BMJ Open

Downstream activities after laboratory testing in primary care: an exploratory outcome of the ELMO cluster randomised trial (Electronic Laboratory Medicine Ordering with evidence-based order sets in primary care)

Veerle Piessens ,[1] Nicolas Delvaux ,[2] Stefan Heytens ,[1] Bert Aertgeerts,[2] An De Sutter[1]

To cite: Piessens V, Delvaux N, Heytens S, *et al*. Downstream activities after laboratory testing in primary care: an exploratory outcome of the ELMO cluster randomised trial (Electronic Laboratory Medicine Ordering with evidence-based order sets in primary care). *BMJ Open* 2022;**12**:e059261. doi:10.1136/bmjopen-2021-059261

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-059261).

Received 12 November 2021
Accepted 01 March 2022

[1]Public Health and Primary Care, Ghent University Faculty of Medicine and Health Sciences, Gent, Belgium
[2]Public Health and Primary Care, KU Leuven, Leuven, Belgium

Correspondence to
Dr Veerle Piessens;
veerle.piessens@ugent.be

ABSTRACT

Objective To estimate the rate and type of downstream activities (DAs) after laboratory testing in primary care, with a specific focus on check-up laboratory panels, and to explore the effect of a clinical decision support system (CDSS) for laboratory ordering on these DAs.

Design Cluster randomised clinical trial.

Setting 72 primary care practices in Belgium, with 272 general practitioners (GPs), randomly assigned to the intervention arm or the control arm.

Participants The study included 10 270 lab panels from 9683 primary care patients (women 55.1%, mean age 56.5). All adult patients who consulted one of the participating GPs during the trial period and needed a laboratory exam were eligible for participation.

Interventions GPs in the intervention group used a CDSS integrated into their online laboratory ordering system, while GPs in the control arm used their lab ordering system as usual. The trial duration was 6 months, with another 6 months follow-up.

Main outcome measures This publication reports on the exploratory outcome of DAs after an initial laboratory exam and the effect of the CDSS on these DAs.

Results 19.7% of all laboratory panels resulted in further diagnostic procedures (95% CI 18.9% to 20.5%) and 19% (95% CI 18.2% to 19.7%) in treatment changes. Check-up laboratory exams showed similar rates of DAs, with 17.5% (95% CI 13.8% to 21.2%) diagnostic DAs and 18.9% (95% CI 13.9% to 23.9%) treatment changes. Using the CDSS resulted in a significant reduction in downstream referrals (−2.4%; 95% CI −4.2% to −0.6%; p=0008), imaging and endoscopies (−0.9%; 95% CI −1.6% to −0.1%; p=0026) and treatment changes (−5.4%; 95% CI −9.5% to −1.2%; p=0.01).

Conclusion This is the largest study so far to examine DAs after laboratory testing. It shows that almost one in three laboratory exams leads to further DAs, even in check-up panels. Using a CDSS for laboratory orders may reduce the rate of some DAs.

Trial registration number NCT02950142.

Strengths and limitations of this study

- ► A large database on downstream activities after laboratory testing in the real life setting of general practice.
- ► The strong design of an RCT to evaluate the late effects on downstream activities of a clinical decision support system for laboratory ordering.
- ► Possible recollection and attribution bias because data were collected 6 months or more after the initial inclusion.
- ► Post hoc analysis, without predefined thresholds for which effects would be clinically significant.

INTRODUCTION

Downstream activities (DAs) are those medical procedures that occur due to an initial abnormal or unexpected test result. They cover a wide range, from additional laboratory tests, imaging, endoscopies and other medical investigations, over referrals, to starting, stopping or changing treatment. In general practice, laboratory tests are, apart from physical examination, the most frequently performed tests[1] and might be an essential trigger for further DAs.

DAs are part of the diagnostic and therapeutic process and should ideally lead the general practitioner (GP) to the correct diagnosis and appropriate treatment. However, additional investigations or therapeutic interventions inherently carry some risks: they may be uncomfortable or painful, there may be a (low) risk of severe complications, and they may lead to anxiety during the diagnostic process and beyond. In addition, they require a patient's time and resources and always put a burden on healthcare expenditures. Finally,

each new investigation may result in another unexpected result or reveal abnormalities with unclear clinical significance (incidental findings), increase diagnostic uncertainty and start a cascade of further investigations and therapeutic interventions, often with unknown added value, the so-called Ulysses syndrome.[2–5]

Not all lab tests are necessary or appropriate for the indication they are ordered for, and the value of abnormal results of such tests may be questionable. Moreover, inappropriate overuse of lab tests increases the risk of false positive results considerably.[6] DAs triggered by false-positive results threaten the quality of care because they never result in health benefits, always come with a financial cost, and sometimes cause discomfort, anxiety or complications.[7–10] This is especially true in 'general' laboratory panels for 'health check-ups'. Performing health check-ups in otherwise healthy adults is widespread despite the lack of evidence on its effectiveness. A recent systematic review confirmed the findings of the 2019 Cochrane Review that it is unlikely that periodic health check-ups would be beneficial in terms of mortality or morbidity, despite improved intermediate procedural outcomes, like higher uptake of some preventive services and better management of some risk factors.[11 12]

This so-called 'low-value care' has been brought to attention since the beginning of the 21st century by initiatives like 'too much medicine', 'less is more' and 'choosing wisely'.[13 14] There is an increasing interest in primary care research on low-value testing and associated DAs.[7–9 15]

However, data are scarce on the type of follow-up activities, how frequent they occur, and whether specific patient or laboratory panel characteristics make them more plausible. Moreover, almost no research exists on interventions to reduce unnecessary DAs.[16] However, withholding further follow-up after an abnormal laboratory test goes against good medical practice and feels unethical. Once the initial tests are performed and bring abnormal results, further cascade activities are unavoidable. Therefore, it is preferable to reduce avoidable triggers of these cascades, such as redundant or inappropriate laboratory tests, which was precisely the main objective of the Electronic Laboratory Medicine Ordering (ELMO) study. This study was a cluster randomised clinical trial that introduced a clinical decision support system (CDSS) for online laboratory test ordering in primary care. Using the CDSS significantly reduced the proportion of inappropriate tests and the number of tests per lab panel during the study period. Meanwhile, the intervention did not increase the risk for diagnostic error.[17] The ELMO study included effects on subsequent DAs as an exploratory outcome and provided us with an extensive database and an opportunity to gain insight into this phenomenon.[18]

RESEARCH QUESTION

We present two exploratory research questions. First, how often and which DAs do GPs undertake after an initial laboratory exam, especially after laboratory exams for a health check-up? Second, what could be the effect of a CDSS in laboratory test ordering on DAs, for all lab panels and for a subgroup of lab panels ordered for a check-up only?

METHOD

Setting, intervention and data collection

This is a post hoc exploratory analysis of a subset of data of the ELMO study, a cluster randomised clinical trial, in which 72 primary care practices (PCPs) with 272 GPs, in Flanders (Belgium), were randomly assigned to an intervention or control arm. The methods of this study have been reported previously[17 18] and can be consulted in online supplemental appendix 1. We also briefly summarise them here.

GPs recruited adult patients who needed laboratory testing for at least 1 of the 17 study indications. In the intervention arm, GPs used a CDSS integrated into their online laboratory ordering system. First, GPs selected the indications for which they wanted to have their patient tested. Then the CDSS suggested order sets with appropriate lab tests for each selected indication. The study indications are common reasons for laboratory testing in primary care, like anaemia, fatigue or diabetes. Some of the predefined indications had sub-indications, for example, for diabetes, GPs could specify whether it was for screening, 3-monthly or annual follow-up. (We provide an overview of all study indications and terminology in online supplemental appendix 2). GPs could order tests for multiple indications in one laboratory panel and were free to add or remove laboratory tests at will from the proposed order sets. Thus, a single patient encounter results in one laboratory panel, containing all tests ordered during that encounter, for one or more indications.

In the control arm, GPs used their online laboratory ordering system as usual, without the integrated CDSS, and specified for which of the 17 indications they ordered the laboratory exam.

The trial included a 6-month intervention period and a 6-month follow-up period.

For each included laboratory panel GPs were asked 6 months later to send an electronic case report form (eCRF) through a secured data system to a trusted third party with information about DAs and new diagnoses in the 6 months following the initial laboratory panel. The eCRF automatically retrieved all new diagnoses from the patient's electronic health record (EHR), and the GPs manually added whether they ordered any further laboratory exams, other investigations or referrals due to abnormalities in the laboratory result or if they changed anything to the patient's treatment plan due to the results of the laboratory tests. We encouraged GPs to interpret treatment changes broadly, including iron or vitamin supplements, dietary advice, drug treatment.

After pseudonymising by the trusted third party, this information was linked to the data of the initial laboratory order, allowing the researchers to investigate any effect or relationship between the laboratory panel and DAs. The researchers were blinded for the allocation of each patient until all data were collected and processed.

Outcomes

We checked for all lab panels (the set of lab tests ordered during one patient encounter) whether or not they resulted in DAs.

We calculated the overall proportions of lab panels with DAs and distinguished two different categories of DAs. Diagnostic DAs (DDAs) are DAs with a diagnostic purpose (additional labs, imaging, endoscopies, functional tests and referrals) and therapeutic DAs (TDAs) are those with a therapeutic purpose (changes in the patient's treatment plan).

For the analysis of the subgroup 'check-up panels', we defined a subgroup of lab panels that had 'check-up' as the only indication and compared those to panels having one or more other indications but not 'check-up'.

Patient and public involvement

Patients or the public were not involved in the design or conduct of this study.

Statistical analysis

We estimated overall proportions of panels with DAs and proportions in both study arms, using a generalised estimating equation with the PCP as the clustering variable and allocated group as a factor. Given the known age differences between allocated groups, we also added patient age as a factor in the model. We performed these analyses on the full dataset and on a subset of lab panels with 'check-up' as the only indication.

We calculated all proportions and differences together with their 95% CIs. Statistical significance was assessed at a significance level of 5%.

Sample sizes were calculated for the primary outcomes of this randomized clincial trial (RCT).[18] Therefore, we do not provide post hoc power calculations for this exploratory outcome but will provide 95% CIs of the differences to allow estimates on the power of this study to find any significant difference.[19]

All statistical analyses were performed with SAS statistical software (V.9.4, SAS Institute).

RESULTS

General

A total of 272 GPs from 72 PCPs included 10 270 eligible laboratory panels from 9683 patients. There was only one panel for most of the participating patients (9163), and 520 patients had two or more panels included. We received information on DAs for 90.7% of the included laboratory panels, evenly distributed over the intervention and control group (figure 1). The main reasons for missing information were the non-response of GPs and technical problems that prevented sending the eCRF through the secured system to the trusted third party. Detailed information on recruitment, inclusion and characteristics of lab panels with missing follow-up data can be consulted in online supplemental appendix 3.

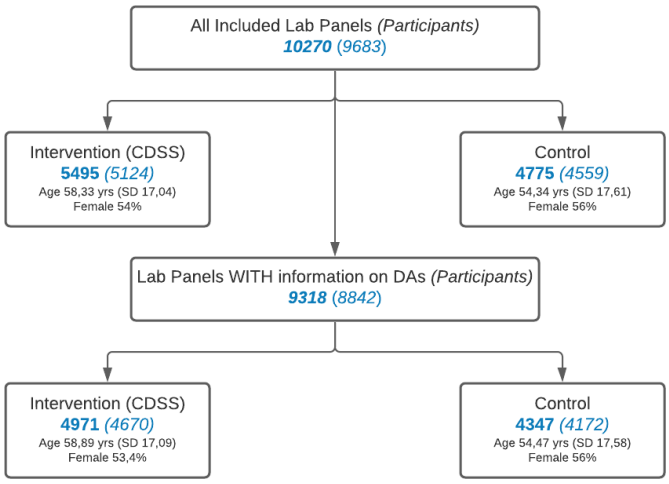


Figure 1 Included lab panels and information on downstream activities in intervention and control arm. CDSS, clinical decision support system; DA, downstream activities.

Downstream activities

Nearly 32% of the lab panels results in DAs. For 19.7% of lab panels, GPs order additional diagnostic procedures (DDAs), and for 19%, they change the patient's treatment plan (TDAs). The majority of DDAs consists of laboratory exams (14.6% of all laboratory panels), followed by referrals (6.3%), whereas TDAs mostly mean 'starting a new medication' (13.6%) (see table 1 for detailed results). Lab panels for which GPs selected 'check-up' as the only indication lead to similar rates of DAs, with 17.5% of check-up labs resulting in DDAs and 18.9% to TDAs.

Intervention versus control

Overall there is no statistically significant difference between both arms in the total number of DDAs. However, there are small but significant reductions in referrals (−2.4%) and in 'imaging and endoscopies' (−0.9%), (p values and 95% CIs are presented in table 2).

We observed a significant reduction in treatment changes (TDA) in the intervention arm. In this arm, 16.5% of laboratory panels result in modifying the patient's treatment plan, mainly the initiation of a new medication (11.7%). In the control arm, this is 21.9% and 15.7%, respectively. This means a significant absolute difference of −5.4% for any change in the treatment plan and −3.9% for starting a new medication.

Effect of the CDSS on check-up laboratory panels

There are twice as many panels for check-up only (22.7%) in the control arm compared with the intervention arm (10.1%) (table 3). GPs in the intervention arm more

Table 1 Number and proportion of laboratory panels resulting in downstream activities in all 10 270 laboratory panels and in a subgroup of 1640 panels for check-up only

	All lab panels		Lab panels for check-up only	
	N=10 270		N=1640	
	No of panels	% (CI 95%) of panels	No of panels	% (CI 95%) of panels
Diagnostic downstream activities (DDA)				
Any DDA	2022	19.7% (18.9% to 20.5%)	287	17.5% (13.8% to 21.2%)
Additional labs	1499	14.6% (13.9% to 15.3%)	213	13% (9.2% to 16.8%)
Referrals	649	6.3% (5.8% to 6.8%)	80	4.9% (3.0% to 6.7%)
Imaging & Endoscopies	248	2.4% (2.1% to 2.7%)	34	2.1% (1.4% to 2.9%)
Functional Tests*	70	0.7% (0.5% to 0.8%)	18	1.1% (0.5% to 1.7%)
Therapeutic downstream activities (TDA)				
Any TDA	1950	19% (18.2% to 19.7%)	310	18.9% (13.9% to 23.9%)
Start medication	1393	13.6% (12.9% to 14.2%)	256	15.6% (11.2% to 20.0%)
Change medication	390	3.8% (3.4% to 4.2%)	26	1.6% (1.1% to 2.2%)
Stop medication	67	0.7% (0.5% to 0.8%)	5	0.3% (0.0% to 0.6%)
Advice on healthy habits†	171	1.7% (1.4% to 1.9%)	38	2.3% (1.1% to 3.7%)

Information on downstream activities was missing for 952 panels. These panels were considered to have no downstream activities.
*Functional tests.
†Nutrition, physical activity, alcohol and tobacco.

often choose (a combination of) specific indications rather than the umbrella indication 'check-up' as the only indication.

In both study arms, GPs order more tests for check-up panels compared with panels without check-up as an indication. Check-up panels have a lower proportion of appropriate tests, result in more abnormal tests and are performed in an older population with more male patients. Although the intervention improved the appropriateness of check-up panels and reduced the number of tests, the differences with panels without check-up remain clear.

When considering laboratory panels with check-up as the only indication, the CDSS is associated with

Table 2 Proportions of lab panels resulting in downstream activities in both study arms, with the absolute differences between study arms, associated 95% CI and p values

	CDSS arm (N=5495)	Control arm (N=4775)	Difference	
	% (95% CI)	% (95% CI)	% (95% CI)	P value
Diagnostic downstream activities (DDA)				
Any DDA	18.4% (12.9% to 23.9%)	21.1% (17.6% to 24.6%)	−2.7% (−9.0% to +3.7%)	0.40
Additional labs	14.3% (8.6% to 20.1%)	14.9% (11.8% to 18.0%)	−0.5% (−6.8% to +5.6%)	0.85
Referrals	5.2% (4.0% to 6.3%)	7.6% (6.2% to 9.0%)	−2.4% (−4.2% to −0.6%)	0.008
Imaging and endoscopies	2.0% (1.5% to 2.5%)	2.9% (2.3% to 3.4%)	−0.9% (−1.6% to −0.1%)	0.026
Functional tests*	0.4% (0.2% to 0.7%)	0.9% (0.5% to 1.4%)	−0.4% (−0.9% to 0.04%)	0.052
Therapeutic downstream activities (TDA)				
Any TDA	16.5% (14.2% to 18.8%)	21.9% (18.5% to 25.3%)	−5.4% (−9.5% to −1.2%)	0.01
Start medication	11.7% (9.9% to 13.6%)	15.7% (12.8% to 18.6%)	−3.9% (−7.3% to −0.6%)	0.02
Change medication	3.6% (2.7% to 4.5%)	4.0% (3.2% to 4.9%)	−0.4% (−1.7% to +0.8%)	0.50
Stop medication	0.63% (0.32% to 0.93%)	0.67% (0.45% to 0.89%)	+0.04% (−0.9% to +0.4%)	0.81
Healthy habits advice†	1.2% (0.7% to 1.7%)	2.3% (1.3% to 3.2%)	−1.1% (−2.1% to −0.05%)	0.05

Information on downstream activities was missing for 524 panels in the CDSS arm and 428 panels in the control arm. Therefore, these panels were considered to have no downstream activities.
*Functional tests.
†Nutrition, physical activity, alcohol and tobacco.
CDSS, clinical decision support system.

Table 3 Lab panel and patient characteristics in laboratory panels for 'check-up' only compared with mixed panels and panels without check-up as an indication

	CDSS arm (N=5495)	Control arm (N=4775)
Laboratory panel characteristics		
Panels (N, %)		
No check-up	3773 (68.7%)	2839 (59.5%)
Check-up only	554 (10.1%)	1086 (22.7%)
Check-up +other indications	1168 (21.3%)	850 (17.8%)
Mean # tests/panel (mean, CI 95%)		
No check-up	19.99 (15.57 to 22.41)	24.33 (21.96 to 26.70)
Check-up only	32.51 (29.37 to 35.66)	41.75 (39.27 to 44.22)
Check-up +other indications	32.99 (31.61 to 34.37)	40.48 (37.89 to 43.08)
Mean # abnormal tests/panel (mean, CI 95%)		
No check-up	3.98 (3.58 to 4.37)	4.38 (3.92 to 4.84)
Check-up only	4.98 (4.39 to 5.57)	5.82 (5.32 to 6.33)
Check-up +other indications	5.50 (5.09 to 5.90)	5.96 (5.42 to 6.49)
Proportion of appropriate tests per panel (mean, CI 95%)		
No check-up	0.74 (0.69 to 0.77)	0.61 (0.59 to 0.63)
Check-up only	0.23 (0.13 to 0.32)	0.16 (0.14 to 0.17)
Check-up +other indications	0.67 (0.64 to 0.71)	0.42 (0.39 to 0.45)
Patient characteristics		
Patients (N)		
No check-up	3447	2660
Check-up only	530	1075
Check-up +other indications	1147	824
Patient age (years, SD)		
No check-up	55.94 (18.25)	49.58 (18.47)
Check-up only	60.75 (13.54)	60.37 (13.44)
Check-up +other indications	62.19 (13.3)	59.03 (15.46)
% Female		
No check-up	55.3	60
Check-up only	47.2	49.8
Check-up +other indications	53.4	53

CDSS, clinical decision support system.

fewer TDAs and referrals and no effect on other DDAs (table 4).

DISCUSSION

Main findings

Nearly 20% of the laboratory panels in the ELMO trial led to DDAs and 19% to TDAs, mainly starting a new medication. A subgroup analysis of check-up laboratory panels showed similar results. The intervention with a CDSS for lab test ordering was associated with a significant overall reduction of downstream treatment changes, referrals and imaging, and referrals and treatment changes in the subgroup of laboratory exams for 'check-up only'.

We believe that the direct effect of the intervention did not extend beyond the moment that GPs ordered the lab tests, so we hypothesise that any effect on further DAs was mediated by the initial reduction in low-value lab tests through the CDSS.

We have no details on what GPs understood by a 'check-up' lab. We assume that there will be a range from real check-up labs (ie, laboratory panels in people who are otherwise healthy and have no complaints or symptoms) over targeted screening tests (eg, screening for diabetes) to broad lab panels for patients who present with multiple, vague complaints. In the intervention arm, there were also fewer panels with 'check-up' as the only indication and more panels with one or more specific

Table 4 Proportions of lab panels for a check-up only resulting in downstream activities in both study arms, with the absolute differences between study arms, associated 95% CI and p values

	CDSS (N=554)	Control (N=1086)	Difference	
	% (CI 95%)	% (CI 95%)	% (CI 95%)	P value
Diagnostic downstream activities (DDA)				
Any DDA	18.6% (12.6% to 24.7%)	16.9% (12.5% to 21.3%)	+1.7% (–5.8% to +9.2%)	0.65
Additional labs	16.6% (9.2% to 24.0%)	11.2% (8.0% to 14.4%)	+5.4% (–2.7% to +13.5%)	0.19
Referrals	2.9% (1.0% to 4.7%)	5.9% (3.6% to 8.2%)	–3.0% (–6.0% to –0.04%)	0.047
Imaging and endoscopies	1.6% (0.2% to 3.0%)	2.4% (1.7% to 3.2%)	–0.9% (–2.4% to +0.7%)	0.28
Therapeutic downstream activities (TDA)				
Any TDA	12.3% (8.5% to 16.0%)	22.3% (16.7% to 27.9%)	–10.1% (–16.7% to –3.4%)	0.003
Start medication	9.9% (6.1% to 13.8%)	18.5% (13.7% to 23.3%)	–8.6% (–14.7% to –2.5%)	0.006

Proportions for other types of downstream activities could not be calculated due to the low number of events.
CDSS, clinical decision support system.

indications. When GPs in the intervention arm selected 'check-up' as an indication, the CDSS only suggested a minimalist set of tests (glucose and cholesterol), probably considerably fewer tests than GPs usually order for a check-up lab. We assume that this might have prompted GPs to reconsider the indications for which they wanted to have the patient tested, which probably led to a shift from lab panels for 'check-up only' to panels with several specific indications. Nevertheless, GPs in both the CDSS and the control group ordered remarkably more tests for check-ups than for other indications, and the CDSS reduced the number of tests and improved the appropriateness of check-up laboratory panels.

The CDSS had no effect on further laboratory testing, which was the largest subcategory of DDAs. We have no further details on the content of these downstream lab tests and whether they were reflex tests on the same blood sample (stepwise ordering of additional tests based on the results of the previous tests, for instance only ordering free thyroxine when thyroid-stimulating hormone levels are abnormal), repeated laboratory panels, or scheduled controls. The CDSS encouraged reflex testing, with a limited set of appropriate, mostly first-line tests. If reflex testing explained the observed difference in downstream laboratory tests, it would represent a shift of certain lab tests from the initial lab panel to follow-up tests rather than actual additional tests.

We only investigated the initial DAs after laboratory exams. Therefore, we have no information on further cascade investigations or treatments resulting from abnormalities that may have been detected in these first-line DAs. The term cascades refers to 'a chain of events initiated by an unnecessary test, an unexpected result, or patient or physician anxiety, which results in ill-advised tests or treatments that may cause avoidable adverse effects and morbidity'.[5] This study shows that a CDSS can mitigate the first two stages of a potential cascade (the number of inappropriate lab tests and subsequent DAs), and a further downstream effect can be assumed, but has yet to be confirmed.

Strengths and weaknesses

The major strength of this study is the strong design of an RCT to examine the late effects of a CDSS for laboratory ordering on DAs beyond its immediate effect on appropriateness and test volume of the lab panels. In addition, the pragmatic real-life setting and the magnitude of the dataset contribute to the reliability of the results. However, this study has some limitations.

First, our data may not be entirely complete. We had to rely on what GPs registered in their EHR, which is sometimes scarce, and on what they actively remembered 6 months later, which could have introduced a recollection bias. However, as health records are digitised, most information is automatically available in the EHR (electronic drug prescriptions, lab reports or specialist referrals). Second, the quality of the data might be suboptimal. At times it was unclear whether the mentioned DAs were due to the laboratory test results or instead related to the complaints the laboratory panel was performed for in the first place. This implies that the decision to initiate a DA was already taken before the lab tests were ordered. This risk of 'assignment bias' is confirmed by the observation that almost 10% of laboratory panels with no abnormal results led to further investigations or referrals. These biases may result in both over and under-reporting of DAs. However, we can assume that they occurred evenly and randomly in all subcategories and that these biases did not affect observed differences between subgroups. Third, one might fear that fewer DAs could mean that necessary DAs for quality patient care are no longer performed, which could lead to missed diagnoses and undertreatment. Unfortunately, we do not have detailed information on the exact contents of the DAs, such as which lab tests were performed or which treatments were started. Therefore, we are unable to distinguish justified DAs from unnecessary ones. However, the risk

of diagnostic error was a crucial secondary endpoint in the ELMO trial, for which it was adequately powered.[18] It was primarily assessed through a systematic chart review of all included patients and completed with data from telephone interviews with a random subset of patients.[18] Results show that the CDSS reduced test volume and inappropriate tests and did not lead to more missed diagnoses or other diagnostic errors.[17] Given this equal health outcome, we could state that the additional reduction in DAs saves healthcare costs and protects patients against unnecessary medical interventions. Fourth, as mentioned before, this is a post hoc analysis without prespecified statistical thresholds. Therefore, the calculated differences between the intervention and control groups should be interpreted with caution. No threshold was set for which effect would be clinically significant. This would have been impossible, as we had no information on the baseline rate of DAs after laboratory exams. Furthermore, it is only after we have established that using a CDSS is safe and does not increase the risk of diagnostic error that we can assume that fewer DAs are not associated with missed diagnoses. Given the now established safety of the intervention, we believe that even a minor decrease in DAs can be clinically significant. For instance, there is an absolute reduction of 0.9% in imaging and endoscopies. One can doubt the clinical significance of such a slight reduction. However, even this tiny reduction would mean 82 000 fewer imaging procedures per year for the small Belgian population (9 million adult inhabitants) if we assume that an adult patient has on average one blood test every year.[1] The net impact of such an intervention will depend on the a priori risk of low value laboratory testing, which may vary between healthcare settings. Belgian GPs lab testing practices are situated mid-range of several European countries.[20]

Comparison with existing literature

There is only limited research on DAs in primary care, and to our knowledge, this is the first study that examines the effect of an intervention for laboratory test ordering on further DAs. Watson *et al* studied British GPs' DA practices after ordering inflammatory marker tests.[9] They found that patients who had their inflammatory markers checked, compared with untested control patients, were 2–3 times more likely to have a referral in the 6 months following the lab test and 1.5–3.5 times more likely to have a new blood test. Overall, 2.4% of patients with an inflammatory marker test were referred, and 23% were scheduled for a new blood test. Houben *et al* observed Dutch GP's DDAs in routine practice and found that GPs ordered further investigations or referrals for 17% of patients with labs ordered for diagnostic reasons and with a low pretest probability. When the pretest probability was high, and the lab results came back abnormal, GPs took further action in up to 77% of patients.[15 21] This is partly in line with our findings, despite our substantially higher number of tests per panel (27 vs 9.9) and a higher proportion of lab panels with at least one abnormal result (90.6% vs 64%).

Other research on TDAs is scarce, so we could not compare our findings with other research on this topic. Almost 14% of all lab panels in this study resulted in the initiation of a new medication, which seems high. However, we believe that this is a reliable reflection of current prescribing behaviour in Belgium. In general, the prescription rate in Belgium is high. Matthys *et al* observed 613 GP consultations, and in 69.8% of all contacts, medication was prescribed.[22] Moreover, most lab panels in this study were not for screening but for specific indications, which probably resulted in a preselection of patients with a higher need to start drug treatment. Finally, the category 'starting medication' was broadly defined and also included iron or vitamin supplementation or adding another diabetes medication to an existing treatment.

A subgroup analysis of check-up laboratory panels showed that only 16% of the tests ordered in these panels was appropriate. This is probably not without consequence. A large retrospective cohort study in patients who underwent an annual check-up exam found that inappropriate, low-value screening tests (ECG, chest radiograph, …) were associated with more, sometimes invasive, downstream procedures, without evidence of health benefits.[7] We observed important rates of treatment changes and DDAs after check-up laboratory exams. This is surprising, given that, in theory, check-up labs should only be performed in healthy people. This could be a confirmation of what other researchers have warned about, that screening healthy people generates multiple DAs with no tangible health benefits.[11] On the other hand, as already mentioned before, this finding might also be due to documentation bias: patients may present with multiple complaints, and the GP decides to do a 'lab for a general check-up' rather than ticking all the relevant indications.

What do these findings mean

Further diagnostic and therapeutic procedures after laboratory exams are standard in the clinical process in primary care. Using a CDSS, which affects lab test ordering behaviour at the beginning of this process, may also impact DAs. This could be an encouragement for GPs who strive to provide a high quality of care and avoid unnecessary medical interventions, to use such CDSS, and likewise for policy-makers to support the implementation of CDSS in lab test ordering systems.

Need for further research

This was the first trial to study the effect of a CDSS on DAs. However, it was only an exploratory outcome with a post hoc statistical analysis. Further research is needed to confirm and elaborate on these results, with a specific focus on gaining insight into what prompts further (low value) DAs: specific lab abnormalities, the indications the tests were ordered for, or rather patient or GP characteristics? Since research indicates that patients appreciate general check-ups and improve their well-being,[12]

we recommend integrating nested qualitative research to capture both the patient's perspective and GP's views on observed differences in DAs.

CONCLUSION

Our findings suggest that one-third of laboratory exams in primary care lead to further diagnostic procedures or changes in the patient's treatment plan. Lab panels for a general check-up in otherwise healthy patients have similar rates of DAs. Using a CDSS for laboratory test ordering in primary care decreased downstream referrals and treatment changes but did not affect additional laboratory testing.

Acknowledgements Special thanks for their contribution to concept, design and obtaining funding: Dirk Ramaekers, data collection monitoring: Tine De Burghgraeve, guidance and support in the statistical analyses: Pavlos Mamouris, professor Dr Els Clays, professor Dr Steffen Fieuws, and the Ghent University Biostatistics Unit for their guidance and support in the statistical analysis. We thank all the participating patients and their GPs, as well as Mario Berth, Eric De Schouwer, An De Vleeschauwer, and all other clinical laboratory personnel involved in the technical support of the ELMO Study.

Contributors VP, ND, BA and ADS are guarantors of this publication, they had full access to all data and take responsibility for the integrity of the data and the accuracy of the analyses. Concept, design and obtaining funding: ND, BA and ADS. Development of the data collection tool: ND. Data collection monitoring: ND and VP. Data cleaning, processing and analysis: VP, ND, ADS and SH. Statistical analyses: VP and ND. Drafting of the manuscript: VP. Critical revision of the manuscript: all authors. All authors read and approved the manuscript.

Funding The ELMO Study was funded through the Belgian Health Care Knowledge Centre (KCE) Trials, Programme agreement KCE16011.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Research Ethics Committee UZ/KU LeuvenB322201733217. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The data can be obtained by a motivated request to the first author.

ORCID iDs
Veerle Piessens http://orcid.org/0000-0002-2305-1762
Nicolas Delvaux http://orcid.org/0000-0002-7335-4075
Stefan Heytens http://orcid.org/0000-0003-1097-4987

REFERENCES

1 Hickner J, Thompson PJ, Wilkinson T, *et al*. Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. *J Am Board Fam Med* 2014;27:268–74.
2 Nguyen LT, Sullivan CT, Makam AN. The diagnostic cascade of incidental findings: a Teachable moment. *JAMA Intern Med* 2015;175:1089–90.
3 Rang M. The Ulysses syndrome. *Can Med Assoc J* 1972;106:122–3.
4 Reznik OI. Ulysses Syndrome and the Case of the Endless Search for Reassurance. In: *The secrets of medical decision making, how to avoid becoming a victim of the health care machine*. Loving Healing Press, 2006.
5 Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health* 2002;23:23–44.
6 Naugler C, Ma I. More than half of abnormal results from laboratory tests ordered by family physicians could be false-positive. *Can Fam Physician* 2018;64:202–3.
7 Bouck Z, Calzavara AJ, Ivers NM, *et al*. Association of low-value testing with subsequent health care use and clinical outcomes among low-risk primary care outpatients undergoing an annual health examination. *JAMA Intern Med* 2020;180:973.
8 Bhatia RS, Bouck Z, Ivers NM, *et al*. Electrocardiograms in low-risk patients undergoing an annual health examination. *JAMA Intern Med* 2017;177:1326–33.
9 Watson J, Salisbury C, Whiting P, *et al*. Added value and cascade effects of inflammatory marker tests in UK primary care: a cohort study from the clinical practice research Datalink. *Br J Gen Pract* 2019;69:e470–8.
10 Whiting D, Croker R, Watson J, *et al*. Optimising laboratory monitoring of chronic conditions in primary care: a quality improvement framework. *BMJ Open Qual* 2019;8:e000349.
11 Krogsbøll LT, Jørgensen KJ, Gøtzsche PC. General health checks in adults for reducing morbidity and mortality from disease. *Cochrane Database Syst Rev* 2019;1:CD009009.
12 Liss DT, Uchida T, Wilkes CL, *et al*. General health checks in adult primary care: a review. *JAMA* 2021;325:2294–306.
13 Moynihan R, Smith R. Too much medicine? *BMJ* 2002;324:859–60.
14 Born KB, Levinson W. Choosing wisely campaigns globally: a shared approach to tackling the problem of overuse in healthcare. *J Gen Fam Med* 2019;20:9–12.
15 Houben PHH, van der Weijden T, Winkens RAG, *et al*. Cascade effects of laboratory testing are found to be rare in low disease probability situations: prospective cohort study. *J Clin Epidemiol* 2010;63:452–8.
16 Fried JG, Andrew AS, Ring NY, *et al*. Changes in primary care health care utilization after inclusion of epidemiologic data in lumbar spine MR imaging reports for uncomplicated low back pain. *Radiology* 2018;287:563–9.
17 Delvaux N, Piessens V, Burghgraeve TD, *et al*. Clinical decision support improves the appropriateness of laboratory test ordering in primary care without increasing diagnostic error: the ELMO cluster randomized trial. *Implement Sci* 2020;15:100.
18 Delvaux N, De Sutter A, Van de Velde S, *et al*. Electronic laboratory medicine ordering with evidence-based order sets in primary care (ELMO study): protocol for a cluster randomised trial. *Implement Sci* 2017;12:147.
19 Levine M, Ensom MH. Post hoc power analysis: an idea whose time has passed? *Pharmacotherapy* 2001;21:405–9.
20 Leurquin P, Van Casteren V, De Maeseneer J. Use of blood tests in general practice: a collaborative study in eight European countries. Eurosentinel Study Group. *Br J Gen Pract* 1995;45:21–5.
21 Houben PHH, van der Weijden T, Winkens B, *et al*. Pretest expectations strongly influence interpretation of abnormal laboratory results and further management. *BMC Fam Pract* 2010;11:13.
22 Matthys J, Elwyn G, Van Nuland M, *et al*. Patients' ideas, concerns, and expectations (ICE) in general practice: impact on prescribing. *Br J Gen Pract* 2009;59:29–36.

APPENDIX 1
METHODOLOGY SECTION FROM
Clinical Study Report

Electronic Laboratory Medicine ordering with evidence-based Order sets in primary care (ELMO) Study

The effect of evidence-based order sets within a CPOE (computerised physician order entry) system on the quantity and quality of laboratory test ordering in family practice: a cluster randomised trial

KCE16011

03 March 2020

CONFIDENTIAL

Signature pages for clinical study report

I have read this report and confirm that to the best of my knowledge it accurately describes the conduct and results of the study.

Signed: Date: ____/____/_____

Print name: Bert Aertgeerts

Affiliation: Academic Center for General Practice, KU Leuven

Address: Kapucijnenvoer 33, Blok J, PB 7001, B-3000 Leuven

Signed: Date: ____/____/_____

Print name: An De Sutter

Affiliation: Department of Public Health and Primary Care, Ghent University

Address: C. Heymanslaan 10, building K3, B-9000 Ghent

Signed: Date: ____/____/_____

Print name: Nicolas Delvaux

Affiliation: Academic Center for General Practice, KU Leuven

Address: Kapucijnenvoer 33, Blok J, PB 7001, B-3000 Leuven

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

1 TITLE PAGE

Study title: Electronic Laboratory Medicine ordering with evidence-based Order sets in primary care (ELMO Study): a cluster randomised trial

Short title: ELMO Study

Indication studied: Laboratory test ordering in primary care

Study description: Evidence-based order sets for laboratory test ordering

Sponsors: KU Leuven Research & Development

Protocol: Version 3.6 date 07/08/2017, Amendment 23/11/2017

Study dates: 01/12/2017 – 01/03/2020

Investigators: Bert Aertgeerts (KU Leuven), An De Sutter (Ghent University)

Trial Manager: Nicolas Delvaux

GCP Statement: This study was performed in compliance with ICH Good Clinical Practice (GCP) including the archiving of essential documents

Date of report: 03 March 2020

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

2 SYNOPSIS

KU Leuven **Academic Center of General Practice**		**Individual Study Table**	
Title of Study	Electronic Laboratory Medicine ordering with evidence-based Orders sets in primary care (ELMO Study): a cluster randomised trial		
Investigator(s)	Prof Dr Bert Aertgeerts (CI), prof Dr An De Sutter (co-CI)		
Study centre(s)	Academic Center for General Practice, KU Leuven Department of Public Health and Primary Care, Ghent University		
Publication	N/A		
Study period	From: 01/12/2017 To: 01/03/2020	Diagnostic study	Phase IV
Objectives	<u>Primary Objectives:</u> 1. Proportion of appropriate tests per indication according to guidelines <u>Secondary Objectives:</u> 2. Number of missed diagnoses at end of trial 3. Number of ordered tests at end of trial		
Methodology	Cluster randomised trial		
Number of patients	Planned: 12 000 Analysed: 10 663		
Diagnosis and main criteria for inclusion	Patients with laboratory tests for 1 or more of the following indications: cardiovascular disease, hypertension, diabetes mellitus, anaemia, liver pathology, medication monitoring, gout, chronic kidney disease, lung embolism, acute coronary syndrome, acute diarrhoea, chronic diarrhoea, thyroid disease, unexplained fatigue, sexually transmitted infections, rheumatoid arthritis, general check-up.		
Test product, dose and mode of administration	Clinical decision support system (CDSS) in the form of evidence-based order sets integrated into the computerized physician order entry (CPOE)		
Duration of treatment	N/A		
Criteria for evaluation	Primary: 1. Appropriateness of each ordered test based on predefined guidelines Secondary: 2. Diagnoses identified as potential diagnostic error through consensus procedure 3. Number of tests per panel		
Statistical methods	Logistic generalized estimating equation (GEE) model		

KU Leuven Academic Center of General Practice	Individual Study Table	

SUMMARY CONCLUSIONS

RESULTS

The CDSS investigated in this study significantly improved appropriateness of laboratory testing. The percentage of appropriate laboratory tests was 38% in the control arm and 58% in the CDSS arm. CDSS improved appropriateness of laboratory testing with 21% points.

We demonstrated that the CDSS investigated in this study was non-inferior compared to control with regards to diagnostic error. In the control arm 3.04% of the patients had a possible diagnostic error and 2.40% of the patients in the CDSS arm. The absolute difference was a decrease of 0.66% in possible diagnostic error.

The CDSS reduced volume of laboratory testing from 31.17 tests per panel in the control arm to 24.02 tests per panel in the CDSS arm (difference of 7.15 tests per panel).

In this study 19,7% of all laboratory panels resulted in extra downstream investigations and 19% led to changes in the patient's therapy plan. Laboratory panels in de CDSS arm seemed to generate less downstream activities than those in the control arm. Differences were greatest for referrals and changes in treatment plan.

CONCLUSION

CDSS significantly improved appropriateness and reduced volume of laboratory test ordering without increasing diagnostic error. Baseline appropriateness of laboratory test ordering was lower than expected. The incidence of diagnostic error was low despite a very sensitive approach to defining this outcome.

DATE OF THE REPORT: 03 March 2020

3 TABLE OF CONTENTS

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

4 LIST OF ABBREVIATIONS & DEFINITION OF TERMS

4.1 ABBREVIATIONS

Abbreviation	Definition
AML	Algemeen Medisch Laboratorium
ARR	Absolute risk reduction
CDSS	Clinical decision support system
CG1/CG2	Code beneficiary (billing status)
CI	Confidence interval
CKD	Chronic kidney disease
CPOE	Computerized physician order entry
CRA	Clinical research assistant
DE	Diagnostic error
DMP	Data management plan
DVT	Deep venous thrombosis
EC	Ethics committee
EHR	Electronic health record
eCRF	Electronic Case Report Form
GCP	Good clinical practice
GEE	Generalized estimating equations
GP	General physician
ICC	Intra-cluster correlation
ICD	International Classification of Diseases
ICF	Informed consent form
ICH	International Conference on Harmonisation
ICPC	International Classification for Primary Care
IFOBT	Immunochemical faecal occult blood test
KCE	Belgian Health Care Knowledge Centre
LIS	Laboratory information system
LOINC	Logical observation identifiers names and codes
MCH	Medisch Centrum Huisartsen
MCHC	Mean corpuscular haemoglobin concentration
MCV	Mean corpuscular volume
NIHDI	National Institute for Health and Disability Insurance
NICE	National Institute for Health and Care Excellence
PCP	Primary care practice
PSA	Prostate-specific antigen
SAP	Statistical analysis plan
SC	Steering committee
SD	Standard deviation
STI	Sexually transmitted infection
TSH	Thyroid stimulating hormone
UTI	Urinary tract infection
XML	Extensible mark-up language

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

4.2 DEFINITIONS

General practitioner (GP): in this study, all participating GPs were also investigators. We will use the term GP to indicate GP investigators.

Primary care practice (PCP): GPs collaborate and work together in a PCP.

Laboratory test panel: also referred to as the laboratory panel, this is the set of ordered laboratory tests ordered by the GP. This panel consists of a series of laboratory tests ordered for one or more indication at one time by a single GP.

Indication: the reason for ordering a series of laboratory tests. A single panel may have more than one indication and sometimes a single test can be ordered for more than one indication.

Study indication: one of the indications included in the study protocol.

Order sets: a set of laboratory tests suggested for a given indication.

Figure 1 illustrates the relationships between all the concepts defined above.

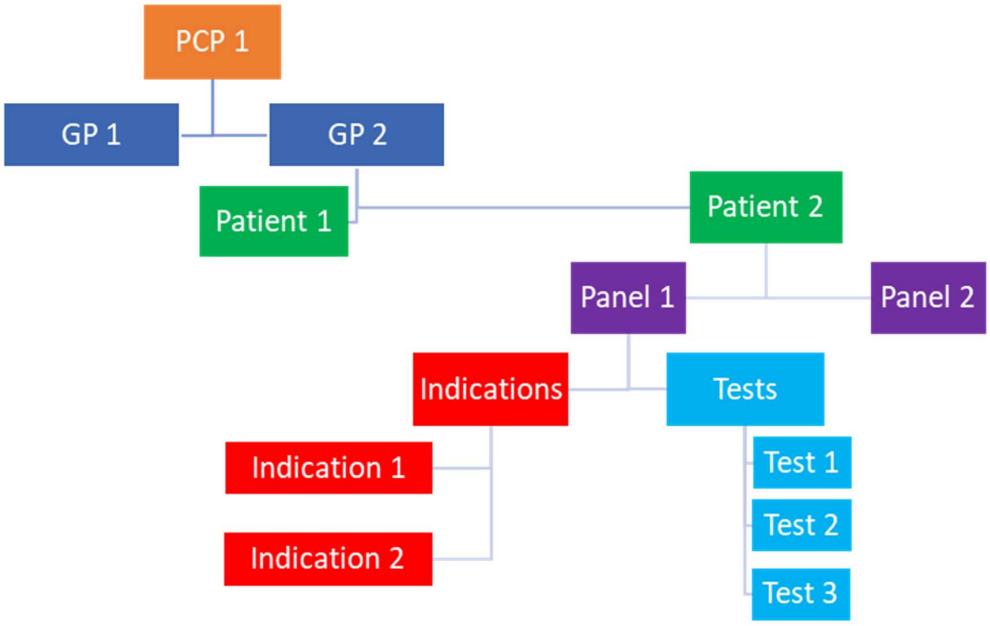


Figure 1: The relationship and clustering between PCP, GP, patient, laboratory test panel, indications and laboratory tests.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

5 ETHICS AND REGULATORY APPROVAL

5.1 INDEPENDENT ETHICS COMMITTEE APPROVAL

The study protocol and all its amendments, and the patient information sheet(s) were reviewed and approved by the appropriate independent ethics committees as detailed in table one below. The study gained full approval from the Ethics Committee (EC) Research UZ/KU Leuven on 25/08/2017 and a copy can be found in Appendix **Fout! Verwijzingsbron niet gevonden.**.

Centre name and study number	KU Leuven	S59472
Investigator	Bert Aertgeerts	
Ethics committee	EC Research UZ/KU Leuven	B322201733217
Chairman	Minne Casteels	
Date of approval of the final protocol	25/08/2017	

Table 1: Details of the EC approval.

5.2 ETHICAL CONDUCT OF THE STUDY

The study was performed in accordance with the current version of the declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil, October 2013). The trial was conducted in agreement with the International Conference on Harmonisation (ICH) guidelines on Good Clinical Practise (GCP).

5.3 PATIENT INFORMATION AND CONSENT

All patients provided written informed consent to participate in the study prior to being screened.

The patient information sheet detailed the procedures involved in the study (aims, methodology. potential risks, anticipated benefits) and the GP investigator (further referred to as GP) explained these to each patient. The patient was then allowed time to consider the information presented before signing and dating the informed consent form to indicate that they fully understood the information, and willingly volunteered to participate in the study. The patient was given a copy of the informed consent form for their information. The original copy of the informed consent was kept in a confidential file stored at the participating laboratories. A sample of the patient information sheet and consent form can be found in the protocol (Appendix **Fout! Verwijzingsbron niet gevonden.**) and a Dutch version of the ICF in Appendix **Fout! Verwijzingsbron niet gevonden.**.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

5.4 REGULATORY APPROVAL

The study was performed in compliance with the requirements of the National Privacy Commission's Sector Committee for eHealth (currently replaced by the Data Protection Authority). The study gained full regulatory approval on 21/11/2017 (under SCSZG number SCSZG/18/174) and was amended on 3/7/2018. A copy can be found in Appendix **Fout! Verwijzingsbron niet gevonden.** and a copy of the amendment to the regulatory approval in Appendix **Fout! Verwijzingsbron niet gevonden.**.

6 INVESTIGATORS AND STUDY ADMINISTRATIVE STRUCTURE

Table 2 shows the principal study personnel involved in the study. Additional study personnel who were involved with specific aspects of the study are mentioned in Appendix **Fout! Verwijzingsbron niet gevonden.**, including their specific role.

Title	Name and affiliation
Principal Investigator	Bert Aertgeerts, Academic Center for General Practice, KU Leuven
Co-principal investigator	An De Sutter, Department of Public Health and Primary Care, Ghent University
Sponsor	KU Leuven Research & Development
Project Manager	Nicolas Delvaux, Academic Center for General Practice, KU Leuven
Project Leader	Veerle Piessens, Department of Public Health and Primary Care, Ghent University
Clinical Research Associate(s)	Tine De Burghgraeve, Academic Center for General Practice, KU Leuven Bart Verheyden, Academic Center for General Practice, KU Leuven
Statistician	Pavlos Mamouris, Academic Center for General Practice, KU Leuven
Laboratory clinical biologist(s)	Eric De Schouwer, Medisch Centrum Huisartsen (MCH) Lisbeth Patteet, Algemeen Medisch Labo (AML) An De Vleesschouwer, Anacura
Data Manager	Roel Heylen, Sciensano

Table 2: Principal study personnel

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

The Steering Committee (SC) consisted of representatives of GPs, academic centers for primary care, clinical biologists, clinical pharmacologists, internists and statisticians (see Table 3). The SC convened 9 times throughout the study duration.

Name	Role	Affiliation
Hanne Cloetens	GP	Domus Medica, Flemish College of Family Physicians
Josse Thomas	Clincal pharmacologist	Independent, Ethics Committee UZ Leuven
Steffen Fieuws	Statistician	Leuven Biostatistics and Statistical Bioinformatics Centre (L-BioStat), KU Leuven
Bert Vaes	GP	Academic Center for General Practice, KU Leuven
Alain Verstraete	Clinical biologist	Faculty of Medicine and Health Sciences, Ghent University
Dirk Ramaekers	Internist, Chief Medical Officer	Leuven Institute for Healthcare Policy, KU Leuven; Jessa Hospital
Robert Vander Stichele	Clinical pharmacologist, GP	Department of Pharmacology, Ghent University

Table 3: Steering Committee (SC) members, their roles and affiliations

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

7 INTRODUCTION

Laboratory testing is an important clinical act with a valuable role in screening, diagnosis, management and monitoring of diseases or therapies. Thirty percent of patient contacts in primary care result in ordering of laboratory tests [1, 2]. In Belgium, more than 370 million tests are ordered annually implying that for each person about 31 laboratory tests are ordered each year [3]. Primary care has seen a continuous increase in the use of laboratory tests over the last decade [4]. Despite the frequency with which laboratory tests are ordered, there is a large variation in the appropriateness of these orders [5–8].

Inappropriate laboratory test ordering has been estimated to be as high as 30% [9]. This seems not to be different in Belgium, where 30 to 50% of tests requested by primary care physicians for the five most common indications were found to be inappropriate in a 2007 KCE study [10]. Besides the burden this poses on health care spending, it may also result in false-positive results and potentially cause excessive downstream diagnostic examinations. The true extent of these downstream diagnostic examinations in primary care have never been thoroughly investigated [11].

Several interventions influence the test ordering behaviour of GPs, including developing evidence-based guidelines, providing feedback, introducing computerized decision support, limiting the number of tests on the order form, and providing financial incentives [10]. Education-based interventions, feedback-based interventions and clinical decision support systems (CDSS) have shown promising results to influence the test ordering behaviour of GPs and to improve appropriateness [1]. However, these findings tend not to be generalizable because many studies either focus on very limited indications or measure testing volume rather than appropriateness.

Indications for ordering laboratory tests include all the reasons why a physician chooses to order a laboratory test such as diagnosis of complaints, the follow-up of medical conditions, the follow-up of drug or other therapies, preventive care and early detection of adverse effects of a condition or therapy. We will refer to all these reasons for ordering of laboratory tests as *indications*. We suggested that computerized CDSS applicable for multiple indications were more effective than those aimed at a limited number of indications in influencing laboratory testing behavior, but conclusive evidence is still lacking [13]. *Order sets*, a form of decision support where a limited set of evidence-based tests are proposed for a series of indications, has been shown to be effective in reducing the volume of ordered laboratory tests [12, 14]. However, good evidence that the use of order sets aimed at multiple indications improves the appropriateness of laboratory test ordering is still lacking. The primary aim of this study is to measure the effect of order sets on the quality and quantity of laboratory test orders by GPs.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

7.1 DIAGNOSTIC AREA

This study was aimed at laboratory test ordering in primary care. Laboratory test ordering is a common procedure in primary care and is considered the single most performed technical procedure by GPs [9].

We chose to study 17 common indications for which laboratory tests are ordered in primary care. The rationale for choosing these order sets is discussed in 9.4.3 and in section **Fout! Verwijzingsbron niet gevonden.** of the protocol (Appendix **Fout! Verwijzingsbron niet gevonden.**).

7.2 RATIONALE FOR THE STUDY

It has been demonstrated that CDSS, in the form of order sets aimed at laboratory test ordering, has the potential to improve appropriateness of laboratory test ordering [13, 14]. However, besides evidence showing that order sets can reduce the volume of laboratory test ordering by 20% [12], no sound evidence exists that these interventions improve appropriateness. Moreover, most studies have evaluated the effect of decision support for one or a limited number of indications. Our review could not detect sufficient evidence of effectiveness and suggested that studies with a more comprehensive intervention are necessary.

A barrier to adhering to evidence-based policy is the fear for missing important pathology and the liability this may create [2]. There is currently no evidence showing that increasing appropriateness of laboratory testing influences morbidity through diagnostic errors or delay. To evaluate the effect of order sets on diagnostic errors or delay, there was need for a large study that assessed the effect of CDSS for laboratory test ordering on the incidence of potentially missed diagnoses.

Pre-test probability and abnormal test results have shown to influence downstream or cascade activities [11]. Downstream or cascade activities are those medical acts which result from altered or deviant tests. For instance, an elevated liver test in an asymptomatic person has a very high probability of being false positive, but may result in additional testing such as repeat laboratory testing, radiology testing, other technical evaluations or specialist consultations. In general practice, where pre-test probabilities of disease are often low, abnormal test results are often false positives, especially in case of inappropriate testing where the risk of false positives is more than 50% [15]. It is generally assumed that the effects of inappropriate test ordering are larger on the downstream activities than on the tests themselves. This phenomenon is often referred to as the Ulysses effect [16]. To date, little research has been done on these cascades in primary care and the size of this Ulysses effect is largely unknown [11]. More insight in these downstream activities is needed.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

8 STUDY OBJECTIVES

Primary Objective

To compare the effect of evidence-based order sets versus control on the proportion of appropriate laboratory tests ordered by primary care GPs on 17 common indications for ordering laboratory tests.

Secondary Objectives

To demonstrate non-inferiority in the effect of evidence-based order sets versus control on the incidence of missed or delayed diagnoses (diagnostic error) for 17 common indications by primary care GPs.

To compare the effect of evidence-based order sets versus control on the number of laboratory tests ordered by primary care GPs with no restriction on the indications.

Exploratory Objective(s)

To assess the effect of our intervention on the downstream activities arising from abnormal results of inappropriate tests.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

9 INVESTIGATIONAL PLAN

9.1 OVERALL STUDY DESIGN AND PLAN

Our trial was a cluster randomised controlled trial and powered for two outcomes. It was powered as a superiority trial for our primary outcome. For the secondary outcome, the trial sought to establish non-inferiority. The trial included a six-month intervention period and a six-month follow-up period.

Six months after the end of the intervention period, all patients received our intervention and we continued to measure appropriateness and volume of testing in a prospective observational design in the original intervention group as a measure of sustainability.

We randomized participating PCPs to the intervention or to a control group. The unit of allocation is the PCP. This meant that all GPs in the same practice were allocated to the same intervention and that either all or no GPs in the PCP were included in the trial. All patients cared for by the same primary care practice were exposed to the same intervention.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

9.1.1 STUDY TIMING

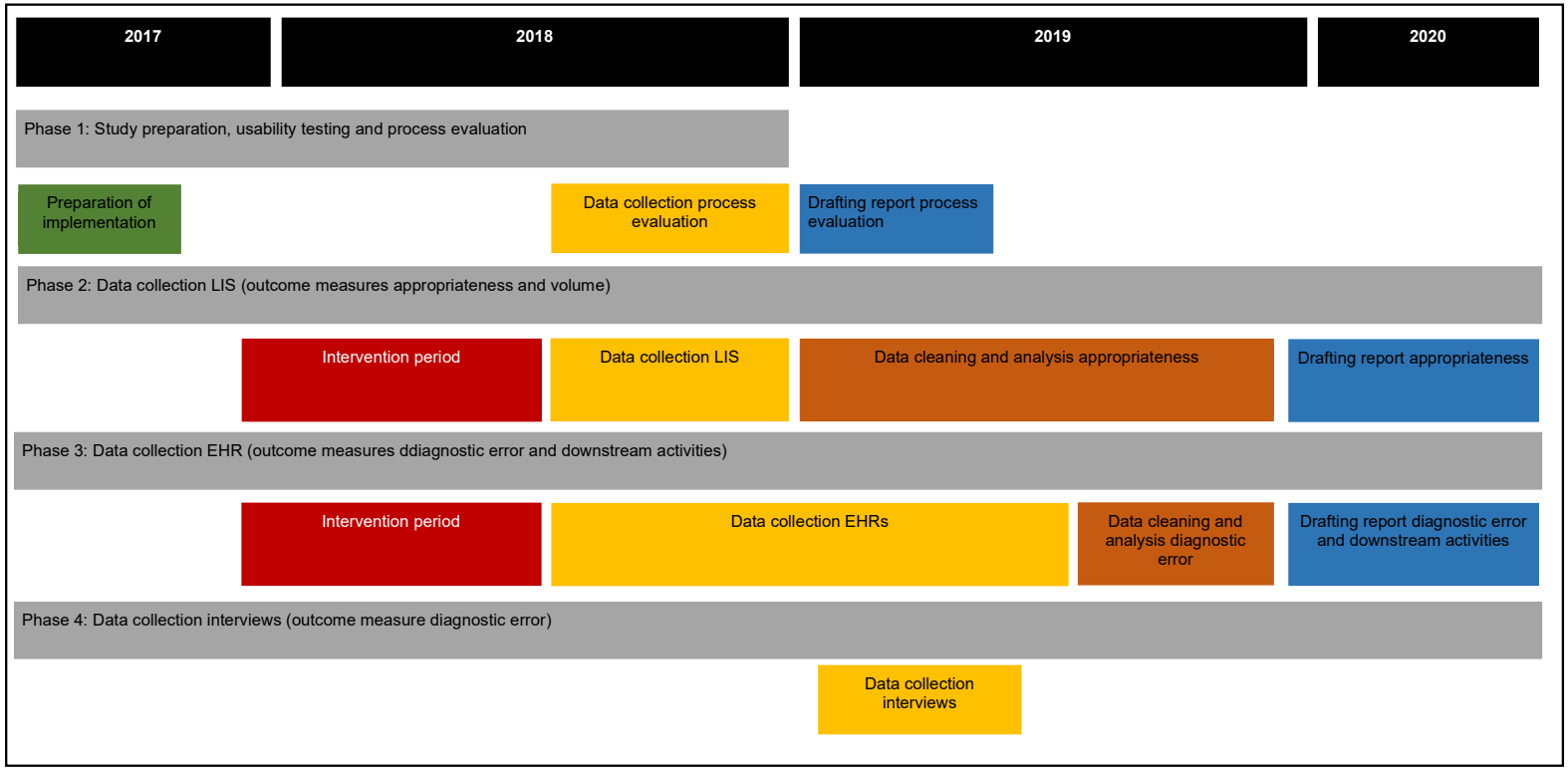


Figure 2: Overview of study timing.

Figure 2 illustrates the timings of the different phases of the ELMO Study. The study was divided into 4 phases, each with a different finality and data source. Phase 1 was aimed at preparing the study and the intervention, and at evaluating the process. Phase 2 was aimed at data collection, analysis and reporting for laboratory test volume and appropriateness. In phase 3, data on diagnostic error and downstream activities was collected, analysed and reported on. Finally, in phase 4 data was collected from patients which was used for the assessment of diagnostic error.

Compared to the trial procedures in the study protocol (see section 8 in Appendix **Fout! Verwijzingsbron niet gevonden.**), some data collections and analyses were performed later than planned. Technical issues in the data collections from GP EHRs (phase 3) delayed this data collection. This also delayed the patient interviews (phase 4) which were dependent on the data from the previous phase. Hence the patient interviews were conducted almost 1 year after inclusion in the trial.

KU LEUVEN
ACADEMISCH CENTRUM HUISARTSGENEESKUNDE

9.1.2 STUDY LOCATION

This study was conducted at 72 different PCPs. Table 4 indicates the name and location of each PCP included in the study.

	Name of PCP	Location
1	Aan de Lieve	Evergem (Ghent)
2	Balansstraat	Antwerp
3	Bartholomeus	Antwerp
4	Bonaventure	Jette (Brussels)
5	Dr Cornelissen	Boechout
6	De Medische Hoek	Hombeek
7	De Pretlei	Brasschaat
8	De Ring	Bruges
9	Dr Symons	Evergem (Ghent)
10	Duopraktijk 180	Heverlee (Leuven)
11	Eksaarde	Lokeren
12	Fruithof	Berchem (Antwerp)
13	GP Blauwput	Kessel-Lo (Leuven)
14	GP De Doenders	Hoeilaert
15	GP De Vest	Heverlee (Leuven)
16	GP Tempelhof	Leuven
17	GP Van 't Sestich	Leuven
18	HA Praktijkhuis 94	Veltem (Leuven)
19	Dr Haemels	Boortmeerbeek
20	HAG Park Noord	Antwerp
21	HAP Seghers Vandenberghe	Mechelen
22	HAP Zwijndrecht	Zwijndrecht
23	Horizon	Ganshoren (Brussels)
24	Huis van Emma	Antwerp
25	Huisartsen De Kaai	Burcht (Antwerp)
26	Huisartsen Klein Antwerpen	Antwerp
27	Huisartsen Koraalberg	Antwerp
28	Kemnet	Beveren-Waas
29	Langeleem 385	Antwerp
30	Dr Martens	Antwerp
31	Netwerk Haasdonk	Haasdonk
32	Nieuwenhoven	Sint-Pieters-Leeuw
33	Drs Op de Beeck	Antwerp
34	Dr Peeters	Berchem (Antwerp)
35	Plantijn	Antwerp
36	Praktijk Blom	Heverlee (Leuven)
37	Praktijk Brugberg	Leuven
38	Praktijk De Midgaard	Wezemaal
39	Praktijk De Vossensteert	Bruges
40	Praktijk De Wijngaard	Diest
41	Praktijk Dr Bruynbroeck	Zaventem
42	Praktijk Dr Christiaens	Veltem-Beisem
43	Praktijk Dr De Groote	Kessel-Lo (Leuven)
44	Praktijk Dr Mestdagh	Haacht
45	Praktijk Dr Van Boxstael	Betekom
46	Praktijk Dr Van Deun	Leefdaal
47	Praktijk Dr Van Overmeire	Kortenberg
48	Praktijk Dr Vandevelde	Veltem
49	Praktijk Keizersberg	Wilsele (Leuven)
50	Praktijk Korte Nieuwstraat	Antwerp
51	Praktijk Lourdes	Oostakker

52	Praktijk 't Zwaantje	Tildonk
53	Praktijk Taragola/Van Mol	Melle
54	Praktijk Twee Waters	Leuven
55	Dr Raes	Antwerp
56	Regenboog	Deurne (Antwerp)
57	Rotonde	Wilrijk (Antwerp)
58	Schaliestraat	Vlezenbeek
59	Dr Slootmaeckers	Bornem
60	Sorghvliedt	Hoboken (Antwerp)
61	Sterrestraat	Lokeren
62	Stuivenbergvaart	Mechelen
63	Ter Linden	Edegem (Antwerp)
64	Universitaire Groepspraktijk (UGP)	Leuven
65	Dr Vanbeveren	Borgerhout (Antwerp)
66	Dr Veraart	Essen
67	Wel en Wee	Mechelen
68	WGC De Brugse Poort	Ghent
69	WGC De Central	Kessel-Lo (Leuven)
70	WGC De Ridderbuurt	Leuven
71	WGC De Sleep	Ghent
72	Wijkpraktijk	Antwerp

Table 4: Overview of PCPs involved in the study.

9.2 DISCUSSION OF STUDY DESIGN

For the design of the ELMO Study, we chose to conduct a cluster randomized trial. The type of intervention was the main motivation for this choice. The intervention was primarily aimed at GPs (see 9.4) and not at patients, hence a design where patients would have been randomized and causing GPs to be exposed to both the control and the intervention would have created contamination bias. One of the main limitations of choosing this design was that by randomizing patients in clusters a much larger sample was required. Despite recruiting a very large number of patients, we were unable to reach the goal set in the sample size calculation.

9.3 SELECTION OF STUDY POPULATION

9.3.1 INCLUSION AND EXCLUSION CRITERIA PCP

PCPs were considered eligible if all GPs in the PCP agreed to participate in the trial. All GPs were eligible if they:

- Collaborated with one of three laboratories (Medisch Centrum Huisartsen (MCH), Algemeen Medisch Laboratorium (AML) or Anacura);
- Agreed to use the online CPOE for their laboratory tests;
- Used an EHR for documenting routine healthcare;
- Had little or no prior experience in the use of order sets within a CPOE;
- Agreed to the terms in the clinical study agreement.

No GPs or PCPs were excluded based on other criteria such as age, demographics, size, prior use of a CPOE (without order sets), prior laboratory test ordering behaviour, etc.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

9.3.2 INCLUSION AND EXCLUSION CRITERIA PATIENTS

Patients were eligible for inclusion if laboratory tests were ordered for at least one of the 17 study indications. Patients were excluded if:

- They were younger than 18 years;
- Laboratory tests were ordered outside of the study period (01/12/2017 until 31/05/2018). A small subset of GPs was allowed to include patients until 30/06/2018 in order to recruit at least 10 patients throughout the study period.

9.3.3 WITHDRAWAL OF PATIENTS FROM THERAPY OR ASSESSMENT

Patients were free to withdraw from the study at any time without giving a reason. Patients were advised that if they requested to withdraw from the study, at any time during the trial, then this would have no negative consequences.

9.4 INTERVENTION

9.4.1 COMPUTERISED PHYSICIAN ORDER ENTRY (CPOE)

A more in-depth description of the CPOEs used in this study is available in the study protocol.

Prior to the start of the study, most laboratory test orders were done with a paper-based system. GPs requested laboratory tests by ticking boxes next to each wanted test on a paper form, manually added the patient contact detail to the form, and sent both the form and the test tubes in a plastic bag to the laboratory. This paper form was an important obstacle to the integration of decision support.

Increasingly, ambulatory laboratories in primary care have started adopting CPOEs for ordering laboratory tests. We used two different CPOEs in our study:

1. LabOnline (Moonchase) implemented at AML and MCH, and
2. E-Lab implemented at Labo Anacura.

Both systems were online platforms that allowed the ordering of laboratory tests and the review of laboratory results through a web-based interface. They were linked to the EHR and integrated patient contact details through an XML message. No other patient-specific medical data was shared between the EHR and the CPOE. When a GP initiated a laboratory test order through the EHR, a web browser opened which allowed the GP to order laboratory tests.

9.4.2 DESCRIPTION OF THE CLINICAL DECISION SUPPORT SYSTEM (CDSS)

Our CDSS was a rule-based system that suggested appropriate laboratory tests based on the indication(s) entered by the GP. The CDSS did not query the EHR for existing conditions but relied on the GP to enter the correct indication(s) into the CDSS. For each condition, several order sets were developed for distinct clinical situations. For instance, for the condition type 2 diabetes, order sets were developed for screening, diagnosis, and follow-up of the condition. For the follow-up of type 2 diabetes, separate order sets were developed for the follow-up of patients with or without diabetic nephropathy. These order sets were based on clinical practice guidelines available through the EBPracticeNet

platform [17, 18]. Included in this platform are recommendations on laboratory test ordering developed by the Flemish College of Family Physicians [19, 20].

Upon opening the CPOE, GPs were prompted to enter the indication(s) for which laboratory tests were ordered, through a searchable drop-down menu of common indications or a list of indications which could be selected through tick-boxes. Selecting one or more of these indications prompted a new window where the appropriate tests for these indications were shown as being ordered. In this window, the user was then able to accept the suggested panel, to cancel one or more of the ordered tests, or to add additional tests. The user was not restricted in ordering any tests, but was 'nudged' in the direction of ordering only the appropriate tests.

9.4.3 SELECTION OF STUDY INDICATIONS

The selection of study indications was based on four criteria: frequency in primary care, baseline inappropriateness, availability of trustworthy guidelines for primary care, and the potential for diagnostic error. The rationale for using these criteria was discussed in the protocol for the study (See Appendix **Fout! Verwijzingsbron niet gevonden.**).
After user testing and review of the CDSS functionalities, we chose to exclude obesity as a study indication. Clinical practice guidelines suggested screening for diabetes in patients with obesity and user testing informed us that it was more practical to include this order set as part of the indication type 2 diabetes. In addition to this change, user testing also informed us that a distinction was necessary in the indication diarrhoea, more specifically between chronic and acute diarrhoea. Finally, we developed our CDSS to include 17 study indications: cardiovascular disease, hypertension, type 2 diabetes, anaemia, liver pathology, medication monitoring, gout, chronic kidney disease, suspected lung embolism, suspected acute coronary syndrome, acute diarrhoea, chronic diarrhoea, thyroid disease, unexplained fatigue, sexually transmitted infections, rheumatoid arthritis, and general check-up.

9.4.4 FOLLOW-UP OF RECRUITMENT

During the study, monthly statistics on trial recruitment and CPOE use were monitored. All GPs received recruitment updates at three time points: at three months, at four months, and at five months. During these updates, GPs received a progress report regarding the number of patients they had included in the study. At the end of the trial recruitment period, a final report on the number of recruited patients was sent to each GP.

9.5 ASSESSMENTS

9.5.1 OUTCOME MEASUREMENTS

Outcome measurements were not performed on patients directly, but were collected from primary sources. In this sense, almost all outcome measurements in this study were all outcomes which were routinely collected clinical data, so-called 'real world data'.
Some measurements were not collected during routine practice and were assessed specifically for this study. Not all data were recorded by the GPs; some data were collected directly from primary sources, such as the clinical laboratories or through interviews with patients or GPs.

For the outcome measurements, we developed four distinct data collections:

1. Data collection from the LIS. This data collection contained information on the laboratory panels, including laboratory tests, indications, results, costs, etc.

2. Data collection from the EHR: this data collection contained information from the EHR, such as new diagnoses, downstream activities, and additional outcome measurements that were not routinely registered, such as therapy changes, GP opinion on diagnostic error, etc. An example of the CRF and the guide to its use can be found in Appendix **Fout! Verwijzingsbron niet gevonden.**.

3. Data collection from patient interviews: this data collection contained the results of structured patient interviews on new diagnoses and downstream activities. The CRF used for this data collection can be found in Appendix **Fout! Verwijzingsbron niet gevonden.**.

4. Data collection from GP interviews: this data collection contained the results of the interviews for the process evaluation.

Appropriateness

Appropriateness was measured based on two variables: 1. the indication(s) for the laboratory test panel, and 2. the panel of actual tests ordered by the GP (ordered panel). Data on indications and results of laboratory tests were collected one month after the end of the intervention period. This month was required to allow the laboratories to generate data on cost of the laboratory test panels, which required information from invoicing.

This data collection was initiated in months 8 to 9 of the trial, between 01/07/2018 and 31/08/2018. After resolving some queries, the full data set was finalised on 1/12/2018.

Data Collection Definition Lab
1. Data concerning patient identification
• Internal patient ID • Name • First name • Date of birth • Sex • Deceased • Date of death • Place of residence • CG1/CG2 code (billing status)
2. Data concerning laboratory test panel
• Name of GP • GP NIHDI number • Data of laboratory test order • Total cost for laboratory test panel

3. Data concerning indications
• Study indications • Selected order sets
4. Data concerning laboratory tests
• Laboratory test name • Laboratory test LOINC code • Laboratory test result • Laboratory test reference value • Normal value for laboratory test

Table 5: Data elements included in the eCRF from the LIS. Elements in red were extracted from the primary source, but were not visible for the researchers who analysed these data.

Based on the indication(s) for the laboratory test panel, a list of laboratory tests that could be expected to be ordered was generated (expected panel). This list included all possible appropriate tests.

Among the tests in the expected panel, we identified several tests that should always have been ordered for this indication. This evaluation was done based on the National Institute for Health and Care Excellence (NICE) quality standards [21].

This list was then compared with the tests in the ordered panel. Tests present in the ordered panel that were not included in the expected panel were considered inappropriate (over-utilization). Moreover, tests in the expected panel considered imperative for the given indication, but not present in the ordered panel were also considered inappropriate (under-utilization). Table 6 illustrates appropriateness for tests in a specific panel.

Appropriateness was determined based on indication and not on the used order set. Despite our use of a restrictive definition of appropriateness, this allowed for certain leniency in determining appropriateness.

Expected panel	Ordered panel	Appropriateness
Test 1	Test 1	Appropriate
Test 2	Test 2	Appropriate
Test 3 (imperative)	Missing	Inappropriate (under-utilization)
Test 4	Test 4	Appropriate
Test 5	Missing	Not ordered
Test 6	Missing	Not ordered
	Test 7	Inappropriate (over-utilization)
	Test 8	Inappropriate (over-utilization)
	Test 9	Inappropriate (over-utilization)

Table 6: definition of appropriateness based on a specific indication. In this example, six (6) tests were ordered. One (1) test was inappropriately not ordered. Three (3) tests were inappropriately ordered. Hence, this resulted in four (4) inappropriate tests on seven (7).

Laboratory test volume

Data on laboratory test volume was collected together with the data on appropriateness. One month after the end of the intervention period, data on all ordered laboratory tests were collected. Laboratory test volume was assessed as the number of test results per panel and per patient.

Tests were not clustered, which implied that for some tests, more results were reported than the number of tests that were ordered. For instance, an order for red blood cells commonly generates several results, such as red blood cell count, mean corpuscular volume (MCV), mean corpuscular haemoglobin concentration (MCHC), but in some cases also the results of a microscopic evaluation of these red blood cells. This microscopic evaluation was only performed if an automated red blood cell count was altered and triggered a microscopic evaluation. These additional results were also included in the evaluation.

Diagnostic error (DE)

To assess DE, we used multiple data sources.

1. For each patient, we collected all new diagnoses recorded by the GPs in the EHR, up to 6 months after the laboratory test order, using an eCRF. These diagnoses were automatically extracted from the EHR, including the free text label and (if available) the coding. Coding of diagnoses included an International Classification of Primary Care (ICPC-2) code and/or an International Classification of Diseases (ICD-10) code.

2. Additionally, for each new diagnosis, GPs were asked to indicate in the eCRF whether, in their opinion, this was a case of DE.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

Data Collection Definition General Practitioners
1. Data concerning patient identification
• Name • First name • Date of birth • Sex • Place of residence • Data of laboratory test order • Deceased • Date of death • CG1/CG2 code (billing status) • Phone number
2. Data concerning GP identification
• GP name • GP NIHDI number
3. Data concerning the reason for ordering tests

• *Reason for ordering laboratory test*
  ○ *To exclude disease*
  ○ *To confirm diagnosis*
  ○ *At patient's request*
  ○ *To reassure patient*
  ○ *Physician's uncertainty*
  ○ *To determine treatment*
  ○ *Check-up for known disorder or screening*
  ○ *Other reason, specify*
• *Pre-test estimate of disease*
  ○ *Certainly not*
  ○ *Probably not*
  ○ *Maybe*
  ○ *Probably yes*
  ○ *Certainly yes*

4. Data concerning new diagnoses
• New diagnosis • Date of new diagnosis • *Relation to laboratory test order* • *Relation to laboratory test results* • *Possible diagnostic error*
5. Data concerning downstream activities

• *Downstream activities*
  ○ *Downstream investigations pick list*
  ▪ *Follow-up laboratory tests*
  ▪ *Imaging*
  ▪ *Function tests*

○ *Downstream activities name* • *Referrals* ○ *Specialty* ○ *Investigations after referral* • *Treatment changes* ○ *Drug treatment – start* ○ *Drug treatment – stop* ○ *Drug treatment – change (e.g. other posology)* ○ *Blood transfusion* ○ *Surgery* ○ *Oncologic treatment (chemotherapy, radiotherapy, immunotherapy, …)* ○ *Physical therapy or occupational therapy* ○ *Psychotherapy* ○ *Other: (free text)*

Table 7: Data elements included in the eCRF from the EHR. Elements in red were extracted from the primary source, but were not visible for the researchers who analysed these data. Elements in italic were unstructured data that needed to be manually added by the GPs.

This eCRF-data collection was initiated on 01/08/2018. Due to technical difficulties in the transfer of the eCRFs, this data collection was extended with six (6) months to 12 months after the end of the intervention period. The data collection was finalised in July 2019.

3. Finally, as a double-check, we assessed potential cases of DE by interviewing a subset of patients. These patients were selected based on the new diagnoses recorded by GPs and whether, in the opinion of the GP, this was a case of DE. We selected a random sample from those patients where no new diagnoses were recorded in the EHR in the six (6) months after the laboratory test order. We also selected a random sample from those patients where the GP had documented a potential case of DE. This data collection was initiated almost one year after the laboratory test order, in November 2018. The data collection lasted four (4) months.

Data Collection Definition Patient
1. Data concerning patient identification
• Internal patient ID • Name • First name • Date of birth • Sex • Deceased • Date of death • Place of residence
2. Data concerning laboratory test order
• Date of laboratory test order • Indication for laboratory test order

3. Data concerning new diagnoses

• *Diagnosis name*
• *Date of new diagnosis*
• *Relation to laboratory test order*
• *Diagnosis result of laboratory test order*
• *Additional investigations*
• *Diagnostic error*
• *Type of diagnostic error*
  ○ *No diagnosis was made*
  ○ *Diagnosis made too late*
  ○ *Wrong diagnosis was made*

4. Data concerning downstream activities

• *Downstream investigations*
  ○ *Investigation type*
  ▪ *Follow-up laboratory tests*
  ▪ *Imaging*
  ▪ *Function tests*
  ○ *Specification of investigation type*
• *Downstream referrals*
  ○ *Referral specialty*
  ○ *Investigations after referral*
  ○ *Investigation type*
  ▪ *Follow-up laboratory tests*
  ▪ *Imaging*
  ▪ *Function tests*
• *Treatment changes*
  ○ *Drug treatment – start*
  ○ *Drug treatment – stop*
  ○ *Drug treatment – change (e.g. other posology)*
  ○ *Blood transfusion*
  ○ *Surgery*
  ○ *Oncologic treatment (chemotherapy, radiotherapy, immunotherapy, …)*
  ○ *Physical therapy or occupational therapy*
  ○ *Psychotherapy*
  ○ *Other: (free text)*

Table 8: Data elements included in the eCRF for the patient interviews. The eCRF was a web-based application. Elements in red were visible to the interviewer, but were not visible for the researchers who analysed these data. Elements in italic were unstructured data that needed to be manually added by the interviewer.

We used the diagnoses extracted from the EHRs as our principal data source to evaluate the risk for possible diagnostic error.

We developed an algorithm to detect cases of possible DE by using a combination of the indication(s) for the preceding laboratory panel and the ICPC-2 code of the registered diagnosis. The flow of this algorithm is described in Figure 3.

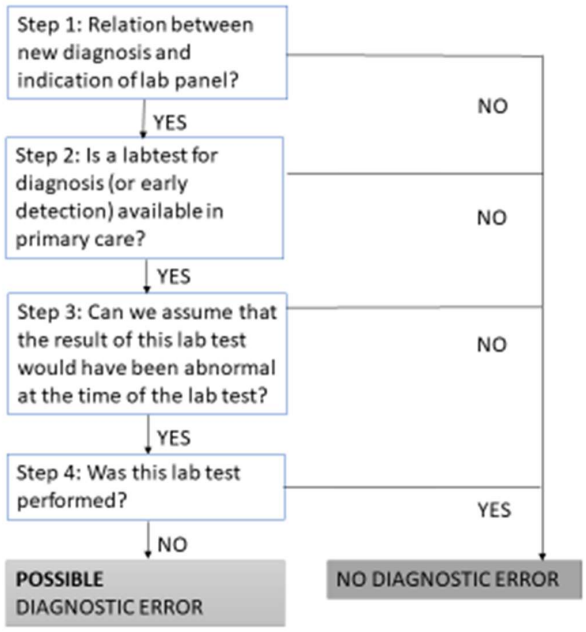


Figure 3: Algorithm for detecting possible DE.

This algorithm was used by a panel of academic clinicians for each possible combination of a new diagnosis and indication. This meant that if a laboratory panel was ordered for three different indications (i.e. type 2 diabetes, chronic kidney disease and fatigue) and the patient developed two new diagnoses in the six months after the laboratory test order (i.e. viral bronchitis and acute kidney failure), then for this patient, 6 different combinations of indication and diagnosis were assessed.
In a preparatory phase we provided each diagnosis with an ICPC-2-code. Diagnoses that had not yet been coded with an ICPC-2-code in the EHR, were given a code by converting the available ICD-code into the corresponding ICPC-2-code. If no code was available at all, free text labels of the diagnoses were evaluated and an ICPC-2-code was assigned. See Appendix **Fout! Verwijzingsbron niet gevonden.** for the ICPC-2-coding system.

Based on their ICPC-2-code some diagnoses were excluded for evaluation with the algorithm. More specifically, ICPC-2-codes indicating psychological disorders, social problems, symptom diagnoses (e.g. headache, cough…) and the codes A98 and A97 for 'prevention' and 'no disease respectively.

In the first step of the algorithm, a relationship between the new diagnosis and the indication for the laboratory test order was established. In a second step, the general availability in primary care of a laboratory test to detect or suspect the new diagnosis was assessed. In a third step the timeliness of the laboratory test was evaluated. For this

evaluation, the academic clinicians determined how long before the clinical manifestation of the new diagnosis, the laboratory test would have been altered. We then checked whether the time between the laboratory panel and the detection of the new diagnosis was within the limits determined in the previous step. In a final step, we assessed whether the test that would have detected the new diagnosis was actually performed.

This algorithm was used for each combination of new diagnosis and indication by two academic clinicians independently. Conflicts were resolved in group.

DE was assessed as a binary variable. All combinations of a diagnosis and an indication were assessed as either: 1. 'possible DE' or 2. 'no DE'. A diagnosis was only considered as a 'possible DE' once, even if there were multiple combinations that were evaluated as 'possible DE'.

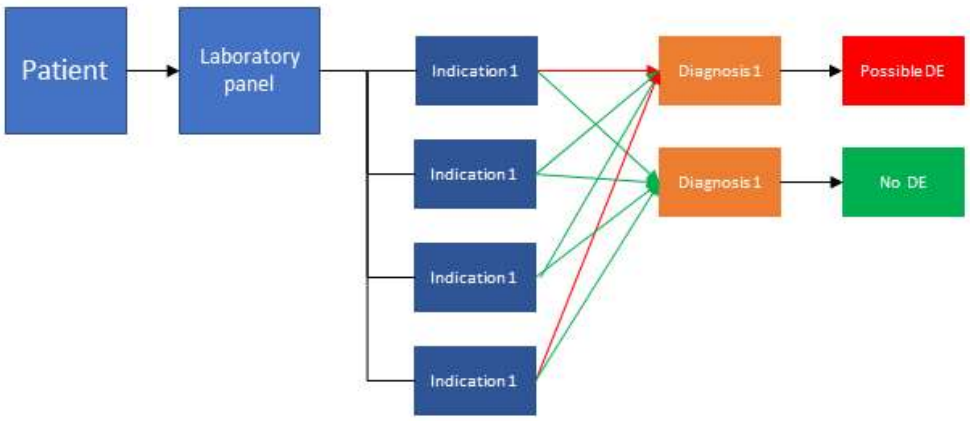


Figure 4: Illustration of the assessment of DE. Red arrows indicate combinations of indication and diagnosis that were assessed as a possible DEs by academic clinicians. Green arrows indicate combinations that were not considered possible DEs.

Downstream activities

Originally, we planned to collect data on this topic through patient interviews only, but we later decided to include the data on downstream activities in the data collection for new diagnoses. GPs were asked to record in the eCRF all additional laboratory tests, functional tests, radiographic tests and referrals that were ordered as a result of the initial laboratory test results. In addition, GPs were asked to indicate whether the laboratory test results led to the initiation, change or stop of a treatment.

9.5.2 PROCESS EVALUATION

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

At the end of the intervention period, a process evaluation was performed to inform further implementation strategies and improvements of the intervention. After collecting an informed consent (ICF can be found in Appendix **Fout! Verwijzingsbron niet gevonden.**), semi-structured interviews with intervention GPs were performed. Interviews were audio recorded and a verbatim transcript of all interviews was used for the data coding. These interviews were initiated at the end of the study intervention period and lasted 3 months.

9.5.3 SCHEDULE OF EXAMINATIONS AND DATA COLLECTIONS

Table 9 gives an overview of the schedule of examinations and data collections.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

Schedule of data measurements and collections

		Study month															
		0	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**	**20**	**22**	**24**	**26**	**28**	**30**
Study		Intervention				Data collections							Analysis & Reporting				
Data appropriateness						■	■	■									
Data diagnostic error	EHR					■	■	■	■	■	■	■					
Data diagnostic error	Patient							■	■								
Data volume						■	■	■									
Data downstream activities						■	■	■	■	■	■	■					
Data process evaluation						■	■										

Table 9: Overview of data measurements and collections. Diagnostic error was assessed through chart review and by patient interviews. Study start date was 01/12/2017.

9.6 DATA QUALITY ASSURANCE

There are challenges to the assurance of data quality when using routinely collected data in clinical trials. We chose to use data collected within the EHR and LIS for our outcome measurements as opposed to subjecting patients to additional visits specifically for this trial. To minimize potential sources of bias in the reuse of routine collected clinical data, we installed several data quality assurance measures.

The main concerns on data quality were:

1. Data completeness and data quality of the GP EHR data.

   Data in the EHR may not be complete due to several reasons [22]:

   a. The patient did not visit the GP, so that the event was unknown to the GP.
   b. The event was known to the GP but not recorded in the EHR.
   c. The event was recorded but in a manner that is was not meaningful and hence unextractable by the eCRF in an automated manner.
   d. The event was recorded in a structured manner but did not correctly represent the longitudinal character of the event. Not all EHRs facilitated episode-oriented registration which posed a problem for chronic conditions. Chronic conditions were often registered multiple times, each time a patient visited his GP for this condition. EHRs that did not facilitate episode-oriented registration caused an over-inflation of the incidence of this condition, because the chronic condition was recorded multiple times as if it were a new diagnosis.
   e. Data on downstream activities in the EHR do not always include information on what triggered the investigation or referral. It may be a true follow-up investigation due to an altered laboratory test result, however it may also be that the investigation was planned anyhow because of the symptoms the patient presented with, regardless of the laboratory test results.

   In addition to these concerns that are inherent to the nature of the data in the EHR, there was an additional concern regarding the outcome assessment for DE. There are reasons to believe that GPs may not readily report on cases of DE for reasons of liability [23]. Hence, self-reporting of DE may not be a valid method for assuring high quality data for this outcome assessment.

2. Data completeness of the LIS data.

   Concerns regarding the quality of the LIS data were less substantial. Laboratories have a responsibility to report all results of ordered test, hence concerns regarding data completeness and quality of laboratory test results were limited. Due to lack of interoperability, it was not possible to provide the LIS with clinical information regarding the indications for laboratory test ordering directly from within the EHR. This required the GP to manually record the indications for which laboratory tests were ordered. Concerns on data completeness were restricted to concerns regarding the recorded indications. It is possible that GPs failed to record all the relevant indications in the CPOE, hence influencing outcome assessments.

Alternatively, it is possible that some GPs recorded too many indications, again, influencing outcome assessments.

To limit concerns on data completeness for the clinical data collected from the GP EHR, we chose to organise the data collection from the EHR in a uniform manner. CRAs were trained in the use of the five EHR systems used by the GPs in the trial. More specifically, they were trained in how to recognize elements that were relevant to the outcome measurements of interest for our study within the structure of the different EHR systems. They understood how the eCRF interacted with the EHR system and which structured fields were automatically queried by the eCRF. These CRAs then made appointments with each PCP to organise the collection of the eCRFs and to ensure that data that could not be automatically queried and needed input from the GPs was prepared beforehand. The CRAs: 1. opened each patient's EHR record; 2. reviewed the information in the EHR record; 3. initiated the eCRF which automatically extracted structured outcome measurements; and 4. manually added outcome measurements that were recorded in the EHR but could not be automatically extracted by the eCRF. This method ensured that unstructured data was maximally recorded into the eCRF and that concerns regarding completeness of data were minimized as much as possible.

Although the patient interviews are only a minor data source, merely intended as double-check, there are some concerns on the added value and reliability of the data.
Due to technical problems, the interviews took place almost 1 year after the initial laboratory test. Patients recollection of new diagnoses after this laboratory test might be flawed, moreover because patients might have had multiple laboratory tests in the previous year.

9.7 PLANNED STATISTICAL METHODS & SAMPLE SIZE

All planned statistical methods were outlined in the DMP & SAP, available in Appendix **Fout! Verwijzingsbron niet gevonden.**.

9.7.1 STATISTICAL AND ANALYTICAL PLANS

Appropriateness

For the definition of the primary outcome, three numbers were relevant: (a) the number of requested tests which are appropriate, (b) the number of requested tests which are inappropriate and (c) the number of inappropriately not-requested tests (inappropriate under-utilization). The latter number was only relevant for diabetes mellitus, chronic kidney disease, rheumatoid arthritis and thyroid disease. Per patient, aggregated over panels if multiple panels were available, the primary outcome was defined by the ratio (a)/(a+b+c). We referred to this ratio as the proportion of appropriate tests in the remainder.
To assess differences between the allocated groups in the proportion appropriate tests, a logistic generalized estimating equation (GEE) model was be used: of interest were the marginal proportions, not the proportions on patient, GP or PCP level.

The logistic GEE model included the allocated group and laboratory as factors and PCP as the clustering variable. The effect of the intervention was expressed as the difference in proportions and presented together with its associated 95% confidence interval. The proportion of appropriate tests in the two allocated groups was also estimated from the GEE model and presented with their 95% confidence intervals.
Appropriateness for the composite of all study tests was compared between intervention and control groups. Furthermore, an additional analysis was performed that only included patients who have no indications in addition to the 17 study indications. This additional analysis corrected for an overestimation of inappropriate tests when more than one indication is selected, including indications not under evaluation. These tests would be considered inappropriate even though they could be appropriate according to one of the other indications not being evaluated.
The analyses were performed on all patients from all GPs according to their allocated group.
Appropriateness for each study indication separately was performed as a secondary analysis. In this analysis we included the number of additional indications for the panel as a factor in the analysis.

Test volume

The total number of tests was analysed using a GEE model for count data (Poisson or Negative Binomial to handle potential overdispersion) that includes allocated group and laboratory as factors in the model and PCP as clustering variable. No offset was used. The number of tests per patient for each group was estimated from the model and presented together with their associated 95% confidence intervals. The effect of the intervention was presented as the ratio between the two numbers with its 95% confidence interval. Statistical significance was assessed at a significance level of 5%.

Diagnostic error

The proportion of patients with a missed diagnosis was analysed by means of a logistic GEE model that included allocated group and laboratory as factors and used PCP as the clustering variable. An independent working correlation matrix was used. The proportion of patients with a missed diagnosis and associated 95% confidence intervals was estimated from the model.
The difference in proportions was obtained by subtracting the two proportions. The associated standard error was calculated from the rules for the variance of a difference between two independent estimates. The 95% confidence interval for the difference was also calculated.
The non-inferiority limit for missed diagnoses was 1%, hence the intervention was deemed non-inferior if the difference between the allocated groups (intervention – control) was shown to be less than 1%. Therefore, the intervention was deemed non-inferior if the upper limit of the 95% confidence interval was below 1.
As for the primary endpoint, the analysis was performed for all 17 study indications together.

Downstream activities

The objective of this outcome was merely exploratory. However, the total number of downstream activities following a laboratory panel was also analysed, using the same methodology as for the total number of tests (test volume, see supra).

Process evaluation

The process evaluation used the QUAGOL protocol for qualitative research as methodology [24]. The data was analysed using the GUIDES checklist [25] as framework for identifying themes and concepts in the interviews. Concepts were identified in the transcripts and grouped under overlying themes. These themes were ordered, again, using the GUIDES checklist [25] as guide.

9.7.2 DETERMINATION OF SAMPLE SIZE

For sample size calculations, we refer to the protocol and appendix 3 of the protocol. At the end of the six-month intervention period the planned sample size was verified based on the number of recruited GPs and the average number of patients per GP. With 280 study GPs (clusters), a sample of 12740 patients (45.5 patients on average per GP) would be necessary to have at least 80% power for the secondary outcome (diagnostic error), assuming the original intra-cluster correlation (ICC) and DE rates. At that point we had recruited 11200 patients and it was deemed unfeasible to recruit an additional 1500 patients. Since most recruitments were realised by GPs who had already recruited 50 or more patients, the Steering Committee agreed that attempts to recruit additional patients should target GPs with a low number of recruited patients. Therefore, only those GPs who had not yet recruited 10 patients were allowed another month to recruit additional patients. Specific calculations regarding the sample size and the assumed intra-cluster correlations can be found in the protocol under 9.1.

The observed ICC for appropriateness of laboratory test ordering in the ELMO Study was 0.04629. In our trial, we observed that PCPs had on average 3.89 GPs, that GPs included on average 35.59 patients, and that they ordered 32.6 tests per patient. Using the same methods as in the Study Protocol (see Appendix **Fout! Verwijzingsbron niet gevonden.**), the design effect was 181. With the observed number of tests per patient this design effect would have required a sample of 106 066 tests or 3254 patients to have at least 80% power to detect the assumed difference in the primary outcome of 10 percentage points (70 versus 80% appropriateness).

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

9.8 CHANGES IN THE CONDUCT OF THE STUDY OR PLANNED ANALYSES

After presenting preliminary results to the SC, some ad hoc sensitivity analyses were performed to investigate potential bias.

To assess the effect of age difference between both groups, the planned analysis for the primary outcome was also performed on subgroups of patients younger than 45, between 45 and 65, and older than 65 years. In addition, the analysis for the composite outcome was also performed on the full population, but also including age as a factor in the analysis. The analysis was also performed on a subset of the total population where PCPs with extreme age differences were omitted.

To assess potential documentation bias, a comparison of several signal tests was made between subgroups in both arms. For instance, the results of mean value for TSH was compared in the subgroup of thyroid disease patients in both arms. This allowed us to evaluate whether both subgroups were comparable. The planned analysis for the primary outcome was also performed on the population without patients for which a laboratory panel for general check-up was performed. This analysis was judged important because the SC considered that potential documentation bias would have been most probable in this subgroup of patients.

9.9 PROTOCOL AMENDMENTS

A single change to the protocol was made (See Appendix **Fout! Verwijzingsbron niet gevonden.**). Stratifying PCPs by prior experience in the use of a CPOE was not deemed feasible because GPs had difficulties objectifying this experience. In addition to this, often the experience in the use of a CPOE varied across GPs in the same PCP making it difficult to make an overall judgement for the whole PCP.

All three laboratories were at a different phase in implementing their CPOE. GPs affiliated with AML had limited experience in using a CPOE, GPs affiliated with Anacura had moderate to high experience in the use of a CPOE, and GPs affiliated with MCH had no experience in the use of a CPOE. Hence, we chose to stratify PCPs according to the laboratory with which they were affiliated, rather than self-reported experience in the use of a CPOE.

Piessens V, *et al. BMJ Open* 2022; 12:e059261. doi: 10.1136/bmjopen-2021-059261

10 REFERENCES

1. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. Implementation Science. 2015;10:167.

2. Hickner J, Thompson PJ, Wilkinson T, Epner P, Shaheen M, Pollock AM, et al. Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results. J Am Board Fam Med. 2014;27:268–74.

3. Statistieken terugbetaalde bedrage en akten van artsen en tandartsen. RIZIV. http://www.riziv.fgov.be/nl/statistieken/geneesk-verzorging/2015/Paginas/terugbetaalde_bedragen_akten_arts_tandarts.aspx#.WecMTlu0O71. Accessed 18 Oct 2017.

4. O'Sullivan JW, Stevens S, Hobbs FDR, Salisbury C, Little P, Goldacre B, et al. Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. BMJ. 2018;363:k4666.

5. Driskell OJ, Holland D, Hanna FW, Jones PW, Pemberton RJ, Tran M, et al. Inappropriate requesting of glycated hemoglobin (Hb A1c) is widespread: assessment of prevalence, impact of national guidance, and practice-to-practice variability. Clin Chem. 2012;58:906–15.

6. Davis P, Gribben B, Lay-Yee R, Scott A. How much variation in clinical activity is there between general practitioners? A multi-level analysis of decision-making in primary care. J Health Serv Res Policy. 2002;7:202–8.

7. Leurquin P, Van Casteren V, De Maeseneer J. Use of blood tests in general practice: a collaborative study in eight European countries. Eurosentinel Study Group. Br J Gen Pract. 1995;45:21–5.

8. O'Kane MJ, Casey L, Lynch PLM, McGowan N, Corey J. Clinical outcome indicators, disease prevalence and test request variability in primary care. Ann Clin Biochem. 2011;48 Pt 2:155–8.

9. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The Landscape of Inappropriate Laboratory Testing: A 15-Year Meta-Analysis. PLoS ONE. 2013;8:e78962.

10. De Sutter A, Van den Bruel A, Devriese S, Mambourg F, Van Gaever V, Verstraete A, et al. Laboratorium testen in de huisartsgeneeskunde. KCE reports. Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2007.

11. Houben PHH, van der Weijden T, Winkens RAG, Grol RPTM. Cascade effects of laboratory testing are found to be rare in low disease probability situations: prospective cohort study. J Clin Epidemiol. 2010;63:452–8.

12. van Wijk MAM, van der Lei J, Mosseveld M, Bohnen AM, van Bemmel JH. Assessment of Decision Support for Blood Test Ordering in Primary Care. A Randomized Trial. Ann Intern Med. 2001;134:274–81.

13. Delvaux N, Van Thienen K, Heselmans A, de Velde SV, Ramaekers D, Aertgeerts B. The Effects of Computerized Clinical Decision Support Systems on Laboratory Test Ordering: A Systematic Review. Arch Pathol Lab Med. 2017;141:585–95.

14. Chan AJ, Chan J, Cafazzo JA, Rossos PG, Tripp T, Shojania K, et al. Order sets in health care: a systematic review of their effects. Int J Technol Assess Health Care. 2012;28:235–40.

15. Naugler C, Ma I. More than half of abnormal results from laboratory tests ordered by family physicians could be false-positive. Can Fam Physician. 2018;64:202–3.

16. Rang M. The Ulysses syndrome. Can Med Assoc J. 1972;106:122–3.

17. Van de Velde S, Vander Stichele R, Fauquert B, Geens S, Heselmans A, Ramaekers D, et al. EBMPracticeNet: A Bilingual National Electronic Point-Of-Care Project for Retrieval of Evidence-Based Clinical Guideline Information and Decision Support. JMIR Res Protoc. 2013;2. doi:10.2196/resprot.2644.

18. Delvaux N, Van de Velde S, Aertgeerts B, Goossens M, Fauquert B, Kunnamo I, et al. Adapting a large database of point of care summarized guidelines: a process description. J Eval Clin Pract. 2017;23:21–8.

19. Avonts M, Cloetens H, Leyns C, Delvaux N, Dekker N, Demulder A, et al. Aanbeveling voor goede medisch praktijkvoering: Aanvraag van laboratoriumtests door huisartsen. Huisarts Nu. 2011;:S1–55.

20. Leysen P, Avonts M, Cloetens H, Delvaux N, Koeck P, Saegeman V, et al. Richtlijn voor goed medische praktijkvoering: Aanvraag van laboratoriumtests door huisartsen - deel 2. Antwerpen: Domus Medica vzw; 2012.

21. NICE indicator menu: new, updated and retired indicators. National Institute for Health and Care Excellence (NICE); 2019. https://www.nice.org.uk/standards-and-indicators.

22. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. Journal of Medical Internet Research. 2018;20:e185.

23. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med. 2009;169:1881–7.

24. Dierckx de Casterlé B, Gastmans C, Bryon E, Denier Y. QUAGOL: a guide for qualitative data analysis. Int J Nurs Stud. 2012;49:360–71.

25. Van de Velde S, Kunnamo I, Roshanov P, Kortteisto T, Aertgeerts B, Vandvik PO, et al. The GUIDES checklist: development of a tool to improve the successful use of guideline-based computerised clinical decision support. Implementation Science. 2018;13:86.

26. van Wyk JT, van Wijk MA, Sturkenboom MC, Mosseveld M, Moorman PW, van der Lei J. Electronic alerts versus on-demand decision support to improve dyslipidemia treatment: a cluster randomized controlled trial. Circulation. 2008;117:371–8.

27. Main C, Moxham T, Wyatt JC, Kay J, Anderson R, Stein K. Computerised decision support systems in order communication for diagnostic, screening or monitoring test ordering: systematic reviews of the effects and cost-effectiveness of systems. Health Technol Assess. 2010;14. doi:10.3310/hta14480.

28. Krogsbøll LT, Jørgensen KJ, Gøtzsche PC. General health checks in adults for reducing morbidity and mortality from disease. Cochrane Database of Systematic Reviews. 2019. doi:10.1002/14651858.CD009009.pub3.

29. Kohn L, Corrigan J, Donaldson M. To err is human: building a safer health system. Washington DC: National Academy Press; 2000.

30. Diagnostic Errors: Technical Series on Safer Primary Care. Geneva: World Health Organization (WHO); 2016.

31. Lippi G, Plebani M, Simundic A-M. Quality in laboratory diagnostics: from theory to practice. Biochem Med. 2010;20:126–30.

32. European IVD Market Statistics Report 2017. Belgium: MedTech Europe; 2017. https://www.medtecheurope.org/wp-content/uploads/2018/12/European-IVD-Market-Statistics-2017.pdf.


KU LEUVEN
ACADEMISCH CENTRUM
HUISARTSGENEESKUNDE

DESCRIPTION OF THE CLINICAL DECISION SUPPORT SYSTEM (CDSS)

Our CDSS was a rule-based system that suggested appropriate laboratory tests based on the indication(s) entered by the GP. The CDSS did not query the EHR for existing conditions but relied on the GP to enter the correct indication(s) into the CDSS. For each condition, several order sets were developed for distinct clinical situations. For instance, for the condition type 2 diabetes, order sets were developed for screening, diagnosis, and follow-up of the condition. For the follow-up of type 2 diabetes, separate order sets were developed for the follow-up of patients with or without diabetic nephropathy. These order sets were based on clinical practice guidelines available through the EBPracticeNet platform. Included in this platform are recommendations on laboratory test ordering developed by the Flemish College of Family Physicians.

Upon opening the CPOE, GPs were prompted to enter the indication(s) for which laboratory tests were ordered, through a searchable drop-down menu of common indications or a list of indications which could be selected through tick-boxes. Selecting one or more of these indications prompted a new window where the appropriate tests for these indications were shown as being ordered. In this window, the user was then able to accept the suggested panel, to cancel one or more of the ordered tests, or to add additional tests. The user was not restricted in ordering any tests, but was 'nudged' in the direction of ordering only the appropriate tests.

SELECTION OF STUDY INDICATIONS

The selection of study indications was based on four criteria: frequency in primary care, baseline inappropriateness, availability of trustworthy guidelines for primary care, and the potential for diagnostic error. The rationale for using these criteria was discussed in the protocol for the study.

After user testing and review of the CDSS functionalities, we chose to exclude obesity as a study indication. Clinical practice guidelines suggested screening for diabetes in patients with obesity and user testing informed us that it was more practical to include this order set as part of the indication type 2 diabetes. In addition to this change, user testing also informed us that a distinction was necessary in the indication diarrhoea, more specifically between chronic and acute diarrhoea. Finally, we developed our CDSS to include 17 study indications:

1. Cardiovascular disease
    a. screening
    b. follow-up
2. Hypertension
    a. Diagnosis
    b. Follow-up general
    c. Diuretic or ACE-I/sartan treatment
    d. Hypertensive nephropathy
3. Type 2 diabetes
    a. screening
    b. Follow-up 3 monthly
    c. Annual follow-up
    d. ACE-I/sartan treatment
4. Anaemia
    a. Clinical suspicion of anaemia
    b. Microcytic or normocytic anaemia
    c. Macrocytic anaemia
5. Liver pathology
    a. Clinical suspicion or in case of risk factors
    b. Follow-up
6. Medication monitoring
    a. Statins

   b. Rheumatoid arthritis treatment (methotrexate, azathioprine, leflumonide, sulfasalazine, cyclofosfamide, chloorambucil)
   c. Diuretic or ACE-I/sartan treatment
   d. Isotretinoin
7. gout
8. chronic kidney disease
   a. Screening (diabetes, hypertension, CVD, family history of stage V CKD)
   b. Monitoring stage I - IIIa (eGFR ≥ 45)
   c. Monitoring stage IIIb (eGFR 30-44)
   d. Monitoring stage IV - V (eGFR ≤ 29)
9. suspected lung embolism
10. suspected acute coronary syndrome
11. acute diarrhoea
   a. Patients at risk
   b. Elderly
12. chronic diarrhoea
13. thyroid disease
   a. Diagnosis
   b. Monitoring after treatment changes
   c. Monitoring stable disease
14. unexplained fatigue
15. sexually transmitted infections
   a. Screening
   b. Diagnosis
16. rheumatoid arthritis
   a. Diagnosis
   b. Follow-up RA treatment (methotrexate, azathioprine, leflumonide, sulfasalazine, cyclofosfamide, chloorambucil)
17. general check-up.

DEFINITIONS

Indication: the reason(s) why the laboratory exam was ordered, e.g. anaemia, STD, diabetes, … . There were 17 predefined indications, but GPs could add other indications in free text. Some of the predefined indications had sub-indications, e.g. for 'diabetes', GPs could specify whether it was for screening, 3-monthly or annual follow-up.

Order set: the set of appropriate lab tests the CDSS proposes when selecting a single indication, e.g. ticking diabetes-annual follow-up would prompt the following tests: cholesterol (total, HDL, LDL, triglycerides, glucose fasting, HbA1c, Creatinine, eGFR and albumin/creatinine ratio (urine)

Lab panel: all lab tests ordered during a single patient encounter

Lab result: the complete report of all test results.

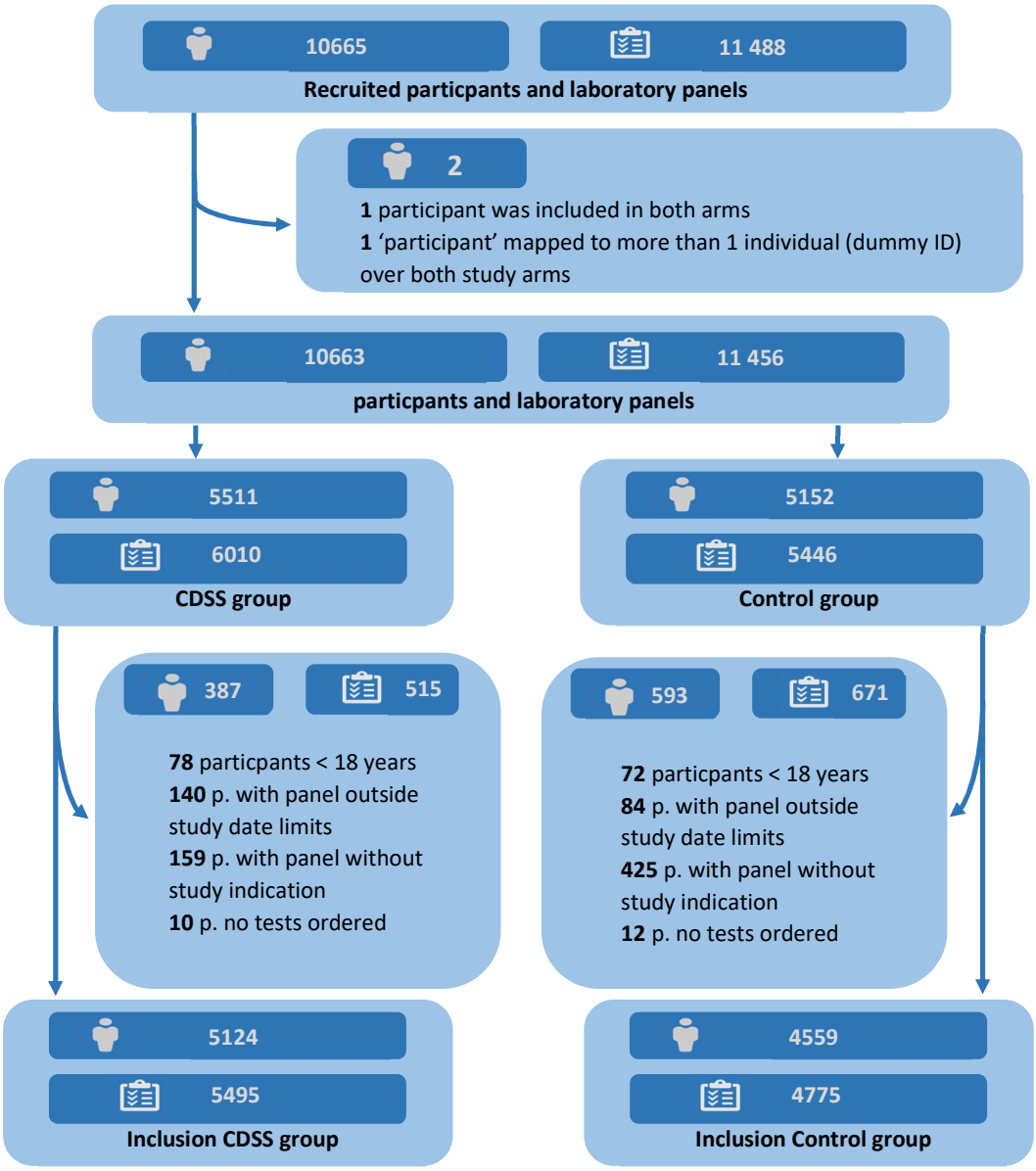
Tests (as outcome measure): all parameters reported in the lab result: e.g. TSH, GPT, RBC. In some cases a single ordered test (e.g. leukocyte formula) would result in multiple reported tests (neutrophiles, eosinophiles, monocytes, …). We used the number of reports tests to calculate the outcome measure (test volume).

Test result: the value reported in the lab result, e.g. 12,5mg/dl (for the test hemoglobine).

Abnormal test: a test of which the result falls outside the reference values. We used the reference values provided by the participating laboratories.

Appendix 2

1/ Flowchart of recruited participants and their lab panels and final inclusion



2/ Lab Panel, Patient and Physician characteristics of those panels without 6 month follow-up data , compared to those with follow-up data.

	Labs without eCRF	Labs with eCRF
Lab panel characteristics (N)	952	9318
Number of tests/panel (mean, CI95%)	26,3 (23,4 - 29,1)	26,3 (23,4 - 29,1)
Proportion appropriate tests/panel (mean, CI95%)	0,58 (0,53 - 0,63)	0,58 (0,53 - 0,63)
Patient characteristics		
Age	56,4	56,3
% female	57,1%	54,6%
Physician characteristics		
Age	43,3	43,8
% female	51,1%	58,5%

eCRF = Electronic Case Report Form (providing 6 month follow-up data on new diagnoses and downstream activitities. CI95% = 95% Confidence Interval