Jeremy R. Huddy, Melody Ni, Shivani Misra, Stella Mavroveli, James Barlow and George B. Hanna*

Development of the Point-of-Care Key Evidence Tool (POCKET): a checklist for multi-dimensional evidence generation in point-of-care tests

https://doi.org/10.1515/cclm-2018-1089
Received October 7, 2018; accepted October 10, 2018

Abstract

Background: This study aimed to develop the Point-of-Care Key Evidence Tool (POCKET); a multi-dimensional checklist to guide the evaluation of point-of-care tests (POCTs) incorporating validity, utility, usability, cost-effectiveness and patient experience. The motivation for this was to improve the efficiency of evidence generation in POCTs and reduce the lead-time for the adoption of novel POCTs.

Methods: A mixed qualitative and quantitative approach was applied. Following a literature search, a three round Delphi process was undertaken incorporating a semi-structured interview study and two questionnaire rounds. Participants included clinicians, laboratory personnel, commissioners, regulators (including members of National Institute for Health and Care Excellence [NICE] committees), patients, industry representatives and methodologists. Qualitative data were analysed based on grounded theory. The final tool was revised at an expert stakeholder workshop.

Results: Forty-three participants were interviewed within the semi-structured interview study, 32 participated in the questionnaire rounds and nine stakeholders attended the expert workshop. The final version of the POCKET checklist contains 65 different evidence requirements grouped into seven themes. Face validity, content validity and usability has been demonstrated. There exists a shortfall in the evidence that industry and research methodologists believe should be generated regarding POCTs and what is actually required by policy and decision makers to promote implementation into current healthcare pathways.

Conclusions: This study has led to the development of POCKET, a checklist for evidence generation and synthesis in POCTs. This aims to guide industry and researchers to the evidence that is required by decision makers to facilitate POCT adoption so that the benefits they can bring to patients can be effectively realised.

Keywords: checklist; diagnostic equipment; diagnostic test approval; point-of-care testing.

Introduction

Point-of-care tests (POCTs) provide rapid results near patient or at the bedside to facilitate real time clinical decision-making. Well-established examples in clinical practice include blood glucose measurement, monitoring of anticoagulation and pregnancy testing. Technological advances have allowed more tests that were once performed in specialised laboratories, to be undertaken at the bedside, in primary care or in the patient home [1]. A POCT approach offers potential advantages such as decreased time to definitive treatment, improved cost-effectiveness, reduced training requirements and increased patient satisfaction [2]. However, barriers to the implementation of POCTs exist including the increased cost on a test-by-test basis, reduced accuracy and the associated maintenance and governance responsibilities. Many such barriers may be overcome if better evidence of their impact to clinical pathways was available. The POCT diagnostic industry is expanding rapidly with an estimated 29% share of the in-vitro diagnostic market and is predicted to be worth US$24 billion by 2018 [3]. It is important that this proliferation is underpinned by a robust methodological pathway that

***Corresponding author: Professor George B. Hanna,** PhD, FRCS, Department of Surgery and Cancer, Imperial College London, Academic Surgical Unit, 10th Floor, QEQM Building, St Mary's Hospital, South Wharf Road, London, W2 1NY, UK,
Phone: +44 203 312 2124, Fax: +44 203 312 6309,
E-mail: g.hanna@imperial.ac.uk
Jeremy R. Huddy, Melody Ni and Stella Mavroveli: Department of Surgery and Cancer, Imperial College London, Academic Surgical Unit, 10th Floor, QEQM Building, St Mary's Hospital, London, UK.
http://orcid.org/0000-0001-7356-848X (M. Ni)
Shivani Misra: Diabetes, Endocrinology and Metabolism, St Mary's Hospital, Imperial College London, London, UK
James Barlow: Imperial College Business School, South Kensington Campus, London, UK

evaluates these devices in a valid, efficient and timely manner.

Quality evidence on POCTs is required for regulation, policy making and implementation. Evidence generation must be centred around the clinical pathway so that healthcare benefits, consequences of misdiagnosis and the economic impact can be properly evaluated. Accumulating appropriate evidence is an expensive, time-consuming process that results in a substantial lead-time from innovation to clinical adoption, with some technologies never making the transition. Many pathways for the evaluation of diagnostic tests have been proposed. Lijmer et al. summarised these in their systematic review [4] highlighting that the process of diagnostic test evaluation may be cyclical in nature and evidence may not need to be acquired in a longitudinal order. Currently, the pathway for evidence generation in diagnostics is fragmented and does not follow the linear sequence of evaluations that has become mainstream for drug evaluation. Diagnostic accuracy is often established with poor methodological quality [5] and alone rarely provides sufficient justification for device adoption. Evidence is required to demonstrate the test inducing a clinical decision that subsequently affects patient outcome. Ensuring devices are easy to use and economically viable is usually performed in isolation or retrospectively, limiting their applicability. Given the considerable overlap in these methodology work streams, a multidimensional concurrent approach to evidence generation may be able to make significant efficiencies to the evidence generation pathway and avoid repetition. Improvements to the evaluation process of new diagnostic technology has been recognised as a priority need internationally. This has led to the creation of several working groups aiming to address this including the NIHR Medtech and *in vitro* diagnostics Co-operatives (MICs) [6] and the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group: Test Evaluation [7].

Tools to improve evidence generation in diagnostic evaluation include the standards for reporting of diagnostic accuracy studies (STARD) [8] initiative to improve accuracy and completeness in diagnostic accuracy studies, the EFLM practical guide for identifying unmet clinical needs for biomarkers [9] and the Consolidated Health Economic Evaluation Reporting Standards statement (CHEERS) [10] for reporting health economic evaluations. This is paralleled in the design literature by standards published by the International Organisation for Standardisation (ISO) and British Standards (BS) including BS EN 62366 [11] that relates to usability engineering in medical devices. The National Institute for Health and Clinical Excellence has published a Diagnostics Assessment Programme Manual [12] outlining what evidence is required for their appraisal process and how this should be synthesised. However, all these standards and guidelines are specific to a particular evidence domain or stakeholder group and there is an absence of an integrated framework to present the multi-dimensional evidence "package" that is required by decision makers when evaluating new POCTs.

The aim of this study was to develop the Point-of-Care Key Evidence Tool (POCKET). This is to be a multi-dimensional checklist to guide the evaluation of POCTs incorporating validity, utility usability, cost-effectiveness and patient experience. The motivation for POCKET is to improve POCT adoption by providing an aid for industry and researchers when developing and evaluating POCTs so that the benefits of these devices to patients can be effectively realised.

Materials and methods

A mixed qualitative and quantitative approach was adopted. Following a literature review a Delphi consensus process was undertaken to gain an expert group consensus as to the ideal evidence requirements to support adoption decisions regarding POCTs. The methodology was modified to include an initial semi-structured interview study that allowed an in-depth exploration of how evidence is currently generated and used by different stakeholders. Emergent themes relating to evidence were then translated into a traditional Delphi questionnaire format with overall results discussed in an expert workshop. The study protocol, including how participants were recruited and interview topic guides, has been previously published [13] and is summarised in Figure 1. The protocol was approved by the Joint Research Compliance Office Imperial College, London, and the Imperial College Hospitals National Health Service (NHS) Trust Research and Development Department (ICREC References 14IC2186 and 14SM2190). The study has been reported

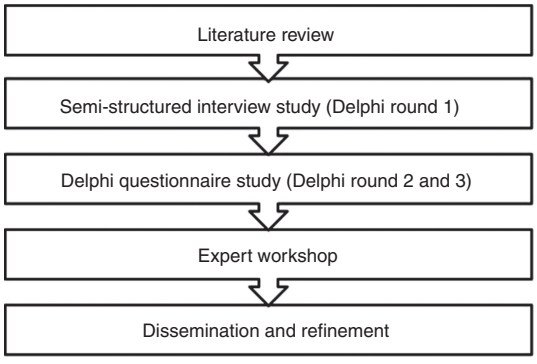

Figure 1: Methodological outline for development of POCKET checklist.

in keeping with the Consolidated Criteria for Reporting Qualitative Studies (COREQ) checklist [14]. Qualitative data was analysed with NVivo V.10.1.1 software (QSR International, Melbourne, Victoria, Australia). Statistics were undertaken using the Statistical Package for the Social Science version 22.0 (SPSS, Chicago, IL, USA).

The study was undertaken as described in the published protocol [10] with the exception of the following amendments. In Delphi rounds two and three the commissioner and regulator groups were amalgamated given their similar approach to evidence utilisation. A further amendment was made to the format for the expert workshop as once the complexity and length of the provisional POCKET checklist were appreciated it was felt that the planned blinded validation exercise would not be feasible. Instead participants were asked to apply the checklist to a POCT they were already familiar with. Therefore, the workshop had three aims: (1) assess the face validity (the degree to which the checklist is subjectively viewed as achieving its intent) of the POCKET checklist; (2) assess the usability of the POCKET checklist; (3) discuss the next steps in refining and disseminating the POCKET checklist. The workshop included a presentation of the aims of the study, the emergent themes from the semi-structured interview phase and Delphi questionnaire results. Usability of the checklist was evaluated by asking the group questions from the Usability Metric for User Experience (UMUX). UMUX is a concise series of questions that has been reported to be a reliable, valid and sensitive metric to measure usability (based on the ISO 9241-11 [15] definition) and user-experience [16]. Minor amendments were made to the tool based upon the group's feedback.

	Evidence Requirement	Cl	C&R	M	I
Technical Description of Test	Size and weight of test device	✓	✓	✓	✓
	Power source of test device (including details about mains supply, battery life, charging time etc)	✓	✓	✓	✓
	Details of equivalent laboratory test (turnaround time/cost/availability)	✓	✓	✗	✓
	Associated equipment required to perform test (device/cartridges/other consumables)	✓	✓	✗	✓
	Turnaround time for a single test	✓	✓	✓	✓
	Maximum throughput of test device (number of tests able to be performed over a given time period)	✗	✓	✓	✗
	Description of the sample collection process	✓	✓	✓	✓
	Description of the test process	✓	✓	✓	✗
	Description of how results are presented to user (including whether units are the same as current laboratory test)	✓	✓	✓	✓
	IT System Interoperability	✓	✓	✓	✓
	Date regulatory approval obtained (CE mark/FDA)	✓	✓	✗	✓
Clinical Pathway	Test Indication and Function (eg. diagnosis/risk prediction/monitoring)	✓	✓	✓	✓
	Clinical need for test	✓	✓	✓	✓
	Description of indicated population	✓	✓	✓	✓
	Intended setting for test	✓	✓	✓	✓
	Description of the intended user	✓	✓	✓	
	Rationale for point-of-care strategy	✓	✓	✓	✓
	A written description or diagram of the current clinical pathway	✓	✓	✓	✓
	A written description or diagram of the clinical pathway incorporating the new test device	✓	✓	✓	✓
	A description of how the clinical pathway is changed by incorporating the test device	✓	✓	✓	✓

Figure 2 The POCKET checklist.
Cl, clinician; C&R, commissioner and regulator; M, methodologist; I, industry; tick, evidence requirement; cross, evidence not required.

	Consequences of the test result to patient. Including description of effectiveness of any treatment instigated as a consequence of test result; Including any patient counselling required	✓	✓	✓	✓
	Consequences of incorrect test result to patient	✓	✓	✓	✓
	Advantages and disadvantages of POC test pathway to the patient	✓	✓	✓	✓
	Advantages and disadvantages of POC test pathway at an institutional or regional level	✓	✓	✗	✓
	Guidelines that incorporate test device	✓	✓	✗	✓
Stakeholders	Description of patient acceptability and their attitudes to test (including how this was determined)	✗	✓	✓	✗
	Stakeholder Analysis (identification of individuals/groups likely to be affected by test adoption, the impact of adoption and their attitudes)	✓	✗	✓	✗
Economic Evidence	Cost of test Including: cost of device, cost of extra, equipment needed to perform, test/store test, consumables, any other costs eg,, including capital, costs, other fixed, costs, variable, costs and professional costs	✓	✓	✓	✓
	Evaluation that compares costs before and after introduction of test to a clinical pathway	✓	✓	✓	✗
	An economic analysis with quality adjusted life years (QALY) and an incremental cost effectiveness ratio (ICER)	✗	✓	✗	✗
	Ability to incorporate local population data into economic model/analysis	✓	✓	✓	✗
Test Performance	Sensitivity and specificity of test device in an optimized or laboratory setting (Sensitivity – proportion of people with disease who have a positive test result; specificity - proportion of people without disease who have a negative test result)	✓	✗	✗	✓

Figure 2: (continued)

Results

Delphi round one: semi-structured interviews

A total of 41 semi-structured interviews were undertaken between August 2014 and December 2015. This included 43 stakeholders (regulators [n = 10], industry [n = 9], commissioners [n = 8], clinicians [including doctors, laboratory staff and point-of-care committee members] [n = 8], patients [n = 8]). Recruitment rates from invitation were 80% for clinicians and 29% for commissioners. Regulators, industry representatives and patients were recruited through bodies external to the study group and recruitment rate is not known. Twenty-three (55%) interviews were undertaken face to face and 19 (45%) by Skype (Microsoft, Seattle, WA, USA) or telephone. Interviews had a median length of 31 min and 21 s (range 11 min and 50 s – 58 min and 20 s). Thematic saturation (the point at which no new concepts emerge from the data) was achieved and revealed a number of common themes relating to

	Sensitivity and specificity of test device in a real world or clinical setting (Sensitivity - proportion of people with disease who have a positive test result; specificity - proportion of people without disease who have a negative test result)	✓	✓	✓	✓
	Negative and positive predictive value of test results (negative predictive value is the probability that people with a negative test don't have the disease, positive predictive value is the probability that people with a positive test have the disease)	✓	✓	✓	✓
	Receiver Operating Curve (ROC) and Area Under Curve (AUC) analysis of continuous diagnostic test results (ROC curve analysis is used to quantify how accurately medical diagnostic tests can discriminate between people with disease and people without disease; the AUC is the average sensitivity at all possible values of specificity and can be used to select optimum diagnostic thresholds)	✓	✗	✗	✗
	Reproducibility of test result (the variation in measurements made in changing conditions eg due to different instruments, different users, different period of time)	✓	✓	✓	✓
	Repeatability of test results (the variation in repeat measurements made on the same subject under identical conditions)	✓	✓	✓	✓
	Biological variation of test results (the natural fluctuation of test biomarker between a person and between different people)	✓	✓	✓	✓
	Evidence that diagnostic/therapeutic thresholds are the same as in laboratory/gold-standard test results	✓	✓	✓	✓
	Overview of alternative/competitor devices available and how they compare	✓	✗	✓	✗
Usability and Training	Suggested standard operating procedure for test device and process	✓	✓	✓	✓

Figure 2: (continued)

the evaluation of POCTs. One theme relating to POCTs in general and six relating to evidence were identified (clinical needs assessment, clinical pathway and utility, validity, usability, cost-effectiveness and efficiency in evidence generation). A summary of all emergent themes with quotations to illustrate each is presented in Table 1. However, only themes relating to device evaluation and evidence were carried forward to round two.

Delphi round two and three: summarising opinions and reaching consensus

The thematic analysis of the semi-structured interviews highlighted 68 different evidence requirements that could be used by policy or decision-makers when considering the adoption of a POCT. Each of these was included as a statement in the Delphi questionnaire. The number of

	Instructions appropriate to end user (written/internet/DVD etc)	✓	✓	✓	✓
	Identification of operator dependent steps	✓	✓	✓	✓
	Training requirements for undertaking sampling procedure (including provider)	✓	✓	✓	✓
	Training requirements for using the test device	✓	✓	✓	✓
	Training requirements to interpret the results	✓	✓	✓	✓
	Suggested method(s) for competency assessment	✓	✓	✗	✓
	Potential risks of test procedure to patient	✓	✓	✓	✓
	Potential risks of test procedure to personnel performing test	✓	✓	✓	✓
	Sample disposal procedure including sharps	✓	✓	✗	✗
	Test device calibration procedure and internal quality control protocol including level of expertise required to perform	✓	✓	✓	✓
	Test device maintenance required and level of expertise necessary to perform	✓	✓	✓	✓
	Support infrastructure provided (eg service, agreements, helpline, website)	✓	✓	✓	✓
	Reported adverse events	✓	✓	✓	✓
	Description of robustness of test device	✓	✓	✗	✗
Clinical Trials	Clinical Trial Results (including funding, limitations and description of potential sources of bias in clinical trials undertaken)	✓	✓	✓	✗
	Linked evidence approach (the synthesis of acquired evidence on test accuracy, impact of decision making and effectiveness of consequent treatment to evaluate overall test effectiveness)	✓	✓	✓	✗
	Diagnostic accuracy study	✓	✓	✓	✓
	National/Regional Dataset (to provide number of tests performed, incidence/prevalence of disease or outcomes)	✗	✓	✗	✓
	List of relevant publications	✓	✓	✗	✓
	Systematic review/meta-analysis	✗	✓	✗	✗
	Evidence that the new device changes practice	✗	✓	✗	✗
	Timeline of any device modifications since evidence obtained and justification that evidence remains reliable	✓	✓	✗	✗
	Local pilot or case study of where test device has been used	✓	✗	✗	✗

Figure 2: (continued)

participants recruited in each group is reported in Table 2. Two rounds of the Delphi questionnaire were undertaken with two statements added to the questionnaire in round three as a result of free text comments from round two.

The number of statements carried forward in each round is shown in Table 3. Consensus, as defined by the *priori* definition of Cronbach's $\alpha \geq 0.8$, was reached in both the clinician group and regulator and commissioner

Table 1: Summary of emergent themes from the semi-structured interview study (Delphi round 1).

Emergent theme	Summary of theme	Quotations to illustrate theme
POCTs	There was a strong sentiment that POCTs were underused and undervalued in healthcare at the present time and that their use will increase. Rather than tests in isolation their use is seen as a service or process highlighting the importance of the clinical pathway. Overall, there was agreement that increased use of POCTs may provide significant solutions to many challenges faced by healthcare systems' particular financial constraints	"*there is an exciting future and the question is how to plan that and pick the right tests at the right time*" (Commissioner) *"significant opportunity to change healthcare"* (Regulator) *"in some instances it's probably easier and quicker just to take a normal blood test than a point-of-care test"* (Clinician)
Clinical needs assessment	The importance of establishing the clinical need for a given POCT was recognised. With developments in technology many tests can now be easily translated to POCT platforms but this does not always affect the clinical pathway. Therefore, establishing whether there is a clinical need for a test to be available at the point of care should ideally precede device development in order to maximise the chances of success. This requires an understanding of the clinical pathway. Industry outlined their approach to evaluating clinical need including expert and patient opinion, evaluating available literature and assessing disease burden. However, it was commented on that this process was not always undertaken with appropriate rigour. Pharmaceutical companies were seen to be ahead in this respect	"*there's too many tests out there that don't require point-of-care platforms*" (Regulator) "*if there is no clinical need there is no point"* (Industry representative)
Clinical pathway and utility	Almost all interviewees agreed that clinical pathway evaluation was vital to the evaluation of POCTs. This introduced the subject of test utility with participants highlighting that the evaluation of diagnostic validity alone was not enough and evidence must be provided to demonstrate that the introduction of a new test does truly affect behaviour and outcomes within a pathway. However, there were examples of clinical pathways that were so complex that complex mapping can sometimes impede the progression of a project An understanding of the clinical pathway was also felt to lead to better clinical trials that evaluate patient outcomes rather than diagnostic accuracy. Some stakeholders felt that too much investment went into demonstrating validity and industry were not investing sufficient resources to demonstrate utility	*"just proving a device works is not good enough" (Regulator)* [what is needed is] "*the translation of that evidence into a convincing story for our clinicians as to why they should be using it, and identifying what are the benefits they're going to accrue to themselves, or to the patients*" (Industry Representative)
Validity	Participants reported POCTs are often less robust than laboratory testing but that accuracy and quality were improving. In most circumstances this led to a trade-off between the test's accuracy and the benefits of an early result to the clinical pathway. However, the level of accuracy required for POCTs is dependent on several other factors that device developers must consider such as the disease being tested, the environment where the test will be used and subsequent effects of inaccuracies to clinical care. There was a recognised discrepancy between the levels of accuracy obtained in a laboratory setting compared to real world use. However, it was commented that analytical performance of POCTs was sometimes chased too much to the detriment of using them, e.g. in circumstances where the level of accuracy is insignificant compared to biological variation	"*what is the minimum performance requirement of a test and how will the test be used in that system*?" (Regulator) "*I'd like to see evidence that real world users were able to get accurate measurements*" (Regulator) *"If you look at the precision on point of care diagnostics, they are rarely as precise as the lab measurement. Now you don't always need the level of precision that you can get with the lab measurement."* (Clinician)

Table 1 (continued)

Emergent theme	Summary of theme	Quotations to illustrate theme
Training, usability and quality control	The importance of usability in the design of POCTs was widely recognised. However, very few interviewees were aware of any recognised assessment of this outside of regulatory approval, which alone was not felt to be enough to guarantee the usability of a device. It was noted that usability was dependent on who the user of a test would be and the level of training provided. Challenges were highlighted in getting "*staff, nurses and doctors, to understand the need for training, certification of practice, and doing quality control*" and this was closely linked to safety and quality assurance (including calibration and maintenance of devices). Standard operating protocols and the identification of operator dependent steps with risk assessment were highlighted as vital to ensuring quality in POCTs programmes	*"it's not just putting a test in the office, it's also about educating your colleagues about how to use it"* (Clinician)
Cost-effectiveness	It was recognised that POCTs frequently lead to an increased cost on a test-by-test basis when compared to clinical laboratory tests given the loss of the economy of scale. However, this had to be balanced against any potential savings to the downstream pathway created by rapid decision-making, e.g. early discharge or a reduction in complications. Commissioners made this point by describing the distinction between the "cost saving" and "resource saving" with the latter being a better way to perceive POCTs In was noted that the costs associated with POCTs are often front-loaded and occur during implementation in the purchasing of equipment, user training and establishing a service. Social costs were also recognised such as fewer days off work for patients and lifestyle benefits	"*I know a lot of people say well, point of care testing is really expensive. But what they do is just compare the cost of a point of care test with the cost of a lab test. And nearly always the lab test is going to be cheaper. But what you've got to do is cost the whole package*" (Clinician) "*point-of-care can sometimes be used inappropriately and there is a cost disadvantage"* (Industry Representative)
Efficiency in evidence generation	Many stakeholders declared that they would want evidence from a UK or local population. Others saw this as inefficiency in evidence generation and industry and researchers highlighted the frustration at the number of repetitive evaluations that are often required to evaluate a POCT In respect to who generates the evidence there was a degree of contention between interviewees, with responsibility being attributed to industry, or academic bodies or even medical colleges. However, the majority of interviewees commented on the benefits of evidence being generated in partnership either with industry-academic or industry-provider collaboration. However, the focus was on quality and there was widespread recognition of the need for appropriate funding to achieve this	"*healthcare systems are different and patient pathways are very different and therefore the use of a point of care test in a pathway in, say, States or Scandinavia might be very different to how its use might be in a UK setting*" (Regulator)

Table 2: Number of stakeholders invited to participate in Delphi rounds two and three and number who subsequently participated in each round.

Round	Methodologists	Clinicians	Regulators and commissioners	Industry
Invited	10	12	21	13
Delphi round two	8	8	8	8
Delphi round three	8	8	8	7

Table 3: Number of statements carried into POCKET checklist by each stakeholder group in each Delphi round and level of consensus (as demonstrated by Cronbach's α).

Stakeholder group	Number of statements taken forward in round two	Number of statements taken forward in round three	Total number of included statements	Cronbach's α after round three
Methodologists	37	11	48	0.74
Clinicians	42	17	59	0.92[a]
Regulators and commissioners	49	13	62	0.90[a]
Industry	31	17	48	0.69

[a]Agreed level of consensus reached.

group. However, this was not achieved in the methodologist or industry group.

Expert workshop

Nine stakeholders participated in the workshop and included representation from the study group, clinicians, the trust point-of-care committee, local clinical laboratories and NICE.

The workshop participants agreed that the checklist was comprehensive and met the objectives set out. Participants had applied the checklist to devices that they were familiar with and commented that it was clear, very easy to use and did not take too much time to complete. The tool did not have any components that participants found frustrating or required repeated correction, although it was commented that industry might be daunted by the amount of evidence required by the checklist. Minor refinements to the wording of the checklist were made during the workshop. The final POCKET checklist is presented in Table 4 and consists of 64 statements that are divided into seven themes.

Discussion

This study has led to the development of POCKET, a checklist tool for evidence generation and synthesis in POCTs. Data were collected from a wide range of stakeholders with face validity, content validity and usability demonstrated through a robust mixed-method methodology. The current version of the POCKET checklist with seven themes and 64 statements is ready for dissemination and real world application. The previous tools to support evidence reporting in diagnostics discussed in the introduction have been limited to a specific focus and an overall framework for the evaluation of POCTs or any diagnostic test has been absent leading to an ineffective and inefficient evidence generation pathway. The POCKET checklist provides an evidence based framework to guide the evaluation process, in order to improve the efficacy of evidence generation and ensure that it is more relevant for decision makers.

In Delphi round three, the two groups who have the responsibility for the adoption of POCTs reached consensus as to what evidence is important to them when making decisions and, therefore, which statements should be included in the tool. However, industry representatives and the group of methodologists with expertise in the evaluation of diagnostics did not. Furthermore, there was discrepancy between the number of statements carried forward by each group into the checklist with the groups separated in the same way. Those responsible for evidence generation (industry and methodologists) put forward 48 statements each in comparison to 59 and 62 statements by the decision-making groups (clinicians, commissioners and regulators). These findings support the large degree of uncertainty regarding what evidence is needed by decision-makers to support adoption and those with a role in providing it are underestimating the breadth and detail that is required. It was expected that different stakeholder groups would have different evidence requirements and the results support his. Use of the checklist can allow the production of a tailored evidence portfolio to meet the specific needs of individual stakeholders and avoid the omission of key requirements.

This study highlights the complexity of the evaluation pathway for POCTs and the breath of evidence required. It has been observed that POCKET will have a role in identifying what evidence is missing in POCT evaluation and provides an argument for bodies such as NHS healthcare trust point-of-care committees being involved at an earlier stage with the evaluation of POCTs. Members of regulatory bodies have feedback that the checklist may provide confidence for industry to approach them for appraisal and could act as a triage tool to enter certain regulatory pathways.

The POCKET checklist will be accessed through an online platform. This will allow a rapid appraisal option

whereby a completion of the checklist can be undertaken to see what evidence has already been collected and where further work is required. The planned website will also generate stakeholder specific reports (i.e. only containing the evidence that is required for each stakeholder as identified by the Delphi study) that can be disseminated to relevant parties. Furthermore, the website will provide a research tool to study the longitudinal process of evidence generation in the development of new POCTs to determine where each component of the checklist can be effectively completed. Over time this will provide a rich data source to prioritise the POCKET statements to guide where investment in evidence generation should be made and at what stage of a device's development particular evidence can be achieved. It is expected that POCKET will require refinement and this will be undertaken periodically.

This study design had limitations. Semi-structured interviews allow for the interviewer to probe the topic of interest, although may be a source of bias and, therefore, the primary researcher was aware of the need to remain open to the ideas of the respondents. It is recognised that the Delphi process may lead to a compromise position rather than a true consensus. In both the semi-structured interviews and Delphi survey rounds the sample size of at least eight participants in each stakeholder group is small and the extent to which included participants will be representative is unknown. Whilst the recruitment protocol aimed to recruit a broad interdisciplinary selection of participants with purposive sampling it is acknowledged that a degree of convenience sampling was unavoidable. Patients' opinions were not included in rounds two or three. This was due to the level of expertise and knowledge required regarding the adoption of healthcare technologies to develop POCKET. In respect to the checklist, given the breadth of the study some of the statements can be seen as relatively broad, particularly relating to clinical trial design and economic modelling. Methodology appraisal was outside the scope of this study and research is underway to address some of these areas by the working bodies previously described. Finally, the majority of included stakeholders were from the UK with experience working with the NHS. Therefore, the checklist may not be sensitive to any nuances relevant to other international healthcare systems and attention will be made to evaluate this once the tool is disseminated.

The POCKET checklist represents 65 evidence requirements that when achieved provides an overall multidimensional evidence package to demonstrate a device's impact to the clinical pathway and, when justified, drive adoption. Future work will be directed at determining the weight of importance each POCKET statement carries to stakeholders so that evaluation resources can be focussed on priority areas. POCKET is now ready for real-world application with the aim of reducing the lead-time for new POCTs to reach clinical practice so that the benefits they can bring to patients and society can be effectively realised.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Transparency statement: The lead author affirms that this manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.
Research funding: This research was supported by the National Institute for Health Research (NIHR) Diagnostic Evidence Co-operative London at Imperial College Healthcare NHS Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and St Mary's Hospital, South Wharf Road, London, W2 1NY, UK.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: All authors have completed the Unified Competing Interest form and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years, no other relationships or activities that could appear to have influenced the submitted work. The funding organisation(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Huddy JR, Ni MZ, Markar SR, Hanna GB. Point-of-care testing in the diagnosis of gastrointestinal cancers: current technology and future directions. World J Gastroenterol 2015;21:4111–20.
2. Price CP. Point of care testing. Br Med J 2001;322:1285–8.
3. Global Point-of-Care Diagnostics Market Outlook. 2018. http://www.researchandmarkets.com/reports/2848147/global-point-of-care-diagnostics-market-outlook. Accessed: 24 Apr 2015.
4. Lijmer JG, Leeflang M, Bossuyt PM. Proposals for a phased evaluation of medical tests. Med Decis Making 2009;29:E13–21.
5. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004;140:189–202.

6. NIHR Medtech and In Vitro Diagnostics Co-operatives (MICs). https://www.nihr.ac.uk/about-us/how-we-are-managed/our-structure/infrastructure/Documents/medtech-and-in-vitro-diagnostic-co-operatives.htm. Accessed: 11 Sep 2018.
7. European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group: Test Evaluation. https://www.eflm.eu/site/page/a/1158. Accessed: 11 Sep 2018.
8. Bossuyt PM. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Clin Chem 2003;49:1–6.
9. Monaghan PJ, Robinson S, Rajdl D, Bossuyt PM, Sandberg S, St John A, et al. Practical guide for identifying unmet clinical need for biomarkers. EJIFCC 2018;29:129–37.
10. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Br Med J 2013;346:f1049.
11. International Organization for Standardization. IEC 62366-1:2015 Medical devices – application of usability engineering to medical devices. https://www.iso.org/standard/63179.html. Accessed: 11 Sep 2018.
12. National Institute for Health and Clinical Excellence. Diagnostics Assessment Programme manual. 2011.
13. Huddy JR, Ni M, Mavroveli S, Barlow J, Williams D-A, Hanna GB. A research protocol for developing a Point-of-Care Key Evidence Tool 'POCKET': a checklist for multidimensional evidence reporting on point-of-care in vitro diagnostics. BMJ Open 2015;5.
14. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19:349–57.
15. International Organization for Standardization. ISO 9241-11. Ergonomics of human-system interaction. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en. Accessed: 11 Sept 2018.
16. Finstad K. Interacting with computers the usability metric for user experience. Interact Comput 2010;22:323–7.