Elwenspoek, M. M. C., Scott, L. J., Alsop, K., Patel, R., Watson, J. C., Mann, E., & Whiting, P. (2020). What methods are being used to create an evidence base on the use of laboratory tests to monitor long-term conditions in primary care? A scoping review. *Family Practice, [cmaa074]*. https://doi.org/10.1093/fampra/cmaa074

Publisher's PDF, also known as Version of record

License (if available): CC BY-NC

Link to published version (if available): 10.1093/fampra/cmaa074

Link to publication record in Explore Bristol Research

PDF-document

This is the final published version of the article (version of record). It first appeared online via Oxford University Press at https://academic.oup.com/fampra/advance-article/doi/10.1093/fampra/cmaa074/5695229 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/
Systematic Review

What methods are being used to create an evidence base on the use of laboratory tests to monitor long-term conditions in primary care?
A scoping review

Martha M C Elwenspoek\textsuperscript{a,b,*}, Lauren J Scott\textsuperscript{a,b}, Katharine Alson\textsuperscript{c,d}, Rita Patel\textsuperscript{a,b}, Jessica C Watson\textsuperscript{b}, Ed Mann\textsuperscript{e} and Penny Whiting\textsuperscript{a,b}

\textsuperscript{a}The National Institute for Health Research Applied Research Collaboration West (NIHR ARC West), University Hospitals Bristol NHS Foundation Trust, Bristol, UK, \textsuperscript{b}Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, \textsuperscript{c}Nightingale Valley Practice, Bristol, UK, \textsuperscript{d}Brisdoc Healthcare Services, Bristol, UK and \textsuperscript{e}Tyntesfield Medical Group, Bristol, UK

*Correspondence to Martha M C Elwenspoek, NIHR Applied Research Collaboration West, 9th Floor, Whitefriars, Lewins Mead, Bristol BS1 2NT, UK; Email: martha.elwenspoek@bristol.ac.uk

Abstract

Background: Studies have shown unwarranted variation in test ordering among GP practices and regions, which may lead to patient harm and increased healthcare costs. There is currently no robust evidence base to inform guidelines on monitoring long-term conditions.

Objectives: To map the extent and nature of research that provides evidence on the use of laboratory tests to monitor long-term conditions in primary care, and to identify gaps in existing research.

Methods: We performed a scoping review—a relatively new approach for mapping research evidence across broad topics—using data abstraction forms and charting data according to a scoping framework. We searched CINAHL, EMBASE and MEDLINE to April 2019. We included studies that aimed to optimize the use of laboratory tests and determine costs, patient harm or variation related to testing in a primary care population with long-term conditions.

Results: Ninety-four studies were included. Forty percent aimed to describe variation in test ordering and 36% to investigate test performance. Renal function tests (35%), HbA1c (23%) and lipids (17%) were the most studied laboratory tests. Most studies applied a cohort design using routinely collected health care data (49%). We found gaps in research on strategies to optimize test use to improve patient outcomes, optimal testing intervals and patient harms caused by over-testing.

Conclusions: Future research needs to address these gaps in evidence. High-level evidence is missing, i.e. randomized controlled trials comparing one monitoring strategy to another or quasi-experimental designs such as interrupted time series analysis if trials are not feasible.

Key words: Chronic disease, common illnesses, continuity of care, diagnostic tests, primary care, scoping review.

Introduction

In primary care, ~50% of laboratory tests are used for monitoring long-term conditions (1). Appropriate monitoring ensures early detection of disease progression and the development of complications, potentially enabling GPs to intervene at an early stage, e.g. by adjusting treatment. The use of laboratory tests among GP practices...
**Key Messages**
- Optimal testing for chronic diseases is an area of uncertainty in primary care.
- The uncertainty causes unwarranted variation in test ordering among GP practices.
- We identified gaps in research on determining the optimal frequency of testing.
- Optimal testing strategies improve patient outcomes and reduce patient harms.
- To optimize testing strategies, high-level evidence is needed.
- Ideally, by running randomized controlled trials or quasi-experimental designs.

and regions varies substantially (2–9), suggesting that chronic disease monitoring is not optimal in many places. Both over- and under-testing can cause patient harm and increase health care costs (10,11).

To avoid under- and over-testing, robust evidence is needed on what optimal monitoring looks like. Our recent review of UK guidelines on monitoring patients with hypertension, type 2 diabetes and chronic kidney disease found that most recommendations were solely based on expert opinion and none had a strong evidence base (12).

Guidance on how to generate this evidence base or a framework to standardize evaluations of testing strategies are lacking. The aim of this scoping exercise is 2-fold. The primary aim is to map the nature, extent and range of research on the use of laboratory tests in monitoring long-term conditions in primary care, e.g. which tests should be used or not used, how frequently and how can this be evaluated. In contrast to a systematic review, we aim to scope the methodology used by these studies, i.e. how have researchers tried to answer these questions, instead of extracting their research findings. The secondary aim is to identify gaps in existing research, i.e. identifying questions that are not being addressed. Because we are addressing a broad topic from different angles, where study design is an outcome rather than an inclusion criterion, we considered a scoping review approach to be most suitable.

**Methods**

A scoping review is a relatively new but increasingly used method for mapping research evidence across broad topics. This scoping review is reported according to the PRISMA Extension for Scoping Reviews reporting guidelines (13). The review protocol is available online (14).

**Sources of evidence**

We searched CINAHL, EMBASE and MEDLINE to April 2019. The search included terms for specific long-term conditions (e.g. hypertension, diabetes, chronic kidney disease), monitoring tests (e.g. glucose, lipids, creatinine, HbA1c) and primary care (see protocol appendix for detailed search strategy) (14). Search terms were adapted for each database. We did not apply any language restrictions.

**Study selection**

Studies that fulfilled the following criteria were eligible for inclusion: primary care setting, adult patients (>18 years of age) with non-communicable long-term conditions (e.g. cardiovascular disease, chronic kidney disease and type 2 diabetes), laboratory tests (e.g. urine and blood tests) and aiming to address any of the following questions (Fig. 1):

- How can testing for chronic disease monitoring be optimized? (i.e. studies that present strategies or methods to define or evaluate optimal testing)
- Which test should be used for monitoring? (i.e. studies on test performance or appropriateness)
- How frequently should patients be monitored?
- What are possible harms to patients of under- or over-testing?
- What are the costs of under- or over-testing?
- How does monitoring vary (e.g. by region, sex or age)?

We did not apply any restrictions on study design or outcome measures. Studies investigating a general population, patients with infectious diseases or patients with mental health disorders were excluded. Publications that did not use a methodology (e.g. editorial) were also excluded, whereas papers that reported on methodology but not on research findings were included (e.g. protocols).

Identified papers were screened independently by at least two reviewers (ME, LS and KA) using Rayyan (15). Discrepancies were resolved through consensus or a third reviewer (PW). Foreign language records were translated and assessed by the review team.

**Results**

The database search yielded 12 061 citations; of which, 8243 were unique. Seven thousand nine hundred and thirty-three citations...
were excluded based on title and abstract. The remaining 310 citations underwent full-text review. At the full-text screening stage, 216 papers were excluded based on publication type (no methods reported, e.g. editorial), setting, monitoring test, population or no abstract available (Fig. 1). Ninety-four studies were included in the scoping review (Fig. 1) (8,18–109), including seven studies for which full texts could not be retrieved, so data were abstracted from the abstract (101,102,104,106–109). Six included non-English papers were in French, German, Hebrew, Hungarian, Polish and Portuguese, which were translated and assessed by the review team (34,40,110–113).

Studies were labelled with one or more of the six prespecified study aims (Fig. 1). The key questions ‘How can testing for chronic disease monitoring be optimized?’ (i.e. studies that presented strategies to optimize testing) and ‘how frequently should patients be monitored?’ were addressed by only 12 studies (18–20,22,35,37,50,51,63,64,72,93) and 11 studies (22,33,46,47,49,52,56,72,82,84,99), respectively. Studies labelled as ‘which test should be used for monitoring’ (n = 34) focussed on the diagnostic performance of specific monitoring tests instead of test appropriateness (e.g. whether performing the test improves patient outcomes). The most common aim was ‘how does monitoring vary’ (n = 38), in which studies described the variation in monitoring between regions, patient subgroups or over time. Three studies investigated the costs of over-testing, but none of the studies reported on the possible harms of over-testing to patients (34,78,105).

Seventy studies (74%) were published within the last 10 years (Fig. 2A, Supplementary Table 1). The selected studies included 62 (66%) original research reports, 19 (20%) conference abstracts or posters, 9 (9%) reviews (including 5 systematic reviews (8,31,47,50,80) and 3 narrative reviews (33,50,63)), 4 protocols (21,27,71,90) and 1 PhD thesis (38) (Fig. 2B). More studies were based in the UK (18%) and the USA (17%) than elsewhere (Fig. 2C, Supplementary Table 1).

The majority of studies focussed on diabetes (51%), especially type 2 diabetes (40%), followed by hypertension (11%) and rheumatoid arthritis (10%) (Fig. 3A). In nine studies, the study population consisted of patients on a certain type of medication, e.g. patients on statins (23,46,51), warfarin (90,104,108,109) or ‘high risk medication’ (55). Population sizes varied from 6 (89) to 2,395,340 (23) patients (Fig. 3B).

Forty-eight studies (49%) employed a retrospective cohort design, using routinely collected data and large sample sizes (>1000). Other study designs included prospective cohort design (23%), reviews (11%), early-stage diagnostic studies (5%) (45,61,89,98,114), randomized controlled trials (2%) (22,108) and case–control design (2%) (44,84) (Fig. 4A). Many studies solely used descriptive statistics to explore the data (28%). These were mainly the studies investigating variation in monitoring. Many studies (39%) used some form of modelling, e.g. regression analysis. Studies looking at test performance, either used correlations (i.e. measuring the correlation between an old and a new test) or formal diagnostic accuracy

Figure 1. PRISMA flow diagram. Some studies had more than one aim and appear more than once in under ‘study aims’ in this diagram.
measures (i.e. estimating sensitivity and specificity of a new test against a reference standard) (Fig. 4B).

The most frequently studied laboratory tests were renal function tests (35%), including estimated glomerular filtration rate, creatinine, urea, potassium and proteinuria, followed by HbA1c (23%) and lipids (17%). Other tests that were less common, were liver function tests (including liver enzymes, 99mTc-Hepida plasma clearance and bilirubin), blood glucose, clotting tests, thyroid and natriuretic peptides (Fig. 5A).

Thirty-six studies (38%) reported ‘rate of monitoring’ as the primary outcome. These were studies aiming to show the variation in monitoring and often compared the actual rate of monitoring to what is recommended in current guidelines (‘compliance with guidelines’) (Fig. 5B). Thirteen studies (14%) reported patient outcomes (18,19,22,30,60,70,71,82,84,90,93,103,115), such as disease progression or incidence of complications. Seven studies (7%) reported costs as the outcome (34,52,76,78,92,99,105), such as cost-effectiveness of a screening program for certain patient subgroups. Two studies
investigated the response to test results (51,72), i.e. whether the clinician acted on the abnormal test results. Although one study reported patient harms associated with not-testing, i.e. incidence of adverse outcomes in patients who were not monitored regularly (77), none of the studies investigated patient harms due to over-testing.

**Discussion**

Most research on laboratory test monitoring of long-term conditions either describes the variation in monitoring or reports the test performance of specific tests. This does not address the fundamental question of whether the test is necessary or beneficial.

The most important reason for monitoring is to improve patient outcomes, especially due to early intervention. Nevertheless, only a few studies reported on this outcome. Another important outcome is whether GPs responded to an abnormal test results because monitoring tests are only useful if the result informs patient treatment. However, only two studies investigated this outcome. Finally, over-testing can cause harms to patients (116), but none of the included studies investigated this.

---

**Figure 4.** Methodology of included studies: study design (A) and statistical methods (B).

**Figure 5.** Tests (A) and reported outcome measures (B) of included studies.
Most research on optimal testing focusses on patients with type 2 diabetes, testing for renal function and HbA1c. The most common study design was retrospective cohort design using routinely collected data, followed by prospective cohort design. Outcomes were often summaries of current practice compared to guideline recommendation or test performance, i.e. whether a new test works just as well as the old test.

This scoping review has identified several gaps in the literature. More evidence is needed on establishing the best testing interval, and what the harms of under- and over-testing are, as well as what strategies or methods can successfully optimize testing to improve patient outcomes. We believe that the best study design to obtain the highest level of evidence on optimal testing is a randomized or cluster randomized controlled trial including an economic evaluation of cost-effectiveness. For instance, patients or GP practices can be randomized to different testing strategies based on comparing different testing strategies comprising different testing sets and different testing intervals. Important outcomes to consider would be (i) testing rates, (ii) adverse outcomes specific for the disease (such as disease progression), (iii) harms to patients due to testing (such as patient anxiety and unnecessary follow-up testing) and (iv) costs (such as health care usage, GP workload). While some of these outcomes would be expected to show changes within short time period (<12 months), for example testing rates, other outcome such as disease progression will require longer follow-up periods.

The second-best approach is to use a quasi-experimental design, comparing the effect of an already implemented policy change (such as the publication of new testing guidelines). Routinely collected primary and secondary care data could then be used to perform an interrupted time series analysis. However, this is problematic because GP practices do not strictly follow the guidelines on testing and the guidelines often do not give clear recommendations (12). Alternatively, a new testing strategy could be implemented in one area and compared to a demographically similar region. These approaches may be cheaper but can still be time consuming because a long follow-up will be needed and are more prone to bias.

Finally, a retrospective comparative cohort study using routinely collected health care data may give some insight on the benefits of using ‘a lot of tests’ versus using ‘a minimal number of tests’ and ‘frequent testing’ versus ‘applying longer testing intervals’. Because of the large variation in testing between GP practices, irrespective of the demographics of the population that they serve, these practices could be divided into practices that tend to test often and practices that tend to test less. Patients from ‘high’ and ‘low’ testing practices could be matched, i.e. on demographic factors, co-morbidities and disease severity. The differences in adverse outcomes specific for the disease and health care usage between both groups could be investigated. Although this approach would be much cheaper and time-saving, it will be difficult to investigate patient harms due to over-testing (especially patient anxiety or unnecessary follow-up testing) and it will be impossible to control for all possible biases.

The main strengths of this scoping review are that we have used a very broad search and inclusion criteria and did not restrict by language or publication date. To our knowledge, this is the first scoping review on methods used to create an evidence base underlying chronic disease monitoring in primary care.

The purpose of this scoping review was to map the main sources and types of evidence available; therefore, the findings of the individual studies were not extracted or analysed. Studies on secondary care populations were excluded, which is a limitation of this review. Gaps identified in the primary care literature may have been addressed in secondary care, although their results may not necessarily be applicable to the primary care setting. Another limitation is that we excluded studies that included a general population, i.e. those investigating how to optimize screening, and focussed on studies investigating populations with long-term conditions. These studies may have used methods that are also relevant to monitoring chronic disease populations.

Future research in this area needs to address these gaps in evidence. High-level evidence is missing, i.e. randomized controlled trials comparing one monitoring strategy to another or quasi-experimental designs such as interrupted time series analysis if trials are not feasible. Methods and reporting guidelines should be developed for studies evaluating optimal testing for long-term conditions. This will improve the quality of evidence, as well as reproducibility and transparency of future research and could serve as a framework to judge the trustworthiness of research findings. For clinicians, appropriate tests and testing frequencies can only be identified if we know whether testing improves patient outcomes and whether the benefits outweigh patient harms. Until better evidence is available, decisions around testing should be a collaborative process between patients and clinicians. These decisions should be informed by the patients’ personal preferences and views in combination with current guidelines.

Supplementary material
Supplementary material is available at Family Practice online.

Declaration
Funding: National Institute for Health Research Applied Research Collaboration West; Bristol, North Somerset and South Gloucestershire Clinical Commissioning Group (to KA); National Institute for Health Research (NIHR-2016-09-034 to JW).

Ethical approval: none.

Conflict of interest: none.

Acknowledgements
We would like to thank Alison Richard for developing and performing the searches.

References


20. Elwenspoek MMC, Scott LJ, Alsop K et al. Methods used to develop the evidence base to optimise chronic disease monitoring in primary care: a scoping review protocol. OSFHome 2019; osf.io/mqtxq/.


72. Lau CC, Sayeed SM, Kennedy A, Skelly J, Cooper S. Are we over-testing for liver enzyme abnormalities in rheumatoid arthritis patients prescribed methotrexate? Arthritis Rheumatol 2017; 69(suppl 10).
96. van Bruggen S, Rauh SP, Kasteleyn MJ. Scoping review of methods used to optimize testing.
104. Gudmundsdottir BR, Onundarson PT. Warfarin anticoagulation variability, testing, and dose-adjustment frequency is reduced in clinical practice after switch from prothrombin time based international normalized ratio (Inr) to Fix-prothrombin time based normalized ratio (Fix-Nr). *Am J Hematol* 2018; 93(9): E23.
115. Shah KB, Rao K, Sawyer R, Gottlieb SS. The adequacy of laboratory testing, and dose-adjustment frequency is reduced in clinical practice after switch from prothrombin time based international normalized ratio (Inr) to Fix-prothrombin time based normalized ratio (Fix-Nr). *Am J Hematol* 2018; 93(9): E23.
122. Gudmundsdottir BR, Onundarson PT. Warfarin anticoagulation variability, testing, and dose-adjustment frequency is reduced in clinical practice after switch from prothrombin time based international normalized ratio (Inr) to Fix-prothrombin time based normalized ratio (Fix-Nr). *Am J Hematol* 2018; 93(9): E23.