Nunn, A. C., Jones, H. E., Morosanu, C. O., Singleton, W. G. B., Williams, M. A., Nagel, S. J., Luciano, M. G., Zwimpfer, T. J., Holubkov, R., Wisoff, J. H., McKhann, II, G. M., Hamilton, M. G., & Edwards, R. J. (2020). Extended Lumbar Drainage in Idiopathic Normal Pressure Hydrocephalus: A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy. *British Journal of Neurosurgery*. https://doi.org/10.1080/02688697.2020.1787948

Extended Lumbar Drainage in Idiopathic Normal Pressure Hydrocephalus: A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy

Adam C. Nunn[a] MRCS, PhD, Hayley E. Jones[b] PhD, Cezar O. Morosanu[a] MD, William G. B. Singleton[a] FRCS(Neuro.Surg), PhD, Michael A. Williams[c,k] MD, Sean J. Nagel[d,k] MD, Mark G. Luciano[e,k] MD, PhD, Thomas J. Zwimpfer[f,k] MD, PhD, Richard Holubkov[g,k] MD, PhD, Jeffrey H. Wisoff[h,k] MD, Guy M. McKhann, II[i,k] MD, Mark G. Hamilton[j,k] MDCM, FRCSC, Richard J. Edwards[a,k,*] FRCS(Neuro.Surg), MD

a Department of Neurosurgery, Southmead Hospital, Bristol, UK
b Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
c Departments of Neurology and Neurological Surgery, University of Washington School of Medicine, Seattle, Washington, USA
d Department of Neurosurgery, Center for Neurological Restoration, Cleveland Clinic, Cleveland, Ohio, USA
e Department of Neurosurgery, The Johns Hopkins School of Medicine, Baltimore, Maryland, USA
f Division of Neurosurgery, Department of Surgery, University of British Columbia, Vancouver, British Columbia, Canada
g Department of Pediatrics, University of Utah, Salt Lake City, Utah, USA
h Division of Pediatric Neurosurgery, Hassenfeld Children's Hospital at NYU Langone Health, New York, USA
i Department of Neurological Surgery, Columbia University School of Medicine, New York, USA
j Department of Clinical Neurosciences, Division of Neurosurgery, University of Calgary School of Medicine, Calgary, Alberta, Canada
k Adult Hydrocephalus Clinical Research Network

***Corresponding Author:-**
Richard Edwards MD
Department of Neurosurgery
Gate 6, Level 3, Brunel Building
Southmead Hospital
Southmead Road
Bristol BS10 5NB
United Kingdom
E-mail: richard.edwards@nbt.nhs.uk

Keywords: Cerebrospinal fluid shunts; extended lumbar drainage; normal pressure hydrocephalus; diagnostic techniques

Word count of main text: 3,292

Running heading: Extended lumbar drainage in iNPH

Biographical Notes:-

Adam Nunn is a Neurosurgery Registrar (ST4) at Southmead Hospital in Bristol.

Hayley Jones is a Senior Lecturer in Medical Statistics at the University of Bristol, with an interest in meta-analysis methods, diagnostic test accuracy and Bayesian analysis.

Cezar Morosanu is a Neurosurgery SHO (Clinical Fellow) at Southmead Hospital.

Will Singleton is a Neurosurgery Registrar (ST8) at Southmead Hospital.

Michael Williams is a Neurologist, Director of the Adult & Transitional Hydrocephalus Program and of the Intracranial Hypertension Program, and Professor of Neurology and Neurological Surgery at University of Washington Medical Center, Seattle.

Sean Nagel is an Assistant Professor of Neurological Surgery at the Cleveland Clinic in Ohio.

Mark Luciano is a Professor of Neurosurgery, Biomedical Engineering and Paediatrics, and Director of the Johns Hopkins Cerebral Fluid Center in Baltimore.

Thomas Zwimpfer is a Neurosurgeon at the Vancouver General Hospital and Professor of Neurosurgery at the University of British Columbia.

Richard Holubkov is an Applied Biostatistician at the University of Utah.

Jeffrey Wisoff is a Neurosurgeon at Hassenfeld Children's Hospital in New York and Professor of Neurosurgery and Paediatrics at the New York University School of Medicine.

Guy McKhann is an Associate Professor of Neurological Surgery at Columbia University Medical Center in New York.

Mark Hamilton is a Professor of Neurosurgery and Director of the Adult Hydrocephalus Program at the Hotchkiss Brain Institute, University of Calgary. He is also President of the Hydrocephalus Society.

Richard Edwards is a Neurosurgeon in Bristol and Chairman of the Cerebrospinal Fluid Disorders Group of the Society of British Neurological Surgeons.

Extended Lumbar Drainage in Idiopathic Normal Pressure Hydrocephalus: A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy

Background: When appropriately selected, a high proportion of patients with suspected idiopathic normal pressure hydrocephalus (iNPH) will respond to cerebrospinal fluid diversion with a shunt. Extended lumbar drainage (ELD) is regarded as the most accurate test for this condition, however, varying estimates of its accuracy are found in the current literature. Here, we review the literature in order to provide summary estimates of sensitivity, specificity, positive- and negative predictive value for this test through meta-analysis of suitably rigorous studies.

Methods: Studies involving a population of NPH patients with predominantly idiopathic aetiology (>80%) in which the intention of the study was to shunt patients regardless of the outcome of ELD were included in the review. Various literature databases were searched to identify diagnostic test accuracy studies addressing ELD in the diagnosis of iNPH. Those studies passing screening and eligibility were assessed using the QUADAS-2 tool and data extracted for bivariate random effects meta-analysis.

Results: Four small studies were identified. They showed disparate results concerning diagnostic test accuracy. The summary estimates for sensitivity and specificity were 94% (CI 41-100%) and 85% (CI 33-100%), respectively. The summary estimates of positive and negative predictive value were both 90% (CIs 65-100% and 48-100%, respectively).

Conclusion: Large, rigorous studies addressing the diagnostic accuracy of ELD are lacking, and little robust evidence exists to support the use of ELD in diagnostic algorithms for iNPH. Therefore, a large cohort study, or ideally an RCT, is needed to determine best practice in selecting patients for shunt surgery.

Keywords: Cerebrospinal fluid shunts; extended lumbar drainage; normal pressure hydrocephalus; diagnostic techniques

Introduction

Idiopathic normal pressure hydrocephalus (iNPH), originally described by Salomon Hakim in 1964,[1] is a progressive neurological condition characterised by ventricular enlargement combined with the clinical triad of gait impairment, cognitive problems and urinary dysfunction. The condition is highly responsive to treatment with a CSF shunt system,[2–5] and the benefit derived from this can persist for many years.[6] Prompt diagnosis and management is crucial, since the beneficial effect of shunt surgery declines with increasing duration of symptoms prior to intervention.[7,8] Idiopathic NPH can be diagnostically challenging due to the non-specific nature of the Hakim triad, which overlaps with many other neurodegenerative disorders. This is also the case for ventricular enlargement, which may be a consequence of cerebral atrophy in patients with cognitive impairment. Limited understanding of the pathophysiological mechanism underlying the condition has hampered attempts to devise diagnostic tests that identify patients who will respond to CSF diversion. The diagnostic test currently favoured involves drainage of a certain volume of CSF from the lumbar subarachnoid space, with clinical assessment before and after (known as ‘extended lumbar drainage’ or ‘ELD’). While successful at predicting patients who will respond to shunt insertion, some authors have suggested that ELD may have a low negative predictive value,[9,10] and hence there is potential for diagnostic algorithms relying upon it to exclude patients from an intervention that might have provided significant benefit.

Here, we have completed a systematic review and meta-analysis of the accuracy of ELD (the index test) as a means of diagnosing iNPH. ‘Definite iNPH’ is defined by the Japanese Society for Neurosurgery as a diagnosis that can only be made retrospectively after response to CSF diversion is demonstrated.[11] Therefore, the reference standard test was defined as improvement following insertion of a CSF shunt system.

Methods

The following question was formulated to guide the review: "How accurate is ELD as a diagnostic test for iNPH in patients referred to a regional hydrocephalus clinic?" In order to answer this question, a set of inclusion/exclusion criteria and a search strategy were designed. A protocol was written in advance of the review, and the project was prospectively registered on the PROSPERO database (ID: CRD42018110518; available at https://www.crd.york.ac.uk/prospero/).

To be included in the review and meta-analysis, a study must: i) have been performed on a population of suspected NPH patients in which either >80% were idiopathic or with iNPH patient data extractable from the report; ii) involve a population of patients undergoing ELD prior to insertion of a ventriculo-peritoneal (VP), ventriculo-atrial (VA) or lumbo-peritoneal (LP) shunt; iii) have been designed such that the intention of the study was to shunt all patients (rather than just those who improved with ELD); iv) allow the extraction of numbers of true positives, false positive, true negative and false negatives, or sensitivity and specificity directly. 'Suspected' NPH must have been defined as involving at least one feature of the Hakim triad combined with imaging findings of ventriculomegaly. Studies in which a 'qualifying' test was applied to a population of suspected NPH patients prior to ELD, such that the pre-test probability of a successful ELD outcome was increased, were excluded.

The following electronic databases were searched on 5th September 2018: Medline; Embase; Cochrane Library; Latin American and Caribbean Health Sciences Literature (LILACS); Aggressive Research Intelligence Facility (ARIF); Database of Abstracts of Reviews of Effects (DARE); and ClinicalTrials.gov. The following keywords were used in the search strategy: 'Normal pressure hydrocephalus', 'normotensive hydrocephalus' and various permutations of 'cerebrospinal' or 'CSF' together with 'lumbar drain' or 'diversion'. Equivalent Spanish terms were used in the search of the LILACS database. No language or date limits were applied. Full details of the search strategies for each database are outlined in the Supplementary Appendix.

Screening of titles/abstracts, and full-text reports, was performed in parallel by two independent assessors using the Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia; available at www.covidence.org). Disagreements between the screening authors were resolved by dialogue, and if this was not possible, a third author held the casting vote. A PRISMA flow chart was created to detail the outcome of the screening process.[12] The QUADAS-2 tool was then used to assess the risk of bias in included studies.[13] Diagnostic test accuracy metrics were extracted independently by two authors and fidelity confirmed, to ensure accurate transcription of data.

Statistical Analysis

Bivariate random effects meta-analysis of sensitivity and specificity was performed in WinBUGS.[14–16] We assumed the following vague prior distributions: Normal(0,100) for means, Half-Normal(0,3) for between-study standard deviations and Uniform(-1,0) for the between-study correlation in sensitivity and specificity. The model was also fitted in a frequentist framework in Stata (StataCorp LLC, College Station, TX) using the 'Metandi' package as a sensitivity analysis.[17]

Summary estimates of sensitivity and specificity were generated with a 95% credible ellipse and a 95% prediction region. These results were used to generate estimates, with uncertainty, of positive predictive value (PPV) and negative predictive value (NPV) across a range of prevalences. A summary Receiver Operating Characteristic (ROC) curve was also generated, drawing on the equivalence of the bivariate model with the hierarchical summary ROC (HSROC) model.[18–20]

Subgroup analyses and/or meta-regression using the following variables were planned if there was sufficient data: criteria for regarding the ELD to have demonstrated improvement, duration of lumbar drainage, volume drained per hour. Sensitivity analyses were also planned if there was a controversial study (about which the screening authors disagreed) or if QUADAS-2 identified one or more studies as being at high risk of bias.

Results

The PRISMA diagram showing flow through the study is presented in Figure 1. Two hundred and seventy-three abstracts were identified for screening after duplicates were removed. Of these, 247 were excluded during abstract screening. Twenty-six reports progressed to full-text screening. Ten were excluded on the basis that the authors' intention was to shunt only the ELD responders. A further 9 studies were excluded because their reports presented insufficient data to allow calculation of sensitivity and/or specificity. The reasons for exclusion of the other 7 are detailed in the PRISMA diagram.

Four small studies (84 patients in total) passed screening and eligibility, and their characteristics are summarised in Table 1. All included studies were of a prospective cohort design and contained predominantly iNPH patients (range 88-100%). In terms of inclusion criteria, all studies required gait impairment as a presenting symptom and ventriculomegaly on CT or MRI. One study (Chen et al.[21]) required both gait and dementia as presenting symptoms. Most (3 out of 4 studies) also required a 'normal' opening pressure on lumbar puncture, and Haan et al.[22] required cisternography evidence of abnormal CSF flow. All the studies excluded patients with severe cortical atrophy (or any cortical atrophy in the case of Chen et al. and Haan et al.). Panagiotopoulos et al.[10] and Walchenbach et al.[9] also excluded patients with minimal gait disturbance in the context of 'severe' dementia. Of note, both Haan et al. and Walchenbach et al. excluded patients who were improving after high volume lumbar puncture.

Most of the included studies conducted ELD for 5 days (range 4-5) and the mean CSF drainage rate was 11.6ml/hr (range 5.3-16.5). The criteria for judging the ELD test to have been successful was gait improvement in Walchenbach et al. and Panagiotopoulos et al.; whereas Chen et al. accepted improvement in any domain of the Hakim triad and Haan et al. required improvement in either cognitive function or urinary function in addition to gait.

The reference standard test was a VP shunt in Panagiotopoulos et al. and Walchenbach et al., a VA shunt in Haan et al., and randomisation to a VP or LP shunt in Chen et al. The only study to use programmable valves was Walchenbach et al. The criteria used to judge the success of the

ELD test and shunting were identical in all cases, except Walchenbach et al., where improvement in gait following shunting had to be confirmed by the patient, their family or the nursing team responsible for the patient's care. Mean follow-up of shunted patients prior to a decision being made regarding shunt outcome was 4.3 months (range 0.25-12).

Risk of bias assessment

The results of the QUADAS-2 assessments are summarised in Figures 2 and 3. Only Chen et al. was deemed to have low risk of bias and low applicability concerns across all domains of the QUADAS-2 tool. The only issue of potential concern with Chen et al. was that not all patients received the same reference standard (they were randomised to a VP or an LP shunt). However, this was not felt to be sufficient to regard the study as high risk in the 'Flow and Timing' domain as the equivalence of these two interventions has been established by a recent trial.[23]

Haan et al. was felt to have unclear risk of bias in the 'Patient Selection' and 'Reference Standard' domains and high applicability concerns in the 'Patient Selection' domain. The risk of bias concerns arose because of the following issues: i) lack of clarity as to whether a consecutive or random method of patient selection was employed or if a more biased strategy was used; ii) the absence of any reference to blinding the assessor of shunt outcome to the result of the ELD index test. The applicability concerns were based on the use of radioisotope cisternography in order to select patients. This test is regarded as relatively insensitive for NPH, and in one study, 55% of patients with a normal cisternogram improved following shunting (compared to 73% with a typical NPH pattern, which was not statistically significant).[24]

Panagiotopoulos et al. was also classified as 'unclear' risk of bias in the 'Patient Selection' domain and 'Reference Standard' domain because of the following issues: i) unclear means of patient selection, and no evidence to suggest a consecutive or random method; ii) no reference to blinding of shunt assessor to ELD outcome.

Finally, Walchenbach et al. was deemed high risk of bias and high applicability concerns in the 'Patient Selection' domain because patients were excluded if they initially responded to a CSF tap test.

Statistical results

The observed sensitivity, specificity, PPV and NPV from each of the studies are presented in Table 2. Sensitivity and specificity derived from these studies are also presented graphically as a coupled forest plot in Figure 4, and in ROC space in Figure 5. Two of the studies (Chen et al. and Haan et al.) report the ELD test as 'perfect', i.e. no false negatives or false positives, however, the sample size was small in both cases (7 and 17 patients, respectively). Panagiotopoulos et al. also reported a high sensitivity (94%) but a significantly lower specificity (40%), whereas Walchenbach et al., the largest included study, suggested the reverse: a low sensitivity (50%) and a moderately high specificity (80%). All studies support a high PPV (88-100%), and two studies also supported a high NPV (100%). However, Panagiotopoulos et al. and Walchenbach et al. suggest a lower value for NPV (67 and 36%, respectively).

The summary estimates of sensitivity and specificity from the meta-analysis were 94% (95% credible interval [CI] 41-100%) and 85% (95% CI 33-100%), respectively (see Table 3; plotted in ROC space in Figure 5). The very wide CIs associated with these estimates reflect the small and heterogeneous nature of the included studies. The 95% prediction region for a new study (Figure 5) encompasses the entire ROC space. The sensitivity analysis performed in Stata gave very similar summary results (sensitivity 95% [CI 36-100%], specificity 86% [CI 35-99%]).

Since the PPV and NPV depend on the true prevalence of NPH within the population who might conceivably receive the test, observed PPV and NPV are plotted against prevalence in Figure 6. The prevalence in the included studies varied from 53-77% which is in line with the senior author's own experience, in which the prevalence of shunt responsive NPH is around 60%. Figure 6 also shows summary estimates of PPV and NPV calculated from the meta-analysis results, across all possible prevalences. At a prevalence of 60%, the summary PPV and NPV are both estimated to be 90%, but with wide CIs of 65-100% and 48-100%, respectively.

Subgroup analyses and meta-regression to explore heterogeneity were not performed due to the small number of studies. Following the QUADAS-2 assessments, three of the four studies were deemed to be high risk of bias overall, and consequently a sensitivity analysis including only low-risk studies could not be performed (as it would have included only one study). There were no disagreements between the screening authors regarding the inclusion of a study that was not resolved by discussion between the authors; therefore, no sensitivity analysis excluding 'controversial' studies was performed.

Discussion

Many diagnostic tests have been proposed for iNPH, and most have been shown to have poor diagnostic accuracy.[4,24–26] ELD is widely regarded as the most robust test for iNPH and several well-conducted studies have supported a high PPV for the test,[10,21,22,26] as have observational cohorts in which only ELD responders were shunted.[27–30] Furthermore, improvement following ELD has been shown to correlate well with improvement following shunt surgery.[31] Consequently, ELD may perform several roles beyond a pure diagnostic test. For instance, the response to ELD is useful in aiding patients' understanding of the potential benefits of shunt surgery and is also useful in patients with cardiac or respiratory co-morbidities or those taking anticoagulant drugs who may be exposed to higher operative risks. However, although the PPV of ELD is high, it has been suggested that its NPV may be low,[9] raising the possibility that ELD may not be an appropriate test to deny or deter patients from surgery.

Our systematic review identified only 4 sufficiently rigorous studies addressing the diagnostic test accuracy of ELD. The commonest reason for exclusion of a study was failure to shunt patients regardless of ELD outcome, an essential characteristic for a study to report accurate sensitivity and specificity. Despite stringent inclusion criteria, the studies passing screening and eligibility reported disparate results and many had significant methodological issues, such as unclear selection methods and lack of blinding.

Overall, the summary estimates of sensitivity and specificity suggest that ELD is a sensitive and relatively specific test for iNPH (sensitivity 94%, specificity 85%), but with a large amount of statistical uncertainty around these estimates. The only study deemed to be at low risk of bias and low applicability concerns in all QUADAS-2 domains (Chen et al.) estimated 100% sensitivity for the test, but this was based on only 7 patients. Of the three remaining studies, two found sensitivity to be high (94-100%), and only one (Walchenbach et al.) estimated it to be low (50%). The true value probably lies closer to the former estimate because Walchenbach et al. excluded patients who responded positively to a CSF tap test, which is likely to have removed true positives and hence under-estimated the sensitivity of the test. Conversely, in excluding

patients with any form of cortical atrophy, Chen et al. and Haan et al. are likely to have over-estimated sensitivity in their studies.

Unlike sensitivity and specificity, which are fixed variables, PPV and NPV depend on the prevalence of shunt-responsive iNPH within the population who might conceivably receive the test. For instance, if a larger population with more condition-negative patients is exposed to the test, the PPV falls and the NPV increases. At an estimated prevalence of 60% (which is that observed in our senior author's practice), both PPV and NPV were estimated at 90%, but again with a large amount of uncertainty due to the sparsity of evidence. The NPV rises slightly to 93% (CI 55-100%), with PPV unchanged, if the true prevalence is at the lower end of that observed in our included studies (53%); if the prevalence is at the higher end of this range (77%), the estimated PPV rises to 100% but the NPV falls to 81% (CIs 81-100% and 29-100%, respectively).

A limitation common to all the included studies, except Walchenbach et al., is that none employed programmable valves. This reflects the historical treatment of iNPH and the belief that any CSF diversion is enough to treat the condition. However, clinical experience has demonstrated that patients may manifest a clinical response at different valve pressure settings.[32] Our current treatment paradigm is more nuanced: Programmable shunts are used, and if patients do not improve following shunt surgery, the setting is gradually adjusted down until improvement of the iNPH symptoms has reached a plateau or symptoms of over-drainage are encountered. Furthermore, none of the included studies confirmed the functioning of the shunt in patients deemed to be non-responders. These aspects of their methodology may have resulted in true positives being missed, and a falsely low estimate of sensitivity and PPV. A further methodological difference between the included studies and current practice is the duration of the ELD test, which was 5 days in 3 of the studies, and 4 in the other. Although practice varies between institutions, most would conduct the test for 3 days.[26,28,29] In addition, the exclusion of patients with cortical atrophy in all studies is a concern, since two of these pre-date the description of DESH, the imaging characteristics of which may increase the probability of response to CSF shunting, but may historically have been mislabelled as cortical atrophy due to enlarged Sylvian subarachnoid spaces.

Although outside the scope of this review, one question raised by the finding of a weak evidence base for ELD is whether supplementary testing (beyond clinico-radiological indicators) is warranted at all. ELD arose in an era of high rates of shunt complications. In the Dutch Normal Pressure Hydrocephalus Study in the 1990s,[33] 16.5% of patients required surgical intervention for infection or hematoma formation during the 12-month period of follow-up, whereas in the European Multicentre Study,[34] published 15 years later, only 1.7% of patients suffered this type of complication. At less than 2%, it may be argued that the risks associated with a lumbar drain (1-5% of meningitis)[9,26] is higher, or at least similar to, those of a VP shunt. The cost effectiveness of this strategy has already been established by an analysis based on a Monte-Carlo simulation,[35] as well as a more recent prospective study,[36] which found a 1.7 quality-adjusted life year (QALY) benefit in suspected iNPH patients shunted purely on clinic-radiological grounds, with an associated cost per QALY of €7,500. The success of a recent randomised, controlled trial of shunt insertion in an iNPH population selected on clinico-radiological grounds (and the fact that no difference in the rate of complications was observed between intervention and control groups) also lends support to the argument for dispensing with supplementary testing.[2] However, it should be noted that this study was performed by experts in iNPH, who are likely to be adept at diagnosing the condition (and distinguishing it from potential mimics) on the basis of history and examination alone. It is also important to note that even if the balance does tip in favour of treatment decisions based on clinico-radiological indicators, it is likely that there will remain a role for ELD in certain circumstances, for instance, in patients in whom surgery would be high risk or those requiring additional persuasion of the benefits.

Conclusion

The results of this meta-analysis suggest a high PPV and NPV for ELD, but this should be interpreted with caution given the limited number of available studies, which are themselves small and heterogenous. Robust studies addressing the question of the diagnostic test accuracy of ELD are lacking and a well-designed, suitably large cohort study is required or, even better, a randomised controlled trial. Only the latter would establish if outcomes are enhanced by prior patient selection with ELD, or if procedural complications associated with shunt insertion are

now so low that the loss of potential responders favours a treatment strategy based upon clinico-radiological indicators alone.

Acknowledgements

The authors would like to acknowledge Katie Barnard (Clinical Librarian, Southmead Hospital) for performing the searches for this review.

Disclosure of Interest

The authors report no conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Hakim S. Some observations on CSF pressure. Hydrocephalic syndrome in adults with "normal" CSF pressure (Recognition of a new syndrome). 1964.
2. Kazui H, Miyajima M, Mori E, Ishikawa M. Lumboperitoneal shunt surgery for idiopathic normal pressure hydrocephalus (SINPHONI-2): an open-label randomised trial. *Lancet Neurol* 2015;14:585–594.
3. Hashimoto M, Ishikawa M, Mori E, et al. Diagnosis of idiopathic normal pressure hydrocephalus is supported by MRI-based scheme: a prospective cohort study. *Cerebrospinal Fluid Res* 2010;7:18.
4. Wikkelso C, Hellstrom P, Klinge PM, et al. The European iNPH Multicentre Study on the predictive values of resistance to CSF outflow and the CSF Tap Test in patients with idiopathic normal pressure hydrocephalus. *J Neurol Neurosurg Psychiatry* 2013;84:562–8.
5. Boon AJ, Tans JT, Delwel EJ, et al. Dutch normal-pressure hydrocephalus study: prediction of outcome after shunting by resistance to outflow of cerebrospinal fluid. *J Neurosurg* 1997;87:687–693.
6. Pujari S, Kharkar S, Metellus P, et al. Normal pressure hydrocephalus: long-term outcome after shunt surgery. *J Neurol Neurosurg Psychiatry* 2008;79:1282–1286.
7. Vakili S, Moran D, Hung A, et al. Timing of surgical treatment for idiopathic normal pressure hydrocephalus: association between treatment delay and reduced short-term benefit. *Neurosurg Focus* 2016;41:E2.
8. Andrén K, Wikkelsø C, Tisell M, Hellström P. Natural course of idiopathic normal pressure hydrocephalus. *J Neurol Neurosurg Psychiatry* 2014;85:806–810.
9. Walchenbach R, Geiger E, Thomeer RTWM, Vanneste JAL. The value of temporary external lumbar CSF drainage in predicting the outcome of shunting on normal pressure hydrocephalus. *J Neurol Neurosurg Psychiatry* 2002;72:503–506.
10. Panagiotopoulos V, Konstantinou D, Kalogeropoulos A, Maraziotis T. The predictive value of external continuous lumbar drainage, with cerebrospinal fluid outflow controlled by medium pressure valve, in normal pressure hydrocephalus. *Acta Neurochir (Wien)* 2005;147:953–8.
11. Mori E, Ishikawa M, Kato T, et al. Guidelines for management of idiopathic normal pressure hydrocephalus: second edition. *Neurol Med Chir (Tokyo)* 2012;52:775–809.
12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
13. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–536.
14. Reitsma JB, Glas AS, Rutjes AWS, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–990.
15. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;59:1331–1333.
16. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* 2000;10:325.
17. Harbord R. METANDI: Stata module to perform meta-analysis of diagnostic accuracy. *Stat Softw Components* 2008:S456932. Available at:

https://ideas.repec.org/c/boc/bocode/s456932.html.
18. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865–2884.
19. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239–251.
20. Arends LR, Hamza TH, van Houwelingen JC, et al. Bivariate random effects meta-analysis of ROC curves. *Med Decis Making* 2008;28:621–638.
21. Chen IH, Huang CI, Liu HC, Chen KK. Effectiveness of shunting in patients with normal pressure hydrocephalus predicted by temporary, controlled-resistance, continuous lumbar drainage: a pilot study. *J Neurol Neurosurg Psychiatry* 1994;57:1430–1432.
22. Haan J, Thomeer RT. Predictive value of temporary external lumbar drainage in normal pressure hydrocephalus. *Neurosurgery* 1988;22:388–391.
23. Miyajima M, Kazui H, Mori E, Ishikawa M. One-year outcome in patients with idiopathic normal-pressure hydrocephalus: comparison of lumboperitoneal shunt to ventriculoperitoneal shunt. *J Neurosurg* 2016;125:1483–1492.
24. Black PM. Idiopathic normal-pressure hydrocephalus. Results of shunting in 62 patients. *J Neurosurg* 1980;52:371–377.
25. Delwel EJ, de Jong DA, Avezaat CJJ. The prognostic value of clinical characteristics and parameters of cerebrospinal fluid hydrodynamics in shunting for idiopathic normal pressure hydrocephalus. *Acta Neurochir (Wien)* 2005;147:1033–1037.
26. Woodworth GF, McGirt MJ, Williams MA, Rigamonti D. Cerebrospinal fluid drainage and dynamics in the diagnosis of normal pressure hydrocephalus. *Neurosurgery* 2009;64:916–919.
27. Chotai S, Medel R, Herial NA, Medhkour A. External lumbar drain: A pragmatic test for prediction of shunt outcomes in idiopathic normal pressure hydrocephalus. *Surg Neurol Int* 2014;5:12.
28. Eide PK, Stanisic M. Cerebral microdialysis and intracranial pressure monitoring in patients with idiopathic normal-pressure hydrocephalus: association with clinical response to extended lumbar drainage and shunt surgery. *J Neurosurg* 2010;112:414–424.
29. Paidakakos N, Borgarello S, Naddeo M. Indications for endoscopic third ventriculostomy in normal pressure hydrocephalus. *Acta Neurochir Suppl* 2012;113:123–127.
30. Williams MA, Razumovsky AY, Hanley DF. Comparison of Pcsf monitoring and controlled CSF drainage diagnose normal pressure hydrocephalus. *Acta Neurochir Suppl* 1998;71:328–330.
31. Li Z, Naugle RI, Wood A, et al. Intermittent lumbar drainage with functional testing in the diagnosis of normal-pressure hydrocephalus. *Eur J Pediatr Surg* 2001;11 Suppl 1:S38-40.
32. Zemack G, Romner B. Adjustable valves in normal-pressure hydrocephalus: a retrospective study of 218 patients. *Neurosurgery* 2002;51:1392.
33. Boon AJ, Tans JT, Delwel EJ, et al. Dutch Normal-Pressure Hydrocephalus Study: randomized comparison of low- and medium-pressure shunts. *J Neurosurg* 1998;88:490–495.
34. Klinge P, Hellstrom P, Tans J, Wikkelso C. One-year outcome in the European multicentre study on iNPH. *Acta Neurol Scand* 2012;126:145–153.
35. Stein SC, Burnett MG, Sonnad SS. Shunts in normal-pressure hydrocephalus: do we place too many or too few? *J Neurosurg* 2006;105:815–822.
36. Tullberg M, Persson J, Petersen J, et al. Shunt surgery in idiopathic normal pressure

hydrocephalus is cost-effective-a cost utility analysis. *Acta Neurochir (Wien)* 2018;160:509–518.

Figure Legends

Figure 1. PRISMA flow diagram indicating flow of studies through screening/eligibility and reasons for exclusion of full-text reports.

Figure 2. Visual representation of the results of the QUADAS-2 analysis (by study).

Figure 3. Visual representation of the results of the QUADAS-2 analysis (by QUADAS-2 domain).

Figure 4. Coupled forest plots for included studies. TP = true positive; FP = false positive; FN = false negative; TN = true negative; CI = confidence interval.

Figure 5. Study-specific estimates of sensitivity and specificity in Receiver Operating Characteristic (ROC) space, together with meta-analysis results. Black circles represent individual studies (diameter of the circle is proportional to the size of the study). Summary estimate of sensitivity/specificity is marked by the red dot. 95% credible region is delineated by a dashed red line, and 95% prediction region by a dotted red line. The blue line shows the hierarchical summary receiver-operating characteristic (HSROC) curve from the meta-analysis.

Figure 6. Plot of summary PPV/NPV against prevalence. Observed PPVs from individual studies are represented by red stars and observed NPVs by blue dots. The ‘predicted’ summary lines for PPV and NPV are in red and blue, respectively, and the corresponding 95% credible intervals are shaded in red or blue as appropriate.

Search Strategies

Search date for all searches: 05.09.2018

Medline

via HDAS (1946 to present)

1	("normal pressure hydrocephalus" OR "normotensive hydrocephalus").ti,ab	2322
2	"HYDROCEPHALUS, NORMAL PRESSURE"/	2046
3	(1 OR 2)	2800
4	((CSF OR "Cerebrospinal fluid" OR "Cerebro-spinal fluid" OR lumbar) ADJ (diversion OR drain*)).ti,ab	2845
5	(3 AND 4)	162

Embase

via HDAS (1974 to present)

1	("normal pressure hydrocephalus" OR "normotensive hydrocephalus").ti,ab	3144
2	"NORMOTENSIVE HYDROCEPHALUS"/	3223
3	(1 OR 2)	3879
4	((CSF OR "Cerebrospinal fluid" OR "Cerebro-spinal fluid" OR lumbar) ADJ (diversion OR drain*)).ti,ab	3763
5	(3 AND 4)	241

Cochrane Library databases

via Wiley (Issue 9, 2018)

1	("normal pressure hydrocephalus" OR "normotensive hydrocephalus"):ti,ab,kw	85
2	MeSH descriptor: [Hydrocephalus, Normal Pressure] explode all trees	43
3	(1 OR 2)	90

4	((CSF OR "Cerebrospinal fluid" OR "Cerebro-spinal fluid" OR lumbar) next (diversion OR drain*)):ti,ab,kw	160
5	(3 AND 4)	13

Latin American Caribbean Health Sciences Literature (LILACS)

(http://lilacs.bvsalud.org/en/)

1	"normotensive hydrocephalus" OR "normal pressure hydrocephalus" OR "hidrocefalia normotensiva"	
2	AND drain OR drainage OR diversion OR drenaje	

Aggressive Research Intelligence Facility (ARIF) database

(https://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/ARIF/databases/index.aspx)

1	("normotensive hydrocephalus" OR "normal pressure hydrocephalus"):ti,ab	1

Database of Abstracts of Reviews of Effects (DARE)

via CRD (https://www.crd.york.ac.uk/CRDWeb/)

1	"normotensive hydrocephalus" OR "normal pressure hydrocephalus"	6

ClinicalTrials.gov

US National Institutes of Health (https://clinicaltrials.gov/)

	Condition or disease: "normal pressure hydrocephalus" OR "normotensive hydrocephalus"	5
	Other terms:	
	drain OR drainage OR diversion	

Figure 1

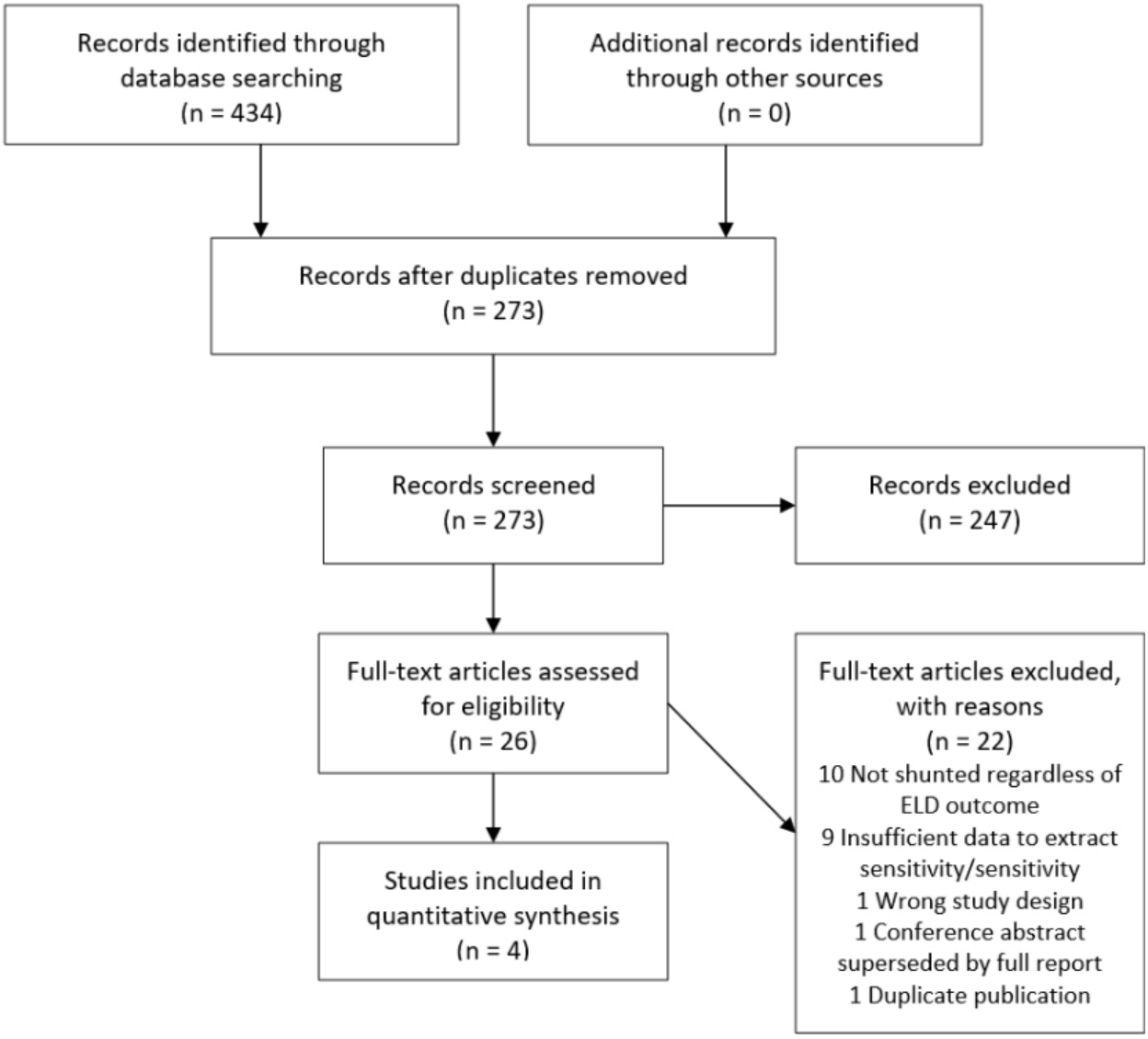

Records identified through database searching (n = 434)
Additional records identified through other sources (n = 0)
Records after duplicates removed (n = 273)
Records screened (n = 273)
Records excluded (n = 247)
Full-text articles assessed for eligibility (n = 26)
Full-text articles excluded, with reasons (n = 22)
10 Not shunted regardless of ELD outcome
9 Insufficient data to extract sensitivity/sensitivity
1 Wrong study design
1 Conference abstract superseded by full report
1 Duplicate publication
Studies included in quantitative synthesis (n = 4)

Figure 2

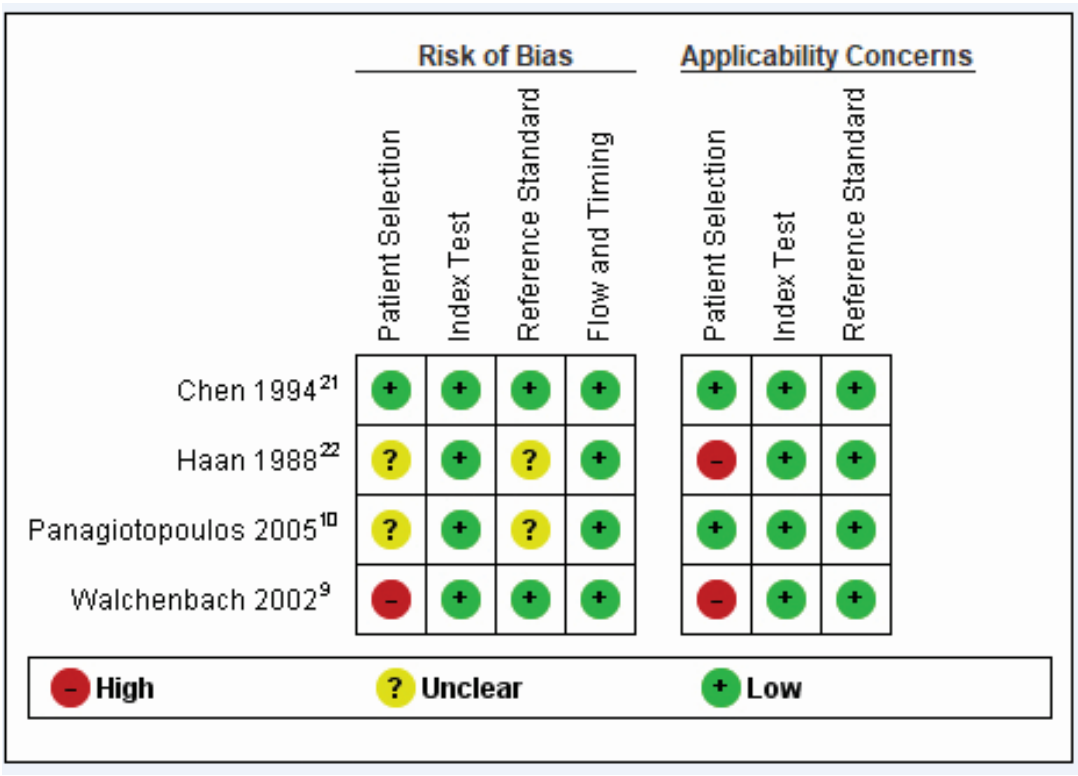


Figure 3

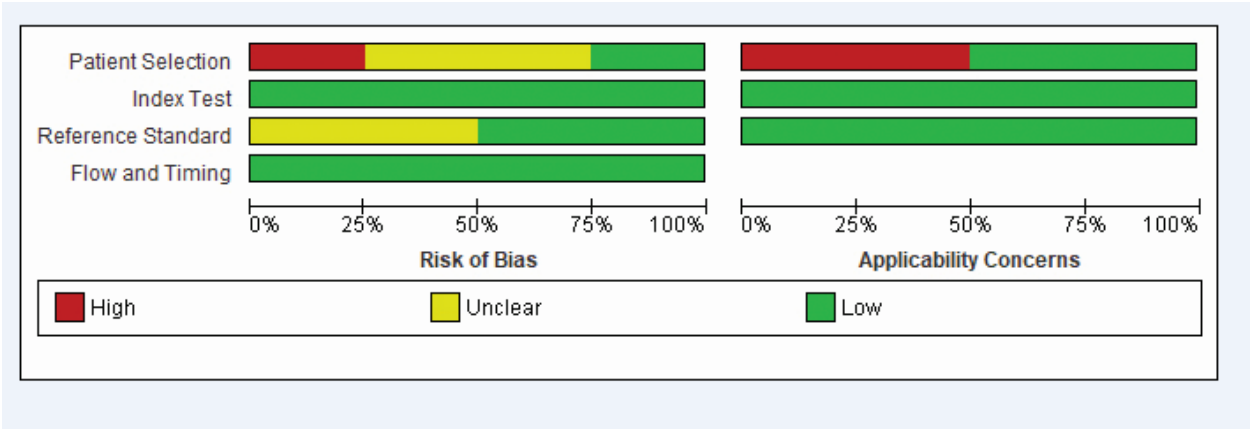


Figure 4

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chen[21]	4	0	0	3	1.00 [0.40, 1.00]	1.00 [0.29, 1.00]
Haan[22]	12	0	0	5	1.00 [0.74, 1.00]	1.00 [0.48, 1.00]
Panagiotopoulos[10]	16	3	1	2	0.94 [0.71, 1.00]	0.40 [0.05, 0.85]
Walchenbach[9]	14	2	14	8	0.50 [0.31, 0.69]	0.80 [0.44, 0.97]

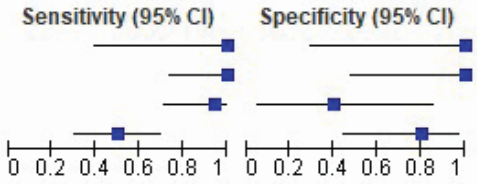


Figure 5

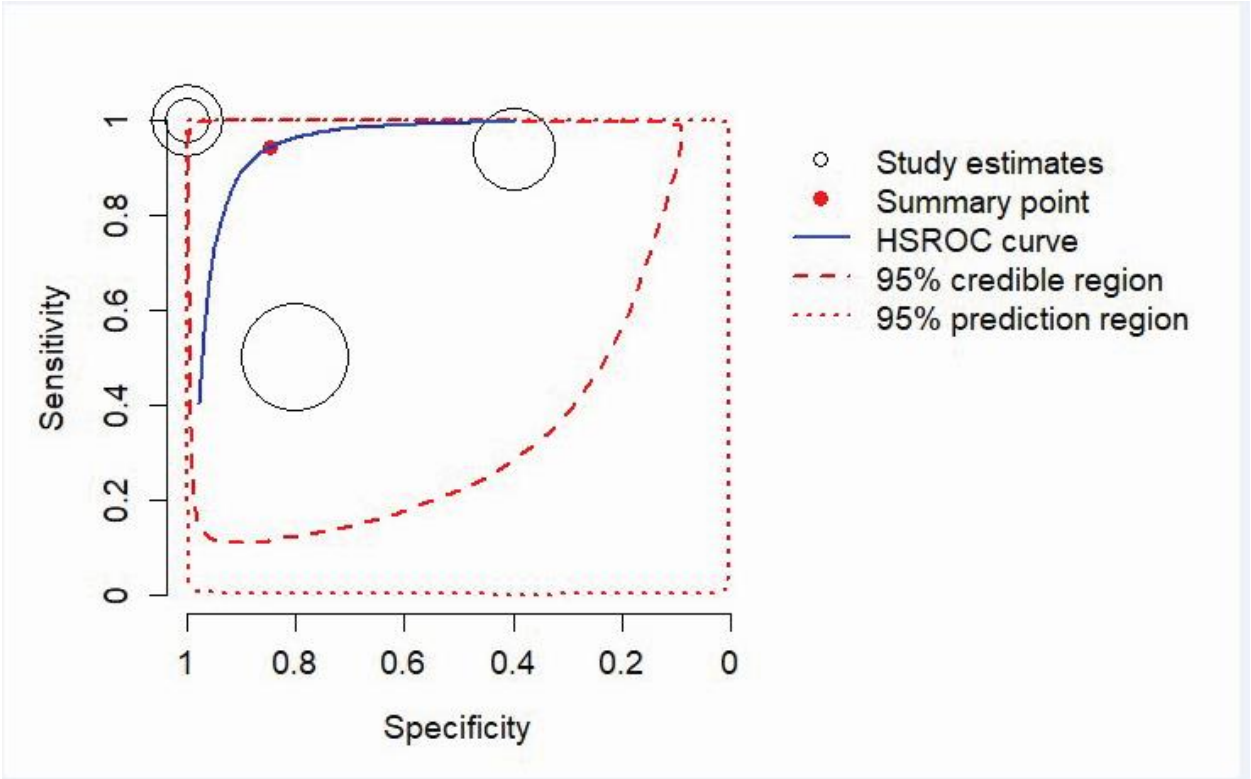
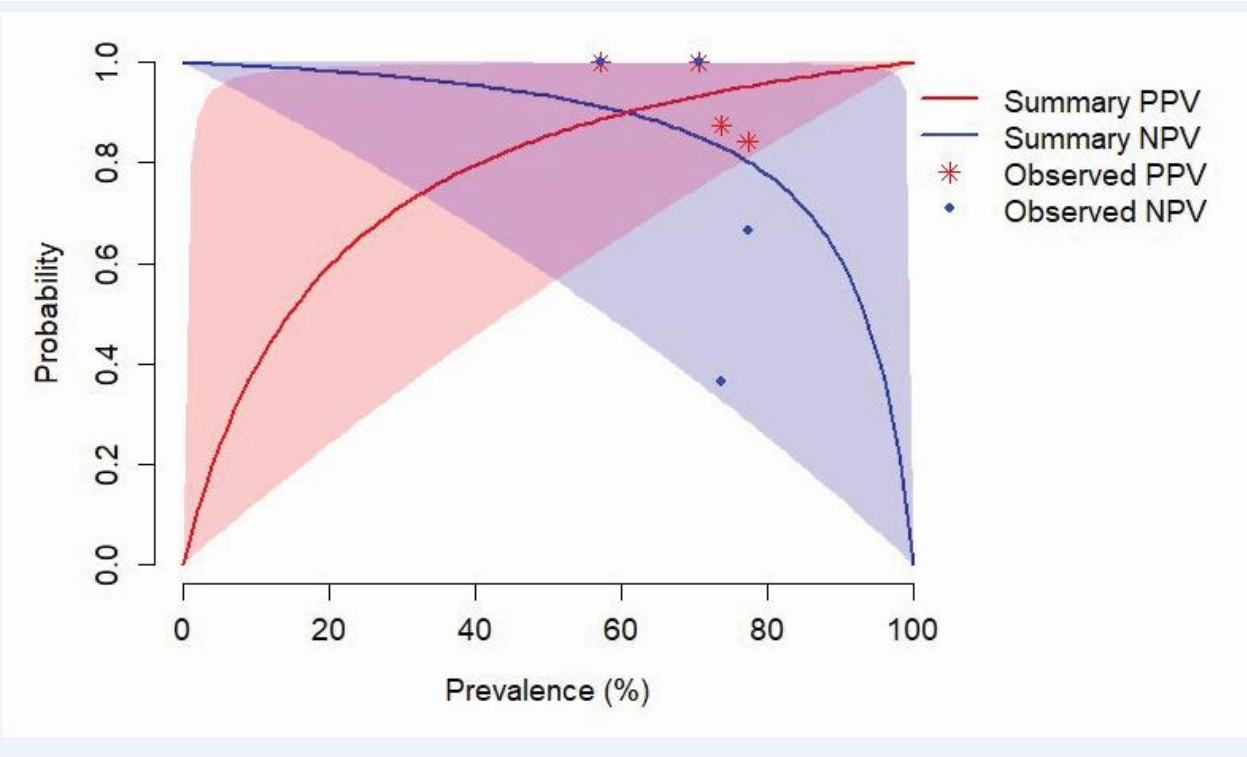


Figure 6


Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Prevalence (%)
Summary PPV
Summary NPV
Observed PPV
Observed NPV

						ELD				Shunt		
Study	**Design**	**N**	**Proportion with idiopathic NPH**	**Inclusion criteria**	**Exclusion criteria**	**Duration (days)**	**Rate (mean or range; ml/hr)**	**Time from removal of drain to assessment of outcome**	**Criteria for positivity**	**Type of shunt implanted**	**Criteria for regarding the shunt to have been successful**	**Follow-up prior to decision upon shunt response (months)**
Chen et al., 1994[21]	Cohort	7	100%	Gait AND dementia AND ventriculomegaly	Cortical atrophy	5	15-18	1, 3 and 5 days	Improvement in cognition OR urinary function OR gait	Randomised to VP / LP (medium-pressure valves)	Not specified	12
Haan et al., 1988[22]	Cohort	17	100%	Gait AND ventriculomegaly AND supportive findings from cisternography AND normal OP on LP	Cortical atrophy OR improving after initial LP	5	12.5	Immediately	Improvement in gait AND (improvement in cognition OR disappearance of urinary incontinence)	VA	As criteria for positive ELD	0.25
Panagiotopoulos et al., 2005[10]	Cohort	22	86%	Gait AND ventriculomegaly AND normal OP on LP	Minimal gait disturbance combined with severe dementia OR severe cortical atrophy	5	13.8	5 days	Improvement in gait	VP (medium-pressure valve)	As criteria for positive ELD	3
Walchenbach et al., 2002[9]	Cohort	38	88%	Gait AND ventriculomegaly AND normal OP on LP	Minimal gait disturbance combined with severe dementia OR severe cortical atrophy OR response to high-volume LP	4	4.2-6.3	Not specified	Improvement in gait	VP (programmable valve)	As criteria for positive ELD, except improvement had to be confirmed by patient, family or nursing staff	2

Table 1 Characteristics of included studies. LP = lumbar puncture/lumboperitoneal (depending on context); OP = opening pressure; VA = ventriculoatrial; VP = ventriculoperitoneal.

Study	Observed sensitivity	Observed specificity	Observed PPV	Observed NPV
Chen et al., 1994[21]	1.00 (CI 0.28-1.00)	1.00 (CI 0.19-1.00)	1.00 (CI 0.28-1.00)	1.00 (CI 0.19-1.00)
Haan et al., 1988[22]	1.00 (CI 0.64-1.00)	1.00 (CI 0.36-1.00)	1.00 (CI 0.64-1.00)	1.00 (CI 0.36-1.00)
Panagiotopoulos et al., 2005[10]	0.94 (CI 0.71-1.00)	0.4 (CI 0.05-0.85)	0.84 (CI 0.60-0.97)	0.67 (CI 0.09-0.99)
Walchenbach et al., 2002[9]	0.50 (CI 0.31-0.69)	0.8 (CI 0.44-0.98)	0.88 (CI 0.62-0.98)	0.36 (CI 0.17-0.59)

Table 2 Diagnostic test accuracy data from included studies.

	Summary estimate with 95% credible interval (Cr-I)	Between study standard deviation on logit scale
Sensitivity	0.94 (0.41, 1.00)	2.80 (0.99, 6.60)
Specificity	0.85 (0.33, 1.00)	1.91 (0.10, 6.06)
Between-study correlation	-0.31 (-0.91, -0.01)	

Table 3 Summary estimates of sensitivity and specificity from meta-analysis.