspect Innov](#). 2006 Oct;3(1):11.
- 5 Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. [BMJ](#). 1995 Feb;310(6977):452–4.
- 6 Citrome L, Ketter TA. When does a difference make a difference? Interpretation of number needed to treat, number needed to harm, and likelihood to be helped or harmed. [Int J Clin Pract](#). 2013 May;67(5):407–11.
- 7 Larner AJ. [Teleneurology by internet and telephone. A study of medical self-help](#). London: Springer; 2011. <https://doi.org/10.1007/978-0-85729-691-7>.
- 8 Habibzadeh F, Yadollahie M. Number needed to misdiagnose: a measure of diagnostic test effectiveness. [Epidemiology](#). 2013 Jan;24(1):170.

- 9 Larner AJ. [Dementia in clinical practice: a neurological perspective. Pragmatic studies in the cognitive function clinic](#). 3rd ed. London: Springer; 2018. <https://doi.org/10.1007/978-3-319-75259-4>.
- 10 Larner AJ. Screening utility of the “attended alone” sign for subjective memory impairment. [Alzheimer Dis Assoc Disord](#). 2014 Oct-Dec;28(4):364–5.
- 11 Williamson JC, Larner AJ. Attended with and head-turning sign can be clinical markers of cognitive impairment in older adults. [Int Psychogeriatr](#). 2018 Apr;1.
- 12 Larner AJ. Head turning sign: pragmatic utility in clinical diagnosis of cognitive impairment. [J Neurol Neurosurg Psychiatry](#). 2012 Aug;83(8):852–3.
- 13 Ghadiri-Sani M, Larner AJ. Head turning sign for diagnosis of dementia and mild cognitive impairment: a revalidation. [J Neurol Neurosurg Psychiatry](#). 2013;84(11):e2.204.
- 14 Bonello M, Larner AJ. Applause sign: screening utility for dementia and cognitive impairment. [Postgrad Med](#). 2016;128(2):250–3.
- 15 Randall A, Larner AJ. La maladie du petit papier: A sign of functional cognitive disorder? [Int J Geriatr Psychiatry](#). 2018 May;33(5):800.
- 16 Bharambe V, Larner AJ: Functional cognitive disorders: demographic and clinical features contribute to a positive diagnosis. 2018;submitted.
- 17 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al.; Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. [Clin Chem](#). 2003 Jan;49(1):7–18.
- 18 Noel-Storr AH, McCleery JM, Richard E, Ritchie CW, Flicker L, Cullum SJ, et al. Reporting standards for studies of diagnostic test accuracy in dementia: The STARDdem Initiative. [Neurology](#). 2014 Jul;83(4):364–73.
- 19 Soysal P, Usarel C, Ispirli G, Isik AT. Attended With and Head-Turning Sign can be clinical markers of cognitive impairment in older adults. [Int Psychogeriatr](#). 2017 Nov;29(11):1763–9.
- 20 Isik AT, Soysal P, Kaya D, Usarel C. Triple test, a diagnostic observation, can detect cognitive impairment in older adults. [Psychogeriatrics](#). 2018 Mar;18(2):98–105.
- 21 Durães J, Tábuas-Pereira M, Araújo R, Duro D, Baldeiras I, Santiago B, et al. The head turning sign in dementia and mild cognitive impairment: its relationship to cognition, behavior, and cerebrospinal fluid biomarkers. [Dement Geriatr Cogn Disord](#). 2018;46(1-2):42–9.
- 22 Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. [Biochem Med \(Zagreb\)](#). 2016 Oct;26(3):297–307.