Title: Diagnostic Test Studies in Nephrology: Quantity, Quality and Scope

Authors and affiliations:

Richard G McGee, MM(Clin Epi)[*†], Brendon L Nguyen[†], Ruth L Mitchell, MA[*†], Jonathan C Craig, PhD[*†], Angela C Webster, PhD[*†‡]

[*]Centre for Kidney Research & Cochrane Renal Group, The Children's Hospital at Westmead, Locked Bag 4001, Westmead, NSW 2145, Australia

[†]Sydney School of Public Health, University of Sydney, NSW 2006, Australia

[‡]Centre for Transplant and Renal Research, Westmead Hospital, Westmead, NSW 2145, Australia

Corresponding author contact information:

Richard McGee

Centre for Kidney Research, Westmead Children's Hospital, Westmead, NSW, 2145 Australia

Phone: +61 2 9845 1466

Fax: +61 2 9351 5049

Email: richard.mcgee@sydney.edu.au

Word count: Abstract: 275, main text word count: 2692, figures: 5

Support and financial disclosure declaration

Support: RGMG is a recipient of a postgraduate research scholarship from the University of Sydney, Australia. BLN was funded by summer research scholarship from the Sydney Medical School, University of Sydney, Australia.

Conflicts of interest: RMG (none) BLN (none) RLM (none) JCC (none) ACC (none)

Abstract

Background: Diagnostic errors are an important source of preventable harm in healthcare, which may be reduced through evidence based choice, use and interpretation of diagnostic tests. We hypothesized that diagnostic errors are reduced through evidence based choice, use and interpretation of diagnostic tests.

Study Design: Retrospective cohort study.

Setting & Population: Diagnostic test studies.

Selection criteria for Studies: Publications from 1966 to 2008 retrieved from Medline.

Intervention: The quality of diagnostic accuracy studies tool (QUADAS) tool.

Outcomes: Number and coverage of diagnostic studies in Nephrology, and the methodological quality of test accuracy subset.

Results: Fewer diagnostic studies were published in Nephrology than other areas of internal medicine, although the proportion of total citations that were diagnostic studies (4.9%, sd±2.8) was not statistically different to other specialties (P=0.2). Within Nephrology, some topic areas (e.g. urinary tract infections) were over represented while others (e.g. acute kidney injury) had relatively few diagnostic studies (Range: 2.7% to 12.5%). Examining the randomly selected subset of studies that were diagnostic test accuracy studies (120) showed variable quality. Ninety-seven percent (116/120) of studies adequately described index test procedure, but only 27% (32/120) of studies adequately blinded investigators to results of index tests and 36% (43/120) to results of reference tests. The quality of nephrology diagnostic test accuracy studies has not substantially improved over the past 30 years.

Limitations: When comparing Nephrology with other specialties, some potential inequalities of scale could not be addressed, which may influence research output results across specialties.

Conclusions: Diagnostic research in Nephrology is published less frequently than most other medical specialties. The quality of diagnostic test accuracy studies that are published is variable and leaves room for improvement.

Introduction

Optimal patient outcomes depend on clinician's use and interpretation of diagnostic tests to trigger appropriate management algorithms for example a primary care physician may monitor a patient's estimated glomerular filtration rate to decide the optimal timing for referral to a nephrology service for assessment and multidisciplinary management of chronic kidney disease. There is empiric evidence to show that poor study design leads to biased results and overestimation of test performance.[1-3] Diagnostic errors are an important source of preventable harm in healthcare, and are often under-recognised.[4,5] Approximately 5% of autopsies reveal lethal conditions which may have been prevented with a correct diagnosis,[6] and in the USA alone approximately 40 000-80 000 deaths occur annually due to misdiagnosis.[7] In one study, harms relating to diagnostic error were more likely to be considered negligent and result in serious disability compared to drug related error.[8] Tort claims for diagnostic errors are nearly twice as common as claims for medication errors and result in the largest payouts.[9]

Diagnostic test accuracy studies provide the best evidence for clinicians to establish the operating characteristics of a test. In contrast to randomised controlled trials of interventions, little is known about the frequency of publication or quality of diagnostic test accuracy studies. What is known is limited to specific medical conditions such as melanoma,[10] breast cancer,[11] lung cancer[12] and bowel cancer,[13] which all demonstrate significant variability in study quality. To our knowledge, no study has examined diagnostic research from a broader perspective such as at a discipline or specialty level. We hypothesized that diagnostic errors are reduced through evidence based choice, use and interpretation of diagnostic tests.

The aims of our study were to estimate the number of diagnostic studies published in Nephrology and to compare this with other specialties of internal medicine, to describe the coverage of diagnostic studies within Nephrology by topic area, and to estimate the quality of design and reporting of diagnostic test accuracy studies in Nephrology over time.

Methods

Firstly, we identified diagnostic studies by searching MEDLINE via the OvidSP platform. We identified appropriate specialty-based medical subject headings (MeSH terms), to search for all citations relevant to thirteen major specialties of internal medicine and then limited results to the years 1966 to 2008. MeSH terms used were: exp Cardiovascular Diseases/, exp Digestive System Diseases/, exp Nervous System Diseases/, exp Immunologic Diseases/, exp Neoplasms/, exp Respiratory Tract Diseases/, exp Endocrine Diseases/, exp Musculoskeletal Diseases/, exp Nutritional and Metabolic Diseases/, exp Skin and Connective Tissues Diseases/, exp Hemic and Lymphatic Diseases/, (exp Bacterial Infections/ OR exp Mycoses/ OR exp Virus Diseases/ OR exp Parasitic Diseases/), (exp Kidney Diseases/ OR exp Urolithiasis/ OR Urinary tract infections/ OR Pyuria/ OR Bacteriuria/ OR Proteinuria/ OR exp Renal Replacement Therapy/). We chose the upper limit of 2008 to reduce any selection bias that might arise due to potential differences among medical speciality journals in time lag from publication to indexing by MEDLINE (last search December 2009). To estimate the number and proportion of the total citations that were diagnostic studies, the filter "diagnosis (optimized)" was applied using MEDLINE's clinical queries option (this filter applies the following search terms and Boolean operators: sensitiv*.mp OR predictive value*.mp OR accurac*.tw).[14] We used a Pearson χ^2 test to determine if the proportion of diagnostic studies was statistically different amongst medical specialties.

To evaluate the coverage of diagnostic studies within Nephrology, we identified the relevant MeSH terms for eleven major areas of Nephrology and used these terms in a new MEDLINE search. MeSH terms used were: exp Renal Tubular Transport, Inborn

Errors/, exp Renal Insufficiency, Acute/, (Renal Insufficiency/ OR Kidney Failure/ OR exp Renal Insufficiency, Chronic/), Diabetic Nephropathy/, (Nephritis OR exp Glomerulonephritis/ OR exp Nephrosis/), Renal Dialysis/, exp Peritoneal Dialysis/, Kidney Transplantation/, exp Urolithiasis/, (Urinary Tract Infections/ OR Bacteruria/ OR Pyuria/), exp Renal Artery Obstruction/. Again, these results were limited to the years 1966 to 2008 and filtered using "diagnosis (optimized)".

Secondly, from the larger pool of diagnostic studies, we randomly selected a subset that were diagnostic test accuracy studies. We defined diagnostic test accuracy studies as those studies that aimed to accurately classify people with and without the target condition, by comparing results of a test or a series of tests with a reference standard. To assess quality of design and reporting of the evidence base for comparative diagnostic test accuracy studies, we randomly selected 600 studies from our search results (approximately 2.5% of the total) and from these two authors (RGMG, BLN) working independently screened the titles and abstracts of diagnostic studies, to identify which of these studies were the subset that were diagnostic test accuracy studies. We discarded other diagnostic studies, which did not attempt to estimate comparative accuracy, including studies of genetic techniques, microbial sensitivities and overview articles. From this sample, the quality of the diagnostic test accuracy studies was assessed by the fourteen components of the quality of diagnostic accuracy studies (QUADAS) tool.[15] QUADAS was developed as a quality assessment tool for diagnostic test accuracy studies; if a diagnostic test accuracy study poorly reports these parameters, it may be associated with a biased estimate of diagnostic test accuracy.[1-3] The quality in each QUADAS domain was graded as high, low or unclear for each study according to the guidance presented in the original QUADAS paper.[15] Change in methodological quality over time (grouped by decade, and stratified into

low/unclear versus high quality) was assessed by ordinal logistic regression and likelihood ratio test statistics using Stata software (Stata11, http://www.stata.com/).

Results

Quantity of diagnostic studies

Figure 1 illustrates the selection of studies for inclusion in this study. **Figure 2** illustrates the number of citations for diagnostic studies, in thirteen internal medicine specialties from 1966 to 2008. The number of diagnostic studies published across all specialties increased over time, but there were fewer diagnostic studies published in Nephrology compared to all other specialties. The mean number of diagnostic studies published per specialty was 59 764 (standard deviation; sd ± 34 855). Oncology published the greatest number of diagnostic studies for each calendar year since 1975, and overall (142 467 total), whereas Nephrology published the fewest diagnostic studies between years 1992 and 2007, and a total of 22 230 studies between 1966 and 2008.

When we allowed for differences in relative size of medical specialties, by calculating the proportion of diagnostic studies compared with total citations, the relative standing of Nephrology improved somewhat, as shown in **Figure 3**. The proportion of total citations that were diagnostic studies increased over time for all specialties. The proportion of total citations that were diagnostic studies in Nephrology was 4.9%, sd±2.8, which was not statistically different to that in other specialties P=0.2, range: Rheumatology 4%, Infectious diseases 7.6%.

Within Nephrology, the topic areas of diabetic nephropathy, renal artery stenosis and urinary tract infection were covered by a greater proportion of diagnostic studies than the overall Nephrology mean 4.9%, range: acute kidney injury 2.7%, urinary tract infections 12.5%. Chronic kidney disease, kidney transplantation, haemodialysis,

glomerulonephritis, peritoneal dialysis, urolithiasis, acute kidney injury and renal tubular transport were under-represented (see **Figure 4**).

Quality of diagnostic test accuracy studies in Nephrology

Of the 600 papers randomly selected for further assessment, 480 studies primarily examined genetic techniques, microbial sensitivities or were overview articles about tests. Only 120 (20%) studies were designed as diagnostic test accuracy studies and were further assessed for design and reporting quality. These 120 studies were published across 75 journals, with only five journals publishing four or more studies: Radiology (6), American Journal of Radiology (5), Transplantation (5), Kidney International (4) and Journal of Clinical Microbiology (4). These studies investigated a range of diagnostic modalities: imaging (39), biochemical (19), pathological (27), risk modelling (12), and other (23).

Of the 120 diagnostic test accuracy studies assessed, quality was mixed. Ninety-seven percent (116/120) of studies adequately described index test procedure, in 95% (114/120) of studies, the index test appropriately did not form part of the reference standard, and 90% (108/120) included a spectrum of patients that was representative of those seen in practice. However, for many other QUADAS domains, quality was suboptimal. Only 27% (32/120) of studies adequately blinded investigators to results of index tests and 36% (43/120) to results of reference tests. For 70% (84/120) and 63% (75/120) of studies respectively, blinding of the index and reference standard tests was unclear. The time interval between administering index and reference tests was unclear in 42% (50/120) of studies. Ten of the fourteen quality parameters of the QUADAS tool did not significantly improve between 1966 and 2008 (see **Figure 5**). Because of the small number of eligible studies, the confidence intervals around these

estimates are large. The four parameters that did improve over time included: Q2 reporting of selection criteria (P< 0.001), Q4 acceptable delay between tests (P= 0.02), Q9 description of reference standard test procedure (P= 0.01), and Q12 same clinical data available as when used in practice (P= 0.01).

Discussion

Overall, the evidence base supporting informed choice of diagnostic tests in Nephrology appears weak. Although the number of diagnostic studies in Nephrology has increased over time, there are fewer diagnostic studies published in Nephrology than any other medical specialty over the past 40 years. This finding is in context given the relatively low publication output of Nephrology in general, across all study types. When adjusted for relative specialty size, using a proportional rather than absolute number of publications, the performance of Nephrology improved, but remained lower than average across all specialties. Within Nephrology, coverage was patchy and did not appear to be related to clinical importance of the topic or disease burden. Some topic areas were over represented such as diabetic nephropathy and renal artery stenosis, but others were under-represented, such as glomerulonephritis and acute kidney injury. Overall design and reporting quality was mixed, with many important methodological features of diagnostic accuracy studies poorly designed and reported, and perhaps most importantly, no strong signal of improvements over time.

This is the first study that we are aware of to have examined the quantity and coverage of diagnostic studies and quality of diagnostic test accuracy studies across a medical specialty. This is important because it reflects the perspective of users of research working in clinical practice, who have to assess and make use of tests in a variety of clinical situations and disease states. Other studies, which have examined the quality and relevance of diagnostic test publications, have done so from a limited perspective of diagnosing a specific disease state.[7-10] A similar study which focussed on interventional research and evaluated the quantity and quality of randomised controlled trials in Nephrology showed comparable results to our study.[16] In 2002, the

number of randomised trials published in Nephrology was lower compared to all other specialties of internal medicine and the quality of those studies published was suboptimal.

It has been demonstrated consistently that studies with suboptimal design and reporting features tend to overestimate the performance characteristics of the index test being evaluated. When used in clinical practice, such studies will therefore tend to lead to unrecognised misclassification of disease by clinicians, leading to additional unnecessary tests and treatment and inappropriately withheld tests and treatment, depending upon the direction and magnitude of the error. Our study has shown that poor design features are highly prevalent in diagnostic test accuracy studies published in Nephrology and are not improving. These findings highlight the importance of high quality reporting and the endorsement of reporting guidelines for specific study designs; for diagnostic test accuracy studies these are the standards for the reporting of diagnostic accuracy studies (STARD) guidelines.[17] Currently only two of the 246 journals that support the use of the STARD statement are Nephrology journals,[18] even though five Nephrology journals support the equivalent guideline for randomised controlled trials – the CONSORT statement.[19]

As a descriptive study, there are potential limitations to our work. Our choice of search terms to identify publications in medical specialty areas was not as comprehensive in scope as a systematic review, but mirrors other similar work.[16] We chose the clinical queries filter "diagnosis (optimized)" to try to achieve a balance of sensitivity and specificity in retrieval of studies. The original study which developed this filter showed that it had a sensitivity of 93% and a specificity of 92%, compared to a gold standard of hand searching,[1] and therefore some true diagnostic test accuracy

studies may have been missed (false negative) or non-diagnostic accuracy studies included (false positive), although this error is likely to be non-differential across the medical specialties. While we adjusted our estimates for the total number of citations within each specialty, we did not address other potential inequalities of scale across specialties, such as number of patients eligible for study participation, and the number of new interventions, tests and devices. Nor did we adjust for the number of tests performed in practice. It may be that if we were able to allow for such differences, we would have found different results when comparing across the specialties. Globally, the range of conditions seen by Nephrologists will vary and there may be some conditions, which we have not considered. Additionally, as we did not apply the QUADAS tool to diagnostic test accuracy studies in other specialties, and as our study is unique, we cannot make comparisons about quality across different medical specialties.

Given the findings of our study, what can be done to improve the quantity and quality of diagnostic test accuracy studies in Nephrology? Greater endorsement of the STARD statement[17] and quality assessment tools such as the QUADAS tool,[15] by journals and peer reviewers would be helpful; as would further efforts to improve awareness of design and reporting quality issues, in the design and analysis of diagnostic tests accuracy studies. One promising initiative, has recently sought to improve clinical research skills amongst Nephrologists.[18] We would also suggest that learned Colleges and Societies promote the uptake of epidemiological and general research methodology training amongst its members, particularly trainees, which could build capacity for the future. Further research should aim to assess the effectiveness of these educational efforts.

In conclusion, we have shown that diagnostic studies in Nephrology are published less frequently than in most other medical specialties. The quality of diagnostic test accuracy studies that are published is variable and leaves room for improvement. Most importantly, these findings should challenge researchers and funding bodies to improve standards of the diagnostic test accuracy evidence base in our specialty.

Acknowledgements

An abstract of this work was presented as an oral presentation at the 2010 Joint Colloquium of the Cochrane & Campbell Collaborations Colloquium, Colorado, USA and as a poster presentation at the Australian and New Zealand Society of Nephrology 2010 Annual Scientific Meeting, Perth, Australia.

References

1. Whiting P, Rutjes A, Reitsma J, Glas A, Bossuyt P, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. *Annals of Internal Medicine.* 2004;140(3):189.

2. Lijmer J, Mol B, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA.* 1999;282(11):1061.

3. Rutjes A, Reitsma J, Di Nisio M, Smidt N, Van Rijn J, Bossuyt P. Evidence of bias and variation in diagnostic accuracy studies. *Can. Med. Assoc. J.* 2006;174(4):469.

4. Newman-Toker D, Pronovost P. Diagnostic Errors--The Next Frontier for Patient Safety. *JAMA.* 2009;301(10):1060.

5. Graber M. Diagnostic errors in medicine: a case of neglect. *Jt. Comm. J. Qual. Patient Saf.* 2005;31(2):106-113.

6. Shojania K, Burton E, McDonald K, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA.* 2003;289(21):2849.

7. Gabel R, Hayward R, Leape L, Berwick D, Bates D. Counting deaths due to medical errors. *JAMA.* 2002;288(19):2404.

8. Leape L, Brennan T, Laird N, et al. The nature of adverse events in hospitalized patients. *N. Engl. J. Med.* 1991;324(6):377-384.

9. Weeks W, Foster T, Wallace A, Stalhandske E. Tort claims analysis in the Veterans Health Administration for quality improvement. *The Journal of Law, Medicine & Ethics.* 2001;29(3-4):335-345.

10. Rosado B, Menzies S, Harbauer A, et al. Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. *Archives of dermatology.* 2003;139(3):361.

11. Flobbe K, Nelemans P, Kessels A, Beets G, Von Meyenfeldt M, Van Engelshoven J. The role of ultrasonography as an adjunct to mammography in the detection of breast cancer:: a systematic review. *European Journal of Cancer.* 2002;38(8):1044-1050.

12. Vansteenkiste J, Fischer B, Dooms C, Mortensen J. Positron-emission tomography in prognostic and therapeutic assessment of lung cancer: systematic review. *The lancet oncology.* 2004;5(9):531-540.

13. de Zwart I, Griffioen G, SHAW M, Lamers C, de Roos A. Barium enema and endoscopy for the detection of colorectal neoplasia: sensitivity, specificity, complications and its determinants. *Clinical radiology.* 2001;56(5):401-409.

14. Haynes R, McKibbon K, Wilczynski N, Walter S, Werre S. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *British Medical Journal.* 2005;330(7501):1179.

15. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC medical research methodology.* 2003;3(1):25.

16. Strippoli G, Craig J, Schena F. The number, quality, and coverage of randomized controlled trials in nephrology. *J. Am. Soc. Nephrol.* 2004;15(2):411.

17. Bossuyt P, Reitsma J, Bruns D, et al. Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy: The STARD Initiative. *Clinical Radiology.* 2003;58(8):575-580.

18. Adopters of STARD. STAndards for the Reporting of Diagnostic accuracy studies. http://www.stard-statement.org/. Accessed February 14, 2011.

19. CONSORT Endorsers - Journals. CONSORT - Transparent reporting of trials. http://www.consort-statement.org/about-consort/supporters/consort-endorsers---journals/. Accessed February 14, 2011.

Figures and Legends

Figure 1. Flow chart of study selection process.

Figure 2. The number of diagnostic studies published in Nephrology and twelve other internal medicine specialties by calendar year.

Figure 3. The proportion of total citations that were diagnostic studies in Nephrology and twelve other internal medicine specialties by calendar year.

Figure 4. The number of diagnostic studies published in eleven topic areas of Nephrology compared to the total number of citations.

Note: The solid line represents the mean proportion of total citations in Nephrology that were diagnostic studies (4.9%).

Figure 5. Proportion of high quality diagnostic test accuracy studies for each QUADAS domain.[15]

Note: Each numbered plot refers to the corresponding QUADAS item number. All P values were calculated using likelihood ratio test statistics.