The future for diagnostic tests of acute kidney injury in critical care: evidence synthesis, care pathway analysis and research prioritisation

Peter S Hall,[1]* Elizabeth D Mitchell,[2] Alison F Smith,[2,3] David A Cairns,[4] Michael Messenger,[3] Michelle Hutchinson,[4] Judy Wright,[2] Karen Vinall-Collier,[2] Claire Corps,[5] Patrick Hamilton,[6] David Meads[2] and Andrew Lewington[5]

[1]Edinburgh Cancer Research Centre, University of Edinburgh, Edinburgh, UK
[2]Academy of Primary Care, Hull York Medical School, Hull, UK
[3]National Institute for Health Research (NIHR) Diagnostic Evidence Co-operative Leeds, Leeds, UK
[4]Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK
[5]Leeds Teaching Hospitals NHS Trust, Leeds, UK
[6]Manchester Institute of Nephrology and Transplantation, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

*Corresponding author p.s.hall@ed.ac.uk

Declared competing interests of authors: Andrew Lewington has received honoraria from Alere, Inc. David Meads is a member of the Health Technology Assessment programme Emergency and Hospital Care panel. Patrick Hamilton has received funding from ChemoCentryx, Inc., outside this work.

Published May 2018
DOI: 10.3310/hta22320

Scientific summary

Future of diagnostic acute kidney injury tests in critical care

Health Technology Assessment 2018; Vol. 22: No. 32
DOI: 10.3310/hta22320

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Acute kidney injury (AKI) is highly prevalent in hospital inpatient populations, leading to significant mortality and morbidity, reduced quality of life and high short- and long-term health-care costs for the NHS. Diagnosis currently relies on serum creatinine concentrations and anuria, which are imperfect biomarkers, leading to delayed detection after initial kidney damage. New biomarker-based in vitro diagnostics (IVDs) offer an opportunity for earlier diagnosis and improved risk stratification, enabling earlier specialist referral, targeted intervention or intensification of therapy when indicated by the test.

Candidate diagnostic tests have been commercially developed and are being marketed to clinicians and commissioners. Their evidence base is variable and there is an urgent need to develop evidence for clinical utility and cost-effectiveness to guide appropriate and suitably informed adoption within routine health services.

The development of evidence for diagnostic tests has historically been inefficient, resulting in uncertain or poor-value adoption decisions. There is a time-limited opportunity to propose an efficient research strategy for AKI diagnostics in the UK.

Aims

- To evaluate the potential for AKI IVD tests to enhance the NHS care of patients admitted to the intensive care unit (ICU)
- To identify the priorities for further research and development.

Design

The AKI-Diagnostics project included a systematic review, evidence synthesis and meta-analysis, care pathway analysis, model-based economic evaluation and value of information (VOI) analysis.

Methodological scope and outputs

- Pre-analytical, analytical and biological measurement properties.
- Clinical validity (sensitivity and specificity) for relevant outcomes identified within the care pathway.
- Clinical efficacy, clinical effectiveness and cost-effectiveness of test-directed care compared with standard care.
- Value of information analysis for research prioritisation.

Methods

The systematic review consisted of three literature searches, prospectively registered on the PROSPERO database (reference number CRD42014013919):

1. A horizon-scanning search of literature published after 2004 to identify current and future biomarkers and diagnostic tests for potential use in the identification and management of AKI in critical care.

A ranking procedure was developed to prioritise tests for further review based on the state of regulatory approvals, number of citations, combined sample size, biological mechanism and expert advice.
2. A more detailed systematic review was then undertaken of the analytical and clinical validity, clinical utility and cost-effectiveness of the highest-priority biomarkers. Two independent reviewers undertook a search of 12 databases and two trials registers in November 2015 using predefined eligibility criteria, with quality assessment carried out using the QUADAS-2 (quality assessment of diagnostic accuracy studies) tool.
3. A third literature review consisted of a series of targeted searches aimed at identifying previous relevant cost-effectiveness analyses and related research to inform the development of an economic model.

The meta-analysis of diagnostic accuracy used a bivariate model to estimate joint pooled means and variances of sensitivity and specificity for each of the diagnostic tests considered for the outcome measure of AKI KDIGO (Kidney Disease: Improving Global Outcomes). Separate meta-analyses were conducted for each diagnostic test, sample media and the two health service settings of ICU and post-cardiac surgery. Approximate estimates of the variance of sensitivity and specificity for use in the economic model were estimated using the delta method.

In the absence of a published tool, a framework for the quality assessment of measurement was developed through a process of expert consensus and review of the published literature. This was applied to a single test as proof of concept.

The economic evaluation relied on a de novo decision model constructed to determine the cost-effectiveness of the biomarkers for the early identification of AKI in the ICU from a NHS and Personal Social Services perspective. Tests were assumed to be used once on entry to the ICU. The primary analysis concerned the use of tests on an all-comer ICU population; a secondary analysis was conducted to explore the impact of the tests in a subgroup of patients in the ICU post cardiac surgery. Costs were reported in 2015 prices and effectiveness was measured in terms of quality-adjusted life-years (QALYs). All future cost and QALY outcomes were discounted at an annual rate of 3.5%.

Model parameters were estimated using individual patient clinical trial and registry data, supplemented when necessary with information from the published literature, which was identified through a series of targeted systematic literature reviews. Uncertainty around parameter estimates was characterised by assigning probability distributions to each of the uncertain parameters according to available variance data. Probabilistic sensitivity analysis was then conducted using 10,000 Monte Carlo simulations. Cost-effectiveness was assessed in terms of the incremental net (health) benefit (INB) and the incremental cost-effectiveness ratio (ICER). The cost-effectiveness threshold was taken as £20,000 per QALY unless otherwise stated.

The VOI analysis used the measures expected value of perfect information (EVPI) and expected value of perfect parameter information (EVPPI), calculated using non-parametric regression meta-modelling techniques including a generalised additive model and multivariate adaptive regression splines. Individual patient- and population-level estimates were calculated. The EVPPI was used to rank groups of model parameters as priorities for further research.

Results

Horizon-scanning search

The scoping search identified 4804 references. After screening by title/abstract, 487 potentially relevant papers remained, relating to 152 individual biomarkers. Those already used in standard care ($n = 11$, including serum creatinine) or with incomplete data related to the dimensions outlined earlier ($n = 19$) were excluded. Ten priority biomarkers/tests were shortlisted: brain natriuretic peptide (BNP), cystatin C, interleukin (IL)-6, IL-18, kidney injury molecule-1 (KIM-1), liver fatty acid-binding protein (L-FABP),

N-acetyl-beta-D-glucosaminidase (NAG), Nephrocheck® (Astute Medical, Inc., San Diego, CA, USA), neutrophil gelatinase-associated lipocalin (NGAL) and tumour necrosis factor alpha (TNF-α).

Systematic review and meta-analysis

Detailed review was undertaken for the top three ranked diagnostic tests: Nephrocheck (urine), which measures a combination of two proteins, tissue inhibitor of metalloproteinases 2 (TIMP-2) and insulin-like growth factor-binding protein 7 (IGFBP-7); NGAL protein; and cystatin C protein. The different media of blood serum, blood plasma and urine were considered separately.

Nephrocheck

Ten studies were included in the review and three were included in the meta-analysis in the critical care setting using urine. The median age of participants in the studies was 64 years and 58% of participants were male. Using the high sensitivity cut-off value (0.3), pooled sensitivity was estimated as 0.90 [95% confidence interval (CI) 0.85 to 0.93] and pooled specificity was estimated as 0.49 (95% CI 0.46 to 0.53). No clinical efficacy, clinical utility or cost-effectiveness studies were identified. One study was included in the cardiac surgery setting using serum, with a sensitivity of 0.92 (95% CI 0.76 to 0.98) and specificity of 0.67 (95% CI 0.47 to 0.82).

Neutrophil gelatinase-associated lipocalin

Thirty-nine studies were included in the review. There was heterogeneity in the outcome definitions, assessment period and threshold used to define a positive test. Eight studies were included in the meta-analysis in the critical care setting using plasma. For plasma, the pooled sensitivity estimate was 0.72 (95% CI 0.65 to 0.79) and the pooled specificity estimate was 0.81 (95% CI 0.75 to 0.86). For serum, one study was included, with a sensitivity of 0.54 (95% CI 0.43 to 0.65) and a specificity of 0.95 (95% CI 0.88 to 0.98). For urine, six studies were included, with a pooled sensitivity of 0.70 (0.59 to 0.80) and a pooled specificity of 0.79 (95% CI 0.71 to 0.86). In the cardiac surgery setting eight studies were included for plasma, with a pooled sensitivity of 0.62 (95% CI 0.49 to 0.74) and a pooled specificity of 0.78 (95% CI 0.75 to 0.81). For serum in the cardiac surgery setting, two studies were included, with a pooled sensitivity of 0.84 (95% CI 0.43 to 0.97) and a pooled specificity of 0.87 (95% CI 0.59 to 0.97). For urine in the cardiac surgery setting, 13 studies were included, with a pooled sensitivity of 0.66 (95% CI 0.54 to 0.76) and a pooled specificity of 0.62 (95% CI 0.41 to 0.79).

Cystatin C

Seventeen studies were included in the review. There was heterogeneity in the outcome definitions, assessment period and threshold used to define a positive test. In the meta-analysis, in the critical care setting, for plasma, three studies were included, with a pooled sensitivity of 0.72 (95% CI 0.59 to 0.82) and a pooled specificity of 0.74 (95% CI 0.65 to 0.81); for serum, four studies were included, with a pooled sensitivity of 0.76 (95% CI 0.57 to 0.88) and a pooled specificity of 0.91 (95% CI 0.85 to 0.95); and, for urine, three studies were included, with a pooled sensitivity of 0.68 (95% CI 0.43 to 0.86) and a pooled specificity of 0.76 (95% CI 0.62 to 0.86). In the cardiac surgery setting there were no suitable studies for plasma; for serum, two studies were included, with a pooled sensitivity of 0.73 (95% CI 0.65 to 0.80) and a pooled specificity of 0.72 (95% CI 0.63 to 0.79); and, for urine, two studies were included, with a pooled sensitivity of 0.52 (95% CI 0.27 to 0.76) and a pooled specificity of 0.72 (95% CI 0.36 to 0.92). Estimates of 95% prediction intervals for each test, medium and setting further demonstrated the heterogeneity in the studies considered.

Quality assessment of measurement

The defining features of ‘quality’ with respect to measurement procedures were agreed as ‘bias’, ‘reproducibility’ and ‘applicability’. Parameters associated with biological within-individual variation, biological pre-analytical factors, technical pre-analytical variation factors and analytical factors were included within the quality assessment framework.

Application of this framework within four Nephrocheck case studies identified several measurement parameters that present a high risk of irreproducibility, including a failure to exclude samples with known interferents; a lack of internal and external quality control; and a complete lack of analytical measurement verification in all studies. It also highlighted several issues that might affect the clinical applicability of test results, including freeze–thawing of samples in the absence of validation data and against the recommendations of the manufacturer, potentially biasing clinical cut-off points and overestimating precision; use of a device in an unvalidated patient population (i.e. aged < 18 years); and reporting the median value of three measurements from different laboratories. Furthermore, it identified several issues that made assessment of the risk of bias uncertain.

Economic evaluation

The economic evaluation assessed seven testing strategies: Nephrocheck, cystatin C in urine, plasma and serum and NGAL in urine, plasma and serum. Based on the mean expected cost and QALY results, lifetime incremental QALYs ranged from 0.012 (cystatin C, urine) to 0.016 (Nephrocheck) and additional costs ranged from £149 (cystatin C, urine) to £301 (Nephrocheck). The ICERs ranged from £11,476 to £13,504 per additional QALY for the cystatin C tests and from £13,372 to £13,828 for the NGAL tests and the ICER was £19,324 for Nephrocheck. For the incremental costs, all of the testing strategies had 95% CIs ranging from –£1000 to +£1400; for the incremental QALYs, all results ranged from –0.16 to +0.19 or +0.20. The overall probability that tests are cost-effective compared with standard care was 48% for Nephrocheck, 51–52% for the NGAL tests and 52–54% for the cystatin C tests (with cystatin C in serum performing the best). Raising the threshold value to £50,000 per QALY only slightly increased these probabilities.

The results of the multiway incremental comparison indicated that, in the base-case analysis, between a threshold of £11,400 and a threshold of £25,400, cystatin C (serum) has the highest probability of cost-effectiveness compared with all other tests. Above a £25,400 threshold, NGAL (serum) is expected to be the most cost-effective strategy. All other strategies are either dominated by cystatin C (serum) or, in the case of Nephrocheck, have an ICER well above £20,000 per QALY.

Similar results were observed in the post-cardiac surgery subgroup analysis. The incremental QALYs ranged from 0.007 (cystatin C, urine) to 0.012 (NGA,L serum) and additional costs ranged from £124 (cystatin C, urine) to £205 (Nephrocheck). The ICERs were £13,051–19,287 per additional QALY for the cystatin C tests, £15,337–20,435 for the NGAL tests and £18,617 for Nephrocheck, with INB values ranging from 0.000 to 0.004 QALYs. Again, there was substantial uncertainty around these results, with a 48–52% probability that the tests would be cost-effective. In the multiway incremental analysis, only NGAL (serum) remained after removal of dominated or extendedly dominated alternatives (ICER £13,051 vs. standard care).

The model results were highly sensitive to changes in key model parameters. Scenarios that led to tests becoming non-cost-effective included shortening the time horizon of the analysis, reducing the incidence of AKI in the ICU, decreasing the impact or increasing the cost of early AKI intervention, applying a mortality risk for patients with false-positive test results, applying a cost saving for patients with negative test results and increasing the mortality rate for false-negative cases and increasing the cost of Nephrocheck.

Value of information analysis

The EVPI was positive for all analyses of the three tests that were studied in detail, suggesting that the current burden of uncertainty is high and that further research into all tests and all settings may be worthwhile. Diagnostic accuracy parameters were not associated with a high EVPPI. The results of the EVPPI analysis suggested that the highest priority areas for further research include obtaining better intelligence on the current incidence of AKI in the ICU; the impact of interventions or changes in care and associated costs in response to test results; and further research on the quality of life experienced by survivors of ICU. Some of this information (e.g. the incidence and progression of AKI) may be obtainable from growing national audits. Determining the comparative impact of patient management changes resulting from different test results, however, is likely to require randomised comparisons.

Conclusions

It is clear that very large numbers of potential biomarkers and diagnostic tests have the potential to contribute to better care for patients at risk of AKI in the critical care setting. The two-stage approach taken in this study – first, horizon scanning and then prioritising biomarkers for in-depth review – proved to be a feasible and effective strategy for dealing with the volume of literature identified, which far exceeded that identified in initial scoping searches. Despite a large volume of literature covering the prioritised tests, the quality of reporting was low, leading to a significant dropout rate between review and meta-analysis. Further efforts to promote the use of reporting standards for diagnostic tests should be undertaken.

The Nephrocheck test appeared to be the best-performing test of the three tests subjected to detailed study. It has high sensitivity and moderate specificity for AKI in adult critical care settings and there was low heterogeneity between studies in the meta-analysis. The NGAL test using plasma has moderate sensitivity and high specificity, but showed greater heterogeneity between studies. Other sample types and the cystatin C test showed evidence of considerable heterogeneity between studies.

As far as we are aware, our study is the first such initiative to attempt a systematic assessment of the quality of measurement procedures used in clinical studies or trials. Our framework proved feasible and provides a foundation for further work in this area. A key finding was that the reporting of critical measurement parameters was very poor in the identified studies and this severely hindered the reviewers' ability to assess the quality of the studies. The major limitation of the meta-analysis was a lack of standardisation in the reference standard definition of AKI as an outcome.

Each of the three tests included in the economic evaluation was found to be cost-effective when compared in two-way analyses against standard care, although the absolute difference in both costs and QALYs is likely to be small. There is substantial uncertainty around these results, with the probability of cost-effectiveness being close to 50% for all tests. The VOI analysis highlighted some of the key parameters that should be the focus of further research in this area. It is apparent that observational studies that aim to better define the current clinical care pathway for patients at risk of AKI in critical care should be a priority, as should further work to understand how the care pathway might change in response to a positive test and the effectiveness of these changes in mitigating against the development of AKI. Such studies would likely be cheap to perform compared with formal randomised controlled trials that seek to directly measure the clinical benefit and observed cost-effectiveness of tests and that may not currently represent good value for research funders. It is also of note that further studies of diagnostic accuracy for the three tests would be unlikely to change the current estimates of cost-effectiveness.

Study registration

This study is registered as PROSPERO CRD42014013919.

Funding

The National Institute for Health Research Health Technology Assessment programme.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.236

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 13/116/13. The contractual start date was in October 2014. The draft report began editorial review in October 2016 and was accepted for publication in August 2017. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).