Multiplex tests to identify gastrointestinal bacteria, viruses and parasites in people with suspected infectious gastroenteritis: a systematic review and economic analysis

Karoline Freeman,[1] Hema Mistry,[1] Alexander Tsertsvadze,[1] Pam Royle,[1] Noel McCarthy,[1] Sian Taylor-Phillips,[1] Rohini Manuel[2] and James Mason[1]*

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK
[2]Public Health England, London, UK

*Corresponding author J.Mason@warwick.ac.uk

Declared competing interests of authors: Sian Taylor-Phillips is partly supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West Midlands at the University Hospitals Birmingham NHS Foundation Trust.

Published April 2017
DOI: 10.3310/hta21230

Scientific summary

Multiplex tests to identify gastrointestinal bacteria, viruses and parasites

Health Technology Assessment 2017; Vol. 21: No. 23
DOI: 10.3310/hta21230

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

Scientific summary

Background

Gastroenteritis is a common, transient disorder usually caused by infection with viruses, bacteria or parasites and is characterised by the acute onset of diarrhoea with or without vomiting. Symptoms are mostly self-limiting, but severe diarrhoea can quickly cause dehydration. Patients may be managed in the community or admitted for observation, symptom management and diagnosis. Hospitalised patients with suspected infectious diarrhoea may be nursed in isolation until an infection has been ruled out or symptoms subside, but treatment is generally not recommended, except in certain situations (e.g. typhoid fever, *Clostridium difficile* infection or immunocompromised patients).

Identifying the infectious agent may aid decision-making in terms of treatment, isolation, management in the community or hospital, and further investigations for non-infectious causes of diarrhoea. Standard laboratory methods include cultures for bacteria, nucleic acid amplification for viruses and microscopy or enzyme immunoassays for parasites. Tests have turnaround times of up to 3 days and, in practice, the number of pathogens tested for is often restricted.

Gastrointestinal pathogen panel (GPP) tests exploit multiplex nucleic acid amplification methodology, testing for a wide range of bacteria, viruses and parasites in a single run, potentially increasing the throughput and volume of information from one test run and decreasing reporting times to ≤ 1 day.

Decision question

What is the clinical effectiveness and cost-effectiveness of the GPP panels xTAG® Gastrointestinal Pathogen Panel (Luminex, Toronto, ON, Canada), FilmArray Gastrointestinal Panel (BioFire Diagnostics, Salt Lake City, UT, USA) and Faecal Pathogens B (AusDiagnostics, Beaconsfield, NSW, Australia) in the identification of gastrointestinal bacteria, viruses and parasites in patients with suspected gastroenteritis presenting in primary or secondary care compared with conventional microbiological methods outlined in the Public Health England (PHE) algorithm?

Objective

To systematically review the evidence for the clinical effectiveness of the GPP tests (xTAG, FilmArray and Faecal Pathogens B), systematically review existing economic evaluations and develop a de novo economic model to assess the cost-effectiveness of GPP tests compared with the current standard of care in England and Wales.

Methods for clinical effectiveness and cost-effectiveness reviews

We searched MEDLINE, EMBASE, Web of Science and the Cochrane Database of Systematic Reviews from inception to November/December 2015 for clinical effectiveness and to January 2016 for cost-effectiveness. Supplementary searches of other online resources were run to check for other published and unpublished studies. Weekly autoalerts for emerging evidence were run in Ovid MEDLINE, Ovid EMBASE and PubMed until 30 April 2016. Reference lists of included studies and information provided by the manufacturers of the intervention tests were checked for additional eligible studies.

Two reviewers independently screened and assessed titles and abstracts of all records for inclusion using the following criteria:

- population – patients with acute diarrhoea with or without vomiting, thought to be a result of infective gastroenteritis, with test referrals from the hospital and community
- interventions – xTAG, FilmArray or Faecal Pathogens B
- comparator – standard microbiology techniques, outlined in the PHE syndromic algorithm for routine testing in cases of gastroenteritis and diarrhoea
- outcomes – any patient-/management-related outcome, test agreement, cost-effectiveness estimates
- study design – test–treat trials, clinical diagnostic test accuracy studies, studies comparing discrepant results between the index and comparator tests using a fair umpire test, studies of agreement and disagreement, studies of head-to-head comparisons of different index tests, full economic evaluations
- health-care setting – clinical laboratory receiving samples from primary and secondary care.

Quality assessment of eligible studies was undertaken using recognised checklists (tailored Quality Assessment of Diagnostic Accuracy Studies-2, Consolidated Health Economic Evaluation Reporting Standards and Philips).

We used the original 2 × 2 table data reported by studies without updating results from discrepant analyses. In the absence of a reference standard, we calculated levels of positive and negative agreements for each pathogen when benchmarked against either the comparator or GPP test, to determine the range of feasible outcomes of test agreement. Test agreement estimates were meta-analysed using a random-effects meta-analysis of proportions for each pathogen at the sample level.

A de novo decision tree model was built in Microsoft Excel® (version 16, Microsoft Corporation, Redmond, WA, USA) to assess the cost-effectiveness of GPP testing compared with conventional care. The base-case economic model included hospitalised adult patients with suspected gastroenteritis. The base-case model was adapted to look at various subgroups: young children, people in the community, immunocompromised patients and people with a recent history of travel. The data for the model included prevalence information from the systematic clinical effectiveness review, published literature and expert opinion. The model estimated the mean total costs and mean total quality-adjusted life-year (QALY) losses for each GPP test compared with conventional care for the initial index episode, and adopted a NHS and Personal Social Services perspective. Costs were in 2014/15 prices. Outcomes are reported as incremental cost-effectiveness ratios expressed in terms of cost per QALY gained. The model was run deterministically and probabilistically with 1000 bootstrapped iterations.

Results

Clinical effectiveness

The search identified 2215 unique records, 23 of which were included in the review. A total of 10 studies contributed 2 × 2 data to the meta-analysis. Studies were heterogeneous in terms of participants (hospital vs. community, risk, comorbidities), country of origin (developing vs. developed), conventional methods used, and number and type of pathogens considered. The review identified 17 studies that evaluated xTAG and four studies that evaluated FilmArray. Two studies compared both tests, but no study that assessed the Faecal Pathogens B assay was identified. The methodological quality of the included studies was poor. None of the studies used a reference standard against which the index tests and conventional methods could be reliably evaluated. Instead, in most studies, the index tests were compared with the conventional methods. Discrepant results between the index test and conventional testing were verified at the pathogen level in only 4 out of the 23 studies, although confirmatory tests were not adequately independent of index/routine tests. In many cases, the routine tests were not performed for all pathogens specified in the GPP test. There were concerns about the applicability and relevance of comparator and verification tests used in the majority of studies.

If conventional methods are an accurate determinant of clinically important disease, then meta-analytic results (benchmarked against conventional testing) suggest that GPP testing is a reliable test and could replace current microbiological methods, although there would be an increase in false-positive reporting with potential overdiagnosis and unnecessary treatment. However, if GPP testing is considered accurate (in the sense that current testing practices are missing clinically important pathology), then it would identify missed infections and could potentially result in more appropriate treatment.

Studies generally included mixed populations and aggregated findings. No subgroup analysis could be undertaken.

A small number of studies attempted to verify samples that did not agree when tested with GPPs and conventional methods. Verification methods were not independent of polymerase chain reaction methods and, broadly, would be expected to resolve discordant results in favour of GPP assays. As anticipated, discordant analyses of GPP-positive/conventional testing-negative samples generally favoured the GPP. No studies investigated discordant results using a 'fair umpire' test (imperfect but unbiased), such as exposure or effectiveness of different treatments.

In head-to-head comparison studies of different GPP tests, no concordance analyses were undertaken regarding whether or not tests identify the same pathogens in the same samples; therefore, no reliable assessment of the comparative performance of the different GPP tests was possible.

Studies reported the following secondary outcomes: GPP run failure ($n = 8$), multiple infections ($n = 23$), patient isolation ($n = 2$), turnaround times ($n = 8$), costs ($n = 3$) and other outcomes ($n = 4$). No study reported on change in management by test outcome, health-related quality of life, morbidity or mortality.

Cost-effectiveness

The systematic review of cost-effectiveness studies identified only one study considered to be a full economic evaluation comparing the xTAG GPP with conventional care in people with suspected infectious gastroenteritis. There were some notable limitations to the study, including use of estimated changes in isolation as a result of GPP testing (rather than observed changes), and no explicit valuation of patient outcomes.

For the base model (adult hospitalised patients) and base-case assumptions, there was considerable uncertainty about the cost-effectiveness of GPP tests. The pattern was similar for the other two hospital-based models involving young children and immunocompromised patients. These models found that varying pathogen prevalence does not significantly affect the cost-effectiveness of testing. Important uncertainties include length of stay and parameters that might influence this, such as false-positive findings. For both the general community model and the more specific recent travellers model, xTAG GPP appeared to be cost-effective, whereas FilmArray did not. In community models, without the potential for changes in hospital length of stay, cost-effectiveness is driven by the cost of the tests themselves and assumptions made when estimating these costs.

Discussion and conclusion

An evidential finding of the review is that GPP testing produces a greater number of pathogen-positive findings than conventional testing, but the clinical importance and consequence of these additional positive findings is uncertain. Evidence retrieved approximates to a mixed population of acute and hospitalised patients, as the studies do not adequately report outcomes by setting or subgroup. The available evidence is heterogeneous in populations studied, design, methods and analysis. The impact of GPP testing on patient management and outcomes, compared with conventional testing, has not been assessed, meaning that the economic modelling involves some key uncertainties and findings should be considered tentative.

Currently, the clinical importance of the additional pathogens identified by GPP testing is uncertain. Discrepancies between GPP tests and conventional methods may result not only from differences in accuracy, but also from differences in their targets. Without validation using a reliable reference standard, the status of additional GPP positives is uncertain; however, there remains a concern about the potential of GPP testing to identify non-viable pathogens.

A further uncertainty is the value of more rapid testing, achieved to a varying extent by each GPP system. In the vast majority of cases, hydration, hygiene and watchful waiting are required. Most cases of gastroenteritis are self-limiting and treatment is usually not indicated.

Strengths and limitations

Strengths of the work include a robust and comprehensive systematic review (literature search, data extraction and analysis) strategy and the building of a de novo decision tree model to assess cost-effectiveness.

No adequate test–treat trials or diagnostic studies using a reference standard were retrieved, and no studies conducted adequate discrepant analysis or applied a fair umpire approach. In the absence of adequate studies, positive and negative agreements were estimated alternating conventional and GPP testing as the benchmark as an aid to explore differences in findings. Agreement measures are not measures of test performance in a conventional sense, and only adequately designed research will resolve uncertainties about the introduction of GPP testing.

Although the economic model represented the clinical care pathway in the NHS, practice and management will vary from site to site. The economic model reflects one pattern of care for which patients are broadly tested in line with PHE guidance, although this practice is not universally followed. Different practices, such as sequential testing, would give rise to different patterns of cost and benefit.

Implications for health care

This review has evaluated GPP systems according to their current specification, but it is anticipated that the coverage of these systems will continue to evolve in response to changing pathogen prevalence; hence, the evaluation problem is a dynamic one. NHS organisation of pathogen testing may evolve in the coming decade and GPP testing technology may further develop, creating challenges for evidence requirements and timing of adoption decisions.

Research priorities

A randomised test–treat trial may be the best option, with patients randomised to conventional and GPP testing, and clinical effectiveness and cost-effectiveness outcomes used to determine the value of GPP testing. Such a trial might include adequate stratification of different populations (e.g. community managed, community acquired and hospital managed, hospital acquired, children, travellers and immunocompromised patients).

Study registration

This study is registered as PROPSPERO CRD42016033320.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

NICE TAR and DAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.058

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI 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 commissioned and funded by the HTA programme on behalf of NICE as project number 15/17/03. The protocol was agreed in December 2015. The assessment report began editorial review in June 2016 and was accepted for publication in October 2016. 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.

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