Using simulation and machine learning to maximise the benefit of intravenous thrombolysis in acute stroke in England and Wales: the SAMueL modelling and qualitative study

Michael Allen,[1*] Charlotte James,[1] Julia Frost,[1] Kristin Liabo,[1] Kerry Pearn,[1] Thomas Monks,[1] Zhivko Zhelev,[1] Stuart Logan,[1] Richard Everson,[2] Martin James[3,4] and Ken Stein[1]

[1]Medical School, University of Exeter, Exeter, UK
[2]Computer Science, University of Exeter, Exeter, UK
[3]Royal Devon and Exeter Hospital, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
[4]University of Exeter, Exeter, UK

*Corresponding author m.allen@exeter.ac.uk

Declared competing interests of authors: Ken Stein has been a member of a number of National Institute for Health and Care Research (NIHR) committees (2011–present) and is currently Programme Director of the NIHR Systematic Reviews Programme and editor-in-chief of the NIHR Journals Library.

Published October 2022
DOI: 10.3310/GVZL5699

Scientific summary

SAMueL modelling and qualitative study
Health and Social Care Delivery Research 2022; Vol. 10: No. 31
DOI: 10.3310/GVZL5699

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

Scientific summary

Background

Stroke

Stroke is a medical condition where blood flow to an area of the brain has been interrupted, causing cell death. Stroke may be broadly categorised into two types: (1) ischaemic (i.e. due to an arterial blockage) and (2) haemorrhagic (i.e. due to bleeding). Stroke is a major cause of adult long-term disability and is a significant burden on health-care services. In England, Wales and Northern Ireland, 85,000 people are hospitalised with stroke each year.

Intravenous thrombolysis

Intravenous thrombolysis is a form of 'clot-busting' therapy developed to treat ischaemic stroke by removing or reducing the blood clot impairing blood flow in the brain. For ischaemic strokes, thrombolysis is an effective treatment for the management of acute stroke if given soon (i.e. within 4 hours) after stroke onset, and is recommended for use in many parts of the world, including Europe.

Based on expert clinical opinion, the Sentinel Stroke National Audit Programme (SSNAP) report, the NHS Long Term Plan and the specification of the Integrated Stroke Delivery Networks all provide a target of 20% of stroke patients receiving thrombolysis. Currently, 11–12% of patients in England and Wales receive thrombolysis, with significant inter-hospital variability (i.e. per-hospital thrombolysis use ranges from 2% to 24%).

Use of thrombolysis is, therefore, lower than target and is highly variable between hospitals.

Clinical audit

Clinical audit seeks to drive quality improvement through the measurement of clinical quality against evidence-based standards. The national audit covering stroke is SSNAP. SSNAP collects longitudinal data on the processes and outcomes of stroke care up to 6 months post stroke for more than 90% of stroke admissions to acute hospitals in England, Wales and Northern Ireland. Every year, data from approximately 85,000 patients are collected. SSNAP publishes quarterly and yearly analysis of results.

Objectives

We sought to enhance the national stroke audit by providing further analysis of the thrombolysis pathway at each hospital, identifying areas that would most improve the clinical benefit of thrombolysis, allowing quality improvement to focus on the most influential aspect.

Modelling of the stroke thrombolysis pathway

We modelled three aspects of the thrombolysis pathway:

1. Pathway speed, that is, what would be the effect of changing time from arrival to scan, or time from scan to treatment?
2. Determination of stroke onset time, that is, what would be the effect of changing the proportion of patients with determined stroke onset time?
3. Clinical decision-making, that is, what would be the effect of treating patients in accordance with decisions made at a benchmark set of hospitals?

Qualitative research

In addition to the modelling work, the project had a qualitative research arm, with the objective of understanding clinicians' attitudes to use of modelling and machine learning applied to the national stroke audit.

Methods

Data

Data were retrieved for 246,676 emergency stroke admissions to acute stroke teams in England and Wales between 2016 and 2018 (i.e. 3 full years). Units were included in the data set if they had at least 300 emergency stroke admissions with at least 10 patients receiving thrombolysis in the 3 years. This study included 132 units.

Modelling clinical decision-making

We modelled clinical decision-making, that is, whether or not a patient is predicted to receive thrombolysis, using machine learning methods (e.g. logistic regression, random forest and neural networks).

Modelling the stroke pathway

We modelled the stroke pathway using Monte Carlo simulation, sampling process times from distributions mimicking process speed at each hospital. Prediction of probability of good outcome was based on a meta-analysis of clinical trials and was based solely on age group and time from onset to treatment. The model was constructed in Python (Python Software Foundation, Wilmington, DE, USA).

Availability of model code

All code used in the project is available at URL: https://samuel-book.github.io/samuel-1/ (accessed 4 May 2022).

Qualitative research

During the project, we undertook interviews with groups and single clinicians (19 physicians in total). We collected data about physicians' backgrounds, their attitudes to thrombolysis and their understanding of variance, their perspectives on machine learning and potential loci for the implementation of machine learning feedback (within and beyond SSNAP). We established the physicians' views on possible unintended consequences that may result from changing the acute stroke pathway and potential means of mitigation.

Results

General descriptive statistics

- 94.7% of patients had an out-of-hospital onset of stroke.
- 11.8% of all arrivals who had an out-of-hospital onset of stroke received thrombolysis.
- 67% of all patients had a determined stroke onset time, 60% of whom arrived within 4 hours of known stroke onset.
- 40% of all arrivals arrived within 4 hours of known stroke onset.

Among those patients with an out-of-hospital onset of stroke who arrived within 4 hours of known stroke onset:

- The mean onset-to-arrival time was 111 minutes.
- Most (95%) received a scan within 4 hours of arrival, with a mean arrival-to-scan time of 43 minutes.
- Thirty per cent of those who received a scan within 4 hours of known stroke onset received thrombolysis.
- The mean scan-to-needle time was 40 minutes, the mean arrival-to-needle time was 63 minutes and the mean onset-to-needle time was 158 minutes.

Copyright © 2022 Allen *et al.* This work was produced by Allen *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Inter-hospital variation

- The proportion of patients receiving thrombolysis varied from 1.5% to 24.3% of all patients and from 7.3% to 49.7% of patients arriving within 4 hours of known stroke onset.
- The proportion of patients with a determined stroke onset time ranged from 34% to 99%.
- The proportion of patients arriving within 4 hours of known stroke onset ranged from 22% to 56%.
- The proportion of patients scanned within 4 hours of arrival ranged from 85% to 100%.
- The mean arrival-to-scan time (for those arriving within 4 hours of known stroke onset and scanned within 4 hours of arrival) ranged from 19 to 93 minutes.
- The mean arrival-to-needle time varied from 26 to 111 minutes.
- The proportion of patients aged ≥ 80 years varied from 29% to 58%.
- The mean patient National Institutes of Health Stroke Scale score (i.e. stroke severity) on arrival ranged from 6.1 to 11.7.

Relationship of time of day and day of week with use of thrombolysis

Nationally, thrombolysis use was significantly lower in patients arriving between 3 a.m. and 6 a.m. (with about 6% of patients arriving during this period receiving thrombolysis, compared with 11–13% of patients arriving at other times of the day); however, only about 3% of patients arrived in this period (in contrast to the 12.5% that would be expected if the arrival rate was uniform).

Nationally, there was a small relationship between day of week and use of thrombolysis, with thrombolysis use ranging from 11.2% to 12.6% by day of week (increasing Monday through to Sunday).

Machine learning for clinical decision-making

Machine learning (to predict the use of thrombolysis) was performed on the 40% of patients arriving within 4 hours of known stroke onset, and these were the patients who had a chance of receiving thrombolysis.

Machine learning performance

Machine learning accuracy was assessed using stratified k-fold cross-validation:

- The accuracy of machine learning models ranged from 81% to 86%, depending on model type. The model with the highest accuracy was a neural network using three embedding layers, one for each of (1) hospital identification, (2) clinical features of the patients and (3) pathway timings.
- There was generally higher agreement between model types (87–93% for any pairwise comparison) than between models and reality (83–86% accuracy for the same model types).
- Combining outputs of the three model types into one combined model (i.e. an ensemble model) did not improve accuracy.
- Accuracy of models increased with training set size up to about a training set size of 30,000 samples.
- All random forest and neural network models were well calibrated, meaning that 9 out of 10 patients with a predicted 90% probability of receiving thrombolysis would receive thrombolysis.

Comparing decisions between hospitals

For most modelling, we chose to use hospital-level random forest models. Although their accuracy is a little lower than that of the best models, these models are easier for people to understand and have strong hospital independence. When comparing predicted decisions between hospitals we found the following:

- It was easier to find majority agreement on who not to thrombolyse than on who to thrombolyse.
- A total of 77.5% of all patients had a treatment decision that was agreed by 80% of hospitals. Of patients who were not given thrombolysis, 84.6% had agreement by 80% of hospitals. Of patients who were given thrombolysis, 60.4% had agreement by 80% of hospitals.

- A comparison of a hospital's likelihood to give thrombolysis was made by passing a standard 10,000-patient cohort set through all hospitals. This evaluated likelihood to give thrombolysis independently from hospitals' own local patient populations.
- A benchmark set of hospitals was created by passing a standard 10,000-patient cohort set through all hospitals and selecting the 30 hospitals with the highest thrombolysis use. If all thrombolysis decisions were made by a majority vote of these 30 benchmark hospitals, then thrombolysis use (in those arriving within 4 hours of known stroke onset) would be expected to increase from 29.5% to 36.9%.
- Models may be used to identify two types of patients (and patient vignettes may be constructed to illustrate particular types of patients). The first group comprises patients for whom the model has high confidence in predicting, but in reality the patient was treated differently (e.g. a patient who appears to have high suitability for thrombolysis, but did not receive it). These patients may be good examples for more careful audit. The second group comprises patients who were treated in accordance with the prediction of the hospital model, but whom the majority of the benchmark hospitals would have treated differently. These patients exemplify where clinical decision-making appears to differ from benchmark hospitals.
- Hospitals may be grouped according to the proportion of patients that would be expected to have the same thrombolysis decision.
- Using embedding neural networks, similar patients (e.g. haemorrhagic stroke patients or patients with severe stroke) are located close together in embedding space. Patient embedding may also be used to rank patients in order of suitability for thrombolysis (by consensus across all hospitals). Similarly, hospital embedding may be used to rank hospitals by likelihood to use thrombolysis, independent of patient and pathway characteristics.

Clinical pathway simulation

The clinical pathway model was used to examine the effect, at each hospital and nationally, of making three key changes (alone or in combination) to the stroke pathway.

Speed

Speed sets 95% of patients as having a scan within 4 hours of arrival, and all patients as having 15 minutes from arrival to scan and as having 15 minutes from scan to needle.

Onset known

Onset known sets the proportion of patients with a known onset time of stroke to the national upper quartile if currently less than the national upper quartile (any hospitals that were greater than the upper national quartile were left at their current level).

Benchmark

The benchmark thrombolysis rate takes the likelihood to give thrombolysis for patients scanned within 4 hours of onset from the majority vote of the 30 hospitals with the highest predicted thrombolysis use in a standard 10,000 cohort set of patients. Benchmark hospitals are identified using hospital-level random forest models.

The results of making these changes were as follows:

- The model predicted current thrombolysis use with high accuracy (i.e. a R^2 of 0.980 and a mean absolute difference in thrombolysis use of 0.5 percentage points).
- Combining the three changes suggested that thrombolysis use could potentially be increased from 11.6% to 18.3% of all emergency admissions. In addition, the clinical benefit increased from 9.4 to 17.6 additional good outcomes per 1000 admissions. The main drivers in improvement in thrombolysis use were benchmark decisions, followed by determining stroke onset and then speed, whereas the main drivers in improvement in outcomes were speed, followed by benchmark decisions and then determining stroke onset.
- The model identified the changes that would have made the most difference at each hospital.

Copyright © 2022 Allen *et al.* This work was produced by Allen *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

In addition, we used the modelling to explore the most significant causes of inter-hospital variation in thrombolysis use (a subtly different question from asking about what will improve thrombolysis use most). The key findings were as follows:

- About half of the variance in current thrombolysis use was due to differences in local patient populations, and the other half was due to differences within hospitals (e.g. pathway speed, determination of stroke onset time, decisions to thrombolyse or not).
- Within the hospitals, the largest contributor to inter-hospital differences in thrombolysis use was differences in decision-making around thrombolysis, followed by speed and then determination of stroke onset time.
- If all hospitals treated a standard set of patients (reflecting national averages in patient population characteristics), then hospitals that had a lower thrombolysis rate would tend to have carried out more thrombolysis and hospitals with a higher thrombolysis rate would tend to have carried out less thrombolysis. However, this explained only about half of the differences between hospitals with low thrombolysis rates and hospitals with high thrombolysis rates.

Qualitative research

Qualitative research demonstrated a varying openness to machine learning and modelling techniques:

- Broadly, those units with higher thrombolysis use engaged more positively with the research, and those hospitals with lower thrombolysis use were more cautious.
- Clinicians from units with lower thrombolysis use tended to emphasise differences in their patients as the reason for lower thrombolysis. Clinicians in units with a middling use of thrombolysis tended to emphasise access to specialist resources as being key in being able to deliver thrombolysis well. Clinicians in units with higher thrombolysis use tended to emphasise the work and investment that had gone into establishing a good thrombolysis pathway.
- Clinicians wanted to see the machine learning models expanded to predict probability of good outcome and adverse effects of thrombolysis.
- Despite this being a small study, physicians engaged with the machine learning process and outcomes, suggesting ways in which the outputs could be modified for feedback to stroke centres and utilised to inform thrombolytic decision-making.

Limitations

Models may only be built using data available in SSNAP. Not all factors affecting use of thrombolysis are contained in SSNAP data, and the model, therefore, provides information on patterns of thrombolysis use in hospitals, but is not suitable for, or intended as, a decision aid to thrombolysis.

Public and patient involvement

Five stroke survivors or carers were involved in the project; however, because of the COVID-19 pandemic, three dropped out during the project.

Public and patient involvement proved useful to the project team (1) to help maintain focus on patients, (2) to help shape ways to explain the work and (3) because the dialectic process involved gave the researchers a deeper understanding of their own work (i.e. explaining things in simpler terms is hugely beneficial for the person doing the explanation, as argued by the late Nobel Prize winner for physics Richard Feynman).

Conclusion

Using modelling and machine learning, we identified potential for reaching close to the 20% target of thrombolysis use and for doubling clinical benefit from thrombolysis. The project is summarised diagrammatically in *Figure a*.

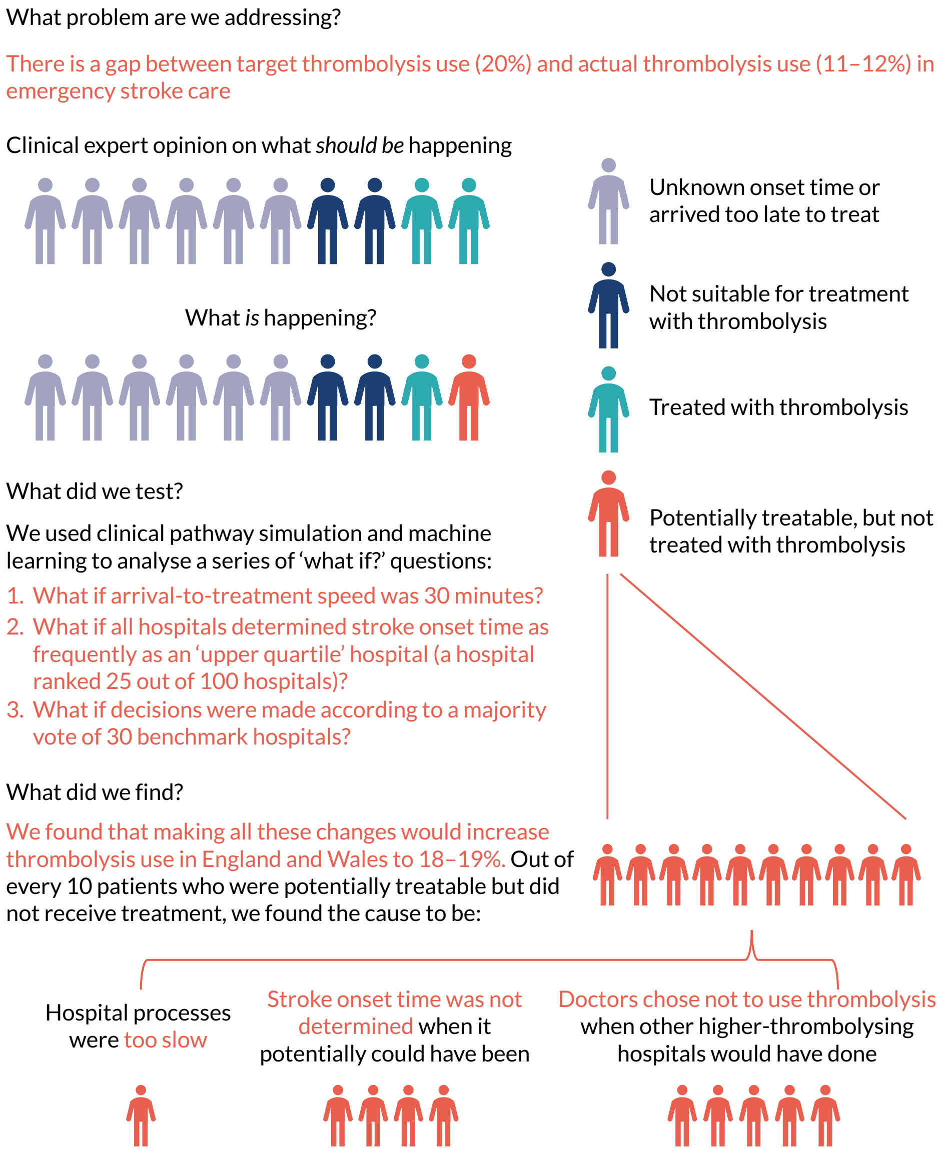

FIGURE a Project summary.

Implications for health care

- Realistically achievable thrombolysis use depends on local patient populations, and so a universal target of 20% across all hospitals may overestimate what is achievable at some hospitals, while also underestimating what is achievable at other hospitals. Local agreed targets may, therefore, be more appropriate.
- The tools developed here have the potential to add further depth of analysis to the national stroke audit outputs and may be transferable to other national clinical audits.

Copyright © 2022 Allen *et al.* This work was produced by Allen *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Recommendations for further research

- Expand machine learning to predict probability of a good outcome, and probability of adverse effects, of thrombolysis.
- Conduct further qualitative research with a more targeted approach to units with lower thrombolysis use or to groups of units (such as Integrated Stroke Delivery Networks) that include units with lower thrombolysis use.
- Expand the outputs of the models to incorporate health economic evaluation of changes. This will demonstrate benefits in health economic terms, such as quality-adjusted life-years, and will allow exploration of the cost-effectiveness of making organisational changes to the care pathway.
- Include organisational features (from the SSNAP Acute Organisational Audit) in machine learning models.
- Develop more methods to explain machine learning models (and the biases that have been learned).

Funding

This project was funded by the National Institute for Health and Care Research (NIHR) Health and Social Care Delivery Research programme and will be published in full in *Health and Social Care Delivery Research*; Vol. 10, No. 31. See the NIHR Journals Library website for further project information.

Health and Social Care Delivery Research

ISSN 2755-0060 (Print)

ISSN 2755-0079 (Online)

Health and Social Care Delivery Research (HSDR) was launched in 2013 and is indexed by Europe PMC, DOAJ, INAHTA, Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and NCBI Bookshelf.

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

This journal was previously published as *Health Services and Delivery Research* (Volumes 1–9); ISSN 2050-4349 (print), ISSN 2050-4357 (online)

The full HSDR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hsdr.

Criteria for inclusion in the *Health and Social Care Delivery Research* journal

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

HSDR programme

The HSDR programme funds research to produce evidence to impact on the quality, accessibility and organisation of health and social care services. This includes evaluations of how the NHS and social care might improve delivery of services.

For more information about the HSDR programme please visit the website at https://www.nihr.ac.uk/explore-nihr/funding-programmes/health-and-social-care-delivery-research.htm

This report

The research reported in this issue of the journal was funded by the HSDR programme or one of its preceding programmes as project number 17/99/89. The contractual start date was in February 2019. The final report began editorial review in August 2021 and was accepted for publication in February 2022. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HSDR 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.

This report presents independent research funded by the National Institute for Health and Care 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, the HSDR 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, the HSDR programme or the Department of Health and Social Care.

Copyright © 2022 Allen *et al.* This work was produced by Allen *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

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

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Professor of Digital Health Care, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HSDR, PGfAR, PHR journals) and Editor-in-Chief of HSDR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Consultant in Public Health, Delta Public Health Consulting Ltd, UK

Dr Peter Davidson Interim Chair of HTA and EME Editorial Board. Consultant Advisor, School of Healthcare Enterprise and Innovation, University of Southampton, UK

Ms Tara Lamont Senior Adviser, School of Healthcare Enterprise and Innovation, University of Southampton, UK

Dr Catriona McDaid Reader in Trials, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Emeritus Professor of Wellbeing Research, University of Winchester, UK

Professor James Raftery Professor of Health Technology Assessment, School of Healthcare Enterprise and Innovation, University of Southampton, UK

Dr Rob Riemsma Consultant Advisor, School of Healthcare Enterprise and Innovation, University of Southampton, UK

Professor Helen Roberts Professor of Child Health Research, Child and Adolescent Mental Health, Palliative Care and Paediatrics Unit, Population Policy and Practice Programme, UCL Great Ormond Street Institute of Child Health, London, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

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