Evaluating Health Information Technologies: A systematic review of framework recommendations

Dr Matthew T Neame1,2, Ms Gerri Sefton2, Dr Matthew Roberts3, Dr David Harkness3, Dr Ian P Sinha3\* PhD, Dr Daniel B Hawcutt2,4\* MD

\*Joint senior author

Affiliations:

1: Department of Information Technology, Alder Hey Children’s Hospital, Liverpool, UK

2: Department of Women’s and Children’s Health, Institute of Translational Medicine, University of Liverpool, UK

3: Department of Respiratory Medicine, Alder Hey Children’s Hospital, Liverpool, UK

4: NIHR Alder Hey Clinical Research Facility, Liverpool, UK

### Corresponding author:

Daniel B Hawcutt, University of Liverpool, Alder Hey Children’s Hospital, Eaton Road, Liverpool, L12 2AP. Email: Dhawcutt@liverpool.ac.uk Tel: 0044 1512284811

Word Count: 3935

ABSTRACT

**Objective**

Evaluating Health Information Technologies (HITs) can be challenging, but studies are necessary so that the most beneficial interventions can be identified. Our objective was to systematically review the available recommendations for improving the methods used in HIT evaluations.

**Methods**

HIT evaluation frameworks were identified from database (MEDLINE, EMBASE, CINAHL) and grey literature searches. Outcome measures included framework recommendations and characteristics. Recommendations were coded and organised using thematic analysis methods. A scoring instrument was used to measure framework quality.

**Results**

The search identified 23 frameworks and 272 recommendations. These were organised into five evaluation domains and 42 themes. The themes included recommendations for improving the evaluation of technical aspects of HITs (e.g. describing aspects of HIT functionality) and suggestions for improving the evaluation of complex factors that may influence the overall effects of HITs (e.g. careful reporting of whether the HIT became integrated into existing working patterns). The frameworks were not generally developed in association with healthcare professionals, or with input from patients. The frameworks tended not to have been developed using systematic methods designed to reduce the risk of bias.

**Discussion**

HIT evaluations are important but they are challenging to conduct and appraise. This review was conducted using systematic methods enabling the organisation of framework recommendations into key themes. These findings may help investigators to successfully plan, conduct and appraise HIT evaluations. The quality appraisal demonstrated that HIT evaluation research may be improved by using more systematic methods and the involvement of participants from a range of differing backgrounds.

INTRODUCTION

Health Information Technologies (HITs) are computing systems used in the storage, retrieval, analysis and communication of health-related data 1-3. Internationally, billions of dollars have been invested in HITs4 5 and they are now routinely used to support the provision of healthcare6.

The aim of using HITs has been to increase the efficiency of health systems and to improve the outcomes experienced by patients and their families 5 7-11. However, it is not clear if these objectives have been achieved 12-15. Because of this uncertainty, ongoing evaluations of HITs are required. These evaluations will help to ensure that the most beneficial HITs are selected for use in clinical practice.

However, it can be challenging to design evaluations of HITs, particularly when they are introduced into hospitals. In these settings HITs may affect numerous processes, including the storage and retrieval of medical records, the scheduling of key interventions, the communication of health information, and the analysis of data to support clinical decision making. These functions may also be affected by other variables including; the integration with, and configuration of, existing software systems; levels of staff training and awareness of new technologies; behavioural interactions (between staff, patients and technologies); and the availability and suitability of the local hardware infrastructure. These factors can all act independently and interdependently to alter the effects of using seemingly similar HITs 16. As a consequence it can be difficult to reach generalisable conclusions about their effects.

Evaluation frameworks are theoretical models that “specify aspects of implementation that could be evaluated to determine implementation success”17 . As such, they may help to overcome some of the challenges associated with conducting HIT evaluations. Evaluation frameworks have previously been used to improve the investigation of complex, non-pharmacological interventions including operative procedures18, public health interventions19 and medical devices20 21 .

Although evaluation frameworks designed to improve the assessment of HITs have also been produced, it is not clear that they contain consistent advice and it is apparent that a variety of methods have been used in their development.

Our aim was to conduct a systematic review to identify and characterise existing HIT evaluation frameworks. Our objectives were to describe and to grade the quality of the methods that were used in their development and to produce a thematic analysis and narrative review of the recommendations included in the frameworks.

By achieving these objectives we hoped to identify key aspects of consensus and divergence in the available recommendations for evaluating HITs. By describing these factors we also hope to provide a useful resource for individuals who are tasked with designing HIT evaluations or conducting further research into methods for evaluating HITs.

METHODS

The systematic review was conducted in accordance with the Preferred Reporting Items for Systematic reviews and Meta‐Analyses (PRISMA) flow and checklist 200922. The protocol was registered with PROSPERO (PROSPERO record CRD42019128786).

**Search Strategy**

Searches were conducted of the EMBASE, MEDLINE and CINAHL databases from January 2000 until March 2019 (see supplementary data file for full details of search strategy). We also conducted a grey literature search by reviewing the websites of organisations including the National Institute for Health and Care Excellence (NICE) (UK), Medicines and Healthcare Regulatory Agency (MHRA) (UK) and the US Food and Drug Administration (FDA) (USA). The search strategy was designed to identify HIT evaluation frameworks.

In keeping with existing descriptions17 evaluation frameworks were defined as:

1. Documents developed with the aim of improving or standardising evaluations
2. That included recommendationsfor standardising or improving the evaluation of the effects of HITs

HITs were defined as computing systems used in the storage, retrieval, analysis or communication of health-related data1-3. We excluded frameworks that had been published before January 2000 and that had not been developed for use in hospital settings.

Title and abstract screening and full text reviews were conducted independently by two reviewers (MN & DH (Dr Harkness)). Where discrepancies were noted, consensus regarding final inclusion or exclusion was reached following discussion. Data were extracted and validated by two reviewers (MN and MR).

**Outcome Measures**

Primary outcome

The recommendations for evaluating the effects of HITs included in each evaluation framework were identified and transcribed onto a data collection sheet.

Secondary outcomes

Additional outcome measures were the characteristics of each evaluation frameworks including; the individuals involved in its development; the patient group for which it was intended to be used; and the methods used to search for evidence during the development of the framework (see table 1 for a full list of framework characteristics).

**Data Analysis**

The identified recommendations were reviewed independently by two investigators (MN & MR) who used a “theoretical thematic analysis” method to code the data23. This approach allows investigators to code data in terms of existing phenomena (e.g. “a Randomised controlled trial”, “report participant characteristics” or “calculate a sample size”). Once the data had been coded by each investigator, rounds of discussion were used to organise the coded data into a collection of themes which were used to inform the narrative review.

**Quality Assessment**

An adapted version of the AGREE II instrument24 25 was used to assess the quality of the included frameworks (see Supplementary Data File for a copy of the adapted instrument). The AGREE II instrument includes measures of guideline quality assessed across six domains including, stakeholder involvement, rigour of development, and clarity of presentation (see supplementary data file for an example of the adapted AGREE II data collection table). Items relating to each domain were scored independently by two reviewers. Scores were aggregated and presented as a percentage score for each domain in accordance with the AGREE II method.

RESULTS

We identified 23 frameworks for inclusion in the review (see figure 1 for PRISMA flow diagram; see supplementary data file for a list of articles/frameworks excluded following a review of the full text).

**Scope and Purpose**

The majority of frameworks (n = 17/23, 73.9%) were developed for use in unspecified healthcare settings (including hospitals). Four frameworks were designed to evaluate specific types of HIT (Electronic Health Records26, Health Information Exchanges27, Digital Behavioural Change Interventions28 and web-based/mobile technologies29), one was specifically designed for use in the United Kingdom30 and one was developed to support evaluation in publically funded, resource limited settings31. None of the frameworks were developed in order to benefit a specific health population.

**Stakeholder Involvement**

Academic researchers were involved in developing all of the identified frameworks (n = 23/23, 100%). Other groups involved in framework development included healthcare professionals (n = 7/23, 30.4%), technology developers/suppliers (n = 4, 17.4%), healthcare commissioners (n = 1, 4.3%), research funders (n = 1, 4.3%) and journal editors (n = 1/23, 4.3%) (see Table1 in the supplementary data file for details). A patient representative was reported to have been involved in the development of one framework (n = 1/23, 4.3%)32.

The majority of framework authors (n = 12/23, 52.2%) described their target audiences as individuals tasked with designing or conducting evaluations. Frameworks were also developed to support individuals tasked with appraising the quality of evaluations conducted by other individuals (n= 3, 13.0%) or with producing evaluation reports (n = 2/23, 8.7%). In some cases the target audience for the framework was not clearly stated (n = 4/23, 17.4%) (see Table 1 in the supplementary data file for details of stakeholder involvement).

**Development Methods**

Sources of evidence to inform framework recommendations included expert opinion (n = 14/23, 60.9%), literature reviews (n = 13/23, 56.5%), reflections on a case study (n = 1/23, 4.3%), a workshop (n = 1/23, 4.3%) and a systematic review (n = 1/23, 4.3%). The source of the evidence was not stated in two cases (n = 2/23, 8.7%) and nine frameworks reported using more than one method (n = 9/23, 39.1%).

Methods for translating the identified evidence into framework recommendations were described in a small majority of frameworks (n = 13/23, 56.5%). The methods used included additional workshops and/or informal discussions (n = 6/23, 26.1%), expert opinion (n = 4/23, 17.4%), a “stakeholder consultation” (n = 1/23, 4.3%), a Delphi process (n = 1/23, 4.3%) and a structured consultation (“saturation method”) (n = 1/23, 4.3%). None of the frameworks utilised a formal method for grading the quality of evidence that they used to inform their recommendations. The majority of frameworks (n = 20/23, 87.0%) were formally peer reviewed prior to publication (see Table 1 in the supplementary data file for details of the development methods used in the framework development process).

**Framework Recommendations**

The frameworks included 272 recommendations (see Tables 2-7 in the supplementary data file for a full list of the identified recommendations).

The scope of the recommendations included in the frameworks was broad; some were related specifically to the evaluation of HITs whilst others could be applicable to a range of health intervention evaluations.

During the thematic analysis, recommendations were grouped into five evaluation domains:

* Planning Evaluations
* Evaluation Methods
* The Selection & Measurement of Outcomes
* Reporting Recommendations
* Interpreting and Appraising Evaluations

**Planning HIT Evaluations**

The frameworks included 43 recommendations relating to the process of planning HIT evaluations and 7 themes were identified from these recommendations.

The themes that were most specifically relevant to planning HIT evaluations included giving careful consideration to the role of technology providers and outlining the expected functionality of HITs during the planning stages.

There were mixed recommendations in relation to the role of technology providers. Some frameworks included recommendations that described the benefits of involving these groups in evaluations whilst others warned of the potential for conflicts of interest.

Many of the identified themes could be characterised as general recommendations for improving the planning phase of healthcare intervention evaluations. Examples of these general recommendations included the need to identify appropriate resources and to carefully consider the objectives of the evaluation during the planning stages.

A full list of the identified themes with examples of the associated recommendations is provided in Table 1.

|  |  |  |
| --- | --- | --- |
| **Planning Themes** | **Number of related recommendations** | **Illustrative Recommendations** |
| Consider involving diverse groups of individuals, including technology providers, during the evaluation planning phase | 3 | Encourage those experiencing or producing health information technology-related changes to participate in designing, and where possible, carrying out evaluation of those changes.32 |
| Evaluation methods should be selected after consideration of the functionality and hypothesised effects of using the HIT | 4 | In order to assess impact, it is first necessary to inventory the functional capabilities that could affect quality or safety33 |
| Identify and allocate appropriate resources to conduct the evaluation (consider timeframes and financial factors) | 4 | Consider the impact of study design on relative cost and feasibility34 |
| Consider the following key factors when designing evaluation objectives:* Develop a clear research question/objective
* Identify a clear health need
* Identify an appropriate comparator group or source of baseline data
* Is it feasible to recruit an appropriate number of participants
 | 21 | [Develop evaluation objectives to] reach large and representative numbers of users, especially those who are most in need35 |
| Consider factors that may introduce a risk of bias into the evaluation at the planning stage, including: * Sources of funding and conflicts of interest
* The selection of evaluators
* Methods of concealment or blinding (where applicable)
 | 5 | Planning and execution of an evaluation should be based on professional expertise and be free from any political, managerial, or other pressure36 |
| Prepare an evaluation protocol prior to conducting an evaluation | 4 | Document all of your decisions and steps in a detailed study protocol. Adhere to this protocol; it is your main tool for a systematic evaluation37 |
| Record planned, ongoing, finalised and terminated evaluations using an appropriate registry | 2 | [Registries/databases of HIT evaluations] should contain information on planned, active and finalised (and also terminated) evaluation studies36 |

**Table 1. Recommendations for planning HIT evaluations**

**Study Design & Methods**

The frameworks included 39 recommendations relating to the study design and methods used in HIT evaluations and four themes were identified from these recommendations. The recommendations were principally themed according to whether they advocated the use of qualitative, quantitative or mixed-methods study designs. An additional theme was related to the observation that HIT interventions are often re-configured or updated during the evaluation process. Because of this, some frameworks included recommendations advising evaluators to be flexible about the methods and timings used during periods of evaluation. The identified themes are summarised in Table 2.

|  |  |  |
| --- | --- | --- |
| **Study Design & Methods Themes** | **Number of related recommendations** | **Illustrative Recommendations** |
| The effects of complex HIT interventions should be evaluated using qualitative methods | 9 | Use narrative as an analytic tool and to synthesise findings38 |
| Where feasible, quantitative/ experimental methods should be used to evaluate the effects of HITs | 11 | [A best practice evidence standard for new HITs would be a] high quality randomised controlled study30 |
| HIT evaluations should be conducted using both qualitative and quantitative (mixed) methods. | 10 | A variety of measures both quantitative and qualitative are required39 |
| HIT evaluations should be designed to be flexible and to adapt to re-configurations, updates and modifications to the HIT intervention being studied | 9 | Particularly needed are evaluation methods that capture how eHealth interventions evolve over time and the impact of these iterations35 |

**Table 2. Recommendations for Study Design and Methods used in HIT evaluations**

**Selection and Measurement of Outcomes**

The frameworks included 62 recommendations relating to the selection and measurement of outcomes and 16 themes were identified from these recommendations. The identified themes included recommendations to consult stakeholders before deciding on outcome measures; defining the outcome measures of interest and agreeing minimally important clinical differences prior to undertaking evaluations; and giving consideration to the appropriate timing of key outcome measurements.

The identified themes included recommendations to consider the measurement of 13 types of outcome in HIT evaluations. These included measures of usability and acceptability, measures of the effects of HITs on the quality of care (e.g. adherence to guidelines), measures of data quality; and the effects of HITs on healthcare related social interactions (see Table3 for a full list of the recommended outcome types). Six recommendations highlighted the importance of identifying adverse events and unintended consequences during HIT evaluations.

|  |  |  |
| --- | --- | --- |
| **Outcome Selection and Reporting Themes** | **Number of related recommendations** | **Illustrative Recommendations** |
| Justify the selection of outcome measures and consider their relevance to key individuals | 12 | The outcome measures reported should reflect best practice for reporting improvements in the specific condition, using validated outcome measures such as those in the COMET core outcome set.30 |
| Consider the timing of effects/outcome measures  | 4 | Consider also if the timing is adequate for measuring [outcomes/effects] as intended40 |
| Define key aspects of outcome measurement (e.g. minimally important clinical differences, proposed statistical analyses) prior to conducting the evaluation | 3 | Describe a detailed plan for each outcome measure (include timing and measurement instruments) and the proposed statistical analysis34 |
| **Recommended Outcome Measures** | **Number of related recommendations** | **Illustrative Recommendations** |
| Health resources (including the costs of utilising the HIT) | 8 | All costs associated with the implementation of ICTs should be recorded and compared with alternative (both existing and new) means of delivering care.41 |
| Patient/healthcare provider satisfaction | 6 | Health care professionals using computerised systems should find them both satisfactory and pleasant to use. Consequently, they should be asked generally about their experiences with the systems41 |
| Quality and process outcomes (e.g. appropriateness of care, adherence to guideline recommendations) | 6 | Evaluations should include an assessment of the effects on quality of care including the components quality of care26 |
| Adverse events and unintended consequences | 6 | Any unintended (positive or negative) side-effects of the system that were not in the focus of the study but that seem remarkable should be reported42 |
| Data quality and reliability (e.g. are data entered, stored and transmitted reliably) | 6 | Evaluations should include an assessment of the "Quality of Information Logistics" which includes the components; completeness/correctness of data26 |
| Changes in social interaction style (e.g. clinical consultations, communication between health professionals) | 2 | The impact of computerised systems on social interaction should be evaluated41 |
| Changes in access to care | 1 | The provision of ICTs [HITs] such as patient information systems and automated appointment systems should provide improved access to care for patients. This should be evaluated41 |
| Acceptability of the HIT (to patients and health professionals) | 1 | The level of acceptance of the technology concerned by both patient and professional should be investigated41 |
| Usability | 1 | [Evaluate] Will this system be easy to use?43 |
| Diagnostic reliability (when HIT systems have clinical decision support functionality) | 1 | Where diagnosis and treatment decision support tools are implemented, diagnostic reliability should be evaluated41 |
| Referral rates | 1 | The use of ICTs in health care may well have an effect on referral rates to other health care providers, pathology, etc. This needs to be evaluated.41 |
| Changes in healthcare provider IT skills | 1 | All health care professionals involved in the development and use of ICTS in health care should be asked whether they consider their involvement to have provided them with new skills41 |
| Configurability of the HIT | 1 | [Evaluate] Can the system be kept up or changed if necessary?43 |

**Table 3. Recommendations for Selecting and Reporting outcome measures in HIT evaluations**

**Reporting Recommendations**

The frameworks included 97 recommendations for improving the quality of reported HIT evaluations, and 7 themes were identified from these recommendations.

Two of the frameworks had been developed with the specific objective of improving the quality of HIT evaluation reports29 42. Both of these frameworks used established reporting guidelines (e.g. CONSORT44 and STARD45) as starting points for developing their recommendations. Consequently, many of the themes identified from recommendations included in these frameworks could be characterised as general recommendations for improving the quality of healthcare intervention reports. Examples of these more general themes included recommendations about the appropriate structure and content of evaluation reports, or suggestions for improving the quality of presentations of statistical data.

Alongside these general recommendations were others that were more specifically related to reporting HIT evaluations. These included the factors that need to be considered when reporting the characteristics of the settings, participants and interventions that were studied during HIT evaluations. Examples of these recommendation themes included; providing detailed descriptions of the overall IT infrastructure; providing descriptions of the technical and functional aspects of the HIT; and carefully reporting of any modifications or re-configurations that were made to the HIT during the evaluation period.

Recommendations for improving the descriptions of participant characteristics in evaluation reports included reporting baseline familiarity with using HITs; describing participant attitudes to the use of HITs; and reporting the type of training offered to and accepted by participants (including patients and healthcare professionals).

There were also recommendations to report the frequency with which the HIT interventions were used by participants, and to describe whether use of the intervention became integrated into standard working patterns in the intended way (intervention fidelity).

A full list of the identified recommendations for improving the quality of evaluation reports is available in the supplementary data file. A summary of the themes related to improving the quality of HIT evaluation reports are summarised in Table 4 below.

|  |  |  |
| --- | --- | --- |
| **Reporting Themes** | **Number of related recommendations** | **Illustrative Recommendations** |
| Consider reporting detailed characteristics of participants who used the intervention, including their attitudes and proficiency relating to the use of HITs. | 12 | [Consider reporting participant] attributes such as specialty, typing skills, EHR [Electronic Health Record] training and experience, and age33 |
| Describe the HIT in the context of the overall health IT infrastructure | 9 | Describe also how the object of study fits within the larger health IT environment of the organization or unit, and where relevant within the wider health sector40 |
| Describe the technical/functional aspects of the HIT in detail | 8 | Clarify and thoroughly describe the information technology which is the object of your evaluation37 |
| Describe any modifications or configurations that were made to the HIT. Report whether these changes were made during the course of the evaluation. | 4 | A description of changes to methods therefore also includes important changes made on the intervention or comparator during the trial (e.g., major bug fixes or changes in the functionality or content)29 |
| Report intervention fidelity (how well the intervention was adopted and used by healthcare providers/patients) | 12 | The description should also include information on (1) how wide-spread the system is used in the facility in which the system is evaluated, for how long and for what purpose and (2) number and professions of the users of the system in that facility42 |
| Describe the training that was provided to participants. | 3 | All technology specific training should be evaluated for appropriateness and applicability as well as the quality of the training provision itself41 |
| Use advice from reporting guidelines to ensure that evaluation reports are well structured and comprehensive and consider using adapted HIT reporting guidelines | 49 | The title should give a clear indication of the type of system evaluated, the study question and the study design.42 |

**Table 4. Recommendations for improving HIT evaluation reports**

**Interpreting and Appraising Evaluations**

Three frameworks30 31 46 included 39 recommendations relating to the interpretation or appraisal of HIT evaluations, and six key themes were identified from these recommendations.

Two of the frameworks presented their recommendations in the form of general prompts or pertinent questions designed to guide the appraisal of evaluations. Examples of the themes identified from these frameworks included recommendations to consider the certainty with which observed effects could be attributed to the evaluated intervention, and advice to appraise whether the results of the evaluation should prompt changes in clinical practice or research priorities.

The “NICE Evidence Standards Framework”30 provided a more structured approach for appraising evaluations. The framework’s authors proposed that appraisals should always be conducted in conjunction with a functional classification and risk assessment of the HIT being evaluated. They suggest that decision makers should seek to identify “higher levels” of evidence before appraising or commissioning HITs that are classified as being high risk, or that are designed to directly affect patient care or clinical decision making.

These recommendations are summarised according to the identified themes in Table 5 below.

|  |  |  |
| --- | --- | --- |
| **Interpretation and Appraisal Themes** | **Number of related recommendations** | **Illustrative Recommendations** |
| Consider whether evaluations answer specific research or clinical questions (e.g. should we use this intervention to improve a specific clinical outcome) | 5 | Evaluations should include an assessment of the question: Should We Change Clinical Practice?31 |
| Consider the confidence with which any observed effects can be attributed to the intervention under evaluation | 18 | [Consider] Was an appropriate study design used to answer study question? Do the chosen methods provide sufficient valid data to answer the study questions?46 |
| Consider how generalizable the results of the evaluation may be | 1 | [Consider] Is the unit of analysis appropriately chosen to answer the study question? Is the setting in which this study took place sufficient representative and appropriate to answer the study question?46 |
| During appraisals decision makers should formally classify a HIT’s functionality and the level of risk associated with its use | 3 | Use the best practice standards for DHTs [Digital Health Technologies] that present a higher potential risk within the tier. Use the minimum evidence standards for DHTs that do not present any specific risks.30 |
| Decision makers should appraise evidence from higher quality “interventional” evaluations (e.g. RCTs or quasi-experimental studies) before commissioning HITs that directly affect patient care or clinical decision making. | 5 | [If HITs are likely to affect treatment or diagnosis, identify] well-conducted meta-analysis of randomised controlled studies if there are enough available studies on the DHT (HIT) |
| Before commissioning low risk HITs that are designed to improve system services without directly affecting patient outcomes commissioners should consider pilot evaluations or case studies that demonstrate:* A plausible mode of action
* Acceptability to User Groups
* Technical Reliability
 | 7 | [A best practice standard for demonstrating] Acceptability with users [is] Published or publically available evidence to show that representatives from intended user groups were involved in the design, development or testing of the DHT30 |

**Table 5. Recommendations for improving the interpretation and appraisal of HIT evaluations**

**Framework Quality**

When the methods used to develop the frameworks were appraised they tended to score most highly on the “Clarity of Presentation” and “Scope and Purpose” domains (Medians scores (Inter Quartile Ranges (IQR)) 59.2% (52.8% - 68.1%) and 53.3% (43.1% - 62.5%) respectively). The frameworks scored least well on the “Rigour of Development” and “Stakeholder Involvement” domains (Median Scores (IQR) 38.9% (31.9% – 44.4%) and 38.9% (34.7% - 48.6%) respectively). The frameworks scored less well on the rigour of development domain because they were generally not developed using systematic methods to identify or select evidence. Stakeholder involvement scores tended to be lower because health professionals and patient and public representatives tended not to have been involved in the development of the frameworks.

The National Institute for Health and Care Excellence (NICE) Evidence Standards Framework30 scored most highly with an average score across the quality domains of 60.8%. The TEAM method for the evaluation of information systems in biomedicine39 scored least well with an average score of 32.3% across the quality domains.

DISCUSSION

HITs should be evaluated so that the most effective interventions can be selected for use in clinical practice. However, HITs arecomplex and evaluations can be difficult to plan, conduct and appraise. This review demonstrates that there are significant variations in the methods that are currently recommended for evaluating HITs. But it also highlights that healthcare practitioners and patients have not tended to be involved in developing the recommended evaluation approaches.

By using systematic review and thematic analysis methods we have constructed a database of the available recommendations for improving the evaluation of HITs in hospital settings. Examples of these recommendations include methods for improving the evaluation of technical aspects of HITs (e.g. measuring data quality or reliability and describing aspects of functionality) and suggestions for improving the evaluation of complex factors that may influence the overall effects of HITs (e.g. careful reporting of participant attitudes towards the use of HITs and measurement and reporting of how well the HIT became integrated into existing working patterns (fidelity)). We would suggest that these results provide a useful and previously unavailable resource for individuals involved in the planning, conduct and appraisal of HIT evaluations.

These results also highlight examples of contradictory advice in published HIT evaluation frameworks. In some cases, framework authors suggested that HITs are so complex that experimental evaluation methods should not be used to study their effects35 38. In other frameworks experimental methods were strongly advocated28 30, or “mixed-methods” approaches were advised34 47. Our results highlight this lack of consensus and demonstrate the need for individuals tasked with evaluating the effects of HITs to carefully consider the methods that they use.

We note that only one of the frameworks included a structured approach for appraising the quality of completed HIT evaluations (the NICE evidence standards framework30). Although general frameworks for appraising the quality of evidence are available (e.g. the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework48), the careful measurement and reporting of a number of HIT specific variables is also required in order for decision makers to draw reliable, generalizable conclusions from HIT evaluations. We would therefore suggest that the development of a structured framework for appraising factors including the characteristics of participants and health settings, the technical specifications of HIT interventions, and the validity of outcome measures may be beneficial.

The framework quality appraisal identified that the HIT evaluation frameworks were not generally developed in association with healthcare professionals, or with input from patients or the public; they also tended not to have been developed using systematic methods for identifying and selecting evidence. This finding may account for some of the variation in the frameworks because both of these approaches may help to reduce the risk of existing biases influencing the development of guideline and framework recommendations.

. This review has a number of limitations. There is no widely agreed definition of a HIT evaluation framework and we used an un-validated, consensus approach to identify publications for inclusion in the review. We also sought to identify frameworks that were designed to improve evaluations of the “effects” of HITs and for this reason we excluded some frameworks that were developed to support the evaluation of wider aspects of HIT implementation, including complexity, and predictors of adoption or “failure” 52 53. . Our use of thematic analysis was a pragmatic approach that enabled us to organise a broad range of data from the included frameworks. However, these methods have been criticised due to concerns that this approach generates results that are affected by the existing biases of researchers23. A limitation of our quality appraisal was that the AGREE-II instrument has not been specifically validated for use with evaluation frameworks25 , potentially limiting the validity of our conclusions about the quality of the frameworks.

Overall, we would suggest that our review highlights the need for further research in a number of areas. One area for particular consideration includes the need to consult closely with the individuals who use HITs (e.g. healthcare practitioners and those with health conditions). One approach may be to present the results of this review to a group of key individuals who use and evaluate HITs and to use a Delphi, consensus approach to identify which of the recommendations most helpfully represent their views on different aspects of evaluation. This approach would help to minimise bias by allowing different individuals to express their views without being influenced by other participants.

These methods could be particularly useful for developing an outcome reporting framework (such as a Core Outcome Set (COS)) or for the development of a guideline for improving the appraisal of HITs.

COS are agreed, standardised, outcomes that should be measured and reported in all trials in a particular research area54-56. COS have been demonstrated to ensure that studies report the most relevant outcomes; to reduce the risk of outcome reporting bias; and to facilitate meta-analysis57 58. But it is recommended that individuals with experience of different aspects of a health condition or setting are involved in their development to make sure that the recommended outcome measures are important to individuals from these key groups.

A guideline to support the appraisal of HIT evaluations could include recommendations relating to the most appropriate choices of study design or may support decision makers with reviews of “directness”. Directness refers to the certainty with which individuals can conclude that the results of an evaluation may also apply to the population or setting that they are interested in. This review highlights a number of key factors that may need to be considered in order to successfully appraise the participant and intervention characteristics that affect directness. However it would be helpful to present these factors to individuals who commission and use HITs in order to establish which features are most relevant to their decision making processes when they review HIT evaluations. We would suggest that the results of our review indicate the need for further research in this area.

**CONCLUSION**

HIT evaluations are important but they are challenging to conduct and appraise. Existing frameworks provide recommendations for improving the methods used in HIT studies but there is still uncertainty about which of the suggested approaches are the most useful. By using systematic methods to organise existing recommendations into key themes, this review provides a useful resource for individuals tasked with conducting and appraising HIT evaluations. This review also highlights the limitations of existing frameworks and helps to identify how future evaluation research may be improved through the use of more rigorous methods, and the involvement of a wider group of stakeholders.

## SUMMARY POINTS

* Evaluating HITs can be challenging; evaluators should use the best methods to meet their objectives
* Evaluations of HITs should report the overall health IT infrastructure, participant characteristics and technology adoption rates
* Research is needed to understand how to make HIT evaluations most useful for the people who use technologies in clinical practice

# FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

# COMPETING INTEREST STATEMENT

The authors have no competing interests to declare

# CONTRIBUTORSHIP STATEMENT

Dr Neame conceptualized and designed the study, collected data, drafted the initial manuscript, and reviewed and revised the manuscript.

Ms Sefton critically reviewed the manuscript for important intellectual content.

Dr Roberts helped to design the data collection instruments, collected data, contributed to the quality scoring process and reviewed and revised the manuscript.

Dr Harkness helped to design the search strategy, jointly conducted the searches and reviewed and revised the manuscript.

Drs Sinha and Hawcutt conceptualized and designed the study, coordinated and supervised data collection, and critically reviewed the manuscript for important intellectual content.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

# ACKNOWLEDGEMENTS

This is a summary of independent research carried out at the National Institute for Health Research (NIHR), Alder Hey Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the, NHS, the NIHR or the Department of Health.

Dr Hawcutt is part funded by the NIHR Alder Hey Clinical Research Facility. Dr Neame is funded by the NHS Digital, Global Digital Exemplar Programme.

No specific funding was received for the completion of this work. There are no conflicts of interest to declare.

1. (ONC) OotNCfHIT. Health IT Terms USA2017 [Available from: <https://www.healthit.gov/topic/health-it-basics/glossary> accessed 03.01.2019 2018.

2. (EPOC) CEPaOoC. Describing interventions in EPOC reviews. EPOC Resources for review authors, 2017 2017 [cited 2017 1.3.18]. Available from: <http://epoc.cochrane.org/resources/epoc-resources-review-authors>. .

3. Hersh W. A stimulus to define informatics and health information technology. *BMC Medical Informatics and Decision Making* 2009;9(1):24. doi: 10.1186/1472-6947-9-24

4. Wachter R. Making IT work: harnessing the power of health information technology to improve care in England. 2016

5. Blumenthal D. Launching HITECH. *New England Journal of Medicine* 2010;362(5):382-85. doi: 10.1056/NEJMp0912825

6. Henry J, Pylypchuk Y, Searcy T, et al. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015. In: Technology OftNCfHI, ed. Washnigton DC, USA: Office for the National Coordinator for Health Information Technology, 2016.

7. Oxtoby K. AI and robots will help NHS deliver better care, claims health secretary [Online]. The Pharmaceutical Journal; 2018 [updated May 2018. Available from: <https://www.pharmaceutical-journal.com/news-and-analysis/news/ai-and-robots-will-help-nhs-deliver-better-care-claims-health-secretary/20204778.article?firstPass=false> accessed 30.08.2018 2018.

8. Darzi A. The Lord Darzi Review of Health and Care: Interim Report London: Institute for Public Policy Research; 2018 [Available from: <http://www.ippr.org/research/publications/darzi-review-interim-report>.

9. Imison C C-CS, Watson R, Edwards N. Delivering the benefits of digital health care. In: Trust N, ed. London, UK, 2016:30-77.

10. Coordinator OotN. Official Website of The Office of the National Coordinator for Health Information Technology (ONC) - About ONC Washington DC, USA: Office of the National Coordinator; 2018 [Available from: <https://www.healthit.gov/topic/about-onc> accessed 19.02.2019 2019.

11. Blumenthal D. Stimulating the adoption of health information technology. *New England journal of medicine* 2009;360(15):1477-79.

12. Black AD, Car J, Pagliari C, et al. The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview. *PLOS Medicine* 2011;8(1):e1000387. doi: 10.1371/journal.pmed.1000387

13. Middleton B, Bloomrosen M, Dente MA, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. *Journal of the American Medical Informatics Association* 2013;20(e1):e2-e8. doi: 10.1136/amiajnl-2012-001458

14. Payne TH. EHR-related alert fatigue: minimal progress to date, but much more can be done. *BMJ Quality &amp; Safety* 2018 doi: 10.1136/bmjqs-2017-007737

15. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. *Annals of internal medicine* 2016;165(11):753-60. doi: 10.7326/M16-0961

16. Shcherbatykh I, Holbrook A, Thabane L, et al. Methodologic issues in health informatics trials: the complexities of complex interventions. *Journal of the American Medical Informatics Association : JAMIA* 2008;15(5):575-80. doi: 10.1197/jamia.M2518

17. Nilsen P. Making sense of implementation theories, models and frameworks. *Implementation science : IS* 2015;10:53-53. doi: 10.1186/s13012-015-0242-0

18. Hirst A, Philippou Y, Blazeby J, et al. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations. *Annals of surgery* 2019;269(2):211-20.

19. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *American journal of public health* 1999;89(9):1322-7. doi: 10.2105/ajph.89.9.1322 [published Online First: 1999/09/04]

20. Neugebauer EA, Rath A, Antoine S-L, et al. Specific barriers to the conduct of randomised clinical trials on medical devices. *Trials* 2017;18(1):427.

21. Smith R. Improving the evaluation and regulation of medical devices. *BMJ Opinion* 2018

22. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine* 2009;6(7):e1000097. doi: 10.1371/journal.pmed.1000097

23. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006;3(2):77-101. doi: 10.1191/1478088706qp063oa

24. Terrace L. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003;12:18-23.

25. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *Cmaj* 2010;182(18):E839-E42.

26. Sockolow P, Crawford P, Lehmann HP. Health services research evaluation principles. *Methods of information in medicine* 2012;51(02):122-30.

27. Dixon BE, Zafar A, Overhage JM. A framework for evaluating the costs, effort, and value of nationwide health information exchange. *Journal of the American Medical Informatics Association* 2010;17(3):295-301.

28. Michie S, Yardley L, West R, et al. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. *J Med Internet Res* 2017;19(6):e232. doi: 10.2196/jmir.7126 [published Online First: 29.06.2017]

29. Eysenbach G, Group C-E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal of medical Internet research* 2011;13(4):e126-e26. doi: 10.2196/jmir.1923

30. (NICE) NIfHaCE. Evidence standards framework for digital health technologies London, UK2019 [cited 2019 09.03.2019]. Available from: <https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies> accessed 09.03.2019 2019.

31. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches: Elsevier, 2016.

32. Darking M, Anson R, Bravo F, et al. Practice-centred evaluation and the privileging of care in health information technology evaluation. *BMC health services research* 2014;14(1):243.

33. Ancker JS, Kern LM, Abramson E, et al. The Triangle Model for evaluating the effect of health information technology on healthcare quality and safety. *Journal of the American Medical Informatics Association* 2012;19(1):61-65.

34. Cusack CM, Poon EG. Health information technology evaluation toolkit. *Rockville, MD: Agency for Healthcare Research and Quality* 2007

35. Glasgow RE. eHealth evaluation and dissemination research. *American journal of preventive medicine* 2007;32(5):S119-S26.

36. Ammenwerth E, Brender J, Nykänen P, et al. Visions and strategies to improve evaluation of health information systems: Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. *International journal of medical informatics* 2004;73(6):479-91.

37. Ammenwerth E, Gräber S, Herrmann G, et al. Evaluation of health information systems—problems and challenges. *International journal of medical informatics* 2003;71(2-3):125-35.

38. Greenhalgh T, Russell J. Why Do Evaluations of eHealth Programs Fail? An Alternative Set of Guiding Principles. *PLOS Medicine* 2010;7(11):e1000360. doi: 10.1371/journal.pmed.1000360

39. Grant A, Plante I, Leblanc F. The TEAM methodology for the evaluation of information systems in biomedicine. *Computers in Biology and Medicine* 2002;32(3):195-207.

40. Nykänen P, Brender J, Talmon J, et al. Guideline for good evaluation practice in health informatics (GEP-HI). *International journal of medical informatics* 2011;80(12):815-27.

41. Shaw NT. ‘CHEATS’: a generic information communication technology (ICT) evaluation framework. *Computers in biology and medicine* 2002;32(3):209-20.

42. Talmon J, Ammenwerth E, Brender J, et al. STARE-HI--Statement on reporting of evaluation studies in Health Informatics. *Int J Med Inform* 2009;78(1):1-9. doi: 10.1016/j.ijmedinf.2008.09.002 [published Online First: 2008/10/22]

43. Integrating Methods to Evaluate Health Information Systems. CSHI; 2015.

44. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine* 2010;8(1):18.

45. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Radiology* 2003;226(1):24-28.

46. De Keizer N, Ammenwerth E. The quality of evidence in health informatics: How did the quality of healthcare IT evaluation publications develop from 1982 to 2005? *international journal of medical informatics* 2008;77(1):41-49.

47. Lilford RJ, Foster J, Pringle M. Evaluating eHealth: How to Make Evaluation More Methodologically Robust. *PLOS Medicine* 2009;6(11):e1000186. doi: 10.1371/journal.pmed.1000186

48. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed)* 2008;336(7650):924-26. doi: 10.1136/bmj.39489.470347.AD

49. Research NIfH. Going the extra mile: Improving the nation's health and wellbeing through public involvement in research. *The final report and recommendations to the Director General Research and Development/Chief Medical Officer Department of Health of the ‘Breaking Boundaries’ strategic review of public involvement in the National Institute for Health Researc h* 2015

50. GREENHALGH T, JACKSON C, SHAW S, et al. Achieving Research Impact Through Co-creation in Community-Based Health Services: Literature Review and Case Study. *The Milbank Quarterly* 2016;94(2):392-429. doi: 10.1111/1468-0009.12197

51. Wilson P, Mathie E, Keenan J, et al. ReseArch with Patient and Public invOlvement: a realisT evaluation: the RAPPORT study. 2015

52. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res* 2017;19(11):e367. doi: 10.2196/jmir.8775 [published Online First: 01.11.2017]

53. Yusof MM, Kuljis J, Papazafeiropoulou A, et al. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). *International Journal of Medical Informatics* 2008;77(6):386-98. doi: <https://doi.org/10.1016/j.ijmedinf.2007.08.011>

54. Neame MT, Chacko J, Kirkham JJ, et al. Patient-focused outcomes are infrequently reported in pediatric health information technology trials: a systematic review. *J Clin Epidemiol* 2020;119:117-25. doi: 10.1016/j.jclinepi.2019.11.017 [published Online First: 2019/12/04]

55. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;13(1):132. doi: 10.1186/1745-6215-13-132

56. Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39-39. doi: 10.1186/1745-6215-8-39

57. Tugwell P, Boers M, Brooks P, et al. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials* 2007;8:38. doi: 10.1186/1745-6215-8-38 [published Online First: 2007/11/28]

58. Kirkham JJ, Davis K, Altman DG, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLOS Medicine* 2017;14(11):e1002447. doi: 10.1371/journal.pmed.1002447