Methods for the evaluation of biomarkers in patients with kidney and liver diseases: multicentre research programme including ELUCIDATE RCT

Peter J Selby,[1,2]* Rosamonde E Banks,[1]
Walter Gregory,[3] Jenny Hewison,[4] William Rosenberg,[5]
Douglas G Altman,[6] Jonathan J Deeks,[7]
Christopher McCabe,[8] Julie Parkes,[9]
Catharine Sturgeon,[10] Douglas Thompson,[2]
Maureen Twiddy,[4] Janine Bestall,[4] Joan Bedlington,[11]
Tilly Hale,[11†] Jacqueline Dinnes,[7] Marc Jones,[3]
Andrew Lewington,[2] Michael P Messenger,[2]
Vicky Napp,[3] Alice Sitch,[7] Sudeep Tanwar,[5]
Naveen S Vasudev,[1,2] Paul Baxter,[12] Sue Bell,[3]
David A Cairns,[1] Nicola Calder,[2] Neil Corrigan,[3]
Francesco Del Galdo,[13] Peter Heudtlass,[3]
Nick Hornigold,[1] Claire Hulme,[4] Michelle Hutchinson,[1]
Carys Lippiatt,[14] Tobias Livingstone,[2] Roberta Longo,[4]
Matthew Potton,[3] Stephanie Roberts,[1] Sheryl Sim,[1]
Sebastian Trainor,[1] Matthew Welberry Smith,[1,2]
James Neuberger,[15] Douglas Thorburn,[16]
Paul Richardson,[17] John Christie,[18] Neil Sheerin,[19]
William McKane,[20] Paul Gibbs,[21] Anusha Edwards,[22]
Naeem Soomro,[19] Adebanji Adeyoju,[23]
Grant D Stewart[24,25] and David Hrouda[26]

[1]Clinical and Biomedical Proteomics Group, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK
[2]Leeds Teaching Hospitals NHS Trust, Leeds, UK
[3]Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK
[4]Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
[5]Institute for Liver and Digestive Health, Division of Medicine, University College London, London, UK

[6]Centre for Statistics in Medicine, University of Oxford, Oxford, UK
[7]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[8]Department of Emergency Medicine, University of Alberta Hospital, Edmonton, AB, Canada
[9]Primary Care and Population Sciences Academic Unit, University of Southampton, Southampton, UK
[10]Royal Infirmary of Edinburgh, Edinburgh, UK
[11]LIVErNORTH Liver Patient Support, Newcastle upon Tyne, UK
[12]Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK
[13]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK
[14]Department of Specialist Laboratory Medicine, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[15]University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[16]Royal Free London NHS Foundation Trust, London, UK
[17]Royal Liverpool and Broadgreen University Hospitals NHS Trust, Liverpool, UK
[18]Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
[19]Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
[20]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[21]Portsmouth Hospitals NHS Trust, Portsmouth, UK
[22]North Bristol NHS Trust, Bristol, UK
[23]Stockport NHS Foundation Trust, Stockport, UK
[24]NHS Lothian, Edinburgh, UK
[25]Academic Urology Group, University of Cambridge, Cambridge, UK
[26]Charing Cross Hospital, Imperial College Healthcare NHS Trust, London, UK

*Corresponding author p.j.selby@leeds.ac.uk
†In memoriam

Declared competing interests of authors: Peter J Selby has an international patent (application no. PCT/GB2014/050768; biomarker; University of Leeds) issued. Rosamonde E Banks reports grants from various non-commercial funding sources during the conduct of the study; in addition, she has a international patent (application no. PCT/GB2014/050768; biomarker; University of Leeds) issued. Jenny Hewison is a panel member of the National Institute for Health Research (NIHR) Clinical Trials Unit Standing Advisory Committee. Grant D Stewart has a patent (1408091.5) pending to the Medical Research Council (MRC) Human Genetics Unit (HGU). James Neuberger is a consultant to Astellas Pharma, Inc., and has received speaker support funding from Astellas and Novartis UK. Claire Hulme has received grants from the NIHR during the conduct of the study and is a panel member of the Health Technology Assessment (HTA) Commissioning Board. Andrew Lewington has received personal fees from Fresenius Medical Care, Baxter Healthcare Ltd, AM-Pharma, BioPorto Diagnostics A/S and GE Healthcare, outside the submitted work; in addition, he has a patent (ACY-1 biomarker) pending. Naveen S Vasudev has received personal fees from Bristol-Myers Squibb and GlaxoSmithKline plc, outside the submitted work. William Rosenberg was an inventor of the enhanced liver fibrosis (ELF) test when he was an employee of the University of Southampton. His rights were transferred to Siemens Healthcare Diagnostics Ltd by the University of Southampton. He does not receive any payment in relation to sales of the test by the manufacturer Siemens Helathcare Diagnostics. He has received grant support and speaker fees from Siemens Healthcare Diagnostics and is a director of iQur Ltd (Southampton, UK), a company that provides ELF testing.

In the context of this NIHR-funded study, all ELF testing provided by iQur Ltd was performed on a not-for-profit, cost-recovery basis. William Rosenberg is married to Julie Parkes, who is a co-investgator on the programme. Julie Parkes reports receiving support from the Speakers' Bureau for Siemens Healthcare Diagnostics, outside the submitted work, and is married to William Rosenberg (co-applicant). Jonathan J Deeks is a panel member of the HTA Commissioning Board.

Published June 2018
DOI: 10.3310/pgfar06030

Scientific summary

Evaluation of biomarkers in patients with kidney and liver diseases

Programme Grants for Applied Research 2018; Vol. 6: No. 3
DOI: 10.3310/pgfar06030

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

Scientific summary

Protein biomarkers in body fluids that have demonstrable associations with the activity and outcomes of a wide range of diseases are now being identified by modern proteomic technologies. They may be simple, accessible, cheap and safe tests that can inform diagnosis, prognosis, treatment selection, monitoring of disease activity and therapy. They may substitute for or augment more complex, invasive and expensive tests. However, their substantial potential to improve patient care and health service provision is not yet being realised because the pathway linking biomarker research to health services research is still quite poorly defined. Liver and renal diseases generate huge and growing patient and service burdens and are amenable to biomarker application.

Our programme consisted of three workstreams that relate to the development pipeline for new biomarkers in renal and liver diseases and aimed to create a framework for research and innovation in this area:

1. workstream 1, methodology – to define current best practice and explore innovations, particularly in relation to the use of biomarkers to monitor disease activity
2. workstream 2, clinical translation – to create and evaluate a framework of practice, samples and clinical data to rapidly identify protein biomarkers with the appropriate analytical and clinical validity and performance characteristics to justify evaluation of their clinical utility in the health service in liver and renal diseases
3. workstream 3, ELF to Uncover Cirrhosis as an Indication for Diagnosis and Action for Treatable Event (ELUCIDATE) randomised controlled trial (RCT) – a RCT on an established biomarker test, the ADVIA Centaur® Enhanced Liver Fibrosis (ELF) test (Siemens Healthcare Diagnostics Ltd, Camberley, UK), for liver fibrosis and cirrhosis, for which clinical evidence for its potential value in chronic liver disease (CLD) is excellent, to determine whether or not its use will sufficiently alter the diagnostic timing and subsequent management of cirrhosis of the liver in order to change the process of care and reduce serious complications and improve outcomes for patients and service provision.

We assembled an outstanding internationally recognised multidisciplinary team of methodologists, clinicians, clinical biochemists, statisticians and marker scientists to deliver these workstreams.

The methodology workstream evaluated published evidence on the quality of recommendations for using prostate-specific antigen (PSA) to monitor patients with prostate cancer, systematically reviewed the use of RCTs to evaluate monitoring strategies, reviewed the monitoring biomarker literature and how monitoring can have an impact on patient outcomes and conducted simulation studies to evaluate monitoring strategies and how monitoring strategies can meet the requirements to improve the value of health-care services. These studies confirmed that the literature on the use of biomarkers in monitoring diseases is modest in scale and robust conclusions are infrequent and we recommend improvements in research practice.

We considered the guidelines that are available for using PSA measurement to monitor patients after they have received either radical surgery or radical radiotherapy for localised disease. The guideline methods were assessed using a formal research evaluation framework, which examined the systematic search methods used in the studies, the selection criteria, the clarity of the formulation of recommendations, the consideration given in the recommendations to relevant issues around monitoring, the explicit nature of the use of evidence, the use of external review and the description of updating procedures. Of the nine main guidelines evaluated using an objective scoring system, the rigour of guideline development varied, with the best score obtained for the 2008 National Institute for Health and Care Excellence (NICE) guidelines. Only one guideline modified its recommendation to reflect the fact that a single PSA measurement may be technically unreliable and it did so by recommending retesting within 2 months. Three guidelines recommended the use of the same assay on every test occasion. Only four guidelines attempted to justify the interval between tests that they recommended. Overall, there was evidence of considerable inconsistency in guideline recommendations for the use of PSA measurement, even when they were published within a few

years of each other. We concluded that general failings in the guideline development process are likely to contribute significantly to the variations between published guidelines. Only the NICE and Australian Cancer Network guidelines cited handbooks on guideline development. It was notable that these guidelines scored relatively well on the evaluation instrument used in our study.

Randomised controlled trials of monitoring regimens are challenging to design and deliver. Such trials are complex and involve serial testing. There are complex interactions between repeated test results, clinical decisions based on these results, the response of clinicians to the results and, of course, the identification through long follow-up times of important patient outcomes. We conducted a methodological review of RCTs of monitoring. Although the target sample size was 60 RCTs, after a comprehensive search 120 titles were selected for further evaluation. Following full-text review, 49 trials published in 58 publications were selected for inclusion. Cancer, followed by cardiovascular disease and renal disease, were the most frequently reported topics. Half of the trials evaluated patient-related primary outcomes, one-third evaluated the impact on mortality and half aimed to report the impact of the monitoring strategy on the detection of new or recurrent disease. Process of care outcomes were evaluated primarily in relation to the number of patients treated in the different trial arms or the time taken to arrive at that treatment. Twelve trials reported statistically significant effects of monitoring on the primary outcome. Only limited attention was given to the test properties and intervention effectiveness in the populations of interest before the trials were undertaken. There was a lack of detailed description of the protocols for trial monitoring and considerable evidence for a lack of compliance to the monitoring strategies. The impact of the monitoring strategies on clinical behaviours, such as whether to administer treatment or withhold treatment, was not always consistent with the test results. It appeared that the monitoring test was treated by clinicians as a guide to possible changes rather than as a definitive indication for a particular change in care. There was an apparent lack of power to detect significant effects in the studies as a whole.

We reviewed the literature on monitoring strategies used to direct the care of patients with recurrent or progressive disease. After a formal search and filtering, the literature was categorised and tabulated. The review identified a limited amount of methodological literature on monitoring strategies. We then focused on the relationship between the monitoring care pathway and the points in that pathway where monitoring might be expected to affect patient outcomes. Three identified frameworks for this were reviewed. Clinical trials of relevance were grouped into three main categories: (1) new monitoring strategies vs. existing strategies; (2) a monitoring strategy vs. immediate treatment; and (3) a monitoring strategy vs. no monitoring. Differences in study design from the use of biomarkers for screening and diagnosis were evaluated. Monitoring strategies considered included (1) detection of significant clinical change earlier than in conventional practice to deploy treatment early; (2) to reduce the invasiveness and cost of testing; (3) to reduce the volume and frequency of testing; (4) to reduce overtreatment; and (5) to delay or avoid treatment. The analysis led to the recommendation that a test validation paradigm be adopted in which a number of methods are used to determine whether or not the results of a test are going to be meaningful in practice and generate benefits for patients. RCTs will be needed in some settings but this level of evidence will not always be essential. Strategic approaches need to be multidisciplinary, involving evaluation of the performance of the tests in the laboratory, rigorous study design and analysis and close collaboration with clinicians and biochemists to determine the appropriate technical and clinical options for evaluation and the probability of changing clinical behaviours with test results.

We describe simulation studies as well as the impact of the simulation studies on the conduct, redesign and extension of the ELUCIDATE trial. Data sources were not always adequate for comprehensive simulation until quite late in the progress of the programme and ELUCIDATE trial. The modelling work allowed the accurate calculation of power based on observed and predicted event rates. This allowed trial recruitment to be completed, the reporting of process of care outcomes and the initiation of long-term follow-up strategies from health-care informatics sources.

Introducing new biomarkers into clinical practice to promote the introduction of a more personalised, precise and stratified approach to patient management requires evaluation of the characteristics of the

tests and their impact on clinical outcomes and the quality and cost of the care delivered. There is tremendous pressure to increase the efficiency of health-care systems by introducing cost-effective new tests. The elimination of unnecessary tests is being explored. We focused especially on the role of monitoring tests and methods to evaluate their health economics. We compared the use of conventional clinical utilities with an approach based on cost-effectiveness, described the framework for characterising personalised medicine technologies and drew on an existing method for optimising diagnostic tests to meet cost-effectiveness targets and extended this to monitoring tests. This work demonstrated, among other things, that the cut-off points used for a test when used repeatedly for monitoring may under some circumstances be different from those used when the test is used for diagnosis.

The findings were formulated to be shared with patients, who strongly endorsed the need for robust and conclusive research in this area and for improved communication about test results between clinicians and patients.

The clinical translation workstream focused on the analytical and clinical validity of tests in renal disease. Prospective cohorts were established for renal cell carcinoma (RCC) and renal transplantation (RT), with samples and patient data obtained from multiple NHS centres and the samples and patient data from workstream 3 (liver disease) curated. The recruitment of patients to obtain high-quality samples and clinical data was challenging but was ultimately completed to target. These resources provide, and will continue to provide, a rapid-access resource for evaluating the validity of biomarkers that are candidates for evaluation to see whether or not they can improve NHS services. To identify tests to be evaluated using this resource, all candidate biomarkers for RCC and RT were identified from the literature, the quality of the studies was evaluated and selected biomarkers were prioritised. Four selected biomarkers were studied further by rigorous evaluation of the validity of the tests and evaluation of their performance within the workstream 2 sample/data cohorts. Systematic evaluation of tests relevant to RCC in the literature suggested that osteopontin (OPN), vascular endothelial growth factor (VEGF), carbonic anhydrase IX (CAIX) and C-reactive protein (CRP) should be prioritised and evaluated further. For RT, the most promising serum biomarkers for the early detection of delayed graft function appeared to be neutrophil gelatinase-associated lipocalin (NGAL), serum cystatin C and serum aminoacylase-1, previously discovered by our group. The performance of available assays for the four prioritised biomarkers (VEGF, CAIX, OPN and NGAL) was rigorously evaluated, including pre-analytical aspects and verification protocols. Therefore, specific biomarker technical evaluations were performed for all of the biomarkers studied within the programme. The important technical aspects of evaluating biomarker assays are illustrated in these studies as well as the critical importance of the principle that all assays must be technically robust before being employed in NHS diagnostics or in clinical trials. Without assay characterisation and validation as part of the early phase of biomarker translation the field will continue to move slowly and waste resources. High-quality biobanking and detailed consideration of pre-analytical factors are essential in this field. The four RCC biomarkers evaluated in the cohorts showed promise but, after multivariate analysis, at this stage we can demonstrate only that CRP has added value to the established panels of tests and clinical data (the Leibovich score) used in RCC practice. More importantly, however, we have demonstrated how to establish a streamlined approach to new biomarker validation. The duration of follow-up was a limitation of the cohorts but we were able to substantiate several existing findings and identify biomarkers that may be taken forward for clinical utility studies.

The ELUCIDATE trial workstream involved the design, conduct and analysis of a trial that registered 1303 patients with CLD and randomised 878 patients out of a target of 1000. The trial started late and recruited slowly initially. However, the trial team identified and opened additional centres, clinicians recruited patients energetically in most centres and new modelling techniques and data collection approaches were introduced by the team so that the trial ultimately recruited an adequate number of patients to provide good statistical power to answer the key clinical questions. Analysis showed that, within the trial, the use of the ELF monitoring strategy altered the patient process of care and led to the introduction of tests that will identify patients who should benefit from the early introduction of interventions to manage serious complications and improve outcomes. The ELUCIDATE trial was an 'exemplar' trial that has demonstrated the challenges of evaluating biomarker strategies in 'end-to-end'

RCTs, in which patients are randomised between a new monitoring strategy and conventional care and are then followed up through to ultimate end points including survival. Its lessons will inform future study design.

There were significant interactions between the three workstreams. Workstream 1 gave the programme investigators a clear insight into the historical, methodological and study design challenges within this field and the scope of previous contributions. Innovations in study design, simulation strategies and the applications of health economic methods to evaluating monitoring tests were developed. These informed the revision of the study design for the ELUCIDATE trial and provided innovative approaches to power calculations based on pre-existing published cohorts and the early trial data. Workstream 2 showed clearly the importance of rigorous assay evaluation. This informed the development of the ELUCIDATE trial and in particular the work to re-evaluate the performance of the ELF test in the context of intra-laboratory variation and inter-laboratory variation. The challenges of delivering the ELUCIDATE trial have informed our recommendations for future methodological approaches. The interactions between workstreams bring out the advantages of developing the clinical cohorts and conducting the RCT within the context of a programme, which included a strong multidisciplinary team of methodologists, clinical biochemists, triallists and clinicians. However, incorporating all three workstreams in a single programme also meant that the work of the programme as a whole was potentially prone to delays. For instance, delays in the recruitment of the cohorts and in the set-up of and recruitment into the ELUCIDATE trial limited our ability to feed back substantial bodies of data, and the outcome evaluation, from the RCT to the methodology workstream.

The findings from the three workstreams were used to synthesise a strategy and framework for future biomarker evaluation by the investigators, with a defined pipeline and innovative contributions to study design, health economics and health informatics, which became the basis of a successful application to become one of four National Institute for Health Research (NIHR) Diagnostic Evidence Co-operatives in England.

Trial registration

This trial is registered as Current Controlled Trials ISRCTN74815110, UKCRN ID 9954 and UKCRN ID 11930.

Funding

Funding for this study was provided by the Programme Grants for Applied Research programme of the NIHR.

Programme Grants for Applied Research

ISSN 2050-4322 (Print)

ISSN 2050-4330 (Online)

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 PGfAR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/pgfar. 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 *Programme Grants for Applied Research* journal

Reports are published in *Programme Grants for Applied Research* (PGfAR) if (1) they have resulted from work for the PGfAR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Programme Grants for Applied Research programme

The Programme Grants for Applied Research (PGfAR) programme, part of the National Institute for Health Research (NIHR), was set up in 2006 to produce independent research findings that will have practical application for the benefit of patients and the NHS in the relatively near future. The Programme is managed by the NIHR Central Commissioning Facility (CCF) with strategic input from the Programme Director.

The programme is a national response mode funding scheme that aims to provide evidence to improve health outcomes in England through promotion of health, prevention of ill health, and optimal disease management (including safety and quality), with particular emphasis on conditions causing significant disease burden.

For more information about the PGfAR programme please visit the website: http://www.nihr.ac.uk/funding/programme-grants-for-applied-research.htm

This report

The research reported in this issue of the journal was funded by PGfAR as project number RP-PG-0707-10101. The contractual start date was in January 2009. The final report began editorial review in February 2016 and was accepted for publication in June 2017. As the funder, the PGfAR programme agreed the research questions and study designs in advance with the investigators. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PGfAR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

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