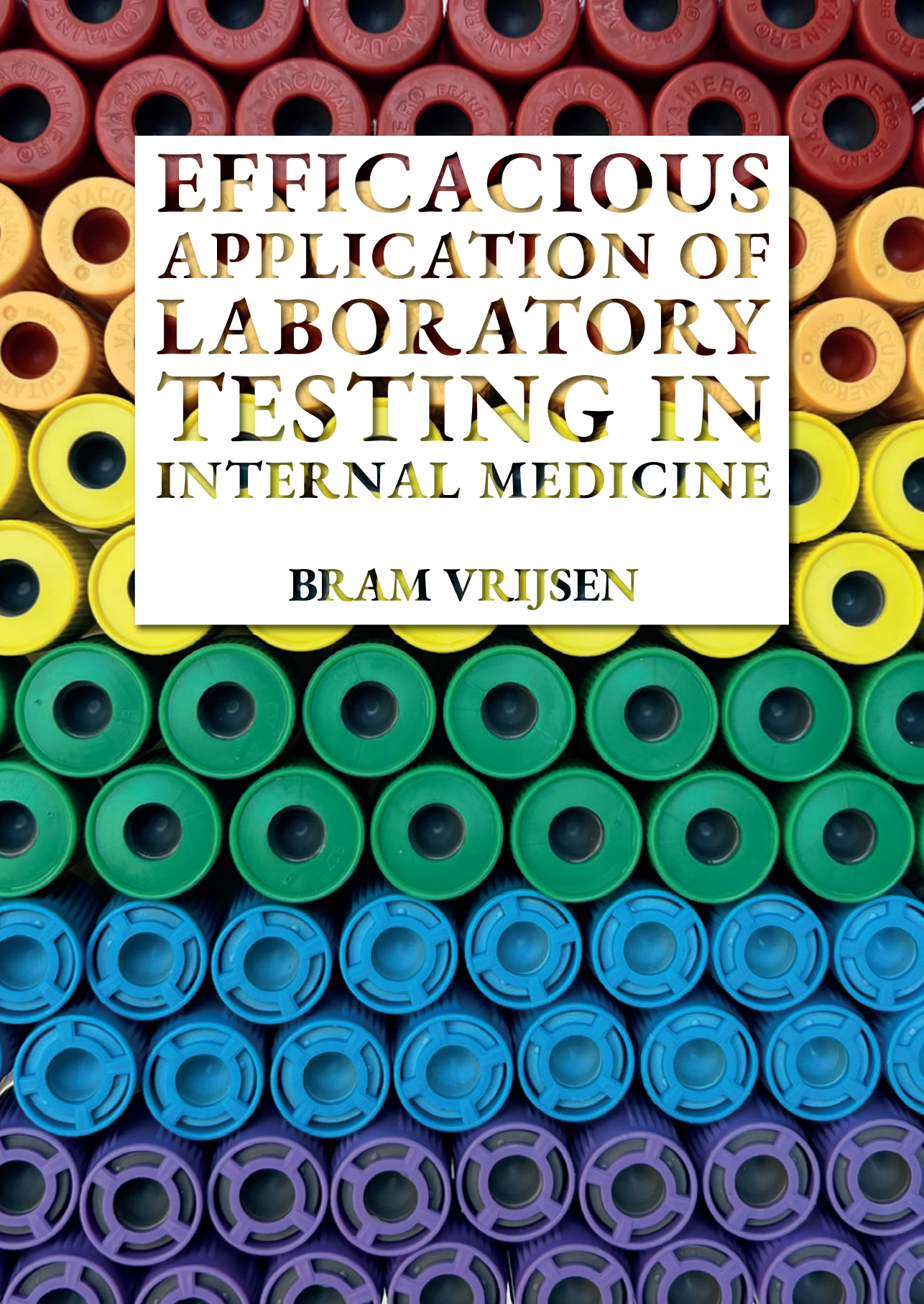


EFFICACIOUS APPLICATION OF LABORATORY TESTING IN INTERNAL MEDICINE

BRAM VRIJSEN

Efficacious application of laboratory testing in internal medicine

BRAM VRIJSEN

Efficacious application of laboratory testing in internal medicine

ISBN 9789039375709

Set in EB Garamond 11

Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Tara Schollema, persoonlijkproefschrift.nl

Efficacious application of laboratory testing in internal medicine

Doeltreffende toepassing
van laboratoriumonderzoek
in de interne geneeskunde
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 26 oktober 2023 des middags te 4.15 uur

door

Bram Evert Luciën Vrijsen
geboren op 5 december 1978
te Tilburg

Promotor:

Prof. dr. W.W. van Solinge

Copromotoren:

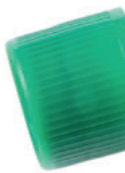
Dr. M.J. ten Berg
Dr. J. Westerink

Beoordelingscommissie:

Prof. dr. J. Frenkel
Prof. dr. P.W.B. Nanayakkara
Prof. dr. J.P. Ruurda (voorzitter)
Prof. dr. F.L.J. Visseren
Dr. D.L.M. Zwart

voor Manu

Table of contents



Chapter 1

Introduction

Laboratory testing has had a prominent place in modern medicine since its advent in the early 20th century. Yet the field of clinical chemistry has roots that go back even further. In the early 19th century, physicians started to apply chemical processes to medicine, inspired by the newly developed experimental orientation of medicine and the emergence of organic chemistry in the late 18th century.[1] Initially, most tests were performed at the bedside of the patient, but improvements in the knowledge of chemical compounds and their analysis led to the establishment of independent clinical laboratories as early as the 1840s.[2] Since then, clinical chemistry has continued to flourish, owing to a plethora of technological advances, such as the Vacutainer, invented in 1946, and the first automated chemical analyzer, developed in 1959.[3]

Since then, clinical chemistry has become indispensable in clinical medicine. A widely cited estimate states that up to 70% of downstream treatment decisions are affected by laboratory testing.[4] This estimate has been criticized for having little evidence to back it up.[5,6] However, the ambiguity about the exact figure may say more about the difficulty in quantifying the value of laboratory testing than about the value per se.[7] Medical decision making is by its nature a multifactorial process, and the individual contributions of its separate components, one of which being laboratory testing, are difficult to untangle.

Setting aside the exact figure, the claim that a large proportion of medical decisions are, at least partially, based on laboratory testing does seem to have merit. In primary care, almost a third of laboratory test leads to downstream activities, the most frequent of which were additional laboratory testing and prescribing new medication.[8] A survey of German and U.S. cardiologists and oncologists found that in vitro testing led to substantial clinical decisions in 66% of patients who had undergone testing.[9] Furthermore, it showed that laboratory testing was used in up to 88% of patients when they were initially diagnosed.

From this viewpoint, the diagnosis or exclusion of diseases is one of the principal roles of laboratory testing. An abundant number of medical conditions are dependent on laboratory testing for their diagnosis, such as coagulopathies, infectious diseases, and endocrine disorders. Besides in diagnosis, laboratory testing also plays an important part in the monitoring of treatments and the follow-up and screening of diseases.[10,11]

Costs of laboratory testing

However, the value of laboratory testing comes at a cost. The costs of laboratory testing differ depending on how one looks at it. In the United States, laboratory

testing constitutes the highest-volume health care service, with over 12 billion individual tests performed annually.[12] Due to the sheer number of tests being ordered, the costs of laboratory testing add up to a sizable amount. One study found that in vitro diagnostics, which not only includes laboratory testing but also tissue biopsies, make up 2.3% of total health care spending in the United States.[9] Health care spending in the U.S. in 2020 amounted to $4.1 trillion, which would imply that the costs of laboratory testing alone exceed $90 billion.[13]

On the other hand, individually most laboratory tests are quite cheap. In the University Medical Center Utrecht in the Netherlands, for instance, the costs of the most commonly ordered tests, such as glucose or hemoglobin, range from €0,52 to €1,12.[14] Of course, there are more expensive tests, but these are generally ordered much less frequently. The prices of these tests do vary across different countries though, with prices generally being higher in the United States than in most European countries.[15-17]

The financial costs of laboratory testing have come under scrutiny due to the enormity of overall health care spending. In Europe, health care spending makes up 5.3 to 11.9% of GDP and in the United States, even 19.7% of GDP is spent on health care.[13,18] This is compounded by the fact that health care spending is increasing over time. In 2019, health care spending has increased by 4.8% in the Netherlands, and 9.7% in the United States. This has already led to intensifying strains on its sustainability over the longer term, forcing many health care systems to investigate ways to lower the costs.

This cost awareness, which has been gaining traction over the last couple of years, has led to several initiatives that aim to reduce health care spending by cutting down on low value health care. A famous example is the international Choosing Wisely campaign, which tries to engage health care professionals and patients in reducing unnecessary tests and procedures.[19] The goal of this campaign is not only to bridle health care spending, but also to improve the quality of patient care. The Choosing Wisely campaign includes several recommendations on laboratory testing. In the Netherlands, the Netherlands Association of Internal Medicine has produced a recommendation that states that laboratory tests should not be ordered more than twice a week in clinically admitted patients.[20] The Netherlands Society for Clinical Chemistry and Laboratory Medicine has produced five Choosing Wisely recommendations, four of which advise against performing certain tests.[21] In addition, a nationwide programme has been set up in the Netherlands, called "Doen of Laten" (To do or not to do) to evaluate and reduce low value health care.[22] This programme includes a project aimed at reducing unnecessary laboratory testing.[23]

Overutilization

These recommendations and projects are based on the implicit idea that a substantial proportion of laboratory testing are redundant, and can therefore be eliminated without having an impact on patient care. When looking at redundant testing, one can distinguish between overdiagnosis and overtesting. Overdiagnosis refers to diagnosing conditions that will not affect patients' health.[24,25] A famous example is the national cancer screening programme in South Korea, which was set up in 1999 and has led to a fifteenfold increase in the incidence of thyroid cancer, while mortality from thyroid cancer has remained the same.[26] This implies that the additional new thyroid cancer diagnoses did not affect clinical outcomes for patients. In overdiagnosis, the testing in itself is appropriate for establishing the diagnosis, even though the diagnosis in question has no value.

In this respect it is different from overtesting, or laboratory test overutilization, which refers to laboratory tests that are performed even though they are not indicated. Overtesting includes both inappropriate initial testing and inappropriate repeat testing. An example of inappropriate initial testing is ordering both prothrombin time and partial thromboplastin time in patients on anticoagulant therapy.[27] Inappropriate repeat testing refers to repeating a laboratory test at a higher frequency than justified, based on the properties of the test and the clinical situation,[28] for instance repeating HbA1c tests within three months of the last test.[29] On average, these overutilized tests comprise over 20% of all laboratory tests.[30]

Many reasons for laboratory test overutilization have been suggested. For one, patients tend to highly value laboratory testing. In fact, their expectations of its capabilities to screen for and diagnose diseases may be unrealistically high.[31,32] At the same time, they tend to underestimate the negative consequences of testing. For instance, in interviews with patients in general practice, the rate of false positive tests was estimated to be one in several thousand tests.[31] In reality, it may be more than half of all positive test results.[33]

Furthermore, laboratory testing can also have additional importance for patients. For instance, patients view tests as a sign that their doctor takes them seriously, and thus additional testing may support the patient-doctor relationship.[34] Clinicians may therefore choose to use laboratory tests, even if strictly there is no good indication for performing the tests. Clinicians are more likely to order laboratory testing when faced with patients whom they perceive as being assertive or wanting reassurance.[35] Additionally, they use laboratory tests to reassure themselves, mostly when experiencing diagnostic uncertaincy. It is therefore not surprising that a risk-taking attitude has been found to be associated with ordering fewer laboratory tests.[36]

However, in the long run, ordering low value diagnostic tests does not reassure patients or reduce their anxiety.[37]

There have been a large number of studies investigating laboratory test overutilization and efforts to reduce it.[38,39] Almost all of these studies have focused on cost savings as the most important outcome measure, whereas patient centered outcomes have been all but neglected.[40] Nevertheless, these patient related outcomes of overtesting are important and warrant further study. First of all, laboratory testing is an invasive procedure, which places a burden on patients. Patients have to travel to the laboratory to get their blood drawn, for which they may need to take time off from work. Venipunctures are unpleasant for patients and in some cases harm them, for instance by leading to anemia.[41] In clinically admitted patients, especially in the intensive care department, where they are subjected to daily venipunctures, the total volume of phlebotomy was independently associated with the need for blood transfusions.[42] Furthermore, laboratory test overutilization leads to additional downstream testing and procedures, which in turn can cause patients harm, as any clinician who has had the experience of going down a garden path of false positive test results can attest.[43] This experience has been dubbed the Ulysses syndrome.[44]

There have been no systematic evaluations of the downstream effects or other patient centered outcomes of laboratory test overutilization. This is a consequence of the methodology of most studies, which use population based data or laboratory databases, where little or no clinical information on those patients involved is available. A more thorough assessment of patient outcomes would require patient level data to assess the actual consequences of inappropriate testing for any given patient. Yet, obtaining this information is difficult as it necessitates measures such as individual chart reviews and questionnaires, which are laborious and costly.

The ample number of different interventions that have been tried to reduce overutilization include changes in the electronical ordering system[45] and feedback to individual health care providers[46] among other things. The efficacy of these measures varies considerably, although very few head to head comparisons have been performed. Furthermore, the majority of studies have only investigated short-term effects of these interventions, which raises questions about the sustainability of the effects.[39] Which intervention or combination of interventions forms the most effective strategy to reduce inappropriate overtesting remains an open question.

Underutilization

Another form of inappropriate laboratory testing is underutilization, in which tests are not ordered even though they are indicated. For instance, even though epididymo-orchitis is most commonly caused by a sexually transmitted pathogen in younger men, a British study in general practice found that only 3% of patients had been tested for chlamydia.[47] Undertesting can lead to delayed or missed diagnoses, which in turn can lead to delays in the commencement of appropriate treatment. Furthermore, if a necessary test is not performed, an incorrect diagnosis could be made, exposing patients to unwarranted treatments.[48] Finally, undertesting can have public health consequences, for instance in the HIV epidemic, given that people who are unaware of their HIV status are more likely to transmit the virus to others.[49]

Evidence for patient harms from undertesting stems mainly from malpractice claims for missed diagnoses.[50,51] The most common cause of missed diagnoses was failing to order an appropriate test. Since malpractice claims typically only showcase the most serious cases, it is likely that undertesting causes more harm than these studies suggest.

Undertesting is more common than overtesting, but it has been studied far less extensively.[30,52] This asymmetrical focus on overtesting may be motivated by the expanding health care spending mentioned earlier and the cost awareness that stems from it. Additionally, it may also reflect the inherent difficulty of determining what constitutes inappopriate testing. The appropriateness of a laboratory test may not be dichotomous, but rather form a continuous spectrum from completely appropriate to completely inappropriate, with many tests falling somewhere in between.[53] Consequently, in areas where no standardized diagnostic protocols exist, it is difficult to determine whether a relevant test has been missed, because that would require determining the clinician's diagnostic reasoning.[40]

Value of the diagnostic process

Overutilization and underutilization of laboratory tests are both undesirable and reducing them will improve the value of the diagnostic process. However, laboratory tests are never performed in a stand-alone fashion, but virtually always in conjunction with other tests. Therefore, thinking about the value of laboratory testing may require a more nuanced approach. Instead of focusing on individual laboratory tests and determining their separate appropriateness, it may be more interesting to look at the value of the diagnostic process in its entirety.

From the perspective of patients, the entire diagnostic process is more relevant than its individual components. Evaluating the entire process also allows for comparisons between the constituent parts, as sometimes the optimum may entail a trade-off between overtesting and undertesting. For example, a study on the diagnosis of anemia in general practice compared an extensive standardized laboratory panel with routine testing. The initial costs of the extended panel were slightly higher, but this strategy led to an earlier diagnosis and was likely to be cost-effective.[54]

Furthermore, the value of the laboratory testing does not only depend on the appropriateness of the tests performed. A diagnostic process can be said to have value if it produces timely and accurate results that are being interpreted correctly, and that contribute to patient management, while resulting in a minimum of incidental findings, and imposing a minimal burden on patients. These factors are closely related. The extent to which diagnostic tests contribute to patient management and lead to incidental findings reflects the issues of overdiagnostics, overutilization and underutilization mentioned earlier, but of course also depends on the timeliness, accuracy and interpretation of the results.

In the case of laboratory testing, the timeliness of the testing is often referred to as the turnaround time. This is the time it takes for the laboratory test results to become available to the ordering physician from being ordered, and it is considered to be an important part of the so-called brain-to-brain loop.[55] Shorter turnaround times mean that the physician has access to the test results sooner, which can shorten the overall diagnostic process. A shorter time to diagnosis has been found to improve patient outcomes in several different settings.[56-58] Also, shorter turnaround times allow for more efficient care for outpatients, as test results can be discussed by the doctor and the patient face-to-face during the visit, without patients having to come to the laboratory on a separate occasion, which is costly and time-consuming.[59]

Laboratory turnaround times can be divided into three separate stages: pre-analytical, analytical, and post-analytical.[60] The pre-analytical stage is the time from the ordering of the laboratory tests to the samples' being ready for analysis. It comprises the logistics of the sample collection, usually through venipuncture, the transportation of the sample to the laboratory, and possible preparation for analysis. The analytical stage is the time from the start of the actual analysis to the confirmation of the test results. The post-analytical stage is the time from confirmation of the results to the moment the test results are evaluated by the ordering physician.[61]

These stages all affect the timeliness and also the accuracy of the diagnostic process, but their relative contributions vary. Because the analytical stage is wholly based in the laboratory, it has been easiest for laboratory specialists to manage. Over

time, increased laboratory automation has improved this aspect of the turnaround time and led to analytical errors becoming relatively rare.[62,63] The relative contribution of the other two stages to both the total turnaround time and the number of diagnostic errors is much higher. For instance, non-analytical delays have been reported to be responsible for up to 96% of turnaround time.[64] There are several quick wins in improving the pre-analytical stage of turnaround time, such as installing a pneumatic tube system for transfer of the samples to the laboratory.[65] However, a large part of the pre-analytical and post-analytical stages have proven to be more difficult to affect, presumably due to their multifactorial nature, in which both the clinical and laboratory medicine are involved.[64,66]

In the pre-analytical stage, there are two main factors that reduce the diagnostic value of laboratory testing. Firstly, and most importantly, physicians sometimes have difficulty correctly ordering the indicated tests.[67,68] For instance, one study in patients with suspected auto-immune conditions found that in more than 60% of cases there had been an error in test selection.[69] This can lead to both over- and underutilization of tests, and subsequently to diagnostic delays and errors. Secondly, incorrect collection or handling of samples can lead to erroneous test results.[70] Most frequently, this is caused by hemolysis, but there are many more potential errors, such as using incorrect tubes or errors in labeling. Not all of these can be readily detected by the laboratory, for example in the case of contamination by infusion fluids.[71,72]

Errors in the post-analytical stage are generally caused by failures to report critical results, and more importantly, errors in the interpretation of the test results by the clinician.[73] As such, it is the stage that is the most outside of the scope of the clinical laboratory, and it has been studied less extensively than the other two.[74] Yet mistaken interpretations of test results can lead to wrong diagnoses with potentially profound consequences, inclusing increased mortality.[75] These worse outcomes reduce patient satisfaction which may in turn prompt patients to seek second opinions.[76]

One study on malpractice claims for diagnostic errors found that in 37% of cases, the error was attributable to an incorrect interpretation of test results.[51] There are many different causes for erroneous interpretations of laboratory tests, such as cognitive errors, a failure in judgment or a lack of knowledge.[77] This is partly due to the intricate nature of diagnostic testing, which involves complex metabolic pathways and extensive interactions with other compounds. For instance, there are approximately 50,000 known drug-laboratory test interactions.[78] This may explain why 28% of urine toxicology screens are interpreted incorrectly by the ordering clinicians.[79] Additionally, laboratory reference ranges do not always reflect the true limits of normal test results, and are often not adjusted for possible differences

between people based on gender, age or ethnicity.[80] Yet, the distribution of normal test results varies significantly between persons of different ethnic backgrounds, with for instance both black and Asian persons having reduced lower limits for most hematological parameters when compared to whites.[81] But even relatively basic principles of diagnostic testing such as sensitivity, specificity and likelihood ratios have proven to be difficult to understand for most physicians.[82] Unsurprisingly, clinicians are quite frequently unsure about their interpretation of laboratory test results.[67]

As health care professionals it is our duty to improve the health and quality of life for patients, while doing as little harm as possible. With regards to laboratory testing, this entails that we should not subject patients to unwarranted testing and excessive blood draws, that patients should not have to wait unreasonably long for the test results to become available, and that the risk of diagnostic error should be minimized.

Outline of the thesis

This thesis addresses several aspects of the value of laboratory testing, such as the timeliness and diagnostic yield of the results and the appropriateness of the ordered tests, in three different clinical settings, which are covered in the three sections of this thesis: the first section, containing chapters 2 and 3, is set in the emergency department, the second section contains chapters 4 to 6 and pertains to the outpatient clinic, and the third section, which consists of chapter 7, concerns the inpatient ward.

In **chapter 2**, two measures of the timeliness of laboratory test results are investigated: the time from arrival of the patient in the emergency department to the ordering of laboratory testing (time to testing), and the time from the laboratory test ordering to the results' becoming available (turnaround time). We performed a retrospective cohort study on the effect of these two measures on the emergency department length of stay, a measure of the quality of care and of emergency department crowding.

Chapter 3 explores the potential of routinely measured hematological parameters in a specific setting: the diagnosis of immune-related adverse events in patients on immune checkpoint inhibitors, which can be challenging. We used machine learning to assess the added diagnostic value of these hematological parameters in addition to standard diagnostic practices.

Making a timely diagnosis is important for newly referred patients to the outpatient clinic. We devised a strategy in which patients had a standardized set of laboratory tests done directly prior to the visit to the internal medicine outpatient

clinic. To assess whether this strategy would lead to more timely diagnoses, we compared it with routine care, in which laboratory tests were ordered after the visit. The results of this study are reported in **chapter 4**.

In cases where the diagnostic process fails or is questioned, patients may seek a second opinion from another health care professional. These second opinions are associated with additional costs and test overutilization, as tests may be unnecessarily repeated. In **Chapter 5** the outcomes and costs of second opinions in the general internal medicine outpatient clinic of our hospital are reported.

Repetitions of laboratory tests also occur when patients are referred to the outpatient clinic. This is a potential source of laboratory test overutilization. **Chapter 6** assesses the extent of laboratory test repetition and whether this yields any new information in referrals by general practitioners to the internal medicine outpatient clinic.

In **Chapter 7** the results of a study on laboratory test overutilization in clinically admitted patients are reported. This study was performed on the general internal medicine ward of our hospital. Both the extent of overutilization and its causes are evaluated, by means of a survey and focus group interviews.

Chapter 8 provides a summary and a discussion of the results, resulting in a conclusion and views on possible further research.

References

1. Mani N. The historical background of clinical chemistry. *J Clin Chem Clin Biochem*. 1981;19(6):311-22.
2. Büttner J. The origin of clinical laboratories. *Eur J Clin Chem Clin Biochem*. 1992;30(10):585-93.
3. Berger D. A brief history of medical diagnosis and the birth of the clinical laboratory. Part 2--Laboratory science and professional certification in the 20th century. *MLO Med Lab Obs*. 1999;31(8):32-4, 6, 8.
4. Forsman RW. Why is the laboratory an afterthought for managed care organizations? *Clin Chem*. 1996;42(5):813-6.
5. Hallworth MJ. The '70% claim': what is the evidence base? *Ann Clin Biochem*. 2011;48(Pt 6):487-8.
6. Ngo A, Gandhi P, Miller WG. Frequency that Laboratory Tests Influence Medical Decisions. *J Appl Lab Med*. 2017;1(4):410-4.
7. Sikaris KA. Enhancing the Clinical Value of Medical Laboratory Testing. *Clin Biochem Rev*. 2017;38(3):107-14.
8. Piessens V, Delvaux N, Heytens S, Aertgeerts B, De Sutter A. 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(4):e059261.
9. Rohr U-P, Binder C, Dieterle T, Giusti F, Messina CGM, Toerien E, et al. The Value of In Vitro Diagnostic Testing in Medical Practice: A Status Report. *PLoS One*. 2016;11(3):e0149856.
10. Wians FH. Clinical Laboratory Tests: Which, Why, and What Do The Results Mean? *Laboratory Medicine*. 2009;40(2):105-13.
11. Regan M, Forsman R. The impact of the laboratory on disease management. *Dis Manag*. 2006;9(2):122-30.
12. Song Z, Lillehaugen T, Wallace J. Out-of-Network Laboratory Test Spending, Utilization, and Prices in the US. *JAMA*. 2021;325(16):1674-6.
13. CMS. National Health Expenditure Data Historical 2021 [Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical].
14. UMC Utrecht Laboratoriumbepalingen LKCH / LTI [Available from: http://lkch.nl/bepalingen].
15. Feldman LS, Shihab HM, Thiemann D, Yeh HC, Ardolino M, Mandell S, et al. Impact of providing fee data on laboratory test ordering: a controlled clinical trial. *JAMA Intern Med*. 2013;173(10):903-8.
16. Robinson JC, Whaley C, Brown TT. Association of Reference Pricing for Diagnostic Laboratory Testing With Changes in Patient Choices, Prices, and Total Spending for Diagnostic Tests. *JAMA Intern Med*. 2016;176(9):1353-9.
17. Sanger-Katz M. They Want It to Be Secret: How a Common Blood Test Can Cost $11 or Almost $1,000. *The New York Times*. 2019 April 30, 2019.
18. CBS. Dutch health expenditure 10th highest in Europe 2020 [Available from: https://www.cbs.nl/en-gb/news/2020/47/dutch-health-expenditure-10th-highest-in-europe#:~:text=The%20Netherlands%20spent%2080.9%20billion,Denmark%20and%20the%20United%20Kingdom].
19. Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf*. 2015;24(2):167-74.

20. Nederlandse Internisten Vereniging. Verstandige keuzes bij interne geneeskunde [Available from: https://internisten.nl/werken-als-internist/beroepsinformatie-divers/publicaties/verstandige-keuzes-bij-interne-geneeskunde].
21. Laboratoriumgeneeskunde NVKC. Verstandig Kiezen [Available from: https://www.nvkc.nl/patient/verstandig-kiezen].
22. To Do Or Not To Do [Available from: https://todoornottodo.nl/#sectie-2].
23. Bindraban RS, van Beneden M, Kramer MHH, van Solinge WW, van de Ven PM, Naaktgeboren CA, et al. Association of a Multifaceted Intervention With Ordering of Unnecessary Laboratory Tests Among Caregivers in Internal Medicine Departments. *JAMA Netw Open*. 2019;2(7):e197577-e.
24. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst*. 2010;102(9):605-13.
25. Brodersen J, Schwartz LM, Heneghan C, O'Sullivan JW, Aronson JK, Woloshin S. Overdiagnosis: what it is and what it isn't. *BMJ Evid Based Med*. 2018;23(1):1-3.
26. Ahn HS, Kim HJ, Welch HG. Korea's Thyroid-Cancer "Epidemic" — Screening and Overdiagnosis. *N Engl J Med*. 2014;371(19):1765-7.
27. Pilsczek FH, Rifkin WD, Walerstein S. Overuse of prothrombin and partial thromboplastin coagulation tests in medical inpatients. *Heart & Lung*. 2005;34(6):402-5.
28. Lang T, Croal B. National minimum retesting intervals in pathology. *The Royal College of Pathologists*; 2021.
29. Laxmisan A, Vaughan-Sarrazin M, Cram P. Repeated Hemoglobin A1C Ordering in the VA Health System. *Am J Med*. 2011;124(4):342-9.
30. 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(11):e78962.
31. van Bokhoven MA, Pleunis-van Empel MCH, Koch H, Grol RPTM, Dinant G-J, van der Weijden T. Why do patients want to have their blood tested? A qualitative study of patient expectations in general practice. *BMC Fam Pract*. 2006;7(1):75.
32. Fallon KE. Blood tests in tired elite athletes: expectations of athletes, coaches and sport science/sports medicine staff. *Br J Sports Med*. 2007;41(1):41-4.
33. 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(3):202-3.
34. Kravitz RL, Callahan EJ. Patients' perceptions of omitted examinations and tests. *J Gen Intern Med*. 2000;15(1):38-45.
35. van der Weijden T, van Bokhoven MA, Dinant G-J, van Hasselt CM, Grol RPTM. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract*. 2002;52(485):974-80.
36. Zaat JO, van Eijk JT. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care*. 1992;30(9):846-54.
37. Rolfe A, Burton C. Reassurance after diagnostic testing with a low pretest probability of serious disease: systematic review and meta-analysis. *JAMA Intern Med*. 2013;173(6):407-16.
38. Müskens JLJM, Kool RB, van Dulmen SA, Westert GP. Overuse of diagnostic testing in healthcare: a systematic review. *BMJ Qual Saf*. 2022;31(1):54-63.
39. Bindraban RS, Ten Berg MJ, Naaktgeboren CA, Kramer MHH, Van Solinge WW, Nanayakkara PWB. Reducing Test Utilization in Hospital Settings: A Narrative Review. *Ann Lab Med*. 2018;38(5):402-12.

40. Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. *BMJ Qual Saf*. 2013;22 Suppl 2(Suppl 2):ii6-ii10.

41. Shander A, Corwin HL. A Narrative Review on Hospital-Acquired Anemia: Keeping Blood where It Belongs. *Transfus Med Rev*. 2020;34(3):195-9.

42. Jackson Chornenki NL, James TE, Barty R, Liu Y, Rochwerg B, Heddle NM, et al. Blood loss from laboratory testing, anemia, and red blood cell transfusion in the intensive care unit: a retrospective study. *Transfusion*. 2020;60(2):256-61.

43. Korenstein D, Chimonas S, Barrow B, Keyhani S, Troy A, Lipitz-Snyderman A. Development of a Conceptual Map of Negative Consequences for Patients of Overuse of Medical Tests and Treatments. *JAMA Intern Med*. 2018;178(10):1401-7.

44. Rang M. The Ulysses syndrome. *Can Med Assoc J*. 1972;106(2):122-3.

45. Lapić I, Rogić D, Fuček M, Galović R. Effectiveness of minimum retesting intervals in managing repetitive laboratory testing: experience from a Croatian university hospital. *Biochem Med (Zagreb)*. 2019;29(3):030705.

46. Minerowicz C, Abel N, Hunter K, Behling KC, Cerceo E, Bierl C. Impact of weekly feedback on test ordering patterns. *Am J Manag Care*. 2015;21(11):763-8.

47. Nicholson A, Rait G, Murray-Thomas T, Hughes G, Mercer CH, Cassell J. Management of epididymo-orchitis in primary care: results from a large UK primary care database. *Br J Gen Pract*. 2010;60(579):e407-22.

48. Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences Engineering and Medicine. Improving Diagnosis in Health Care. Balogh EP, Miller BT, Ball JR, editors. Washington, DC: *The National Academies Press*; 2015. 472 p.

49. Arya M, Kumar D, Patel S, Street RL, Jr., Giordano TP, Viswanath K. Mitigating HIV health disparities: the promise of mobile health for a patient-initiated solution. *Am J Public Health*. 2014;104(12):2251-5.

50. Aaronson EL, Quinn GR, Wong CI, Murray AM, Petty CR, Einbinder J, et al. Missed diagnosis of cancer in primary care: Insights from malpractice claims data. *J Healthc Risk Manag*. 2019;39(2):19-29.

51. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med*. 2006;145(7):488-96.

52. Sarkar MK, Botz CM, Laposata M. An assessment of overutilization and underutilization of laboratory tests by expert physicians in the evaluation of patients for bleeding and thrombotic disorders in clinical context and in real time. *Diagnosis (Berl)*. 2017;4(1):21-6.

53. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390(10090):156-68.

54. Kip MM, Schop A, Stouten K, Dekker S, Dinant GJ, Koffijberg H, et al. Assessing the cost-effectiveness of a routine versus an extensive laboratory work-up in the diagnosis of anaemia in Dutch general practice. *Ann Clin Biochem*. 2018;55(6):630-8.

55. Plebani M, Laposata M, Lundberg GD. The Brain-to-Brain Loop Concept for Laboratory Testing 40 Years After Its Introduction. *Am J Clin Pathol*. 2011;136(6):829-33.

56. Howanitz JH, Howanitz PJ. Laboratory results. Timeliness as a quality attribute and strategy. *Am J Clin Pathol*. 2001;116(3):311-5.

57. Akers IE, Weber R, Sax H, Böni J, Trkola A, Kuster SP. Influence of time to diagnosis of severe influenza on antibiotic use, length of stay, isolation precautions, and mortality: a retrospective study. *Influenza Other Respir Viruses*. 2017;11(4):337-44.
58. Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer*. 2015;112(1):S92-S107.
59. Schnell O, Crocker JB, Weng J. Impact of HbA1c Testing at Point of Care on Diabetes Management. *J Diabetes Sci Technol*. 2017;11(3):611-7.
60. Chung HJ, Lee W, Chun S, Park HI, Min WK. Analysis of turnaround time by subdividing three phases for outpatient chemistry specimens. *Ann Clin Lab Sci*. 2009;39(2):144-9.
61. Bhatt RD, Shrestha C, Risal P. Factors Affecting Turnaround Time in the Clinical Laboratory of the Kathmandu University Hospital, Nepal. *EJIFCC*. 2019;30(1):14-24.
62. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. *Clin Chem*. 2007;53(7):1338-42.
63. Armbruster DA, Overcash DR, Reyes J. Clinical Chemistry Laboratory Automation in the 21st Century - Amat Victoria curam (Victory loves careful preparation). *Clin Biochem Rev*. 2014;35(3):143-53.
64. Hawkins RC. Laboratory turnaround time. *Clin Biochem Rev*. 2007;28(4):179-94.
65. Guss DA, Chan TC, Killeen JP. The impact of a pneumatic tube and computerized physician order management on laboratory turnaround time. *Ann Emerg Med*. 2008;51(2):181-5.
66. Stotler BA, Kratz A. Determination of Turnaround Time in the Clinical Laboratory: "Accessioning-to-Result" Time Does Not Always Accurately Reflect Laboratory Performance. *Am J Clin Pathol*. 2012;138(5):724-9.
67. Hickner J, Thompson PJ, Wilkinson T, Epner P, Sheehan M, Pollock AM, et al. Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results. *J Am Board Fam Med*. 2014;27(2):268.
68. Laposata M, Dighe A. "Pre-pre" and "post-post" analytical error: high-incidence patient safety hazards involving the clinical laboratory. *Clin Chem Lab Med*. 2007;45(6):712-9.
69. Rajendran R, Salazar JH, Seymour RL, Laposata M, Zahner CJ. Overutilization and underutilization of autoantibody tests in patients with suspected autoimmune disorders. *Diagnosis (Berl)*. 2021;8(4):497-503.
70. Nilsson K, Grankvist K, Juthberg C, Brulin C, Söderberg J. Deviations from venous blood specimen collection guideline adherence among senior nursing students. *Nurse Educ Today*. 2014;34(2):237-42.
71. Lippi G, von Meyer A, Cadamuro J, Simundic AM. Blood sample quality. *Diagnosis (Berl)*. 2019;6(1):25-31.
72. Allen TL, Wolfson AB. Spurious values of serum electrolytes due to admixture of intravenous infusion fluids. *J Emerg Med*. 2003;24(3):309-13.
73. Plebani M. The detection and prevention of errors in laboratory medicine. *Ann Clin Biochem*. 2010;47(2):101-10.
74. Asmelash D, Worede A, Teshome M. Extra-Analytical Clinical Laboratory Errors in Africa: A Systematic Review and Meta-Analysis. *EJIFCC*. 2020;31(3):208-24.
75. Hautz WE, Kämmer JE, Hautz SC, Sauter TC, Zwaan L, Exadaktylos AK, et al. Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room. *Scand J Trauma Resusc Emerg Med*. 2019;27(1):54.

76. Payne VL, Singh H, Meyer AND, Levy L, Harrison D, Graber ML. Patient-Initiated Second Opinions: Systematic Review of Characteristics and Impact on Diagnosis, Treatment, and Satisfaction. *Mayo Clin Proc*. 2014;89(5):687-96.

77. Howard J. Cognitive Errors and Diagnostic Mistakes: A Case-Based Guide to Critical Thinking in Medicine 2019.

78. Balveren JAv, Venne WPHGV-vd, Erdem-Eraslan L, Graaf AJd, Loot AE, Musson REA, et al. Diagnostic error as a result of drug-laboratory test interactions. *Diagnosis*. 2019;6(1):69-71.

79. Chua I, Petrides AK, Schiff GD, Ransohoff JR, Kantartjis M, Streid J, et al. Provider Misinterpretation, Documentation, and Follow-Up of Definitive Urine Drug Testing Results. *J Gen Intern Med*. 2020;35(1):283-90.

80. Smellie WS. When is "abnormal" abnormal? Dealing with the slightly out of range laboratory result. *J Clin Pathol*. 2006;59(10):1005-7.

81. Lim E, Miyamura J, Chen JJ. Racial/Ethnic-Specific Reference Intervals for Common Laboratory Tests: A Comparison among Asians, Blacks, Hispanics, and White. *Hawaii J Med Public Health*. 2015;74(9):302-10.

82. Whiting PF, Davenport C, Jameson C, Burke M, Sterne JAC, Hyde C, et al. How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open*. 2015;5(7):e008155.


LOT 2227155

Chapter 2

Shorter laboratory turnaround time lowers emergency department length of stay: a retrospective cohort study

B.E.L. Vrijsen, S. Haitjema, J. Westerink, C.A.R. Hulsbergen-Veelken, W.W. van Solinge, M.J. ten Berg.

BMC Emergency Medicine 2022 Dec 21; 22(1): 207

Abstract

Background

A longer emergency department length of stay (EDLOS) is associated with poor outcomes. Shortening EDLOS is difficult, due to its multifactorial nature. A potential way to improve EDLOS is through shorter turnaround times for diagnostic testing. This study aimed to investigate whether a shorter laboratory turnaround time (TAT) and time to testing (TTT) were associated with a shorter EDLOS.

Methods

A retrospective cohort study was performed, including all visits to the emergency department (ED) of an academic teaching hospital from 2017 to 2020 during which a standardized panel of laboratory tests had been ordered. TTT was calculated as the time from arrival in the ED to the ordering of laboratory testing. TAT was calculated as the time from test ordering to the reporting of the results, and was divided into a clinical and a laboratory stage. The outcome was EDLOS in minutes. The effect of TTT and TAT on EDLOS was estimated through a linear regression model.

Results

In total, 23,718 ED visits were included in the analysis. Median EDLOS was 199.0 minutes (interquartile range [IQR] 146.0 – 268.0). Median TTT was 7.0 minutes (IQR 2.0 – 12.0) and median TAT was 51.1 minutes (IQR 41.1 – 65.0). Both TTT and TAT were positively associated with EDLOS. The laboratory stage comprised a median of 69% (IQR 59 – 78%) of total TAT.

Conclusion

Longer TTT and TAT are independently associated with longer EDLOS. As the laboratory stage predominantly determines TAT, it provides a promising target for interventions to reduce EDLOS and EDLOS crowding.

Background

Emergency department length of stay (EDLOS) is an important benchmark of the quality of care in the emergency department (ED).[1] EDLOS is affected by many factors, both patient-related and organizational. Generally, more complex and acute patients, who generally require more extensive diagnostics, have longer EDLOS.[2,3] Organizational factors that increase EDLOS include a shortage of beds leading to hospital transfer, evaluation by medical students or trainees, and sequential specialist consultations.[4,5] Extended lengths of stay lead to crowding in the ED, which in turn is associated with worse outcomes, including death.[6-8] In a study on patients with severe pneumonias, performed during the COVID-19 pandemic, EDLOS was an independent risk factor for in-hospital mortality (odds ratio 1.84 for EDLOS in hours).[9] Given the increasing age and complexity of patients presenting to the ED, long EDLOS is a growing concern.[10] Therefore it is important to look for interventions that shorten EDLOS.

A potential determinant of the EDLOS that has been relatively understudied is the laboratory testing turnaround time (TAT).[11,12] A shorter TAT results in the clinician's, and thus the patient's, having earlier access to the test results, which play an important role in many medical decisions.[13]

One solution to shorten TAT is point-of-care testing, in which laboratory tests are performed at the patient's bedside.[14] In the ED, point-of-care testing has been shown to lead to shorter length of stay in some studies,[15-18] but not in all.[19-21] Studies also show conflicting results on the effect of point-of-care testing on hospital admission rates in the same patient groups.[18,19] These conflicting results may be caused by point-of-care testing's only being available for a limited number of tests, such as cardiac markers, blood gases and certain electrolytes. Often, additional laboratory tests at the central laboratory are required, which might prolong the EDLOS. In one trial, 94.7% of patients who had point-of-care testing done still required additional testing from the central laboratory.[22]

Therefore, the TAT of regular laboratory tests at the central laboratory may prove a more promising target for shortening the EDLOS. Many studies on TAT have focused on the laboratory's perspective, but from the perspectives of the patient and the clinician, what matters is not only the time the laboratory needs to generate the results, but also the time required to send the samples to the laboratory and even the time it takes the clinician to decide to order laboratory testing in the first place, which we dubbed the time to testing (TTT). To our knowledge the latter has not been studied in relation to EDLOS.

Consequently, this study was set up to investigate whether a shorter TAT and TTT were associated with a shorter EDLOS.

Methods

Patient selection

This is a single center retrospective cohort study performed in the University Medical Center Utrecht (UMC Utrecht), an academic teaching hospital in the Netherlands, with around 20,000 ED visits per year. The study period ran from January 2017 until January 2020. The study population included all ED visits of adult patients for whom a standard panel of laboratory tests was ordered (pre-defined in our electronic health record as the "internal medicine lab") through the order management module of the electronic health record. This standard panel comprises the following fourteen tests: hemoglobin, thrombocyte count, leukocyte count, sodium, potassium, urea, creatinine, alkaline phosphatase, gamma glutamyltransferase, alanine transaminase, aspartate transaminase, lactate dehydrogenase, glucose, and C-reactive protein. Other tests can be added to this panel, for instance cardiac markers in patients with suspected myocardial infarction. Only visits in which these standard tests were ordered were included to prevent confounding by indication.

If the standard panel was ordered more than once during an ED visit, the first order was used for the analysis. Visits in which not all of the ordered tests of the standard panel were actually performed were excluded from the analysis. If patients visited the ED more than once in the study period all eligible visits were included in the study and were seen as individual events.

In the ED, the nurses are responsible for the venipuncture and the transportation of the samples to the laboratory by pneumatic tube.

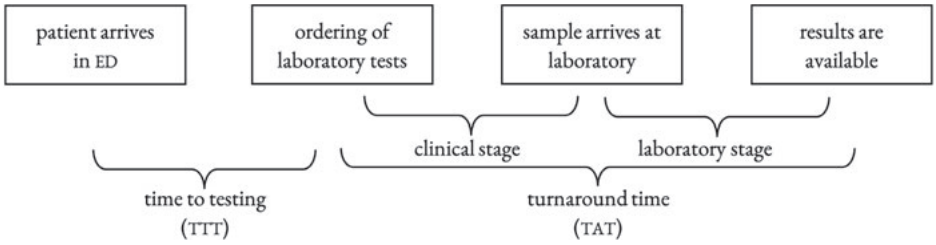
Measures and outcomes

The primary outcome was the EDLOS, defined as the time from arrival in the ED to either discharge, admittance, transfer elsewhere, or death, in minutes.

The TTT was calculated as the time in minutes between the patient's arriving in the ED and the ordering of laboratory tests. The TAT was calculated as the time in minutes between the time the laboratory tests were ordered at the ED and the time the last result of all tests in the standard panel was reported in the electronic health record.

Furthermore, the TAT was divided into a clinical stage (the time from the ordering of the laboratory tests to the arrival at the laboratory) and a laboratory stage (the time from the arrival of the sample at the laboratory to the results becoming available in the electronic health record). The different times are represented schematically in Figure 1.

FIGURE 1: TIME TO TESTING (TTT) AND TURNAROUND TIME (TAT)



All times are reported in minutes.
ED: emergency department.

Additionally, data were collected on the date of the visit, the patient's age and sex, the time of arrival at the ED (recorded as being between 8:00 and 16:59 (day time shift), between 17:00 and 00:59 (evening shift) or between 01:00 and 07:59 (night shift), the Emergency Severity Index level,[23] which is used as the triage system in the ED, the number of out-of-range laboratory test results in the panel, the daily total number of ED visits, the destination after the ED visit (discharge home, admittance to hospital, transfer to another facility, or deceased) and the specialty treating the patient at the ED (grouped into medicine, surgery, and other specialties including psychiatry, rehabilitation medicine, pain management, and radiotherapy).

The time of day and the date were used to account for potential confounding due to variations in the laboratory workload and ED crowding both over time and at different times of the day. The Emergency Severity Index level, the specialty, the destination after the ED visit and the number of out-of-range tests were included as proxies for the acuity level, which was identified as a confounder in previous studies.[11,12]

Data acquisition

All determinants were collected by two of the authors through the Utrecht Patient Oriented Database (UPOD) and checked for quality by a third. UPOD is an infrastructure of relational databases that automatically retrieve data from the hospital information system on patient characteristics, hospital discharge diagnoses, medical procedures, medication orders and laboratory tests for all patients treated at the UMC Utrecht since 2004. UPOD data acquisition and management is in accordance with current regulations concerning privacy and ethics. The structure and content of UPOD have been described in more detail elsewhere.[24] All measures regarding ED arrival and discharge and the logistics of the laboratory tests are automatically time stamped in the hospital information system.

Statistical analysis

Baseline characteristics were reported. Means and standard deviations (SD) were calculated for normally distributed data, and medians and interquartile ranges (IQR) for non-normal data. Visits with data suspected to be incorrect (e.g. incorrect time stamps due to the change from daylight savings time) were excluded. In order to handle erratic outliers caused by administrative errors, the data set was trimmed to exclude the top and bottom 0.5% values for the EDLOS and the clinical and laboratory stages of the TAT.

The effect of the total TAT and TTT on the EDLOS was estimated by using linear regression, both in a univariate model as in a multivariate model, controlling for the other variables mentioned above.

The regression model gives estimates for the increase in EDLOS in minutes per one minute increases in TAT and TTT, as well as for changes in the other variables. A p-value of $< .05$ was considered significant. A sensitivity analysis was performed by performing the linear regression on the full dataset including the trimmed values. Furthermore, we calculated how much of the total TAT consisted of the clinical and laboratory stages. As a post-hoc analysis, we compared the TAT for the different times of day, which was tested using the Kruskal-Wallis test. All analyses were performed in R version 4.0.3.

Results

There were 24,727 eligible visits to the ED during the study period, out of a total of 39,992 visits during which any laboratory tests were ordered. In 290 cases, not all laboratory tests that had been ordered were actually performed, and these were excluded. A further 6 patients were excluded because one of the time points in their ED visit fell in the transition period from summer time to winter time, which led to ambiguous turnaround times.

Trimming the data set to deal with outliers led to the exclusion of 713 cases (2.9%), which left 23,718 visits for the analysis.

The mean age of the included patients was 58.4 years (SD 17.8), and 47% of patients were female. The majority of visits (87%) were treated by medicine specialties. About half of the visits (53%) resulted in admission to the hospital. The median TTT was 7.0 minutes (IQR 2.0 – 12.0) and the median total TAT was 51.1 minutes (IQR 41.1 – 65.0). The median EDLOS was 199 minutes (IQR 146.0 – 268.0).

All the baseline characteristics of the patients in the analysis are provided in Table 1.

Table 1: BASELINE CHARACTERISTICS

total number of ED visits	23,718
age (in years; mean + SD)	58.4 ± 17.8
female sex	11,041 (47%)
specialty	
- medicine	20,551 (87%)
- surgery	3,098 (13%)
- other	69 (< 1%)
time of day	
- 8:00 – 16:59	14,279 (60%)
- 17:00 – 0:59	7,510 (32%)
- 1:00 – 7:59	1,929 (8%)
number of daily patients (mean + SD)	55.9 ± 8.7
number of abnormal tests (median + IQR)	5.0 (3.0 – 7.0)
destination	
- home	10,713 (45%)
- admission	12,524 (53%)
- transfer to another facility	410 (2%)
- death	71 (< 1%)
ESI triage level	
- level 1	644 (3%)
- level 2	6,401 (27%)
- level 3	15,189 (65%)
- level 4	1,434 (6%)
- level 5	23 (< 1%)
- not scored	28 (< 1%)
TAT (in minutes; median + IQR)	51.1 (41.1 – 65.0)
TTT (in minutes; median + IQR)	7.0 (2.0 – 12.0)
EDLOS (in minutes; median + IQR)	199.0 (146.0 – 268.0)

All variables are absolute numbers and percentages (%) except where otherwise specified.
ED: emergency department; EDLOS: emergency department length of stay; ESI: emergency severity index[23]; IQR: interquartile range; SD: standard deviation; TAT: turnaround time; TTT: time to testing

After adjustment for the other co-variables in the model, a one minute increase in the TTT led to an increase in the EDLOS of 0.56 minutes (95% CI 0.50 – 0.62) while a one minute increase in TAT led to an increase in the EDLOS of 0.32 minutes (95% CI 0.28 – 0.37). Furthermore, age, female sex, the number of out-of-range laboratory tests, ED visits during office hours, the daily total number of patients in the ED, all but the highest Emergency Severity Index level, and being admitted to hospital or transferred to another facility were also associated with a longer EDLOS (Table 2).

Table 2: determinants of EDLOS

	Change in EDLOS in minutes[a] (univariate) (95% CI)	Change in EDLOS in minutes[a] (multivariate) (95% CI)	p-value (multivariate)
TAT (in minutes)	0.50 (0.45 – 0.54)	0.32 (0.28 – 0.37)	< .001
TTT (in minutes)	0.50 (0.43 – 0.56)	0.56 (0.50 – 0.62)	< .001
time from the start of the study period (in days)	0.003 (-0.001 – 0.007)	0.002 (-0.002 – 0.006)	.26
age (in years)	0.29 (0.22 – 0.36)	0.12 (0.05 – 0.19)	< .001
female sex	4.1 (1.7 – 6.5)	4.9 (2.6 – 7.2)	< .001
specialty			
- medical	–[b]	–[b]	–
- surgical	16.7 (13.1 – 20.2)	14.0 (10.6 – 17.4)	< .001
- other	-36.4 (-58.7 – -14.0)	-27.6 (-49.1 – -6.1)	.01
time of day			
- 8:00 – 16:59	–[b]	–[b]	–
- 17:00 – 0:59	-24.9 (-27.5 – -22.3)	-20.7 (-23.3 – -18.2)	< .001
- 1:00 – 7:59	-34.6 (-39.1 – -30.2)	-25.7 (-30.1 – -21.4)	< .001
number of daily patients	0.28 (0.15 – 0.42)	0.20 (0.07 – 0.34)	.003
number of abnormal tests	5.2 (4.87 – 5.6)	4.8 (4.4 – 5.3)	< .001
destination			
- home	–[b]	–[b]	–
- admission	11.2 (8.8 – 13.6)	6.2 (3.7 – 8.7)	< .001
- transfer to another facility	97.2 (88.0 – 106.5)	90.0 (81.1 – 98.9)	< .001
- death	13.0 (-8.8 – 34.9)	53.2 (31.5 – 74.9)	< .001
ESI triage level			
- level 1	–[b]	–[b]	–
- level 2	49.9 (42.3 – 57.5)	65.2 (57.7 – 72.7)	< .001
- level 3	66.2 (58.8 – 73.6)	75.4 (68.0 – 82.8)	< .001
- level 4	54.4 (45.6 – 63.1)	63.8 (55.1 – 72.5)	< .001
- level 5	24.3 (-14.7 – 63.3)	22.1 (-15.4 – 59.6)	.25
- not scored	8.6 (-27.5 – 44.7)	12.0 (-22.6 – 46.6)	.50

[a] The change in EDLOS is the change in minutes per one unit increase in the explanatory variable for continuous explanatory variables (minutes for TAT and TTT, days for the time from the start of the study period, years for age, and number for daily patients and abnormal tests. For categorical explanatory variables, it is the change in minutes compared to the reference category.

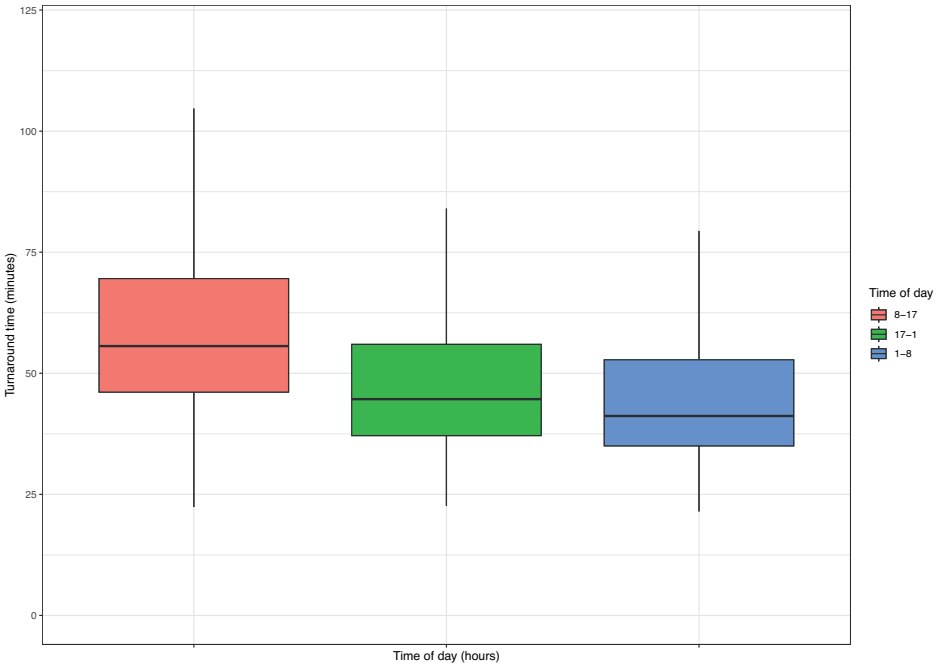
[b] Reference category

EDLOS: emergency department length of stay; ESI: Emergency Severity Index[23]; CI: confidence interval; TAT: turnaround time; TTT: time to testing

There were no relevant interactions between the variables in the model, so these were left out of the final model. The adjusted R^2 of the model was 10%. Performing the linear regression on the untrimmed data set did not change the results.

The total TAT was mostly driven by the laboratory stage, which comprised a median of 69% (IQR 59 – 78%) of the total laboratory turnaround time. TAT was generally shorter in the evenings and nights (Figure 2). The median TAT was 55.6 minutes (IQR 46.1 – 69.5) during office hours, 44.7 minutes (IQR 37.1 – 60.0) in the evenings and 41.2 minutes (IQR 35.0 – 52.8) in the nights ($p < .001$).

FIGURE 2: LABORATORY TURNAROUND TIMES PER TIME OF DAY



Discussion

This study found that a one minute increase in TTT leads to a 0.56 minute increase in EDLOS and that a one minute increase in TAT leads to a 0.32 minute increase in EDLOS.

This confirms the results from previous studies.[11,12] However, the study by Kaushik et al. only investigated the relationship between the TAT and the EDLOS for patients who were being discharged home. We found that the relationship also holds in patients who are admitted to hospital. These patients are probably most likely to benefit from shorter turnaround times, given that they generally have higher acuity levels and their EDLOS was typically longest (data not shown). Furthermore, the TATs in this study were shorter than in other studies, implying that improving the TAT is still worthwhile even if these times are already relatively short.

The low R^2 (10%) of the fitted model suggests that there are other important factors determining the EDLOS. This is not surprising. An important example of such determinants is the patient's acuity level.[3] Our study shows that the Emergency Severity Index level, the rate of out-of-range laboratory tests and the destination of the patient after the ED visit, which can be seen as proxies of the acuity level, are indeed positively associated with the EDLOS. Besides acuity level, other studies have found that diagnostic imaging is associated with a longer EDLOS.[2,25] Even though acuity and the decision to perform diagnostic imaging may have a greater effect on the EDLOS, these factors are not modifiable.

Laboratory turnaround times on the other hand can be modified. For instance, we found that TAT were on average 10.9 and 14.4 minutes shorter in respectively evenings and nights as compared to during the day, most likely due to a lower workload in the clinical laboratory outside of office hours. If this reduction could be realized during office hours, it would lead to an increase in ED capacity of 2.3%. This is in line with the study by Kaushik et al., which found that a 15 minute reduction in TAT would lead to a 3% increase in ED capacity.

This implies that significant reductions in EDLOS can be achieved through improved TATs. In this study the TAT was further divided into a clinical and a laboratory stage. The laboratory stage is the period from the arrival of the sample at the laboratory until the reporting of the results, thus comprising all the internal processes of the laboratory. It is therefore the stage that is most readily modifiable from the laboratory's point of view.

Over the last decade, many medical laboratories, including ours, have made substantial progress in improving the TAT. Examples include installing robot track systems for transport of samples to centrifuges and analyzers, using shorter centrifugation times and analyzers with short analytical procedure times, and

establishing auto-verification procedures for checking out-of-range results. Such improvements have been shown to reduce the TAT.[26-30]

The clinical stage on the other hand has been studied less extensively. This may be due to its multifactorial nature, depending on factors such as the time it takes the ED staff to draw the blood and the logistics of the transport of the sample from the ED to the laboratory, which are influenced by factors as having the tubes labeled with barcodes at the point of care and installing pneumatic tube systems for transport of patient samples.

The TTT is also a potential target for interventions, for instance by pre-ordering laboratory tests and performing the venipuncture immediately after the patient arrives in the emergency department. This is currently standard practice in our hospital, which may be why the median TTT was only 7 minutes in this study.

This study has several limitations. Firstly, this is a single center study in an academic hospital, with a different case mix from general hospitals. Still, this is unlikely to affect the relationship between the TAT and the EDLOS. Secondly, as in all observational studies, there is a risk of unmeasured confounding. For instance, there was no detailed information on the patients' acuity level, other than the Emergency Severity Index level and the other rather crude proxies mentioned above.

A strength of this study is that it includes all patients coming to the ED who had the aforementioned laboratory tests done. Focusing on this group reduces the risk of confounding that would have been introduced by including all patients with any laboratory testing done, as it is likely that more complicated patients both will have had more laboratory testing done and will have had a longer EDLOS. The selected panel of laboratory tests comprised 62% of all visits during which any laboratory testing was ordered. Another strength is that we were able to divide the TAT in a clinical and a laboratory stage, which can help to determine where potential targets for improvement lie.

Conclusion

In conclusion, longer time to testing and laboratory turnaround time are associated with a longer emergency department length of stay. However, a causal effect is difficult to determine in this observational setting. Interventions that improve laboratory turnaround times may lead to shorter emergency department lengths of stay, for instance through increased laboratory automation. Prospective studies are needed to investigate whether such interventions affect the adverse patient outcomes associated with ED crowding.

References

1. Andersson J, Nordgren L, Cheng I, Nilsson U, Kurland L. Long emergency department length of stay: A concept analysis. *Int Emerg Nurs*. 2020;53:100930.
2. Casalino E, Wargon M, Peroziello A, Choquet C, Leroy C, Beaune S, et al. Predictive factors for longer length of stay in an emergency department: a prospective multicentre study evaluating the impact of age, patient's clinical acuity and complexity, and care pathways. *Emerg Med J*. 2014;31(5):361-8.
3. Kreindler SA, Cui Y, Metge CJ, Raynard M. Patient characteristics associated with longer emergency department stay: a rapid review. *Emerg Med J*. 2016;33(3):194-9.
4. Driesen BEJM, van Riet BHG, Verkerk L, Bonjer HJ, Merten H, Nanayakkara PWB. Long length of stay at the emergency department is mostly caused by organisational factors outside the influence of the emergency department: A root cause analysis. *PLoS One*. 2018;13(9):e0202751.
5. Brouns SH, Stassen PM, Lambooij SL, Dieleman J, Vanderfeesten IT, Haak HR. Organisational Factors Induce Prolonged Emergency Department Length of Stay in Elderly Patients--A Retrospective Cohort Study. *PLoS One*. 2015;10(8):e0135066.
6. McCarthy ML, Zeger SL, Ding R, Levin SR, Desmond JS, Lee J, et al. Crowding Delays Treatment and Lengthens Emergency Department Length of Stay, Even Among High-Acuity Patients. *Ann Emerg Med*. 2009;54(4):492-503.e4.
7. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, et al. The Effect of Emergency Department Crowding on Clinically Oriented Outcomes. *Acad Emerg Med*. 2009;16(1):1-10.
8. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. *J Nurs Scholarsh*. 2014;46(2):106-15.
9. Ha JY, Sung WY. Impact of COVID-19 pandemic on emergency department length of stay and clinical outcomes of patients with severe pneumonia: A single-center observational study. *Medicine (Baltimore)*. 2022;101(38):e30633.
10. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS One*. 2018;13(8):e0203316.
11. Kaushik N, Khangulov VS, O'Hara M, Arnaout R. Reduction in laboratory turnaround time decreases emergency room length of stay. *Open Access Emerg Med*. 2018;10:37-45.
12. Li L, Georgiou A, Vecellio E, Eigenstetter A, Toouli G, Wilson R, et al. The effect of laboratory testing on emergency department length of stay: a multihospital longitudinal study applying a cross-classified random-effect modeling approach. *Acad Emerg Med*. 2015;22(1):38-46.
13. Hawkins RC. Laboratory turnaround time. *Clin Biochem Rev*. 2007;28(4):179-94.
14. Florkowski C, Don-Wauchope A, Gimenez N, Rodriguez-Capote K, Wils J, Zemlin A. Point-of-care testing (POCT) and evidence-based laboratory medicine (EBLM) – does it leverage any advantage in clinical decision making? *Crit Rev Clin Lab Sci*. 2017;54(7-8):471-94.
15. Murray RP, Leroux M, Sabga E, Palatnick W, Ludwig L. Effect of point of care testing on length of stay in an adult emergency department. *J Emerg Med*. 1999;17(5):811-4.
16. Lee-Lewandrowski E, Corboy D, Lewandrowski K, Sinclair J, McDermot S, Benzer TI. Implementation of a Point-of-Care Satellite Laboratory in the Emergency Department of an Academic Medical Center. *Arch Pathol Lab Med*. 2003;127(4):456-60.

17. Lewandrowski K, Flood J, Finn C, Tannous B, Farris AB, Benzerx TI, et al. Implementation of Point-of-Care Rapid Urine Testing for Drugs of Abuse in the Emergency Department of an Academic Medical Center: Impact on Test Utilization and ED Length of Stay. *Am J Clin Pathol*. 2008;129(5):796-801.

18. Goodacre S, Bradburn M, Cross E, Collinson P, Gray A, Hall AS. The RATPAC (Randomised Assessment of Treatment using Panel Assay of Cardiac Markers) trial: a randomised controlled trial of point-of-care cardiac markers in the emergency department. *Health Technol Assess*. 2011;15(23):1-102.

19. Kendall J, Reeves B, Clancy M. Point of care testing: randomised controlled trial of clinical outcome. *BMJ*. 1998;316(7137):1052-7.

20. Asha SE, Chan ACF, Walter E, Kelly PJ, Morton RL, Ajami A, et al. Impact from point-of-care devices on emergency department patient processing times compared with central laboratory testing of blood samples: a randomised controlled trial and cost-effectiveness analysis. *Emerg Med J*. 2014;31(9):714-9.

21. Lee EJ, Shin SD, Song KJ, Kim SC, Cho JS, Lee SC, et al. A point-of-care chemistry test for reduction of turnaround and clinical decision time. *Am J Emerg Med*. 2011;29(5):489-95.

22. Parvin CA, Lo SF, Deuser SM, Weaver LG, Lewis LM, Scott MG. Impact of point-of-care testing on patients' length of stay in a large emergency department. *Clin Chem*. 1996;42(5):711-7.

23. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med*. 2000;7(3):236-42.

24. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-9.

25. Gardner RL, Sarkar U, Maselli JH, Gonzales R. Factors associated with longer ED lengths of stay. *Am J Emerg Med*. 2007;25(6):643-50.

26. Ialongo C, Porzio O, Giambini I, Bernardini S. Total Automation for the Core Laboratory: Improving the Turnaround Time Helps to Reduce the Volume of Ordered STAT Tests. *J Lab Autom*. 2016;21(3):451-8.

27. Singh G, Savage NM, Gunsolus B, Foss KA. Requiem for the STAT Test: Automation and Point of Care Testing. *Lab Med*. 2020;51(2):e27-e31.

28. Jensen K, Haniff R, Kamarinos A, Rosenberg A, Santiago M, Laser J. Improving Turnaround Times through a Process Improvement Initiative Involving Barcoding, Floorplans, Dual Measuring Cells, Chemistry Analyzers, and Staff Shifts. *J Appl Lab Med*. 2019;4(3):311-22.

29. Ellison TL, Alharbi M, Alkaf M, Elimam S, Alfaries M, Nounou RA, et al. Implementation of total laboratory automation at a tertiary care hospital in Saudi Arabia: effect on turnaround time and cost efficiency. *Ann Saudi Med*. 2018;38(5):352-7.

30. Angeletti S, De Cesaris M, Hart JG, Urbano M, Vitali MA, Fragliasso F, et al. Laboratory Automation and Intra-Laboratory Turnaround Time: Experience at the University Hospital Campus Bio-Medico of Rome. *J Lab Autom*. 2015;20(6):652-8.

Chapter 3

Added diagnostic value of routinely measured hematology variables in diagnosing immune checkpoint inhibitor mediated toxicity in the emergency department

M.S.A. Niemantsverdriet*, B.E.L. Vrijsen*, T. Visser 't Hooft, K.P.M. Suijkerbuijk, W.W. van Solinge, M.J. ten Berg, S. Haitjema
* These authors contributed equally to the manuscript

Cancer Med 2023; 12(11): 12462-9

Abstract

Background

Immune checkpoint inhibitors (ICI) show remarkable results in cancer treatment, but at the cost of immune related adverse events (irAE). irAE can be difficult to differentiate from infections or tumor progression, thereby challenging treatment, especially in the emergency department (ED) where time and clinical information are limited. As infections are traceable in blood, we were interested in the added diagnostic value of routinely measured hematological blood cell characteristics in addition to standard diagnostic practice in the ED to aid irAE assessment.

Methods

Hematological variables routinely measured with our hematological analyzer (*Abbott CELL DYN Sapphire*) were retrieved from Utrecht Patient Oriented Database (UPOD) for all patients treated with ICI who visited the ED between 2013 and 2020. To assess the added diagnostic value, we developed and compared two models; a base logistic regression model trained on the preliminary diagnosis at the ED, sex and gender, and an extended model trained with lasso that also assessed the hematology variables.

Results

409 ED visits were used in this analysis. The extended model showed an improvement in performance (area under the receiver operator characteristic curve) over the base model, 0.79 (95% CI 0.75 – 0.84) and 0.67 (95% CI 0.60 – 0.73), respectively. Two standard blood count variables (eosinophil granulocyte count and red blood cell count) and two advanced variables (coefficient of variance of neutrophil depolarization and red blood cell distribution width) were associated with irAE.

Conclusion

Hematological variables are a valuable and inexpensive aid for irAE diagnosis in the ED. Further exploration of the predictive hematological variables could yield new insights into the pathophysiology underlying irAE and in distinguishing irAE from other inflammatory conditions.

Introduction

Within the immunotherapeutic field of cancer treatment, multiple new and promising treatment options have emerged over the past years.[1] Among these, immune checkpoint inhibitors (ICI) are increasingly being used as an oncologic treatment strategy for multiple types of cancer and have drastically improved the survival of responding patients. For example, patients with advanced melanoma treated with combined nivolumab and ipilimumab therapy have shown to result in a median overall survival of over 60 months,[2] whereas the median survival of patients with metastatic melanoma used to be less than one year before the introduction of checkpoint inhibitors.[3] The proportion of cancer patients benefitting from ICI is increasing rapidly, with now over 40% of cancer patients qualifying for ICI treatment.[4] However, their use is associated with a wide variety of immune-related adverse events (irAE), such as auto-immune colitis and pneumonitis.[5] Because of overlap in clinical presentation, it can be difficult to differentiate these irAE from progressive disease or other inflammatory conditions, such as infections. Especially in the emergency department (ED) where time and resources are limited, this may lead to diagnostic delay, inappropriate treatment and a considerable amount of (unnecessary) diagnostic testing.[6,7] Accurate and early diagnosis of patients presenting in the ED with irAE is therefore key to start adequate treatment as soon as possible.[8,9]

Currently, there are only a few biomarkers available that can aid in diagnosing irAE.[6,10] A solution to this problem might be found in routinely measured hematological variables. Bacterial infection and viral infections are commonly characterized by high neutrophil and lymphocyte counts respectively, whereas auto-immune diseases and allergies typically show high eosinophil counts. Previous research has found associations between irAE and increased counts of standard hematology measurements (e.g. absolute lymphocyte count and eosinophil count).[6] In addition, changes in B- and T-cell receptor repertoire show associations with irAE onset and prognosis.[6] However, none of these biomarkers have been extensively validated or are used in clinical practice. Most modern hematology analyzers not only provide blood cell counts, but also measure morphologic characteristics, such as cell size, intrinsic properties and cell viability that carry diagnostic and prognostic value. This raises the question whether they may also be of use in the setting of immunological toxicity.[11-13]

To answer these types of questions, scrutinizing complex datasets with conventional statistical methods, such as logistic regression, do not provide stable estimates of the variable's coefficients as models contain too many variables and a low number of samples. New advanced statistical and machine learning (ML) methods

are able to remove irrelevant variables thereby reducing the number of variables. In addition, variables of high importance, also known as predictors, can be identified by evaluating the trained coefficient of the trained model. This way, ML allows for the possible identification of new biomarkers and exploration of new horizons in research to aid irAE diagnosis.

The aim of this study was therefore to determine the added value of routinely measured hematology characteristics, modelled through ML, as compared to the standard diagnostic practice. This may aid in the diagnosis of irAE in the ED and understanding of the pathophysiology.

Methods

Study population

This retrospective observational study included all visits to the ED of the University Medical Center Utrecht (UMC Utrecht) between 2013 and 2020 of patients who were being treated with any type of ICI for any type of cancer, until three months after cessation of treatment. Because irAE can occur even after cessation of treatment, we chose to include ED visits up to three months after treatment with ICI ended.[14] The cut-off of three months was chosen after discussion between the authors. If patients had more than one disease episode (defined as a consecutive period with infection-like symptoms), all patient's ED visits were included separately, whereas for patients with multiple ED visits during one disease episode, only the first visit was included. If patients visited the ED multiple times for the same condition (e.g. due to worsening of symptoms), only the first visit was included.

Data collection

For all ED visits, demographic (age and sex), medication and hematology data were extracted from the Utrecht Patient Orientated Database (UPOD). In brief, UPOD is a relational database combining clinical characteristics, medication and laboratory measurements of patients in the UMC Utrecht since 2004.[15] We used hematological variables measured by the *CELL-DYN Sapphire* hematology analyzer (*Abbott diagnostics, Santa Clara, USA*). The *CELL-DYN Sapphire* is a cell counter equipped with a 488-nm blue diode laser and uses multiple techniques, such as electrical impedance, spectrophotometry and laser light scattering, to measure morphological characteristics of leukocytes (incl. 5-part differential), red blood cells (RBCs), and platelets for both classification and enumeration. Each time a component of a complete blood cell count (CBC) is requested, all data generated by the hematology

analyzer are automatically stored in UPOD, including a substantial number of raw and research-only values and background data on cell characteristics which are made available for research purposes. Only visits with available Sapphire data within the first four hours after ED presentation were included in this study to ensure we only used data from patients with infection-like symptoms during the ED visit. UPOD data acquisition and management is in accordance with current regulations concerning privacy and ethics.

irAE label definition

A manual chart review was done for all ED visits within our study population by two of the authors (TVtH and BV). Visits for evidently unrelated conditions were excluded. We recorded both the preliminary and definite diagnosis. The preliminary diagnosis was defined as the diagnosis made by the treating physician in the ED, and was characterized as either *suspected irAE* or *other*. The definitive diagnosis was defined as the diagnosis made by the treating physician at discharge from the hospital or at the end of treatment, and was characterized as *irAE* or *other*. Ambiguous cases were resolved through consensus.

Model development

Two models were trained to evaluate the added diagnostic performance of the hematology variables for irAE diagnosis. The first model (base) assessed the preliminary diagnosis, sex and age with logistic regression thereby imitating clinical practice at the ED, whereas the second model (extended) also included the 77 additional hematology variables. A quality control protocol was performed to remove variables with no additional predictive value during model development: hematology variables with a Pearson correlation of > 0.80 or with low number of unique (n = 5) values were removed. The extended model was trained using lasso machine learning that can automatically reduce the number of variables, thereby reducing the risk of overfitting and aiding the interpretability of the model. Means and standard deviations are shown for normally distributed variables whereas medians and interquartile ranges (IQR) are shown for non-normally distributed variables.

Model performance was assessed using cross validation (CV). With CV, the data is split in K number of partitions (folds), of which $K - 1$ folds are used for training and 1 for testing. This exercise is .repeated K times resulting in K models with K performance estimates. Contrary to the conventional train-and-test split, multiple models are trained on multiple data splits, thereby using all data to assess the model's performance. The lasso algorithm performs shrinkage of coefficients that

can get as small as 0, thereby removing variables. The lambda hyper-parameter of lasso determines the degree of shrinkage and was optimized in a double loop cross validation (DLCV) scheme, also known as nested cross validation (Supplementary Figure 1)[16]. A K of 10 was used for both the CV and DLCV schemes.

Model evaluation

The discrimination of models was assessed by plotting receiver operator characteristic (ROC) curves. The area under the ROC (AUROC) is a measure of discrimination, an AUROC of 1 indicates a perfect model, whereas an AUROC of 0.5 indicates a random model. The 95% confidence interval (CI) of the AUROC was computed with the R cvAUC package by evaluating the test performances of the two model configurations trained in both CV schemes.[17] Variable coefficients of the ten models trained in the DLCV were evaluated as variable importance (predictors).

The clinical application and value of the trained models were evaluated with both calibration plots and net benefit curves. Calibration plots portray the agreement between predicted probabilities and the observed frequency of irAE. A calibration with an intercept 0 and slope of 1 shows perfect calibration, whereas a slope of > 1 shows a model that overestimates outcome and a slope of < 1 underestimates diagnosis. 80% and 95% CI intervals of the calibration plots were generated with the R givitR package.[18] Net benefit is a measure to evaluate the clinical benefit of a prediction model by comparing the benefit (treating diseased, true positives (TP)) and cost (treating non-diseased, false positive (FP)).[19] Net benefit is assessed by subtracting the cost from the benefit for the complete range of predictions values (p_t). Formula 1 shows that the net benefit increases by the number of TP and is penalized by the number of non-diseased (FP), especially when the prediction threshold value increases ($\frac{p_t}{1-p_t}$). Besides the net benefit, the number needed to treat (NNT) is shown as a comparison to how health care professionals consider whether the patient has a specific illness or that treatment is required. All analyses were performed in R version 4.1.2.[20]

Formula 1: $net\ benefit\ (p_t) = \frac{TP}{n} - \frac{FP}{n}\left(\frac{p_t}{1-p_t}\right)$

Post-hoc subgroup analysis

To assess the independence of the identified biomarkers we adjusted for the baseline clinical variables. We performed a multivariate analysis including the identified biomarkers, age, sex, cancer type, and ICI medication. To reduce the number of coefficients and to remove groups with low prevalence, various cancer types and ICI medications were grouped.

A second post-hoc analysis was performed to check whether the identified biomarkers were associated with disease severity as measured by CTCAE grade.

Results

Patient characteristics

Between 2013 and 2020, 409 ED visits of 257 patients who were treated with ICI and had available blood counts were included in this study (mean ED visits per patient 1.6). The irAE diagnosis of 91 visits were inconclusive from the medical records, of which the diagnosis was later adjusted in 24 cases. In both the *other* (n = 268) and irAE (n = 141) sub-groups there were more males, 63.1% and 64.5%, respectively (Table 1). Mean age did not differ between the *other* (62.2) and irAE group (61.7). The use of both ipilimumab and nivolumab was significantly higher in the irAE group ($p < .01$), whereas the use of nivolumab and pembrolizumab was significantly lower in the irAE group ($p < .01$). An overview of the irAE diagnoses is shown in Supplementary Table 1.

Model performance

After removing variables that did not meet our quality control criteria, 53 of the 77 Sapphire variables were used in the extended model (Supplementary Table 2, Supplementary Figure 2). The base model had an AUROC of 0.67 (95% CI 0.60 – 0.79) and the extended model had an AUROC of 0.79 (95% CI 0.75 – 0.84), a difference of 0.13. The training performance was marginally higher for both the base and extended model as compared to the test performance, 0.74 (95% CI 0.72 – 0.76) and 0.86 (95% CI 0.84 – 0.87), respectively, providing evidence there was no overfitting. In line with the AUROC metrics, the extended model trained on all data showed the best ROC and PRC curves (Figure 1).

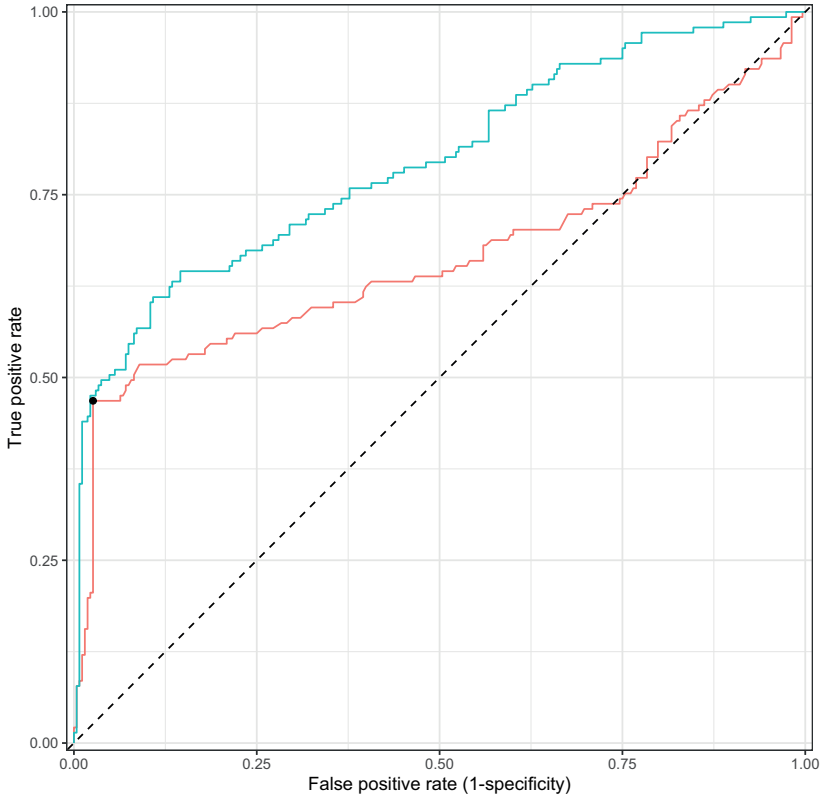
Table 1: baseline characteristics of ED visits treated with ICI

	Other (n = 268)	irAE (n = 141)	p-value
Age (SD)	62.2 (11.9)	61.7 (12.8)	.680
Sex, male count (%)	169 (63.1%)	91 (64.5%)	.851
Cancer diagnosis			
- Central nervous system	4 (1.5%)	3 (2.1%)	
- Gynaecological	6 (2.2%)	2 (1.4%)	
- Head and neck	3 (1.1%)	2 (1.4%)	
- Hematological	7 (2.6%)	4 (2.8%)	
- Hepato-pancreato-biliary	2 (0.7%)	1 (0.7%)	
- Intestinal	14 (5.2%)	0 (0.0%)	
- Lung	88 (32.8%)	23 (16.3%)	
- Skin	119 (44.4%)	94 (66.7%)	
- Urological	25 (9.3%)	12 (8.5%)	
Preliminary diagnosis, count			
- other	261	75	< .001
- irAE	7	66	
CTCAE grade:			
- 1		13 (9.2%)	
- 2		48 (34.0%)	
- 3		69 (48.9%)	
- 4		10 (7.1%)	
- 5		1 (0.7%)	
ICI medication*, count (%)			
- atezolizumab	16 (6.0%)	2 (1.4%)	.060
- durvalumab	4 (1.5%)	5 (3.5%)	.322
- ipilimumab	25 (9.3%)	26 (18.4%)	.013
- nivolumab	89 (33.2%)	24 (17.0%)	.001
- pembrolizumab	86 (32.1%)	26 (18.4%)	.005
- tremelimumab	5 (1.9%)	4 (2.8%)	.778
- ipilimumab and nivolumab	43 (16.0%)	54 (38.3%)	< .001

* not mutually exclusive

CTCAE: common terminology criteria adverse events; ED: emergency department; ICI: immune checkpoint inhibitor; irAE: immune-related adverse events

Figure 1: ROC of the base (red line) and extended (blue line) models test predictions



Predictions on the test folds of the double loop cross validation scheme were concatenated to draw the ROC curves. The black dot denotes the discriminative metrics of the preliminary diagnosis. The diagonal line shows the performance of a random model.
ROC: receiver operating characteristic

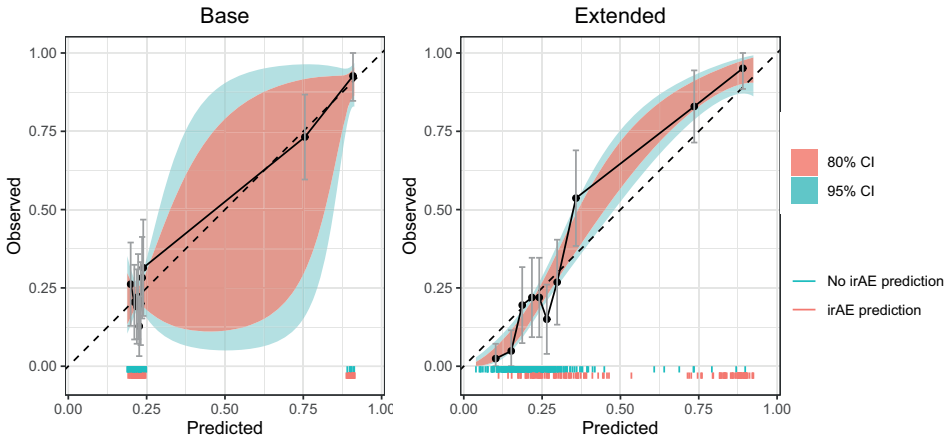
Discriminative metrics

To assess the potential value in clinical practice of the extended model, predictions of the base and extended models were evaluated with both calibration and net benefit plots. The extended model showed better calibration than the base model (Figure 2). The 95% CI of the base model are very wide compared to the extended model and the predictions of the extended model are more equally distributed. In addition, decision curve analysis showed improved net benefit of the extended model as compared to the base model over the complete threshold probability range (Figure 3).

Variable importance

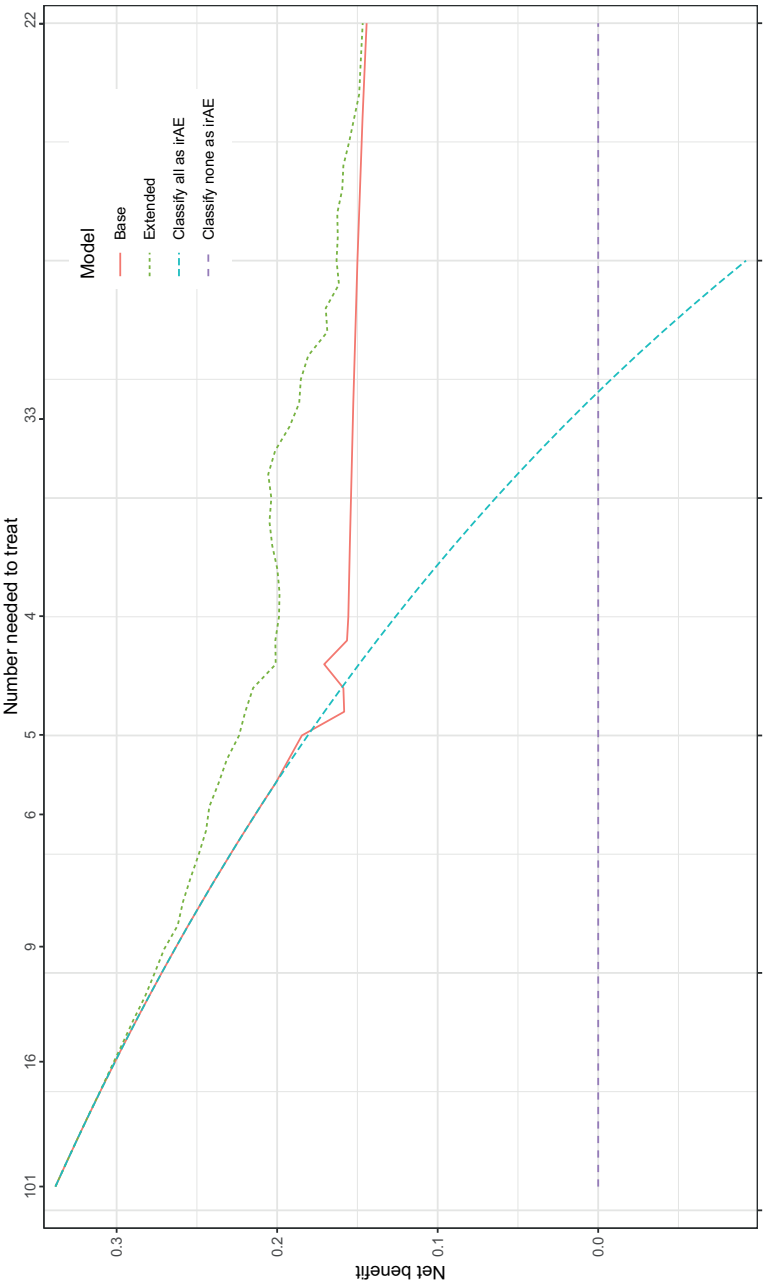
Variables' coefficients, as well as the number of times a variable was selected by the extended model, were documented during training and are shown in Figure 4 and Table 2. The preliminary diagnosis was highly predictive for irAE diagnosis in both the base and extended model with a coefficient of 3.53 ± 0.14 and 2.88 ± 0.18, respectively. The extended model also identified the following Sapphire variables as predictors for irAE diagnosis: number of eosinophils (EOS), red blood cell count measured with impedance (RBCI), coefficient of variance neutrophil depolarization (NDCV) and red blood cell distribution width (RDW), of which the latter was negatively associated with irAE. EOS was highly correlated with percentage of eosinophils (PEOS) and RBCI with other red blood cell measurements variables (RBCO, HGB and HCT) (Supplementary Figure 2). The sex and age variables were not selected by lasso in any of the ten iterations in the DLCV scheme.

FIGURE 2: CALIBRATION PLOTS OF BOTH THE BASE AND EXTENDED MODELS



Both calibration curves computed with the number of expected (model predictions) and observed irAE are shown, as well as the 80% and 90% confidence intervals (CI). The segments on the lower part of both plots indicate the computed predictions for each model.
irAE: immune-related adverse event

FIGURE 3: DECISION CURVE OF THE BASE AND EXTENDED MODELS



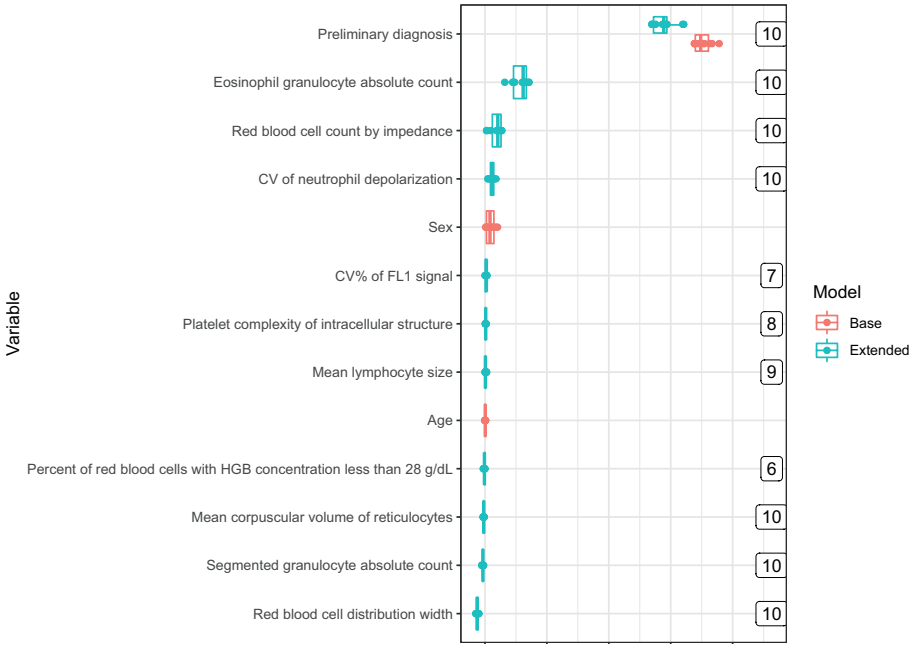
The blue and purple indicate the extreme base strategies of either treating everyone as irAE, or none as irAE, respectively.
irAE: immune-related adverse event

Table 2: variable importance for both the base and extended models estimated in the CV schemes.

Definition	Estimated coefficient	
	Base	Extended
Preliminary diagnosis	3.52 (± 0.14)	2.89 (± 0.19)
Eosinophil granulocyte absolute count		0.57 (± 0.13)
Red blood cell count by impedance		0.18 (± 0.10)
Coefficient of variance of neutrophil side scattering		0.11 (± 0.04)
Red blood cell distribution width		-0.12 (± 0 .02)

Top 5 selected variables based on absolute coefficient mean (± SD) for both models.
CV: cross validation; SD: standard deviation

Figure 4: variable importance of both the base and extended model



Variable importance of both the base and extended model. Only variables selected more than 5 times by lasso in the DLCV are shown. The number positioned on right shows the number of times a variable was selected by the extended model in the 10-fold DLCV scheme.
CV: coefficient of variance; DLCV: double loop cross validation; HGB: hemoglobin

Post-hoc subgroup analysis

After adjusting for age, sex, cancer type (grouped as skin, lung, urological or other) and ICI medication (grouped as ipilimumab, nivolumab, pembrolizumab, ipilimumab and nivolumab, or other) we found that three of the four identified variables were still significantly associated with irAE, namely: EOS ($p = .0144$), RBCI ($p = .0035$), and RDW ($p = .0003$). In this model we did not find a significant association for NDCV ($p = .0781$). Furthermore, we did not find an association between the values of the identified variables and the irAE severity as measured by CTCAE grade (Supplementary Figure 3).

Discussion

Accurate identification of irAE in patients using ICI in the ED is of vital importance to guide treatment decisions. With new statistical methods and ML, we explored the possible added diagnostic value of 77 hematological variables measured by the *CELL-DYN Sapphire* in diagnosing irAE in patients using ICI as compared to standard clinical practice. The extended model showed improvement in discrimination, calibration and net benefit as compared to the base model, indicating that the hematological variables indeed have added value in the diagnostic process of identifying irAE in patients using ICI in the emergency department setting.

Our extended model showed better performance as well as calibration over the base model. However, due to the low number of values of the base model and the good predictive performance of the preliminary diagnosis, the predictions of the base model were not equally distributed. The net benefit of the extended model was better than the base model, especially in the therapeutic range around 25%. The exact threshold for the number needed to treat will vary depending on the characteristics of the individual patient and the severity of the symptoms. A false-positive diagnosis of irAE will lead to cessation of the checkpoint inhibitor, which would possibly withhold a life-saving therapy from the patient. On the other hand, a false-negative diagnosis will lead to a delayed treatment for irAE, which is potentially fatal.[21]

Of all variables, the preliminary diagnosis was deemed highly important by both the base and extended models indicating that the first diagnosis of the physician is a very good proxy for irAE diagnosis. Both age and sex showed low importance in the base model and were not selected by the lasso algorithm in any of the 10 DLCV iterations, which is in line with existing evidence.[22] Interestingly, only a few of the 77 hematological variables were selected by the lasso algorithm in each iteration. This diagnostic study cannot not determine causality. However, a causal relationship can be postulated based on the literature.

Eosinophiles are thought to play a pathogenic role in auto-immune disorders and are known to be associated with irAE.[6] Neutrophil depolarization is a feature of neutrophil activation, which has also been associated with auto-immunity, but this has not been studied extensively.[23] We found the red blood cell distribution width (RDW) to be negatively associated with irAE. Increased RDW is known to be associated with infections, which are arguably the most likely alternative diagnosis when considering irAE.[24]

Our study has some limitations. The population is highly heterogeneous, with multiple types of tumors and treatments. This may have hampered the identification of a specific predictor for a particular subset of patients. Unfortunately, we did not have enough data to stratify patients based on either cancer type or medication. Even though the post-hoc group analysis showed significant results for 3 of the 4 identified variables after adjusting for the baseline characteristics, future research is needed to validate these results. Moreover, the diagnoses were retrospectively defined or changed as our data was collected on routine basis.

To our knowledge, this study is one of the first of its kind in exploring the diagnostic potential of these raw and research-only hematological variables using ML in the emergency department setting. Since the raw data from this type of hematology analyzer are not ubiquitously available, we were not able to externally validate our results. As a result, this study has to be viewed as exploratory and more research is required before these hematological variables, either individually or in a model, can be used in clinical practice. The diagnostic performance of such a model might be improved by combining hematological variables with other new sets of biomarkers, as well as the preliminary diagnosis.

This study raises the question if the hematological variables might also have diagnostic value in the setting of other diseases and treatments.[11-13] As they are inexpensive and relatively easily and rapidly obtained in general blood counts, they could be an interesting new tool in future diagnostic research. As shown here, a clinical diagnostic model may aid the clinical decision-making process of a physician by providing a continuous prediction score that can be combined with the professional interpretation by a clinical chemist to accommodate integral diagnostics of a patient's clinical state.[25] Instead of looking at differences between patients using cross-sectional data, within-patient differences may be a better approximation of a patient's health trajectory potentially allowing for predicting the incidence of irAE at the start of ICI treatment.

Overall, we show that hematological variables show diagnostic performance in the identification of irAE in patients using ICI at the ED and that they have added

value compared to standard diagnostic practice. Our results suggest new directions for further research using (advanced) hematological variables for irAE diagnosis in the emergency setting.

Conflict of interest

MN is employed by SkylineDx, Rotterdam and receives a PhD fellowship from SkylineDx, Rotterdam. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

KS: Consulting/advisory relationship: Bristol-Myers Squibb, Merck Sharp and Dome, Abbvie, Pierre Fabre, Novartis. Honoraria received: Novartis, Roche, Merck Sharp and Dome. Research funding: TigaTx, Bristol Myers Squibb, Philips. All paid to institution and outside the submitted work.

References

1. Farkona S, Diamandis EP, Blasutig IM. Cancer immunotherapy: the beginning of the end of cancer? *BMC Med*. 2016;14(1):73.
2. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Lao CD, et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med*. 2019;381(16):1535-46.
3. Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, et al. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med*. 2015;372(4):320-30.
4. Haslam A, Prasad V. Estimation of the Percentage of US Patients With Cancer Who Are Eligible for and Respond to Checkpoint Inhibitor Immunotherapy Drugs. *JAMA Netw Open*. 2019;2(5):e192535-e.
5. Martins F, Sofiya L, Sykiotis GP, Lamine F, Maillard M, Fraga M, et al. Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. *Nat Rev Clin Oncol*. 2019;16(9):563-80.
6. Hommes JW, Verheijden RJ, Suijkerbuijk KPM, Hamann D. Biomarkers of Checkpoint Inhibitor Induced Immune-Related Adverse Events-A Comprehensive Review. *Front Oncol*. 2020;10:585311.
7. von Itzstein MS, Khan S, Gerber DE. Investigational Biomarkers for Checkpoint Inhibitor Immune-Related Adverse Event Prediction and Diagnosis. *Clin Chem*. 2020;66(6):779-93.
8. Puzanov I, Diab A, Abdallah K, Bingham CO, 3rd, Brogdon C, Dadu R, et al. Managing toxicities associated with immune checkpoint inhibitors: consensus recommendations from the Society for Immunotherapy of Cancer (SITC) Toxicity Management Working Group. *J Immunother Cancer*. 2017;5(1):95.
9. La-Beck NM, Jean GW, Huynh C, Alzghari SK, Lowe DB. Immune Checkpoint Inhibitors: New Insights and Current Place in Cancer Therapy. *Pharmacotherapy*. 2015;35(10):963-76.
10. Jia XH, Geng LY, Jiang PP, Xu H, Nan KJ, Yao Y, et al. The biomarkers related to immune related adverse events caused by immune checkpoint inhibitors. *J Exp Clin Cancer Res*. 2020;39(1):284.
11. Kofink D, Muller SA, Patel RS, Dorresteijn JAN, Berkelmans GFN, de Groot MCH, et al. Routinely measured hematological parameters and prediction of recurrent vascular events in patients with clinically manifest vascular disease. *PLoS One*. 2018;13(9):e0202682.
12. Bindraban RS, Ten Berg MJ, Haitjema S, Hoefer IE, de Regt M, Kramer M, et al. Exploring the value of routinely measured hematology parameters for identification of elderly patients at high risk of death at the Emergency Department. *Acute Med*. 2018;17(4):188-202.
13. Gijsberts CM, den Ruijter HM, de Kleijn DPV, Huisman A, Ten Berg MJ, van Wijk RHA, et al. Hematological Parameters Improve Prediction of Mortality and Secondary Adverse Events in Coronary Angiography Patients: A Longitudinal Cohort Study. *Medicine (Baltimore)*. 2015;94(45):e1992.
14. Postow MA, Sidlow R, Hellmann MD. Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *N Engl J Med*. 2018;378(2):158-68.
15. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-9.
16. Wessels LF, Reinders MJ, Hart AA, Veenman CJ, Dai H, He YD, et al. A protocol for building and evaluating predictors of disease state based on microarray data. *Bioinformatics*. 2005;21(19):3755-62.

17. LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat*. 2015;9(1):1583-607.

18. Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration Belt for Quality-of-Care Assessment Based on Dichotomous Outcomes. *PLoS One*. 2011;6(2):e16110.

19. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-74.

20. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

21. Wang DY, Salem J-E, Cohen JV, Chandra S, Menzer C, Ye F, et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. *JAMA Oncol*. 2018;4(12):1721-8.

22. Verheijden RJ, May AM, Blank CU, van der Veldt AAM, Boers-Sonderen MJ, Aarts MJB, et al. Lower risk of severe checkpoint inhibitor toxicity in more advanced disease. *ESMO Open*. 2020;5(6):e000945.

23. Bissenova S, Ellis D, Mathieu C, Gysemans C. Neutrophils in autoimmunity: when the hero becomes the villain. *Clin Exp Immunol*. 2022;210(2):128-40.

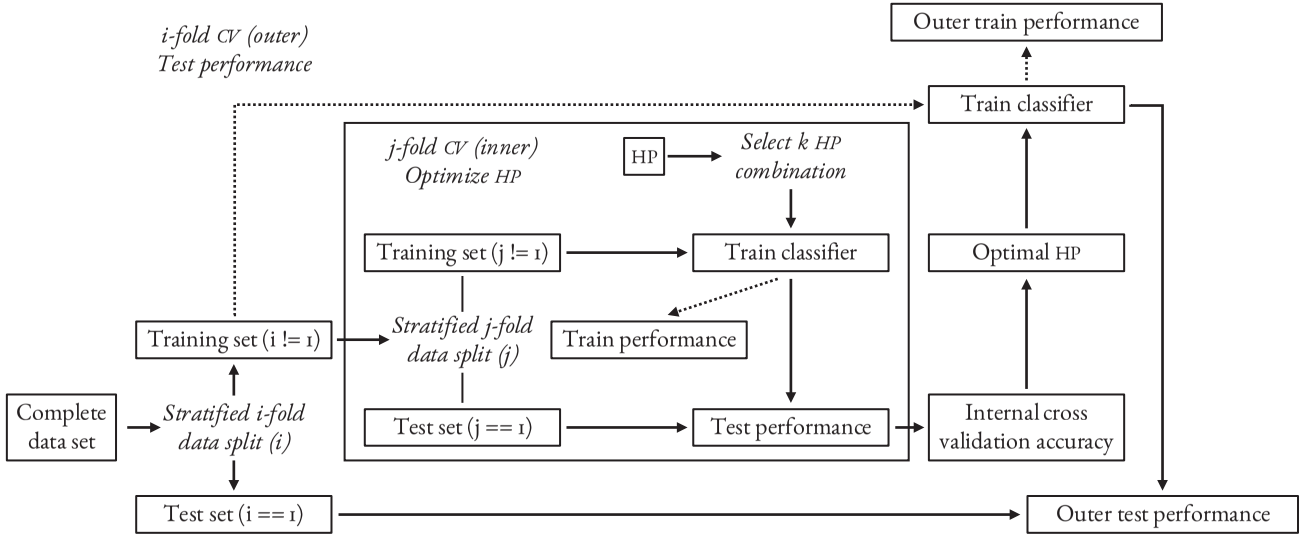
24. Uffen JW, Oomen P, de Regt M, Oosterheert JJ, Kaasjager K. The prognostic value of red blood cell distribution width in patients with suspected infection in the emergency department. *BMC Emerg Med*. 2019;19(1):76.

25. van Solinge WW, Ten Berg MJ, Haitjema S. [Data-driven integrated diagnostics: the natural evolution of clinical chemistry?]. *Ned Tijdschr Geneeskd*. 2019;163.

Supplementary material

Supplementary Figure 1: scheme of the double loop cross validation (DLCV)
Supplementary Figure 2: dendrogram of all 77 hematological variables (CELL-DYN Sapphire) computed with Euclidean distance
Supplementary Figure 3: distributions of the four identified variables
Supplementary Table 1: total number of diagnoses of immune related adverse events
Supplementary Table 2: description of all 77 CELL-DYN Sapphire variables

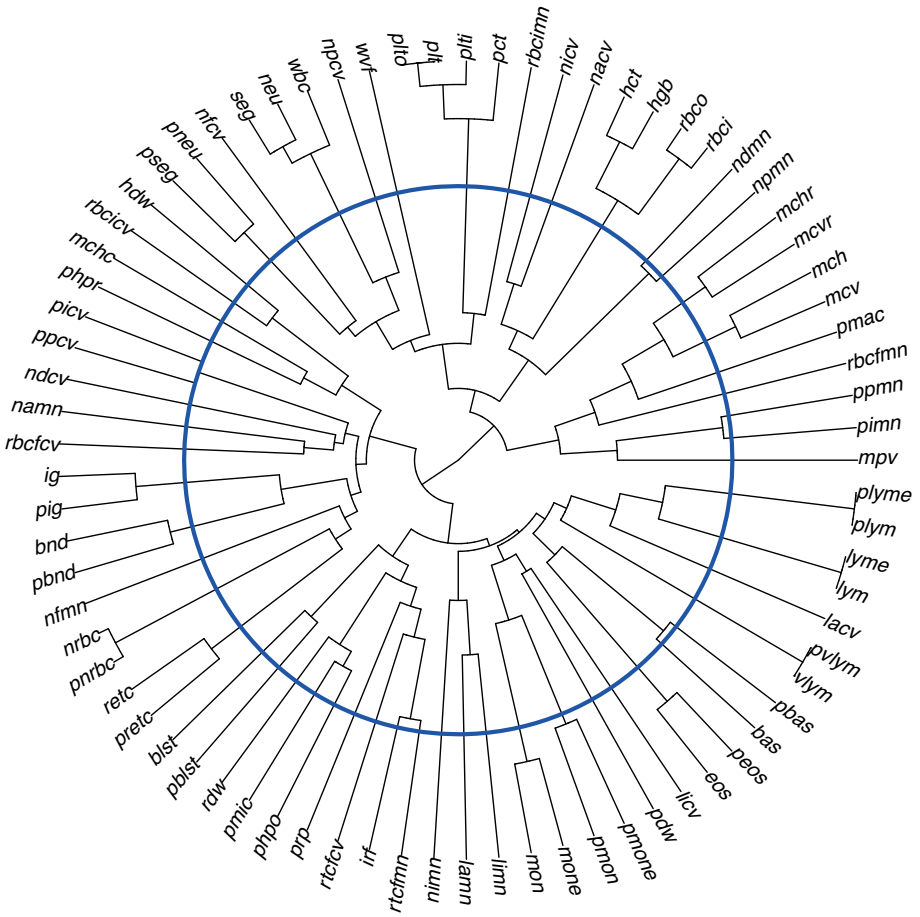
Supplementary Figure 1: scheme of the double loop cross validation (DLCV)



* data was stratified on both label and patient ID.

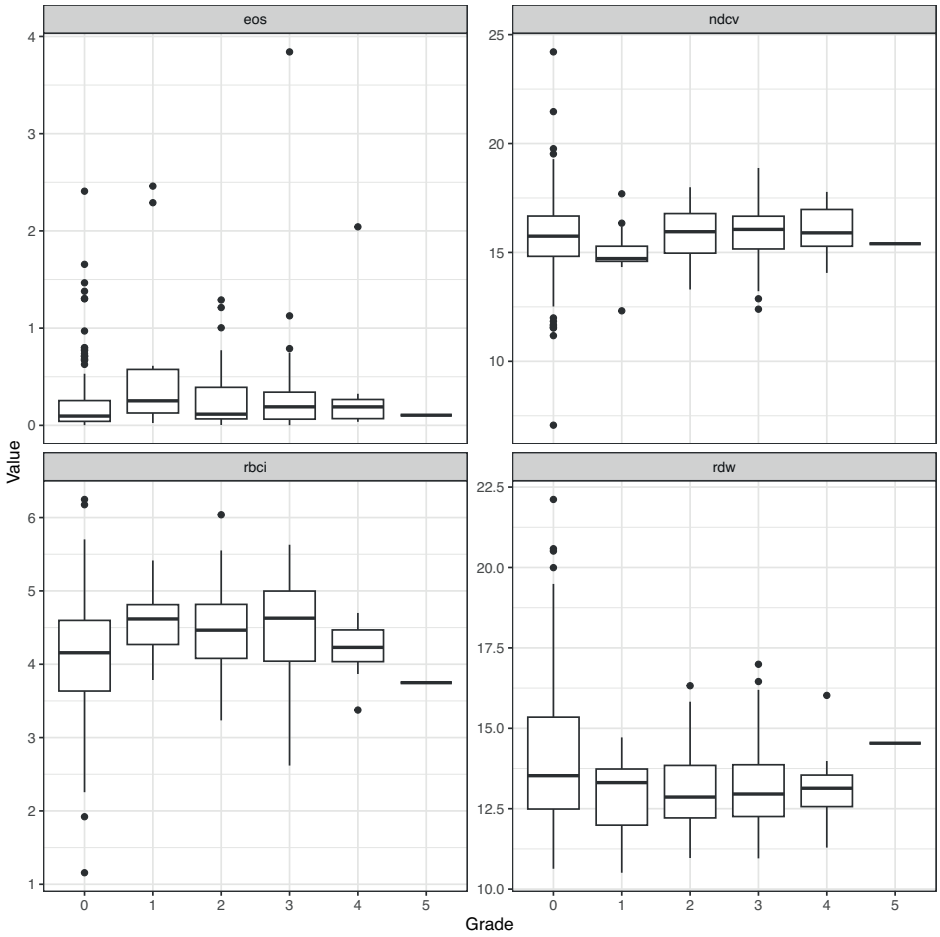
CV: cross validation; HP: hyperparameter.

Supplementary Figure 2: dendrogram of all 77 hematological variables (*cell-dyn sapphire*) computed with euclidean distance



Length of each node depicts the similarity between variables. The blue circle represents the 0.80 cut-off. Only one of the clustered variables outside the circle was used in training.

Supplementary Figure 3: distributions of the four identified variables



Visits with irAE were graded between 1 – 5, visits without irAE were assigned with a CTCAE grade of 0. CTCAE: common terminology criteria adverse events; irAE: immune-related adverse events

Supplementary Table 1: total number of diagnoses of immune related adverse events

Medical diagnosis	Count (%)
arthritis	2 (0.9%)
cholangitis	2 (0.9%)
colitis	61 (27.2%)
dermatitis	15 (6.7%)
diabetes	4 (1.8%)
duodenitis	2 (0.9%)
encephalitis	5 (2.2%)
fever (unspecified)	1 (0.4%)
gastritis	4 (1.8%)
hepatitis	29 (12.9%)
hypophysitis	24 (10.7%)
immune thrombocytopenia (ITP)	1 (0.4%)
meningitis	7 (3.1%)
myocarditis	1 (0.4%)
myositis	2 (0.9%)
nephritis	4 (1.8%)
pancreatitis	2 (0.9%)
pericarditis	1 (0.4%)
pleuritis	1 (0.4%)
plexopathy	1 (0.4%)
polymyalgia rheumatica	2 (0.9%)
pneumonitis	40 (17.9%)
radiculitis	1 (0.4%)
sarcoid like reaction	3 (1.3%)
thyroiditis	9 (4.0%)

(n = 224), not mutually exclusive.

Supplementary Table 2: description of all 77 cell-dyn Sapphire variables

Variable	Description
BAS	Basophilic granulocyte absolute count
BLST	Blast absolute count
BND	Banded granulocyte absolute count
CHCR	Mean corpuscular HGB concentration per reticulocyte
EOS	Eosinophil granulocyte absolute count
HCT	Hematocrit
HDW	Hemoglobin distribution width
HGB	Hemoglobin USA units
IG	Immature granulocyte absolute count
IRF	Immature reticulocyte fraction
LACV	CV of axial light loss
LAMN	Mean lymphocyte size
LICV	CV of intermediate angle scattering
LIMN	Intermediate angle scattering
LYM	Lymphocyte absolute count
LYME	Lymphocyte (excluding atypical lymphocytes)
MCH	Mean corpuscular hemoglobin (USA units)
MCHC	Mean corpuscular hemoglobin concentration (USA units)
MCHR	Mean corpuscular HGB per reticulocyte
MCV	Mean corpuscular volume
MCVR	Mean corpuscular volume of reticulocytes
MON	Monocyte absolute count
MONE	Monocytes (excluding blasts) absolute count
MPV	Mean platelet volume
NACV	Coefficient of variance of neutrophil size
NAMN	Mean neutrophil size
NDCV	CV of depolarized side scattering
NDMN	Neutrophil lobularity / granularity and nuclear lobularity
NEU	Neutrophilic granulocyte (segments, banded and immature granulocytes) absolute count
NFCV	CV of fluorecent channel 3
NFMN	Fluorecent channel 3
NICV	CV of intermediate angle scattering
NIMN	Intermediate angle scattering

Supplementary Table 2: continued.

Variable	Description
NPCV	CV of polarized side scattering
NPMN	Polarized side scattering
NRBC*	Nucleated red blood cells absolute count
PBAS	Percentage of basophilic granulocytes
PBLST	Percentage of blasts
PBND	Percentage of banded granulocytes
PCT	Plateletcrit
PDW	Platelet distribution width
PEOS	Percentage of eosinophilic granulocytes
PHPO	Percent of red blood cells with HGB concentration less than 28 g/dL
PHPR	percent of red blood cells with HGB concentration more than 41 g/dL
PICV	CV of intermediate angle scattering
PIG	Percentage of immature granulocytes
PIMN	Platelet complexity of intracellular structure
PLT	Platelet count
PLTI	Platelet count by impedance
PLTO	Platelet count by optics
PLYM	Percentage of lymphocytes
PLYME	Percentage of lymphocytes (excluding atypical lymphocytes)
PMAC	Percent of red blood cells with volume greater than 120 fL
PMIC	Percent of red blood cells with volume less than 60 fL
PMON	Percentage of monocytes
PMONE	Percentage of monocytes (excluding blasts)
PNEU	Percentage neutrophilic granulocytes
PNRBC*	NRBC percentage count per 100 WBC
PPCV	CV of polarized side scattering
PPMN	Polarized side scattering
PRETC	Percentage of reticulocytes
PRP	Percentage of reticulated platelets / enumeration of reticulated platelets
PSEG	Segmented granulocyte percentage count
PVLYM*	Percentage of atypical lymphocyte
RBCFCV	CV% of FL1 signal
RBCFMN	Mean of FL1 signal
RBCI	Red blood cell count by impedance

Supplementary Table 2: continued.

Variable	Description
RBCICV	CV of intermediate angle scattering
RBCIMN	Intermediate angle scattering
RBCO	Red blood cell count by optics
RDW	Red blood cell distribution width
RETC	Reticulocyte absolute count
RTCFCV	CV% of Reticulocyte population during reticulocyte measurement on FLI signal
RTCFMN	Position of Reticulocyte population during reticulocyte measurement on FLI signal
SEG	Segmented granulocyte absolute count
VLYM*	Atypical (variant) lymphocyte absolute count
WBC	White blood cell count
WVF	White blood cell viability fraction

* discarded variables with less than five unique values.
CV: coefficient of variance


455007
A221038N
Ward

Chapter 4

The Impact of a Standardized Pre-visit Laboratory Testing Panel in the Internal Medicine Outpatient Clinic: a Controlled "On-Off" Trial

B.E.L. Vrijsen, M.J. ten Berg, C.A. Naaktgeboren, J.Y. Vis, H.M. Dijstelbloem, J. Westerink, D. Dekker, I.E. Hoefer, S. Haitjema, C.A.R. Hulsbergen-Veelken, W.W. van Solinge, H.A.H. Kaasjager

J Gen Intern Med 2021; 36(7): 1914-20

Abstract

Background

In several settings, a shorter time to diagnosis has been shown to lead to improved clinical outcomes. The implementation of rapid laboratory testing allows for pre-visit testing in the outpatient clinic, meaning that test results are available during the first outpatient visit.

Objective

To determine whether pre-visit laboratory testing leads to a shorter time to diagnosis in the general internal medicine outpatient clinic.

Design

An "on-off" trial, allocating subjects to one of two treatment arms in consecutive alternating blocks.

Participants

All new referrals to the internal medicine outpatient clinic of a university hospital were included, excluding second opinions. 595 patients were eligible; one person declined to participate, leaving data from 594 patients for analysis.

Intervention

In the intervention group, patients had standardized pre-visit laboratory testing before the first visit.

Main measures

The primary outcome was the time to diagnosis. Secondary outcomes were the correctness of the preliminary diagnosis on the first day, health care utilization and patient and physician satisfaction.

Key results

There was no difference in time to diagnosis between the two groups (median 35 days vs 35 days; hazard ratio 1.03 [0.87 – 1.22]; p = .71). The pre-visit testing group had higher proportions of both correct preliminary diagnoses on day 1 (24% vs 14%; p = .003) and diagnostic workups being completed on day 1 (10% vs 3%; p < .001). The intervention group had more laboratory tests done (50.0 [interquartile range (IQR) 39.0 – 69.0] vs 43.0 [IQR 31.0 – 68.5]; p < .001). Otherwise there were no differences between the groups.

Conclusions

Pre-visit testing did not lead to a shorter overall time to diagnosis. However, a greater proportion of patients had a correct diagnosis on the first day. Further studies should focus on customizing pre-visit laboratory panels, to improve their efficacy.

Introduction

A shorter time to diagnosis has been shown to lead to improved clinical outcomes in some, but not all health care settings.[1-4] Given that the majority of medical decisions rely on laboratory testing,[5] speeding up access to laboratory results may lead to earlier diagnoses. For instance, faster turnaround times of microbiology tests have been shown to lead to faster initiation of adequate antibiotic therapy and shorter inpatient length of stay.[6-7]

One way to speed up the diagnostic process is through pre-visit laboratory testing, in which patients have laboratory testing done directly prior to their doctor's appointment, and the tests are performed with a short turn-around time at a routine clinical laboratory. By necessity, pre-visit testing makes use of standardized laboratory test panels, as the patient has not been examined yet so the information from the history and physical examination is not yet available to guide the selection of laboratory tests. Standardized laboratory panels invariably include unnecessary tests, which potentially lead to downstream overutilization.[8]

In the Netherlands, pre-visit testing is not currently standard practice in the outpatient clinic, due to the often long turnaround times in the laboratory.

In 2014, the Central Diagnostic Laboratory of the University Medical Center Utrecht (UMC Utrecht) introduced rapid testing for a broad panel of routine laboratory tests for clinical chemistry, hematology, coagulation and endocrinology. This service guarantees that the results of these tests are available within 60 minutes after the sample arrives at the laboratory. This rapid testing enables the implementation of pre-visit testing in the outpatient clinic, making the test results available to the treating physician at the time of the visit, as opposed to usual care, in which the treating physician orders laboratory tests during the visit and the test results are only available afterwards, so the patient has to return for another visit.

Consequently, this "on-off" trial was set up to evaluate whether pre-visit laboratory testing benefits the patient and physician alike, shortening the time to diagnosis in newly referred outpatients, and to evaluate the downstream consequences of potential overutilization.

Methods

Trial design

This is a single-center controlled "on-off" trial, in which patients were alternately allocated to the intervention group or the usual care group in three-months blocks.[9] Subjects were not randomized individually, because it was not feasible to incorporate

the two different testing strategies in the laboratory order management system simultaneously.

In the intervention group, patients had laboratory tests performed directly prior to their first visit to the outpatient clinic, and in the usual care group laboratory tests were done afterwards.

Patient selection

All adult patients newly referred to the general internal medicine outpatient clinic of the University Medical Center Utrecht (UMC Utrecht), a large university hospital in Utrecht, the Netherlands, were eligible for inclusion. Referrals for second opinions were excluded because these patients often already have a diagnosis.

In the Netherlands, patients require a referral from their primary care physician before they visit an outpatient clinic. This referral consists of a letter with the reason of referral, sometimes accompanied by laboratory results, but not with the entire patient record due to privacy regulations. The general internal medicine outpatient clinic of the UMC Utrecht receives the primary care physicians' referrals electronically (and by regular mail). New referrals are triaged by an attending physician and are then randomly assigned to either a resident or an attending physician for an outpatient consultation. The residents are required to confer all consultations with their supervising attendings.

Based on an estimated sample size of 460 participants, four three-months blocks were initially planned from April 2015 to April 2017. However, because inclusion went slower than previously anticipated the inclusion period was extended to August 2017 by adding two two-months blocks.

Intervention

In the intervention group, patients were asked to have their blood drawn one hour before their scheduled appointment for their first visit, so that the treating physician had access to the test results during the visit. All patients in the intervention group received the same standard laboratory panel, regardless of the referral reason. This panel had been established before start of the trial by a panel of experienced internists and clinical chemists and comprised the following tests: hemoglobin, cell counts/differential, sodium, potassium, calcium, urea, creatinine, alkaline phosphatase, gamma-glutamyltransferase (GGT), glucose, aspartate transaminase (AST), alanine transaminase (ALT), lactate dehydrogenase (LDH), albumin, C-reactive protein (CRP), thyroid stimulating hormone (TSH), erythrocyte sedimentation rate (ESR), and a urine strip screening.

In the usual care group, laboratory tests were ordered at the discretion of the treating physician during the first visit and test results were only available afterwards.

In both groups all other aspects of medical care, such as imaging tests and the planning of follow-up visits, were at the discretion of the treating physician.

Outcomes and measures

BASELINE CHARACTERISTICS

For all patients, age at referral and gender were retrieved from the Utrecht Patient Oriented Database (UPOD), an infrastructure of relational databases comprising data on patient characteristics, hospital discharge diagnoses, medical procedures, medication orders and laboratory tests for all patients treated at UMC Utrecht. UPOD data acquisition and management is in accordance with current regulations concerning privacy and ethics. The structure and content of UPOD have been described in more detail elsewhere.[10] Furthermore, whether the physician performing the first consultation was a resident or an attending was obtained from the patient charts. Referral reasons and, if available, the results of laboratory testing performed by the referring physician were taken from the referral letter. Referral reasons were grouped into the following categories: anemia, fatigue, weight loss, gastro-intestinal complaints, abnormal laboratory test result(s) (other than anemia), lymphadenopathy/suspected malignancy, and other. These categories were non-exclusive as some patients had more than one referral reason.

PRIMARY OUTCOME

The primary outcome was the time to diagnosis, defined as the number of days between the first visit to the outpatient clinic and the final diagnosis being made. The date of final diagnosis was defined as the date the patient was informed of the diagnosis and after which no further tests or examinations were performed to confirm or refute this diagnosis. If a patient had more than one diagnosis, the date of the last diagnosis was used.

An expert panel of internal medicine physicians assessed the date of the final diagnosis. Assessment was planned two years after the patient's initial visit. Each case was individually assessed by two panelists and a third panelist was consulted when there was a disagreement. If two out of three panelists agreed on a date, this date was chosen. Cases where none of the panelists agreed on a date were resolved by consensus through discussion between the panelists. Inter-rater agreement was evaluated using the one-way random effects intra-class coefficient.[11]

SECONDARY OUTCOMES

Secondary outcomes were the correctness of the preliminary diagnosis at the first visit to the outpatient clinic, utilization of health care resources during the diagnostic process, and patient and physician satisfaction.

Correctness of the preliminary diagnosis

The correctness of the preliminary diagnosis on the first day was assessed using two parameters: firstly, the proportion of patients in whom the treating physician's preliminary diagnosis at the first visit agreed with the final diagnosis, and secondly, the proportion of patients in whom the diagnostic process was completed on the first day.

Utilization of health care resources

Health care utilization was measured by the number of medical procedures during the diagnostic process, including the pre-visit laboratory tests. Specifically, the number of outpatient clinic visits, clinical admissions, laboratory tests, venipunctures, imaging, and endoscopies were collected from the UPOD database. Furthermore, the physicians were asked to report the duration of the first consultation at the outpatient clinic, in minutes.

Patient and physician satisfaction

Patient satisfaction was assessed through a questionnaire, which was handed out to patients at their first visit to the outpatient clinic. Patients were asked about their preferences regarding laboratory testing strategies, as well as their satisfaction with their first visit to the outpatient clinic by an overall grade (1-10) and by using a modified Patient Satisfaction Questionnaire Short-Form (PSQ-18)[12] that excluded questions on financial consequences and accessibility of health care, as these items were not thought to be relevant in this setting.

Physicians were asked three yes-or-no questions regarding satisfaction with testing strategies after the first visit: whether they had a good overview of the patient's problem, whether they were able to help the patient efficiently, and whether the diagnosis was already in sight.

Statistical analyses

A sample of 200 patients per treatment arm was required to detect a seven day difference in the mean time to final diagnosis with a power of 80% and an alpha of .05, assuming a standard deviation of 25 days. To compensate for the non-normality

of the data, the required sample size was increased by 15%, yielding a required total sample size of 230 patients in each group.[13] All analyses were done according to the intention-to-treat principle.

Survival analysis was used to evaluate the primary outcome (time to diagnosis) to account for censoring due to loss to follow-up. A hazard ratio was calculated using Cox proportional hazards analysis. A Kaplan-Meier plot as well as medians and interquartile ranges for the time to diagnosis were also reported to aid with interpretation of the data.

Subgroup analyses of the primary outcome were performed on gender, referral reason, whether the referral letter contained the results of laboratory testing by the referring physician, and on whether the physician who treated the patient at the first visit to the outpatient clinic was a resident or attending physician.

The rates of correct preliminary diagnoses on the first day and diagnostic processes being completed on the first day were tested with Pearson's χ^2 test. The number of medical procedures in the diagnostic process was tested using negative binomial regression. Patient preferences were tested using the Wilcoxon-Mann-Whitney test and patient satisfaction was tested using the Student's t-test. Physician preferences were tested using Pearson's χ^2 test. The duration of the first visit was tested using the Student's t-test.

All statistical analyses were performed in R version 3.5.1.[14]

Ethical considerations

Because both laboratory testing strategies we investigated were already used in clinical practice, and no other burden was imposed on patients other than the questionnaire, the institutional review board waived the requirement for informed consent. Patients had the opportunity to opt out of the study. The study was registered in the Netherlands Trial Register, number NL5009.

Results

Baseline characteristics

In total, 595 patients were eligible for inclusion. One patient in the usual care group declined to participate, which left 594 patients for inclusion: 256 in the intervention group and 338 in the usual care group. 34 patients (13%) in the intervention group erroneously did not have pre-visit laboratory testing done. All 594 patients were included in the analyses. A flowchart of the inclusions is provided in Figure 1. Baseline characteristics are presented in Table 1. The mean time between the initial

visit and the expert panel's assessment was 713 days (95% confidence interval [CI] 701 – 726 days). Loss to follow-up was 6% after a median follow-up time of 138 days (interquartile range [IQR] 43 – 270).

FIGURE 1: INCLUSIONS FLOWCHART

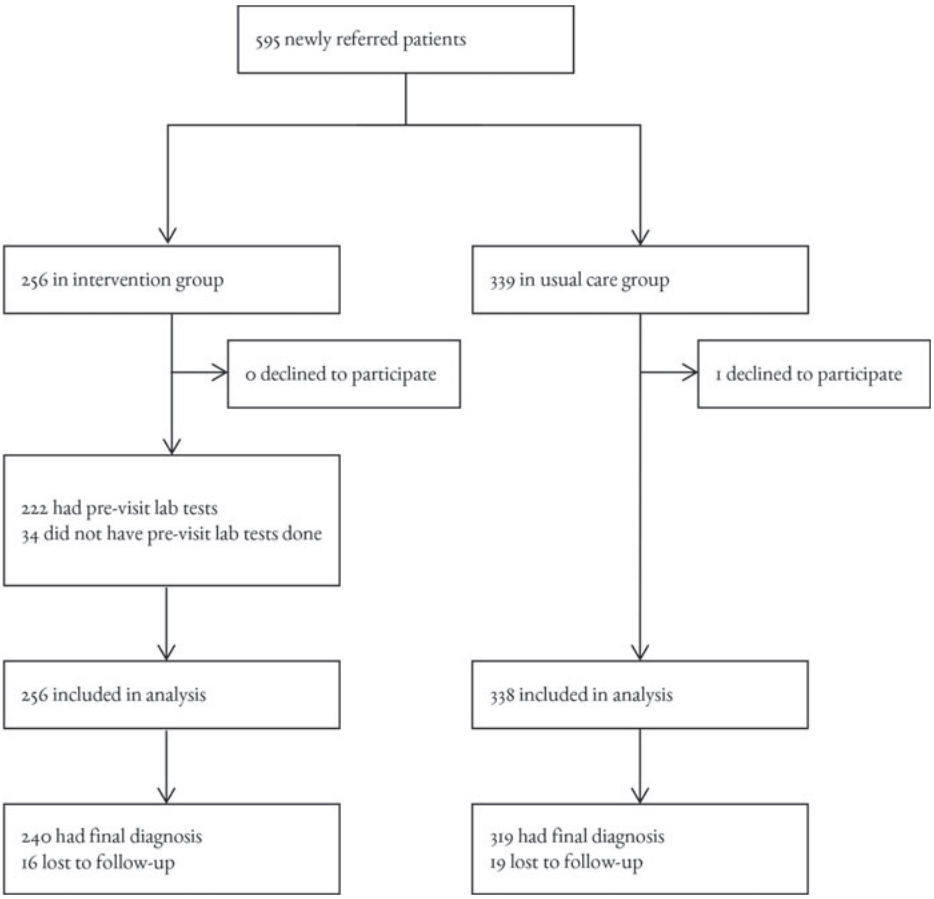


TABLE 1: BASELINE CHARACTERISTICS

	Pre-visit testing (n = 256)	Usual care (n = 338)
Female gender	145 (57%)	207 (62%)
Age	53.9 (51.7 – 56.1)	51.7 (49.8 – 53.7)
Referral reason (grouped)*		
- abnormal lab test	55 (21%)	45 (13%)
- anemia	28 (11%)	43 (13%)
- fatigue	64 (25%)	83 (25%)
- gastro-intestinal complaints	24 (9%)	41 (12%)
- lymphadenopathy / suspected malignancy	15 (6%)	17 (5%)
- weight loss	26 (10%)	42 (12%)
- other	65 (25%)	88 (26%)
Availability of pre-referral laboratory test results	121 (47%)	174 (51%)
Seen by attending physician	27 (11%)	50 (15%)

* categories are non-exclusive

Time to diagnosis

There was no difference in time to diagnosis (in days) between the two groups (Figure 2; hazard ratio 1.03 [0.87 – 1.22]; p = .71). A definitive diagnosis was made in 94% of patients. For these patients, median time to diagnosis was 35.0 days (IQR 14.0 – 77.3) in the intervention group and 35.0 days (IQR 14.0 – 83.0) in the control group. A list of the final diagnoses is provided in Supplementary Table 1.

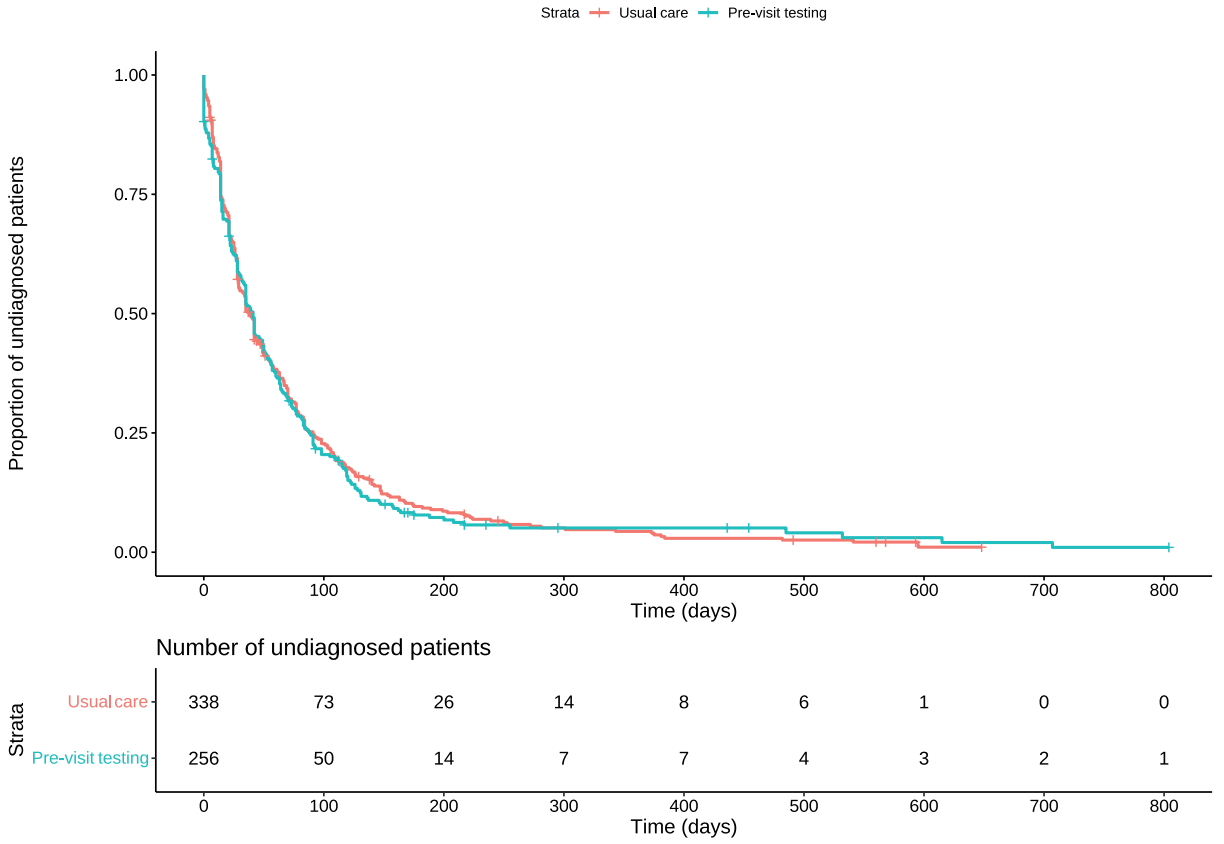
When establishing the time to diagnosis, the two experts agreed in 408 cases (69%). When a third reviewer was necessary to resolve discrepancies, there was agreement between two of the three experts in 551 cases (77%); the remaining 43 cases (23%) were resolved through discussion. The one-way random effects intra-class coefficient was 0.63, indicating moderate inter-rater agreement.

There were no differences in time to diagnosis between the intervention and control groups for subgroups based on gender, referral reason, the availability of the results of pre-referral laboratory testing by the referring physician, or whether the initial consultation was done by a resident or an attending (Supplementary Table 2).

Diagnosis after first visit

The rate of agreement between the treating physician's preliminary diagnosis on the first day and the definitive diagnosis at the end of the diagnostic process was 24% in the intervention group versus 14% in the control group (p = .003). The proportion of patients in whom the diagnostic process was completed on the first day was also significantly higher in the intervention group (10% vs 3%; p < .001).

Figure 2: kaplan-meier plot of time to diagnosis in days



Hazard ratio 1.03 (0.87 – 1.22); p = .71

Healthcare utilization

Patients in the intervention group had more laboratory tests done, both on the day of the first visit (median 22.0 [IQR 21.0 – 26.0] versus 20.0 [IQR 10.0 – 26.0]; p < .001) as during the entire diagnostic process (median 50.0 [IQR 39.0 – 69.0] versus 43.0 [IQR 31.0 – 68.5]; p = .001). There were no other differences in the number of medical procedures during the time to diagnosis. Notably, the total number of visits to the outpatient clinic was similar in both groups (2.0 [IQR 1.8 – 4.0] vs 2.0 [IQR 2.0 – 4.0]; p = .63). The first visit did last slightly longer in the intervention group (48.3 [95% CI 46.6 – 49.9] vs 45.0 [95% CI 43.3 – 46.7] minutes; p = .006). All data on health care utilization are presented in Table 2.

TABLE 2: HEALTH CARE UTILIZATION DURING THE DIAGNOSTIC PROCESS

	Pre-visit testing (n = 256)	Usual care (n = 338)	p-values
	medians (IQR)	medians (IQR)	
Visits to outpatient clinic			
- any	2.0 (1.8 – 4.0)	2.0 (2.0 – 4.0)	.63*
- internal medicine	2.0 (1.0 – 3.0)	2.0 (1.0 – 3.0)	.96*
Teleconsultations	1.0 (0.0 – 3.0)	1.0 (0.0 – 2.0)	.70*
Number of clinical admissions	0.0 (0.0 – 1.0)	0.0 (0.0 – 1.0)	.37*
Clinical admission days	0.0 (0.0 – 1.0)	0.0 (0.0 – 1.0)	.42*
Laboratory tests (total)	50.0 (39.0 – 69.0)	43.0 (31.0 – 68.5)	.001*
Laboratory tests (first day)	22.0 (21.0 – 26.0)	20.0 (10.0 – 26.0)	< .001*
Laboratory test orders	3.5 (2.0 – 5.3)	3.0 (2.0 – 5.0)	.11*
Imaging tests	1.0 (0.0 – 3.0)	1.0 (0.0 – 2.0)	.46*
Number of patients with imaging tests	n (%)	n (%)	
- any imaging	180 (70%)	222 (66%)	.27†
- MRI	20 (8%)	21 (6%)	.55†
- CT	62 (24%)	68 (20%)	.27†
- ultrasound	84 (33%)	108 (33%)	.89†
- nuclear	10 (7%)	18 (5%)	.38†
- X-ray	111 (43%)	144 (43%)	.92†
- endoscopy	34 (13%)	57 (17%)	.28†

* negative binomial regression

† Pearson's χ^2 test

CT: computed tomography; IQR: interquartile range; MRI: magnetic resonance imaging

Additionally, a post-hoc analysis to assess the adequacy of the standard laboratory panel showed that in 66% of subjects in the intervention group additional laboratory tests were ordered on the first day (Supplementary Table 3).

Satisfaction

There was no difference in physician and patient satisfaction between the two groups (Tables 3 and 4). However, there were some differences in patients' preferences (Supplementary Figure 1): patients who had had pre-visit laboratory tests done were more likely to want to learn the diagnosis on the same day (93% vs 91%; $p < .001$), and less likely to want to see the doctor before having laboratory tests done (17% vs 43%; $p < .001$). However, they were more likely to object to more laboratory tests being done than necessary (16% vs 8%; $p = .01$).

Table 3: PHYSICIAN SATISFACTION

	Pre-visit testing (n = 191) (% of respondents answering affirmatively)	Usual care (n = 161) (% of respondents answering affirmatively)	p-value
Good overview of the problem?	98%	98%	> .99
Able to help the patient efficiently?	65%	71%	.28
Diagnosis already in sight?	57%	50%	.24

Differences between the groups were tested using Pearson's χ^2 test

Table 4: PATIENT SATISFACTION

Respondents, n (%)	Pre-visit testing (n = 125 [49%]) (means + 95% CI)	Usual care (n = 131 [39%]) (means + 95% CI)	p-value
Overall grade	8.0 (7.8 – 8.2)	8.1 (7.9 – 8.3)	.54
Modified PSQ-18 questionnaire			
General satisfaction	3.85 (3.71 – 3.98)	3.81 (3.66 – 3.96)	.72
Technical quality	3.86 (3.75 – 3.97)	3.73 (3.62 – 3.84)	.11
Interpersonal manner	4.35 (4.02 – 4.26)	4.20 (4.07 – 4.33)	.07
Communication	4.14 (4.02 – 4.26)	4.00 (3.87 – 4.12)	.10
Time spent with doctor	4.02 (3.90 – 4.15)	3.84 (3.70 – 3.98)	.05

Differences between the groups were tested using Student's t-test
Items of the PSQ-18 are scored on a 1 – 5 scale, with high scores reflecting greater satisfaction with medical care
CI: confidence interval; PSQ-18: Patient Satisfaction Questionnaire Short-Form

Discussion

In our single-center on-off study, performing a standardized laboratory test panel prior to the first visit to the outpatient clinic did not result in a shorter overall time to diagnosis. However, it did increase the chance of obtaining a final diagnosis during the first visit. The number needed to test in order to finish the diagnostic process on the first day was 15. In our questionnaire, the vast majority of patients preferred receiving their diagnosis on the first day, and at the same time did not object to having more laboratory tests done than necessary.

Ordering standardized laboratory panels is in sharp contrast with advice from several guidelines, including Choosing Wisely recommendations, because of potential overutilization.[15] Overutilization leads to increased costs, as well as potentially more false positive test results.[16]

In this study, patients in the intervention group on average had 7 more laboratory tests performed during the diagnostic process. No other differences in health care utilization were found, which implies that the excess laboratory tests did not lead to significant downstream overutilization.

To the best of our knowledge, this is the first time that the effect of pre-visit laboratory testing on the diagnostic process in the outpatient setting has been studied.

Several possible explanations for the study's negative result can be proposed. First of all, laboratory testing may not be as important for establishing a diagnosis as previously hypothesized. This might be especially true in the setting of a tertiary hospital, which typically has a more complex case mix that requires a more extensive diagnostic work up. This might also explain why pre-visit testing would have an effect on the number of correct diagnoses on the first day, as in these cases typically no additional diagnostic tests were performed. It might also be argued that laboratory testing already performed by the referring physician negated the effect of pre-visit testing, although in that case one would have expected to see an effect in the subgroup without pre-referral testing.

Alternatively, given that additional tests were ordered after the first visit in 66% of cases in the pre-visit arm, it might also be argued that the pre-visit panel proved inadequate in those patients, and that a different or more extensive test panel would have made a difference. However, extending the pre-visit panel would lead to greater costs and possibly more downstream overutilization. Tailoring the pre-visit panel to individual patients based on their referral reason might be a more promising alternative.

One of the strengths of this study is that all subjects were comprehensively analyzed through chart review by an expert panel. As a result, loss to follow-up was

limited at 6%. Furthermore, the study population was relatively unselected, as all newly referred patients were included, which increases the study's external validity.

This study has several limitations. Firstly, the on-off trial design may have compromised the comparability of the two groups.

Secondly, this was an open label study. The subjects and their treating physicians were aware of their allocation. The expert panelists who determined the time to diagnosis were not actively informed about the allocation of the subjects they evaluated, but in many cases it could be inferred from the treating physician's chart notes they reviewed.

Thirdly, the response rate of the patient survey was quite low at 49% of patients in the intervention group and 39% in the control group, which could limit the generalizability of the questionnaire's results.

Fourthly, the physician survey comprised only three questions, because we presumed that physicians would not be willing to fill out longer surveys. This may limit the survey's applicability.

Finally, even though we found no differences in health care utilization apart for a modest increase in the number of laboratory tests in the intervention group, not all potentially negative effects of test overutilization, such as anxiety due to false positive test results, were monitored for in this study.

In conclusion, standardized pre-visit laboratory testing did not lead to a shorter time to diagnosis but increased the chance of obtaining the correct diagnosis during the first visit. Further studies should focus on adaptations and differentiations to the standard pre-visit laboratory panel.

References

1. Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer*. 2015;112(1):S92-S107.
2. Nell VPK, Machold KP, Eberl G, Stamm TA, Uffmann M, Smolen JS. Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. *Rheumatology (Oxford)*. 2004;43(7):906-14.
3. Akers IE, Weber R, Sax H, Böni J, Trkola A, Kuster SP. Influence of time to diagnosis of severe influenza on antibiotic use, length of stay, isolation precautions, and mortality: a retrospective study. *Influenza Other Respir Viruses*. 2017;11(4):337-44.
4. Nguyen VQ, Jiang D, Hoffman SN, Guntaka S, Mays JL, Wang A, et al. Impact of Diagnostic Delay and Associated Factors on Clinical Outcomes in a U.S. Inflammatory Bowel Disease Cohort. *Inflamm Bowel Dis*. 2017;23(10):1825-31.
5. Rohr U-P, Binder C, Dieterle T, Giusti F, Messina CGM, Toerien E, et al. The Value of In Vitro Diagnostic Testing in Medical Practice: A Status Report. *PLoS One*. 2016;11(3):e0149856.
6. Trenholme GM, Kaplan RL, Karakusis PH, Stine T, Fuhrer J, Landau W, et al. Clinical impact of rapid identification and susceptibility testing of bacterial blood culture isolates. *J Clin Microbiol*. 1989;27(6):1342-5.
7. Rogers BB, Shankar P, Jerris RC, Kotzbauer D, Anderson EJ, Watson JR, et al. Impact of a Rapid Respiratory Panel Test on Patient Outcomes. *Arch Pathol Lab Med*. 2014;139(5):636-41.
8. Janssens PMW, Staring W, Winkelman K, Krist G. Active intervention in hospital test request panels pays. *Clin Chem Lab Med*. 2015. p. 731.
9. Mathe N, Johnson ST, Wozniak LA, Majumdar SR, Johnson JA. Alternation as a form of allocation for quality improvement studies in primary healthcare settings: the on-off study design. *Trials*. 2015;16(1):375.
10. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-9.
11. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-63.
12. Marshall GN, Hays RD, Rand Corporation. The Patient Satisfaction Questionnaire short-form (PSQ-18). Santa Monica, Calif.: *RAND*; 1994.
13. Lehmann EL, D'Abrera HJM. Nonparametrics: Statistical Methods Based on Ranks: *Prentice Hall*; 1998.
14. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
15. Federatie Medisch Specialisten. Overzicht Verstandige Keuzes 2019 [19-4-2020]. Available from: https://www.demedischspecialist.nl/sites/default/files/FMS%20Overzicht%20Verstandige%20Keuzes.pdf.
16. Korenstein D, Chimonas S, Barrow B, Keyhani S, Troy A, Lipitz-Snyderman A. Development of a Conceptual Map of Negative Consequences for Patients of Overuse of Medical Tests and Treatments. *JAMA Intern Med*. 2018;178(10):1401-7.

Supplementary materials

Supplementary Table 1: list of final diagnoses
Supplementary Table 2: time to diagnosis in subgroups
Supplementary Table 3: laboratory tests performed in addition to the standard pre-visit panel
Supplementary Figure 1: patient preferences

Supplementary Table 1: list of final diagnoses

Diagnosis	Patients (n)	Completed diagnostic workups on the 1st day (n)
CARDIOVASCULAR		
Atrial fibrillation	1	
Coronary artery disease	1	
Hypercholesterolemia	2	1
Hypertension	1	
Pericarditis	1	
Peripheral arterial occlusive disease	3	
Venous insufficiency	3	
Venous thromboembolism	2	2
ELECTROLYTE DISTURBANCES		
Hypomagnesemia, due to proton pump inhibitors	1	1
Hyponatremia, due to diuretics	1	
Hyponatremia, due to reset osmostat	1	
Hyponatremia, due to SIADH	9	5
Hyponatremia, unspecified	1	
Hypophosphatemia, due to seizures	1	
ENDOCRINE		
Diabetes mellitus	3	3
Hyperthyroidism	4	2
Hypogonadism	1	
Hypopituitarism, due to pituitary mass	1	
Hypothyroidism	9	4
Osteoporosis	2	
Tertiary adrenal insufficiency	1	
EXTERNAL CAUSES		
Nickle intoxication	1	
Side effects of medication	17	7
GASTROINTESTINAL		
Bowel obstruction	1	
Celiac disease	2	1
Constipation	6	1
Diverticulitis	1	
Duodenic ulcer	1	
Gastritis	6	1
Hernia diafragmatica	2	
Inflammatory bowel disease	1	
Irritable bowel syndrome	15	1
Reflux esophagitis	8	2

Supplementary Table 1: continued.

Diagnosis	Patients (n)	Completed diagnostic workups on the 1st day (n)
HEMATOLOGICAL		
Anemia, due to hemoglobinopathy	1	
Anemia, due to medication	1	
Anemia, hemolytic	2	1
Anemia, of chronic disease	5	2
Anemia, secondary to chronic kidney disease	4	1
Anemia, unspecified	1	
Chronic lymfatic leukemia	1	
Iron deficiency, of gastrointestinal causes	18	
Iron deficiency, of other causes	19	6
Iron deficiency, unspecified	22	1
Malignant lymphoma	4	
MGUS	4	
Myeloproliferative neoplasm	2	
Reactive lymphadenopathy	2	
Secondary leukocytosis	1	
Secondary polycythemia	8	3
Von Willebrand's disease	1	
HEPATOBILIARY		
Cholecystitis	1	1
Cholelithiasis	2	
Hemangioma of the liver	1	
Liver disease, alcoholic	4	
Liver disease, due to heart failure	2	1
Liver disease, fatty liver disease	1	1
Liver disease, unspecified	1	
Pancreatitis	2	
Primary biliary cirrhosis	1	

Supplementary Table 1: continued.

Diagnosis	Patients (n)	Completed diagnostic workups on the 1st day (n)
INFECTIOUS		
Bartonellosis	2	
Chlamydia infection	1	
Endocarditis	2	
Gastroenteritis	6	1
Lyme's disease	2	1
Pelvic abscess post-prostatectomy	1	
Recurring skin infections	2	1
Respiratory tract infection	4	
Sepsis	1	
Spondylodiscitis	1	
Strongyloides infection	1	
Tuberculosis	1	
Urinary tract infection	8	
Viral infection	9	2
MALIGNANCY		
Breast cancer	3	
Cholangiocarcinoma	1	
Colorectal cancer	3	
Gastric cancer	1	1
Gastrointestinal stromal tumor	1	
Lung cancer	3	
Melanoma	1	
Neuro-endocrine tumor	1	
Cancer, not specified	2	2
Ovarian cancer	3	
Pancreatic cancer	3	1
NEUROLOGICAL		
ACNES	3	1
Autonomic dysregulation	1	
Benign paroxysmal positional vertigo	1	
Essential tremor	1	1
Parkinson's disease	2	1
Polyneuropathy due to spinal stenosis	1	
PSYCHIATRIC		
Alcohol abuse	6	1
Depression	1	
Eating disorder	2	1
Panic attacks	2	1

Supplementary Table 1: continued.

Diagnosis	Patients (n)	Completed diagnostic workups on the 1st day (n)
PULMONARY		
Bronchiectasia	1	1
COPD	1	1
OSAS	1	
Pleural lipoma	1	
Sarcoidosis	2	
RENAL		
Acute kidney injury	5	2
Chronic kidney disease	8	1
Exercise-induced hematuria	1	
Hydronephrosis	1	
Kidney stone	3	
RHEUMATOLOGICAL / IMMUNOLOGICAL		
Axial spondyloarthritis	1	
Anaphylaxis	1	
Bechterew's disease	1	
Cholinergic urticaria	1	
Dermatomyositis	1	
Facioscapulohumeral dystrophia type 1	1	
Familial Mediterranean fever	1	
Fibromyalgia	6	1
Giant cell arteritis	7	1
Leukocytoclastic vasculitis	2	1
Osteoarthritis	4	2
PFAPA syndrome	1	
Raynaud's disease	1	
SADNI	1	
Sjögren's disease	1	
Urticarial dermatitis	1	
VITAMIN / TRACE ELEMENT DEFICIENCIES		
Folate deficiency	6	1
Vitamin B12 deficiency	7	1
Vitamin D deficiency	3	
Zinc deficiency	1	

Supplementary Table 1: continued.

Diagnosis	Patients (n)	Completed diagnostic workups on the 1st day (n)
OTHER		
Chronic fatigue syndrome	21	6
Condition ruled out	35	8
Fatigue, due to cardiopulmonary disease	1	1
Fatigue post radiotherapy	1	
Fatigue post stroke	1	1
Fibroadenoma of the breast	1	
Medically unexplained physical symptoms	86	8
Menstrual complaints	1	
No evidence of disease	65	12
Obesity	3	2
Perimenopausal complaints	1	1
Pharyngeal pain after intubation	1	
Polydipsia	1	1
Post intensive care syndrome	1	1
Post menopausal flushing	1	
Rhabdomyolysis	2	
Vaginwl prolaps	1	1
Weight loss due to decreased caloric intake	3	1
Werner's progeria	1	1

ACNES: abdominal cutaneous nerve entrapment syndrome; BPPV: benign paroxysmal postural vertigo; COPD: chronic obstructive pulmonary disease; MGUS: monoclonal gammopathy of unknown significance; OSAS: obstructive sleep apnea syndrome; PFAPA: periodic fever, aphthosis, pharyngitis, and adenitis; SADNI: selective antibody deficiency with normal immunoglobulins; SIADH: syndrome of inappropriate antidiuretic hormone secretion
45 patients had more than one diagnosis.

Supplementary Table 2: time to diagnosis in subgroups

	Number	Median time to diagnosis (+ IQR)*		Hazard ratio	p-value†
		Pre-visit testing	Usual care		
	594	35.0 (14.0 – 77.3)	35.0 (14.0 – 83.0)	1.03 (0.87 – 1.22)	.71
Gender					
Female		35.0 (14.0 – 83.3)	35.0 (14.0 – 83.5)	0.95 (0.77 – 1.19)	.68
Male		36.5 (14.0 – 68.0)	35.0 (14.0 – 81.5)	1.16 (0.89 – 1.51)	.26
Referral reason‡					
Abnormal lab test	100	28.5 (1.0 – 57.0)	21.0 (7.0 – 78.8)	1.40 (0.92 – 2.13)	.12
Anemia	71	53.0 (20.0 – 105.0)	55.0 (23.5 – 116.0)	0.85 (0.50 – 1.43)	.54
Fatigue	147	42.0 (19.5 – 81.5)	42.0 (20.3 – 73.0)	0.93 (0.67 – 1.31)	.69
Gastro-intestinal complaints	65	43.0 (15.5 – 67.5)	49.0 (26.3 – 90.8)	1.38 (0.82 – 2.34)	.23
Lymphadenopathy / suspected malignancy	32	22.0 (14.0 – 66.5)	27.5 (8.8 – 46.3)	0.83 (0.40 – 1.72)	.61
Weight loss	68	41.0 (16.5 – 84.5)	35.0 (16.0 – 112.0)	1.67 (0.98 – 2.84)	.06
Other	153	34.5 (14.8 – 74.3)	28.0 (14.0 – 77.3)	0.77 (0.55 – 1.08)	.13
Availability of the results of pre-referral laboratory testing by referring physician					
Yes	299	42.0 (14.0 – 88.0)	32.5 (14.0 – 77.0)	0.89 (0.70 – 1.13)	.32
No	295	35.0 (14.0 – 82.5)	40.5 (14.0 – 101.5)	1.19 (0.94 – 1.51)	.15
Patient seen by					
Resident	517	35.0 (14.0 – 73.0)	34.0 (14.0 – 79.0)	1.06 (0.65 – 1.75)	.81
Attending physician	77	63.0 (33.5 – 95.0)	52.5 (19.3 – 94.8)	1.00 (0.84 – 1.20)	.97

* for completed cases

† p-values from Cox regression

‡ categories are non-exclusive

IQR: interquartile range

Supplementary Table 3: laboratory tests performed on the first day in addition to the standard pre-visit panel

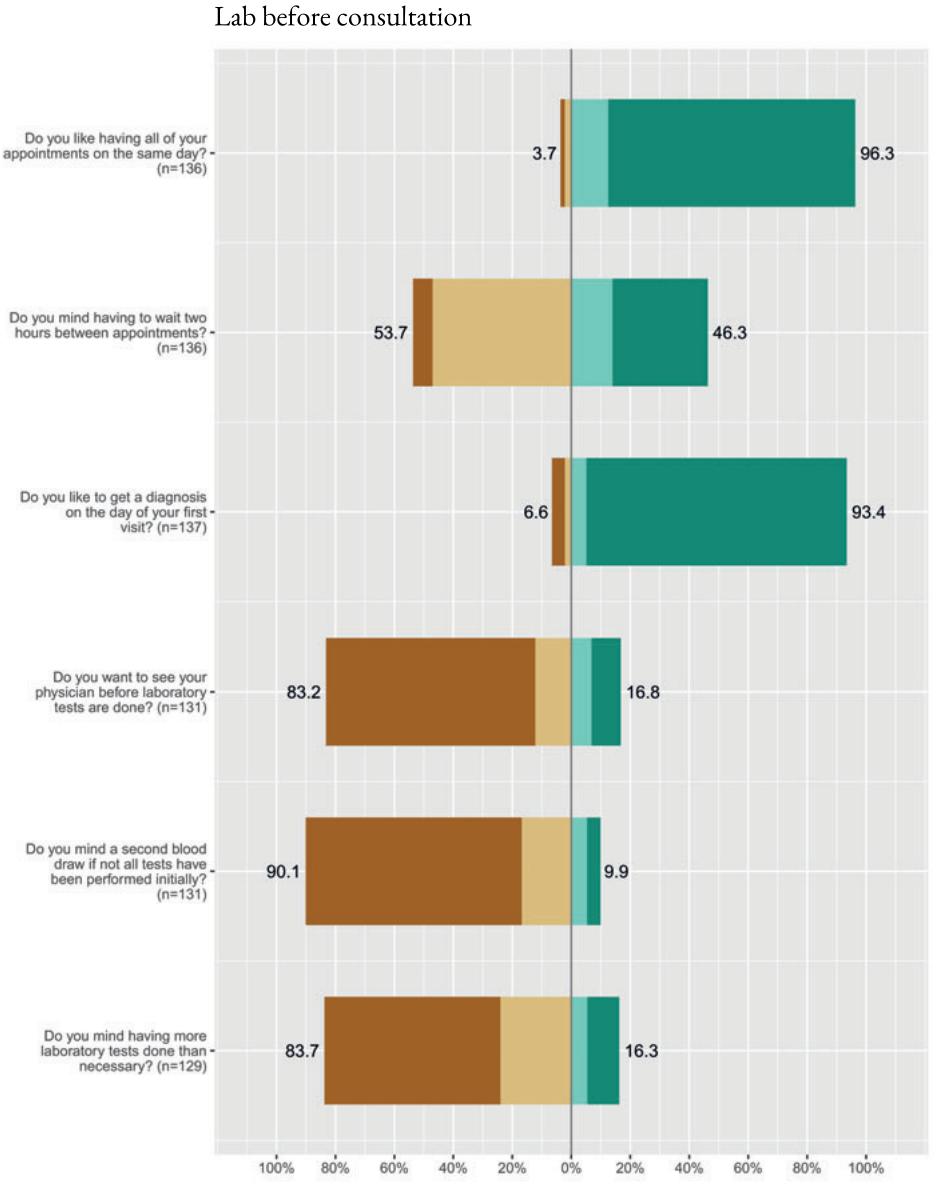
Additional test	number of times performed
Ferritin	70
Vitamin B12	46
Folic acid	42
Free thyroxine (FT4)	30
Creatine kinase	28
Phosphate	22
Bilirubin (total)	19
M-protein	19
Sodium	17
Total protein	17
Iron	15
25-hydroxyvitamin D	14
Osmolality	14
Creatinine (urine)	13
Osmolality (urine)	13
Reticulocytes	13
Magnesium	12
Transferrin	10
Transferrin saturation	10
Glycated hemoglobin (HbA1c)	9
Anti-endomysium	8
Anti-nuclear antibodies	8
Anti-TTG	8
Cholesterol	7
HDL-cholesterol	7
LDL-cholesterol	7
Triglycerides	7
Bicarbonate	6
Haptoglobin	6
Immunoglobulin A	6
Prothrombin time	6
Anti-ds-DNA	5
Cortisol	5
Immunoglobin G	5
Lipase	5
Potassium	5
Urea	5
aPTT	4

Supplementary Table 3: continued.

Additional test	number of times performed
Albumin (urine)	4
Anti-smooth muscle antibodies	4
Erythropoietin	4
INR	4
Rheumatoid factor	4
Amylase	3
ACE	3
Anti-mitochondrial antibodies	3
Citrullin	3
Immunoglobulin M	3
Lactate	3
Parathyroid hormone	3
Red cell distribution width	3
Reticulocytes MCHC	3
Thyroid-stimulating immunoglobulin	3
ACTH	2
Anti-TPO antibodies	2
BNP	2
Complement C3	2
CANCA	2
Dysmorphic erythrocytes (urine)	2
Free light chains	2
Insulin	2
Intrinsic factor	2
PANCA	2
PSA	2
Testosterone	2
Thrombocytes (citrate)	2
Thrombocytes (heparin)	2
Type and screen	2
Uric acid	2
Vitamin B6	2

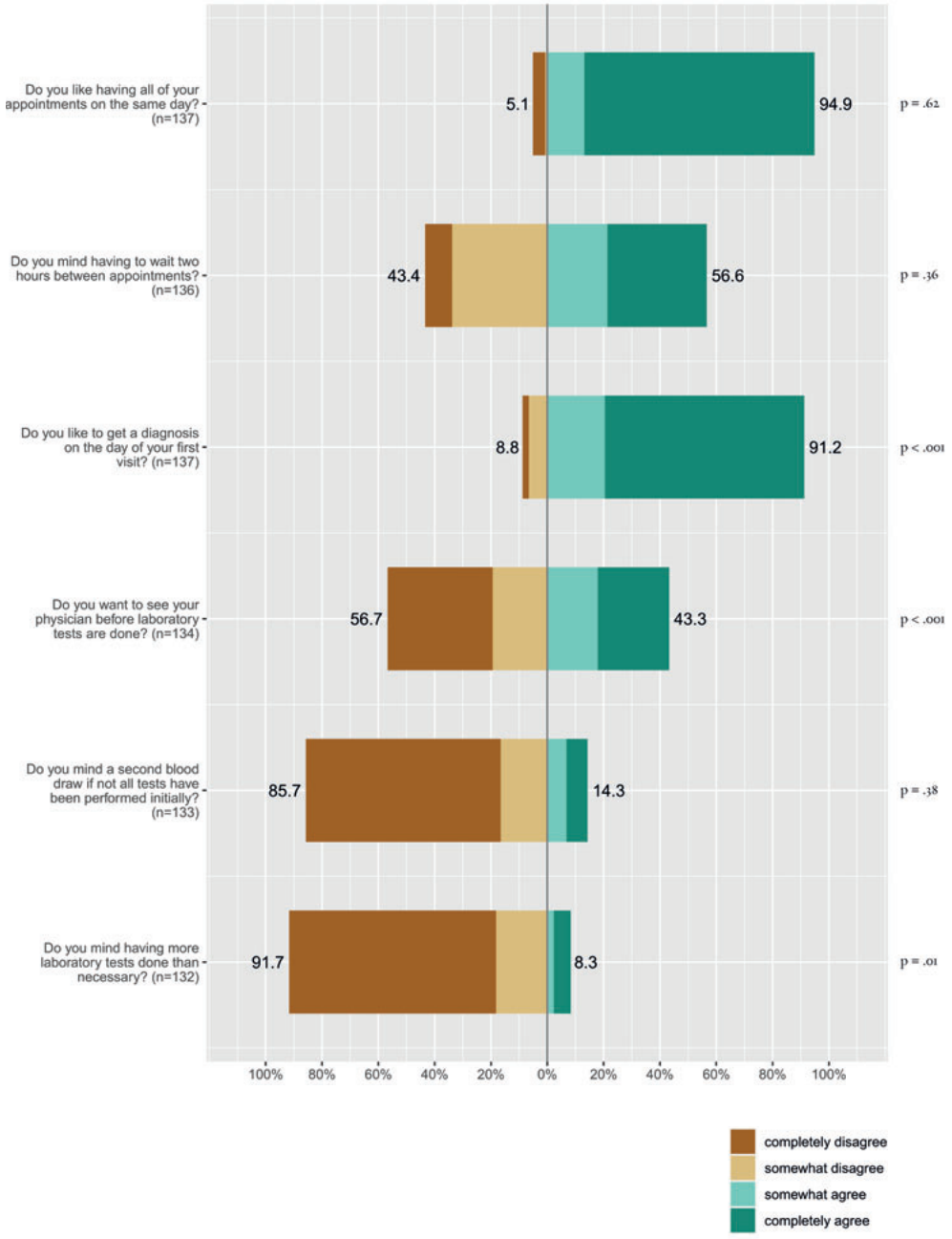
Total group size is 222 subjects. Only tests ordered more than once are included.ACE: angiotensin converting enzyme; ACTH: adrenocorticotropic hormone; aPTT: activated partial thromboplastin time; BNP: B-type natriuretic peptide; cANCA: cytoplasmic anti-neutrophil cytoplasmic antibodies; DNA: deoxyribonucleic acid; HDL: high density lipoprotein; INR: international normalized ratio; LDL: low density lipoprotein; MCHC: mean corpuscular hemoglobin concentration pANCA: perinuclear anti-neutrophil cytoplasmic antibodies; PSA: prostate specific antigen; TPO: thyroid peroxidase; TTG: tissue transglutaminase

Supplementary Figure 1: patient preferences



Comparisons between groups were tested by Mann-Whitney-Wilcoxon test

Lab after consultation




LOT 2227155

Chapter 5

Outcomes of second opinions in general internal medicine

P.M. Burger, J. Westerink, B.E.L. Vrijsen

PLoS One 2020; 15(6): e0236048

Abstract

Background

To date, the outcomes of second opinions in internal medicine in terms of diagnostic yield and patient benefit have not been studied extensively. This retrospective study explores the outcomes of second opinions at a general internal medicine outpatient clinic in an academic hospital.

Methods

A register of all patients referred to the general internal medicine outpatient clinic of the University Medical Center in Utrecht for a second opinion, was kept. All 173 patients referred between June 2016 and August 2018 were selected. Case records were analyzed for patient characteristics, referring doctor, chief complaint, performed investigations, follow-up time and, established diagnosis, additional diagnoses, initiated treatment and reported benefit.

Results

A new diagnosis was established in 13% of all patients. A new treatment was initiated in 56% of all patients: 91% and 51% of patients with and without a new diagnosis respectively ($p < .001$). Of all patients, 19% received an effective treatment (52% vs 14% of patients with vs without a new diagnosis; $p < .001$). Regardless of treatment, resolution or improvement of the chief complaint was achieved in 28% of all patients (52% vs 25% of patients with vs without a new diagnosis; $p = .006$). Regarding diagnostics, 23 – 33% of radiology, endoscopy and pathology tests performed during second opinion were a repetition of previously conducted investigations. Conventional blood tests were a repetition in 89% of cases. Median time to diagnosis was 64 days (interquartile range [IQR]: 25 – 128) and median time to discharge was 75 days (IQR: 31 – 144).

Conclusion

Second opinions in general internal medicine lead to the establishment of a new diagnosis in a small proportion of patients. However, the value of second opinions may not be limited to the establishment of diagnoses, as new treatments are often initiated and overall patients report improved symptomatology in 28% of cases.

Introduction

A second opinion is defined as a reevaluation of the diagnosis and/or treatment given by a doctor, carried out by a second, independent doctor from the same medical field.[1,2] Patients request second opinions for various reasons.[3-6] Mostly, second opinions are requested when no explanation for the patients' complaints is found by the original doctor, or when treatment is ineffective.

Over the years, many studies in various medical specialties and multiple countries have shown that second opinions lead to the establishment of a new diagnosis in 2-60%[7-30] and a change in treatment in 20-60% of patients.[8-19,22,24-27,31] In addition, studies have shown that patients are generally satisfied with the process, even if it has not led to a new diagnosis or treatment.[7,9,12,32] Studies exploring the outcome of second opinions in general internal medicine have shown that a new diagnosis is established in only approximately 10% of patients.[7,8] However, to date, only a limited number of observational studies have been carried out in this field. Moreover, previous studies did not evaluate the establishment of additional diagnoses, treatment initiation and effects, patient-reported symptomatology and relevance of performed investigations in the context of second opinions, in all patients.[7,8]

This raises the question as to what the actual outcomes of second opinions in internal medicine are, when studied extensively. Therefore, the aim of this study is to determine the outcomes of second opinions in a general internal medicine outpatient clinic in an academic hospital. Primarily, this study will assess in how many patients a new diagnosis was established during second opinion. Secondarily, this study will assess established additional diagnoses, initiated treatment and its effects, patient-reported symptomatology, relevance of (repeated) diagnostic investigations and time to diagnosis and time spent in the clinic during second opinion.

Methods

Second opinions in the Dutch health care system

In the Dutch health care system, for every medical issue, a patient's initial consultation is always with a general practitioner. The general practitioner acts as a gatekeeper to hospital and specialist care. The general practitioner can decide to refer patients to a hospital for specialist care. This can be either a regular hospital or an academic hospital. After patients have received specialist care in a regular or academic hospital, they have a legal right to demand a second opinion, and second opinions are covered by basic insurance.[2] If patients demand a second opinion, they are then referred to another hospital by their original physician or their general practitioner. Again, this

can be either a regular or an academic hospital. Physicians from regular and academic hospitals have similar levels of expertise, and have similar diagnostic resources at their disposal. So, second opinions are carried out on the same level of care as the first opinions.

Study Design

This study is a retrospective analysis of data retrieved from case records, stored in the electronic hospital information system at the University Medical Center in Utrecht (UMC Utrecht), an academic hospital in the Netherlands. Due to the retrospective and non-invasive nature of the study, it was not subject to the Dutch Medical Research Involving Human Subjects Act and formal consent was not required. The study was approved by the Medical Ethics Review Committee in the UMC Utrecht before data acquisition. Starting from June 2016, a register of all patients referred to the general internal medicine outpatient clinic of the UMC Utrecht for a second opinion, has been kept for administrative reasons. All patients referred for a second opinion between June 2016 and August 2018 were considered for this study, so there would be at least eight months between time of referral and the start of this study (1 May 2019). Patients who did not visit the clinic or visited the clinic of a different medical specialty were excluded. For all included patients, age at time of referral, gender and the following measures were collected from case records by the first researcher (PB).

A glossary of terms used throughout the manuscript is provided in Supplementary Table 1.

Referral

Case records were screened for referring doctor, dates of last consultation with the previous physician and first consultation with the physician formulating the second opinion, and chief complaint. Referring doctor was based on the referral letter, and divided into three groups: general practitioner, locum general practitioner and medical specialist. Time between consultations was calculated. This was defined as the number of days between the last consultation with a previous physician (which was based on dates specified in the referral letter) and the first consultation with the physician formulating the second opinion. Chief complaint of the patient was based on the main complaint mentioned by the patient during the first visit to the clinic, as documented in the case record by the doctor formulating the second opinion. Chief complaints were divided into the following groups: fatigue, abdominal pain, pain (multifocal), weight loss, edema, fever, and other, based upon observed frequencies.

Diagnosis

Case records were analyzed for diagnosis at time of referral, diagnosis by the doctor formulating the second opinion, diagnosis established during inter-collegial consultation (consultation by a doctor from another medical specialty, requested by the doctor formulating the second opinion) and additional diagnoses established in the context of the second opinion. Whether these types of diagnoses were established and the actual diagnoses were noted for the different types of diagnoses separately. Diagnoses were only included if they were considered definitive diagnoses, using the following definition: a diagnosis is said to be a definitive diagnosis when the treating physician concludes that the diagnosis has been established and that no further investigations to confirm this diagnosis are required. The diagnosis by the referring doctor was based on the referral letter. If a diagnosis was established by the doctor formulating the second opinion, it was documented whether it was a new diagnosis: a new diagnosis was defined as the establishment of a diagnosis different from the diagnosis at the time of referral, or the establishment of a diagnosis in patients without a diagnosis at the time of referral. The same was done for diagnoses established during inter-collegial consultation. Inter-collegial consultation was seen as a part of second opinions, and therefore, diagnoses established during inter-collegial consultation were added to diagnoses established by the doctors formulating the second opinions, when analyzing outcome of second opinions. Finally, for additional diagnoses established during second opinion, the relevance was determined. An additional diagnosis was considered relevant only if it led to treatment for this diagnosis.

Treatment and Patient-reported Symptomatology

Records were analyzed for treatment initiated by the doctor formulating the second opinion and changes made to preexisting management plans, and their effects on the chief complaint. Treatment was divided into the following groups: newly prescribed medication, change in medication (dosage) used at the time of referral, vitamin/iron supplementation (vitamin B11/ B12/D and iron), analgesia (local anesthetics or transcutaneous electric nerve stimulation), physical therapy, cognitive behavioural therapy (CBT), change in diet, surgery (for example gastroenterological surgery) and other (for example radiotherapy). Treatment effects were based on patient opinion as documented by the doctor in the case record, and were divided into four groups: resolution, improvement, unchanged and worsened. An effective treatment was defined as a treatment leading to the resolution or improvement of the chief complaint. In a similar way, we also analyzed case records of all patients for patient-

reported symptomatology at the end of second opinions. The same four groups were used to define the outcome.

Investigations

Investigations performed in the context of the second opinion were collected from case records: blood tests, urinalysis, microbiology tests, radiological tests, endoscopic procedures and pathology tests. Laboratory tests were divided into conventional blood tests (specified in Supplementary Table 2) and additional blood tests. Radiological tests were divided into X-ray, ultrasonography, CT, MRI and PET/SPECT. Microbiology tests were divided into the following groups: viral, bacterial and other (parasites, fungi, protozoa). For every investigation it was noted whether it was a new investigation or a repetition of a previous investigation. An investigation was considered a repeated investigation if the investigation had already been performed by a previous physician before the start of the second opinion and the exact same investigation was then performed again during the second opinion. If the results of previous investigations, or images or tissue obtained by radiology or pathology tests were transferred from a hospital of a previous physician to the UMC Utrecht, and were reassessed by a physician of the UMC Utrecht during the second opinion, this was not considered an investigation or a repeated investigation as the actual investigation was not performed during second opinion. For every investigation was also noted whether it led to any form of relevant information. Relevant information of any form was defined as information not known from previous investigations leading to either the establishment of a diagnosis or additional diagnosis, the initiation of a new treatment or the requirement for another investigation for further assessment. Finally, for every investigation it was noted whether it had shown anomalous results contributing to the establishment of a diagnosis.

Follow-up

For each patient it was noted whether the entire diagnostic process was completed, or the diagnostic process was still ongoing or the patient was lost to follow-up. Time to diagnosis, time to discharge from the clinic and time spent in the clinic were collected from case records. Time to diagnosis was defined as the number of days between the first visit to the clinic and the moment the diagnosis was established and discussed with the patient. Time to discharge from the clinic was defined as the number of days between the first visit to the clinic and the last visit to the clinic, or other departments of the hospital, as part of the diagnostic process or treatment of the chief complaint. If patients had not been discharged by the start of this study (1 May 2019), time to

discharge from the clinic was defined as the number of days between the first visit to the clinic and 1 May 2019. For patients that were lost to follow-up, time to discharge from the clinic was defined as the number of days between the first and the last visit to the clinic (or other departments of the hospital), and was reported separately. Time spent in the clinic was defined as the total amount of time (in minutes) reserved for the patients' appointments at the internal medicine outpatient clinic, as well as for appointments by phone. Total time spent in the clinic was calculated similarly, but all appointments regarding the chief complaint at any outpatient clinic in the UMC Utrecht were included.

Validation of Outcomes

After all data were collected from case records by the first researcher (PB), all established diagnoses and additional diagnoses were evaluated, also based on case record examinations, by two experienced internists (the two other authors: JW-BV). In three cases (1 diagnosis, 2 additional diagnoses) opinions differed between authors, and consensus was reached through group discussion involving all three authors (PB-JW-BV). In all other cases, authors agreed on the validity of the (additional) diagnoses collected from case records by the first researcher. Besides (additional) diagnoses, treatment including treatment effects and time to diagnosis were also checked by a second researcher (BV) in a random sample of 5% of all patients (n = 9). This was done to ensure that outcome definitions were adequately described, so that usage of the definitions by two independent researchers would lead to consistent results. All outcomes of patients from the sample determined by the second researcher were consistent with the outcomes determined by the first researcher.

Statistical Analyses

Descriptive statistics were used to summarize patients' characteristics at baseline and established diagnoses, initiated treatment, follow-up times and performed investigations in the context of the second opinion. Categorical variables were characterized using frequencies and percentages, continuous variables were characterized using means and standard deviations or medians and interquartile ranges (IQR), when appropriate.

In order to compare outcome between groups of categorical/dichotomous variables, such as gender, referring doctor, chief complaint and groups of patients with and without a new diagnosis or treatment, Pearson's χ^2 test was used. To assess the relationship between age at time of referral or time between consultations, and outcomes of second opinion, logistic regression models were used.

Results were considered statistically significant if p-value was < .05. Statistical analyses were performed using spss software, version 25 (IBM SPSS Statistics for Windows, Version 25.0).

Results

Study Population

In total, 196 patients were referred for a second opinion between June 2016 and August 2018. Out of these patients, 23 patients did not visit the clinic or visited the clinic of a different medical specialty. Therefore, 173 patients were included in this study.

Patient Characteristics

Mean age was 42.0 (± 16.4) years and the majority of patients were female (69%) (Table 1). Of 173 patients, 65% were referred by their own general practitioner, 21% by a locum general practitioner and 14% by a specialist. At time of referral, a diagnosis had been established by previous doctors in only 15% of patients. Median time between last consultation by a previous physician and first consultation with the physician formulating the second opinion was 97 days (IQR: 43 – 248). Most prevalent presenting symptoms were fatigue (34%), abdominal pain (28%), pain (multifocal) (11%), weight loss (6%), edema (5%) and fever (3%). A list of all other chief complaints can be found in Supplementary Table 3.

Diagnosis

Out of 173 patients, the diagnostic process was completed in 150 patients (87%). In 23 patients (13%) the diagnostic process was still ongoing (4%) or they were lost to follow-up before the diagnostic process was completed (9%). At the conclusion of the second opinion, a diagnosis was established in 38 of all patients (22%) (Table 2). In 23 of these patients (13% of total population) the established diagnosis was considered a new diagnosis. Specified for patients with and without a diagnosis at time of referral, a diagnosis was established in 17 out of 26 patients (65%) with a diagnosis at baseline, including 2 new diagnoses (8%), and 21 out of 147 patients (14%) without a diagnosis at baseline. Most frequently established new diagnoses were Anterior Cutaneous Nerve Entrapment Syndrome (ACNES) (4 patients) and Irritable Bowel Syndrome (IBS) (3 patients). A complete list of new diagnoses established during second opinion is presented in Supplementary Table 4. Diagnoses of patients with a diagnosis at time of referral and their diagnosis after second opinion are summarized in Supplementary Table 5.

Table 1: baseline characteristics

Characteristics	Study population N = 173
Age, years	42.0 (± 16.4)
Gender	
Male	53 (31%)
Female	120 (69%)
Referring doctor	
General practitioner	112 (65%)
Locum general practitioner	37 (21%)
Specialist	24 (14%)
Diagnosis at time of referral	
Yes	26 (15%)
No	147 (85%)
Time between consultations*, days	
Mean (± SD)	253 (± 456)
Median (IQR)	97 (43-248)
Chief complaint	
Fatigue	59 (34%)
Abdominal pain	48 (28%)
Pain (multifocal)	19 (11%)
Weight loss	10 (6%)
Edema	8 (5%)
Fever	5 (3%)
Other	24 (14%)

Baseline characteristics are presented as mean (± standard deviation) or number (%). Time between consultations is also presented as median with IQR (25 and 75 percentiles).
* Time between consultations was defined as the number of days between the last consultation with a previous physician and the first consultation with the physician formulating the second opinion.
IQR: interquartile range; SD: standard deviation.

ADDITIONAL DIAGNOSES

Furthermore, additional diagnoses were established in 55 patients (32%) (Table 2). In 91% of those patients (29% of total population), established additional diagnoses were considered relevant, as treatment for the condition was initiated. Most prevalent additional diagnoses were vitamin (B11, B12, D) and iron deficiencies, urinary tract infection, hypertension and dyslipidemia. A list of all additional diagnoses established during second opinion and their prevalence can be found in Supplementary Table 6.

INTER-COLLEGIAL CONSULTATION

During second opinion, 62 patients (36% of total population) were referred for inter-collegial consultation, leading to a total number of 92 consultations. Of the 23 new diagnoses established during second opinion, 6 diagnoses were established during inter-collegial consultation. An overview of inter-collegial consultations during second opinions is presented in Supplementary Table 7.

TABLE 2: DIAGNOSES ESTABLISHED DURING SECOND OPINIONS

Outcome measure	n	New diagnosis* n (%)
Diagnosis established		
Total population (N = 173)	38 (22%)	23 (13%)
Complete cases (n = 150)	38 (25%)	23 (15%)
Diagnosis established in patients with a diagnosis at time of referral		
All (n = 26)	17 (65%)	2 (8%)
Complete cases (n = 23)	17 (74%)	2 (9%)
Diagnosis established in patients without a diagnosis at time of referral		
All (n = 147)		21 (14%)
Complete cases (n = 127)		21 (17%)
Additional diagnosis established (number of patients)		
Total population (N = 173)	55 (32%)	
Relevant additional diagnosis	50 (29%)	

Data are presented as number of patients (% of patients in category).

* Diagnosis established during second opinion (by the internist formulating the second opinion or during inter-collegial consultation) different from diagnosis at time of referral, or established in a patient without a diagnosis at time of referral.

Baseline characteristics are presented as mean (± standard deviation) or number (%). Time between consultations is also presented as median with IQR (25 and 75 percentiles).

* Time between consultations was defined as the number of days between the last consultation with a previous physician and the first consultation with the physician formulating the second opinion.

IQR: interquartile range; SD: standard deviation.

Treatment

A new treatment was initiated or a change was made in a preexisting management plan in 97 patients (56%) (Table 3). New treatment mainly involved the prescription of new medication (56%) or the supplementation of vitamins or iron (28%). In 6% of patients receiving new treatment, a change in medication (change in dose or discontinuation) was made. Of 97 patients, 7% received analgesia, 5% received physical therapy, a diet was prescribed in 5%, 4% underwent surgery and 3% received cognitive behavioural therapy.

TREATMENT EFFECTS

Regarding treatment effects, resolution of the chief complaint was observed in 6% of patients receiving new treatment and improvement in 28%. Chief complaint had remained unchanged in 38% of patients with a newly initiated treatment, and worsened in 2%. In 26% of patients, treatment effects were unknown as follow-up of the symptoms attributable to the chief complaint were not documented, or patients were discharged or lost to follow-up shortly after the treatment was initiated. Considering the total study population, 19% of all patients received an effective treatment (resolution or improvement of chief complaint) during second opinion.

TABLE 3: TREATMENT INITIATED DURING SECOND OPINIONS

	n	% of total population
Treatment initiated		
Number of treatments	116	–
Number of patients	97	56%
Treatment type		
Medication (new)	54 (56%)	31%
Change in medication	6 (6%)	3%
Supplementation*	27 (28%)	16%
Analgesia	7 (7%)	4%
Physical therapy	5 (5%)	3%
Diet	5 (5%)	3%
Surgery	4 (4%)	2%
CBT	3 (3%)	2%
Other†	5 (5%)	3%
Treatment effects‡		
Resolution	6 (6%)	4%
Improvement	27 (28%)	16%
Unchanged	37 (38%)	21%
Worsened	2 (2%)	1%
Unknown	25 (26%)	15%

Data are presented as number (% of patients receiving treatment) and % of total population (N = 173). A patient can receive multiple treatments.

* This includes supplementation of vitamin B11/B12/D and/or iron.

† Other types of treatment included: cyst drainage (n = 1), enteral tube feeding (n = 1), chemoradiation therapy (n = 1), avoidance of sternal pressure (n = 1) and fecal transplantation (n = 1).

‡ Treatment effects were determined per patient. If a patient received multiple treatments, the overall effect of the treatments combined was used for this analysis.

CBT: cognitive behavioural therapy.

DIAGNOSIS AND TREATMENT

Treatment was initiated significantly more frequently in patients with a new diagnosis, established during second opinion (91% vs 51%; $p < .001$). Also, patients with a new diagnosis more frequently received an effective treatment (52% vs 14%; $p < .001$).

Patient-reported Symptomatology

Overall, the chief complaint improved or resolved in 28% of all patients referred for a second opinion (Table 4). Resolution or improvement of the chief complaint was more frequently observed in patients who received a new treatment compared to patients who did not (34% vs 21%), although this difference was not statistically significant ($p = .06$). Patients with a new diagnosis more frequently reported improvement or resolution of symptoms (52% vs 25%; $p = .006$). Patients with neither a new diagnosis nor a new treatment still reported improved symptomatology in 22% of cases.

TABLE 4: PATIENT-REPORTED OUTCOME OF CHIEF COMPLAINT AFTER SECOND OPINION

	All patients	With new treatment	Without new treatment	With new diagnosis	Without new diagnosis
	N = 173	n = 97	n = 76	n = 23	n = 150
resolution	9 (5%)	6 (6%)	3 (4%)	3 (13%)	6 (4%)
improvement	40 (23%)	27 (28%)	13 (17%)	9 (39%)	31 (21%)
unchanged	56 (32%)	37 (38%)	19 (25%)	3 (13%)	53 (35%)
worsened	9 (5%)	2 (2%)	7 (9%)	1 (4%)	8 (5%)
unknown	59 (34%)	25 (26%)	34 (45%)	7 (30%)	52 (35%)

Data are presented as n (% of patients in group), for all patients and for patients with or without a new treatment or diagnosis.

* Diagnosis established during second opinion (by the internist formulating the second opinion or during inter-collegial consultation) different from diagnosis at time of referral, or established in a patient without a diagnosis at time of referral.

Investigations

BLOOD TESTING

Conventional blood testing was performed in 86% of all patients (Table 5). In 89% of these cases, conventional blood testing had already been carried out by the previous physician, but was repeated during second opinion. Conventional blood testing led to relevant information in 23% of all cases in which conventional blood testing was performed: 23% of cases in which conventional blood testing was repeated, and 24% of cases in which conventional blood testing was performed for the first time. It showed anomalous results contributing to the establishment of a diagnosis in 4% of all cases. Additional blood tests were performed in 72% of all patients, leading to relevant information in 17% and leading to anomalous results contributing to the establishment of a diagnosis in 2% of these patients.

URINALYSIS AND MICROBIOLOGY

Urinalysis was carried out in 53% of patients (repetition rate 42%). Relevant information was discovered in 14% of these patients (13% of new investigations, 15% of repeated investigations). Anomalous results discovered by urinalysis contributed to the establishment of a diagnosis in only 1% of patients it was performed in. In 50% of all patients, microbiology tests were performed, mostly focused on bacterial and viral pathogens. Repetition rates were low (4-6%). Overall relevant information rates ranged from 6% (viral) to 20% (other: parasites, fungi, protozoa). Noticeably, anomalous results of microbiology tests did not once contribute to the establishment of a diagnosis.

RADIOLOGY

Radiological tests were performed in 49% of all patients. X-ray was most frequently performed (32%), MRI and PET-CT/SPECT were each only performed in 5% of patients. Repetition rates ranged from 23% to 33%, except for MRI, which was never a repetition of a previously conducted investigation. When repeated, ultrasonography and CT lead to relevant information in 29% and 30% of cases respectively, while repeated PET-CTs or SPECTs always, and repeated X-rays never lead to relevant new information. Regarding new investigations, relevant information rates were 11% for X-ray, 22% for ultrasonography, 27% for CT, 75% for MRI and 67% for PET-CT/SPECT. When all performed tests, repeated or new, were considered, relevant information rates were high for MRI and PET-CT/SPECT (75%), intermediate for ultrasonography and CT (23% and 28%), and low for X-ray (7%). When PET-CT or SPECT was performed, it led to the discovery of anomalous results relevant to the diagnosis in 38% of patients.

Table 5: performed investigations during second opinions

Investigation	Performed	Repeated*	Relevant information†			Anomalous result contributing to diagnosis‡
		% of performed	% of performed	% of new	% of repeated	% of performed
Blood tests						
Conventional§	148 (86%)	89%	23%	24%	23%	4%
Additional‖	124 (72%)	2%	17%	17%	–	2%
Urinalysis	92 (53%)	42%	14%	13%	15%	1%
Microbiology	87 (50%)					
Viral	69 (40%)	6%	6%	5%	25%	–
Bacterial	72 (42%)	8%	17%	15%	33%	–
Other	25 (15%)	4%	20%	21%	–	–
Radiology	84 (49%)					
X-ray	55 (32%)	33%	7%	11%	–	–
Ultrasonography	30 (17%)	23%	23%	22%	29%	10%
CT	32 (19%)	31%	28%	27%	30%	3%
MRI	8 (5%)	–	75%	75%	–	–
PET-CT/SPECT	8 (5%)	25%	75%	67%	100%	38%
Endoscopy	16 (9%)	31%	25%	27%	20%	6%
Pathology	25 (15%)	28%	32%	33%	29%	24%

Data are presented as number of patients (% of total population) with % of performed investigations that were repeated, led to relevant information and contributed to the diagnosis. Relevant information rates are also specified for new and repeated investigations.

* Investigation performed during second opinion, which had already been performed by a previous physician before the start of the second opinion. Reassessments of results, images or tissue obtained by previous investigations and transferred from a previous hospital to the UMC Utrecht were not considered investigations as the actual investigations were not performed during second opinion, so reassessments were also not considered repeated investigations.

† Information not known from previous investigations leading to either the establishment of a diagnosis or additional diagnosis, the initiation of a new treatment or the requirement for another investigation for further assessment.
‡ Anomalous results discovered by an investigation performed during second opinion, contributing to the establishment of a diagnosis.
§ Blood tests regularly performed during second opinions (specified in Supplementary Table 2).
‖ Blood tests not included in conventional blood testing.
CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; SPECT: single photon emission computed tomography.

Anomalous results of ultrasonography and CT led to a diagnosis in 10% and 3% of cases respectively. X-ray never showed anomalous results contributing to the establishment of a diagnosis.

Table 6: time spent during second opinions.

Measure	Mean (± SD)	Median	IQR (25 – 75 percentiles)	Total
Time to diagnosis, days				
All* (n = 38)	96 (± 130)	64	25 - 128	
New† (n = 23)	117 (± 153)	68	35 - 153	
Time to discharge, days				
Complete cases (n = 143)	109 (± 108)	75	31 - 144	
Not yet discharged‡ (n = 14)	499 (± 193)	433	335 - 619	
Lost to follow-up§ (n = 16)	99 (± 94)	49	20 - 185	
Time in clinic, minutes				
Internal medicine	80 (± 31)	70	60 - 90	230 hours
All outpatient clinics	114 (± 93)	80	60 - 135	330 hours

Data are presented as mean (± SD), median and IQR (25 – 75 percentiles).

* All diagnoses established by the doctor formulating the second opinion plus relevant diagnoses established during consultation by another specialist.

† Diagnosis established during second opinion (by the internist formulating the second opinion or during inter-collegial consultation) different from diagnosis at the time of referral, or established in a patient without a diagnosis at the time of referral.

‡ Patients not yet discharged at the start of the study: 1 May 2019 was used as time of discharge.

§ Patients lost to follow-up: last visit was used as time of discharge.

IQR: interquartile range; SD: standard deviation

ENDOSCOPY AND PATHOLOGY

Endoscopic procedures were carried out in 9% of patients (repetition rate 31%), leading to the discovery of relevant information in 25% (27% of new investigations, 20% of repeated investigations), and to anomalous results contributing to the establishment of a diagnosis in 6% of these patients. In 15% of all patients, pathology tests were performed (repetition rate 28%), leading to relevant information in 32% (33% of new investigations, 29% of repeated investigations), and anomalous results contributing to the diagnosis in 24% of patients.

Table 7a: potential determinants of outcomes of second opinions

Determinant	New diagnosis (n = 23)	p-value	New treatment (n = 97)	p-value	Effective treatment* (n = 33)	p-value
Gender						
male (n = 53)	5 (9%)	.320	29 (55%)	.812	8 (15%)	.376
female (n = 120)	18 (15%)		68 (57%)		25 (21%)	
Referring doctor						
general practitioner (n = 112)	16 (14%)	.858	59 (53%)	.409	20 (18%)	.389
locum general practitioner (n = 37)	4 (11%)		22 (60%)		6 (16%)	
specialist (n = 24)	3 (13%)		16 (67%)		7 (29%)	
Chief complaint†						
fatigue (n = 59)	5 (9%)	.108	31 (53%)	.632	9 (15%)	.715
abdominal pain (n = 48)	11 (23%)		30 (63%)		10 (21%)	
pain (multifocal) (n = 19)	3 (16%)		9 (47%)		3 (16%)	
other (n = 47)	4 (9%)		27 (57%)		11 (23%)	
Diagnosis at time of referral						
yes (n = 26)	2 (8%)	.361	15 (58%)	.856	5 (19%)	.983
no (n = 147)	21 (14%)		82 (56%)		28 (19%)	

Data are presented as number (% of subcategory). P-values for differences in outcome within categories are given.

* Initiated treatment leading to improvement or resolution of the chief complaint.

† Chief complaints with a prevalence of ≥ 10% were used as separate groups, remaining chief complaints were placed in 'Other'.

Follow-up

Median time to diagnosis was 64 days (IQR 25 – 128) for patients in whom a diagnosis was established (Table 6). When regarding new diagnoses only, median time to diagnosis was 68 days (IQR 35 – 153). Median time to discharge was 75 days (IQR 31 – 144) for patients whose second opinions were completed. Median time spent at the internal medicine outpatient clinic was 70 minutes (IQR 60 – 90), and median time spent at any outpatient clinic of the UMC Utrecht (in the context of the second opinion) was 80 minutes (IQR 60 – 135).

Determinants of Outcome

Statistically significant differences in outcome between patient groups based on gender, referring doctor, chief complaint or presence of a diagnosis at time of referral were not found (Table 7a). Neither was there a significant relationship between age or time between consultations and chance of a new diagnosis or (effective) treatment

(Table 7B/C). However, when specifically comparing patients with abdominal pain to patients with fatigue, a new diagnosis was more frequently established in patients with abdominal pain (23% vs 9%, p = .037).

Table 7B. OTHER POTENTIAL DETERMINANTS OF OUTCOMES.

	New diagnosis		New treatment		Effective treatment	
	Yes (n = 23)	No (n = 150)	Yes (n = 97)	No (n = 76)	Yes (n = 33)	No (n = 140)
Age	42.1 (± 17.5)	42.0 (± 16.2)	41.9 (± 15.6)	42.1 (± 17.4)	41.0 (± 16.5)	42.2 (± 16.4)
Time between consultations	226 (± 339)	258 (± 472)	256 (± 481)	251 (± 425)	267 (± 564)	250 (± 428)

Data are presented as mean (± standard deviation)

Table 7C: LOGISTIC REGRESSION MODELS.

	New diagnosis (n = 23)		New treatment (n = 97)		Effective treatment (n = 33)	
	B coefficient	p-value	B coefficient	p-value	B coefficient	p-value
Age	0.001	.960	-0.001	.931	-0.004	.722
Time between consultations	0.000	.765	0.000	.946	0.000	.852

Data are reported as B coefficients and p-values from logistic regression models.

Discussion

During second opinions in a general internal medicine outpatient clinic of an academic hospital, a new diagnosis was established in 13% of patients, while overall, resolution or improvement of the chief complaint was achieved in 28% of patients. In approximately one third of patients a relevant additional diagnosis was established, and in over half of all patients, a new treatment was initiated. Treatment, whether a new diagnosis was established or not, led to improvement or resolution of the chief complaint in 34% of patients. Many investigations were carried out, often repeating previously performed investigations. Anomalous results from investigations rarely contributed to the establishment of a diagnosis.

Regarding the establishment of a new diagnosis, results presented in this study are very similar to findings of the two previous studies exploring the outcomes of second opinions in internal medicine (both in Dutch academic hospitals), with a new diagnosis being established in approximately 10% of patients in these studies.[7,8]

We add to the body of evidence by detailing the diagnostic process and outcome. When compared to second opinions in other medical specialties, the diagnostic value of second opinions in internal medicine is low. Studies in other medical specialties have shown that a new diagnosis is established in approximately 30 – 60% of patients, ranging from 30% in surgical oncology to 60% in orthopedic surgery.[10,12,13,18,20,23,24] This difference might be explained by the fact that, in this study and in previous studies,[7,8] up to 85% of patients referred for a second opinion in internal medicine had poorly defined conditions without a diagnosis at the time of referral. This usually concerns patients with a high suspicion of medically unexplained physical symptoms, in whom a diagnosis cannot be easily established. Also, part of the diagnoses that were established in this study are diagnoses without objective criteria and for which treatment options are lacking. One could question the value of the establishment of these kinds of diagnoses. The same applies to additional diagnoses, such as iron and vitamin deficiencies, which were frequently established in this study.

In this study, a new diagnosis was more frequently established in patients presenting with abdominal pain compared to fatigue. This was also observed in small numbers of patients presenting with abdominal pain or fatigue in a previous study exploring outcomes of second opinions in internal medicine.[7] In another previous study, new diagnosis rates of these groups were very similar.[8] The fact that more new diagnoses were established in patients with abdominal pain in this study, is most likely related to the fact that ACNES and IBS were the most frequently established new diagnoses in this study (Supplementary Table 4). ACNES is known to be a poorly recognized and commonly underdiagnosed cause of abdominal pain.[33-36] Therefore, it is likely that ACNES is sometimes not recognized by the original physician, but the diagnosis is established by the physician formulating the second opinion, as internists in our center are aware of the fact that ACNES is a commonly underdiagnosed cause of abdominal pain. Diagnosis of IBS is not based on objectifiable findings from diagnostic tests, and is often not established before other diagnoses have been ruled out. So, potentially, physicians are reluctant to establish this diagnosis when they are aware a patient wishes to be referred for a second opinion, as they know that during second opinion, a different diagnosis might be established. If a different diagnosis cannot be established, potentially, the diagnosis of IBS could be established by the physician formulating the second opinion, contributing to the higher number of new diagnoses in patients with abdominal pain.

This is the first study exploring treatment initiation and patient-reported symptomatology during second opinions in internal medicine to date. Noticeably, this study showed that the proportion of patients who received a new treatment

was substantially larger than the proportion of patients in whom a new diagnosis was established. This indicates that, even though a new diagnosis was strongly related to the initiation of a new treatment in this study, a new treatment is also frequently initiated in patients in whom no diagnosis was established. In addition, resolution or improvement of symptoms was also frequently achieved in patients who did not receive a new treatment, nor a new diagnosis. This means that the yield of second opinions is not limited to the establishment of diagnoses. However, it is commonly known that a placebo effect can play a substantial role in patient-reported symptomatology and treatment effects. In the absence of a control group, it is hard to determine what part of the treatment effects and reported improved symptoms in this study are attributable to a placebo effect. It is likely that information and reassurance provided by the physician carrying out the second opinion can lead to improved symptoms, as it was shown to increase patient satisfaction in a previous study.[7] One could argue that this is part of the value of second opinions, whether it is based on a placebo effect or not.

This is the first study that thoroughly analyzed investigations performed in the context of second opinions. One previous study in an internal medicine outpatient clinic reported that, depending on the type of investigation (for example blood testing, urinalysis or radiology), approximately 40 – 90% of investigations performed by original physicians were repeated by the physician formulating the second opinion.[8] Our study focused on how many of the performed investigations during second opinion were in fact a repetition of investigations already carried out by the original physicians. Repetition rates for radiological tests seemed to be lower in our study, possibly caused by the fact that, nowadays, information from case records is more easily transferred between hospitals. Also, a radiological second opinion of investigations performed in other hospitals can be easily obtained. Repetition of investigations could be seen as waste. However, noticeably, in this study, repeated investigations led to the discovery of relevant information relatively frequently, which is in contrast with the aforementioned study.[8] Partly this is due to the fact that in a considerable number of patients conventional blood testing was considered repeated, while in fact a small share of the tests had not been performed before. Conventional tests that had not been performed before mostly included vitamin and iron tests, which often led to relevant information. Thus, conventional blood testing often showed relevant information when repeated, but relevant information was often only found in the share of tests that were actually not a repetition. The relatively high relevant information rates for repeated microbiology, radiology, endoscopy and pathology tests are remarkable. These results suggest that, when doctors formulating

second opinions believe it is necessary, repeating investigations can be useful. Time lapsed between original and repeated investigations was not assessed in this study, so a statement on the possible relationship between amount of time between investigations and relevance of results cannot be made.

One of the strengths of this study is the fact that it is the most extensive research on second opinions in general internal medicine to date. Additional diagnoses, treatment effects, patient-reported symptomatology, relevance of all performed investigations and follow-up time had never been assessed before. Also, this is the first study exploring the value of second opinions in internal medicine in ten years' time. Finally, strengths of this study include the large population size and the fact that all patients who visited our clinic for a second opinion in the given time frame were included in the study, so selection bias was avoided.

A limitation of this study is the retrospective design and the fact that all outcomes were based on case records. Nevertheless, most important outcome measures, such as diagnoses and treatment, are generally carefully documented by physicians in case records, including correspondence, so a considerable impact of the study design on end points is unlikely. However, treatment effects were not always accurately documented, and one could question whether improvement of symptoms after the initiation of treatment is always caused by the treatment. It is likely that in some patients symptoms resolve due to a placebo effect. So, treatment effects might be overestimated in this study. Furthermore, the fact that in some patients the diagnostic process was incomplete could be seen as a limitation of this study. However, by reporting our outcome (new diagnosis) as percentage of the total of referred patients, we describe current practice and thus approximate the real benefit of referral for second opinion in general internal medicine. Finally, there was a limitation in the way relevance of information discovered by investigations was determined. Information was only considered relevant in case of anomalous results leading to the establishment of a diagnosis or additional diagnosis, the initiation of treatment, or the requirement for another investigation for further assessment, while normal results or negative tests might be relevant in establishing or ruling out a diagnosis as well.

Conclusion

In conclusion, this extensive research on the outcomes of second opinions in general internal medicine, has shown that a new diagnosis is established in 13% of patients. Patients in whom a new diagnosis is established benefit more from second opinions, but the value of second opinions is not limited to the establishment of

diagnoses, as patients without a new diagnosis also frequently receive treatment and report improvement of symptoms. Overall, at least 28% of patients benefit from second opinions, as resolution or improvement of symptoms is achieved. However, remarkably, a large number of investigations are performed and repeated during second opinions, while these investigations rarely contribute to the establishment of a diagnosis. Despite that, this study has shown that second opinions in internal medicine are valuable in terms of the establishment of diagnoses, initiation of treatment and improvement of symptoms in a considerable number of patients.

References

1. Zorginstituut Nederland. Second opinion in de Zorgverzekeringswet. 2015. [Available from: https://www.zorginstituutnederland.nl/publicaties/standpunten/2015/01/06/second-opinion-in-de-zorgverzekeringswet].
2. Wet op de geneeskundige behandelingsovereenkomst. Burgerlijk Wetboek, Boek 7.
3. van Dalen I, Groothoff J, Stewart R, Spreeuwenberg P, Groenewegen P, van Horn J. Motives for seeking a second opinion in orthopaedic surgery. *J Health Serv Res Policy*. 2001;6(4):195-201.
4. Mellink WA, Dulmen AM, Wiggers T, Spreeuwenberg PM, Eggermont AM, Bensing JM. Cancer patients seeking a second surgical opinion: results of a study on motives, needs, and expectations. *J Clin Oncol*. 2003;21(8):1492-7.
5. Moumjid N, Gafni A, Bremond A, Carrere MO. Seeking a second opinion: do patients need a second opinion when practice guidelines exist? *Health Policy*. 2007;80(1):43-50.
6. Sutherland LR, Verhoef MJ. Why do patients seek a second opinion or alternative medicine? *J Clin Gastroenterol*. 1994;19(3):194-7.
7. Westland GJ, Nanayakkara PW, Kramer MH, Smulders YM. [Diagnostic second opinion: what does it add? Patient satisfaction, but diagnosis is rarely altered]. *Ned Tijdschr Geneeskd*. 2009;153:A777.
8. Mustafa M, Bijl M, Gans R. What is the value of patient-sought second opinions? *Eur J Intern Med*. 2002;13(7):445.
9. Maaskant JM, van Muilekom HA. Eerste arts zit er zelden naast. Second opinion: patiënt zoekt vooral info, hoop en geruststelling. *Medisch Contact*. 2009;64:605-7.
10. van Leeuwen R, van der Zaag-Loonen H. Second opinion wel degelijk zinnig. *Medisch Contact*. 2014:882-3.
11. Schook R, ter Avest M, van Setten C, E. S. Longkanker: retrospectief onderzoek ondersteunt concentratiegedachte. Second opinion in expertcentrum zinvol. *Medisch Contact*. 2012;67:1119-21.
12. Wieske L, Vergouwen MD, Wijers D, Stam J, Smets EM, Richard E. [Limited benefit of diagnostic second opinions in a neurology department]. *Ned Tijdschr Geneeskd*. 2011;155(32):A3542.
13. Wieske L, Wijers D, Richard E, Vergouwen MD, Stam J. Second opinions and tertiary referrals in neurology: a prospective observational study. *J Neurol*. 2008;255(11):1743-9.
14. Mellink WA, Henzen-Logmans SC, Bongaerts AH, Ooijen BV, Rodenburg CJ, Wiggers TH. Discrepancy between second and first opinion in surgical oncological patients. *Eur J Surg Oncol*. 2006;32(1):108-12.
15. Meyer AN, Singh H, Graber ML. Evaluation of outcomes from a national patient-initiated second-opinion program. *Am J Med*. 2015;128(10):1138.e25-33.
16. Payne VL, Singh H, Meyer AN, Levy L, Harrison D, Graber ML. Patient-initiated second opinions: systematic review of characteristics and impact on diagnosis, treatment, and satisfaction. *Mayo Clin Proc*. 2014;89(5):687-96.
17. Tattersall MH, Dear RF, Jansen J, Shepherd HL, Devine RJ, Horvath LG, et al. Second opinions in oncology: the experiences of patients attending the Sydney Cancer Centre. *Med J Aust*. 2009;191(4):209-12.
18. Mellink WA, Wiggers T. [Second consult outpatient clinic Surgical Oncology in the Daniel de Hoed Clinic: analysis of the first 245 patients]. *Ned Tijdschr Geneeskd*. 1999;143(50):2547-8.
19. Hillen MA, Medendorp NM, Daams JG, Smets EMA. Patient-Driven Second Opinions in Oncology: A Systematic Review. *Oncologist*. 2017;22(10):1197-211.

20. Chan TY, Epstein JI. Patient and urologist driven second opinion of prostate needle biopsies. *J Urol*. 2005;174(4 Pt 1):1390-4; discussion 4; author reply 4.

21. Schook RM, ter Avest MJ, van Setten CH, de Man FF, Smit EF, Postmus PE. Lung cancer patients benefit from second opinions by improvement of diagnosis and therapy. *Cancer Clin Oncol* 2014;3:43-57.

22. Ruetters D, Keinki C, Schroth S, Liebl P, Huebner J. Is there evidence for a better health care for cancer patients after a second opinion? A systematic review. *J Cancer Res Clin Oncol*. 2016;142(7):1521-8.

23. Garcia D, Spruill LS, Irshad A, Wood J, Kepecs D, Klauber-DeMore N. The Value of a Second Opinion for Breast Cancer Patients Referred to a National Cancer Institute (NCI)-Designated Cancer Center with a Multidisciplinary Breast Tumor Board. *Ann Surg Oncol*. 2018;25(10):2953-7.

24. Lenza M, Buchbinder R, Staples MP, Dos Santos OFP, Brandt RA, Lottenberg CL, et al. Second opinion for degenerative spinal conditions: an option or a necessity? A prospective observational study. *BMC Musculoskelet Disord*. 2017;18(1):354.

25. Shmueli L, Davidovitch N, Pliskin JS, Balicer RD, Hekselman I, Greenfield G. Seeking a second medical opinion: composition, reasons and perceived outcomes in Israel. *Isr J Health Policy Res*. 2017;6(1):67.

26. Benson WE, Regillo CD, Vander JF, Tasman W, Smith AF, Brown GC, et al. Patient-initiated second medical opinions: their necessity and economic cost. *Retina*. 2001;21(6):633-8.

27. Shmueli L, Davidovitch N, Pliskin JS, Hekselman I, Balicer RD, Greenfield G. Reasons, perceived outcomes and characteristics of second-opinion seekers: are there differences in private vs. public settings? *BMC Health Serv Res*. 2019;19(1):238.

28. Cook IS, McCormick D, Poller DN. Referrals for second opinion in surgical pathology: implications for management of cancer patients in the UK. *Eur J Surg Oncol*. 2001;27(6):589-94.

29. Staradub VL, Messenger KA, Hao N, Wiley EL, Morrow M. Changes in breast cancer therapy because of pathology second opinions. *Ann Surg Oncol*. 2002;9(10):982-7.

30. Westra WH, Kronz JD, Eisele DW. The impact of second opinion surgical pathology on the practice of head and neck surgery: a decade experience at a large referral hospital. *Head Neck*. 2002;24(7):684-93.

31. Clauson J, Hsieh YC, Acharya S, Rademaker AW, Morrow M. Results of the Lynn Sage Second-Opinion Program for local therapy in patients with breast carcinoma. *Cancer*. 2002;94(4):889-94.

32. Wijers D, Wieske L, Vergouwen MD, Richard E, Stam J, Smets EM. Patient satisfaction in neurological second opinions and tertiary referrals. *J Neurol*. 2010;257(11):1869-74.

33. Chrona E, Kostopanagiotou G, Damigos D, Batistaki C. Anterior cutaneous nerve entrapment syndrome: management challenges. *J Pain Res*. 2017;10:145-56.

34. Applegate WV. Abdominal Cutaneous Nerve Entrapment Syndrome (ACNES): A Commonly Overlooked Cause of Abdominal Pain. *Perm J*. 2002;6:20-7.

35. Towfigh S, Anderson S, Walker A. When it is not a Spigelian hernia: abdominal cutaneous nerve entrapment syndrome. *Am Surg*. 2013;79(10):1111-4.

36. Koop H, Koprdova S, Schürmann C. Chronic Abdominal Wall Pain. *Dtsch Arztebl Int*. 2016;113(4):51-7.

Supplementary Materials

Supplementary Table 1: glossary of terms used throughout the manuscript
Supplementary Table 2: list of all blood tests regarded as conventional blood tests
Supplementary Table 3: other chief complaints
Supplementary Table 4: new diagnoses established during second opinions
Supplementary Table 5: diagnosis after second opinion in patients with a diagnosis at time of referral
Supplementary Table 6: additional diagnoses established during second opinions
Supplementary Table 7: intercollegial consultation during second opinions

Supplementary Table 1: glossary of terms used throughout the manuscript

Term	Definition
Referring doctor	The doctor who referred the patient for the second opinion and wrote the referral letter.
Time between consultations	Number of days between the last consultation with a previous physician and the first consultation with the physician formulating the second opinion.
Chief complaint	Main complaint mentioned by the patient during the first visit to the clinic.
Diagnosis at time of referral	Diagnosis established by referring doctor or another physician before second opinion, as documented by the referring doctor in the referral letter.
Diagnosis by doctor formulating the second opinion	Diagnosis established by the internist carrying out the second opinion.
Diagnosis established during inter-collegial consultation	Diagnosis established by a doctor from another medical specialty during inter-collegial consultation requested by the internist formulating the second opinion.
New diagnosis	Diagnosis established during second opinion (by the internist formulating the second opinion or during inter-collegial consultation) different from diagnosis at time of referral, or established in a patient without a diagnosis at time of referral.
Additional diagnosis	Diagnosis established during second opinion, concerning a condition which cannot cause the chief complaint.
Relevant additional diagnosis	Additional diagnosis leading to the initiation of treatment for that condition.
New treatment	The initiation of (a change in) medication, vitamin/iron supplementation, analgesia, therapy, dietary prescriptions or surgery during second opinion.
Treatment effects	Effects of new treatment initiated during second opinion on the chief complaint as reported by the patient and documented by the doctor in the case record.
Effective treatment	New treatment leading to the resolution or improvement of the chief complaint.
Patient-reported symptomatology	Outcome of chief complaint at the end of second opinion as reported by the patient and documented by the doctor in the case record.
Conventional blood testing	Blood tests regularly performed during second opinions (specified in Supplementary Table 2).
Additional blood tests	Blood tests not included in conventional blood testing.

Supplementary Table 1: continued.

Term	Definition
New investigation	An investigation performed during second opinion, which had not been performed by a previous physician.
Repeated investigation	An investigation was considered a repeated investigation if the investigation had already been performed by a previous physician before the start of the second opinion, and the exact same investigation was then performed again during the second opinion. Reassessments of results, images or tissue were not considered investigations or repeated investigations.
Relevant information	Information not known from previous investigations leading to either the establishment of a diagnosis or additional diagnosis, the initiation of a new treatment or the requirement for another investigation for further assessment.
Overall relevant information rate	Percentage of investigations that lead to relevant information for new and repeated investigations combined.
Anomalous results contributing to diagnosis	Anomalous results discovered by an investigation performed during second opinion, contributing to the establishment of a diagnosis.
Time to diagnosis	Number of days between the first visit to the clinic and the moment the diagnosis was established and discussed with the patient, as documented in the case record.
Time to discharge	Number of days between the first visit to the clinic and the last visit to the clinic, or other departments of the hospital, as part of the diagnostic process or treatment of the chief complaint.
Time spent in the clinic	Total amount of time (in minutes) reserved for the patients' appointments at the internal medicine outpatient clinic, as well as for appointments by phone.
Total time spent in the clinic	Total amount of time (in minutes) reserved for the patients' appointments, regarding the chief complaint, at any outpatient clinic in the UMC Utrecht, as well as for appointments by phone.

Supplementary Table 2: list of all blood tests regarded as conventional blood tests

Blood chemistry	Hematology	Endocrinology
Natrium	Hemoglobin	Thyroid-stimulating hormone
Kalium	Hematocrit	Free thyroxine (free T4)
Calcium	Erythrocytes	25-hydroxyvitamin D
Phosphate	Mean corpuscular volume (MCV)	
Uric acid	Mean corpuscular hemoglobin (MCH)	
Creatinine	Mean corpuscular hemoglobin concentration (MCHC)	
Estimated Glomerular Filtration Rate (eGFR)	Thrombocytes	
Bilirubin	Leukocytes	
Alkaline phosphatase	Neutrophilic granulocytes	
Gamma glutamyltransferase (GGT)	Basophilic granulocytes	
Aspartate transaminase (AST)	Eosinophilic granulocytes	
Alanine transaminase (ALT)	Lymphocytes	
Lactate dehydrogenase (LD)	Monocytes	
Creatine kinase (CK)	Erythrocyte sedimentation rate (ESR)	
Amylase		
Lipase		
Albumin		
C-reactive protein (CRP)		
Cholesterol		
Triglycerides		
High-density lipoprotein (HDL)		
Low-density lipoprotein (LDL)		
Ferritin		
Transferrin		
Transferrin iron saturation		
Iron		
Folic acid (vitamin B11)		
Vitamin B12		
Glucose		

Supplementary Table 3: other chief complaints

Chief complaint	Prevalence (Total: n = 24)
Headache	3
Syncope	3
Ascites	2
Flushes	2
Joint complaints	2
Pruritus	2
Anaphylaxis	1
Diarrhea	1
Dyspnea	1
Hiccups	1
Hyperhidrosis	1
Hypothermia	1
Nausea	1
Skin bumps	1
Weight gain	1
Wounds	1

Prevalence of chief complaints is reported as number.

Supplementary Table 4: new diagnoses established during second opinions

New diagnosis*	Prevalence (Total: n = 23)
Anterior Cutaneous Nerve Entrapment Syndrome (ACNES)	4
Irritable Bowel Syndrome (IBS)	3
Abdominal angina	1
Chronic Fatigue Syndrome (CFS)	1
Erythromelalgia	1
Facet syndrome	1
Familial Mediterranean Fever (FMF)	1
Fibromyalgia	1
Hyperventilation syndrome	1
Iron deficiency	1
Metastatic mammary carcinoma	1
Morbus Castleman	1
Morbus Crohn	1
Non-alcoholic steatohepatitis (NASH)	1
Postural Orthostatic Tachycardia Syndrome (POTS)	1
Schnitzler syndrome	1
Supragastric belching	1
Tietze syndrome	1

Prevalence of diagnoses is presented as number. Diagnoses established (partly) based on objectifiable findings from biochemical, radiological or pathological examinations are in bold.

* Diagnosis established during second opinion (by the internist formulating the second opinion or during inter-collegial consultation) different from diagnosis at time of referral, or established in a patient without a diagnosis at time of referral.

Supplementary Table 5: diagnosis after second opinion in patients with a diagnosis at time of referral

Diagnosis at time of referral (n = 26)	Diagnosis after second opinion (n = 17)
Anterior cutaneous nerve entrapment syndrome (ACNES)	Anterior cutaneous nerve entrapment syndrome (ACNES)
Chronic fatigue syndrome (CFS)	Chronic fatigue syndrome (CFS)
Chronic idiopathic urticaria	Chronic idiopathic urticaria
Erythema nodosum	Erythema nodosum
Fibromyalgia	Fibromyalgia
Fibromyalgia	Fibromyalgia
Fibromyalgia	–
Fibromyalgia + chronic obstructive pulmonary disease (COPD) + hypothyroidism	Fibromyalgia + chronic obstructive pulmonary disease (COPD) + hypothyroidism
Graves' disease	–
Iron deficiency anemia	Iron deficiency anemia
Iron deficiency anemia	Iron deficiency anemia
Irritable bowel syndrome (IBS)	**Anterior cutaneous nerve entrapment syndrome (ACNES)**
Irritable bowel syndrome (IBS)	**Supragastric belching**
Irritable bowel syndrome (IBS)	Irritable bowel syndrome (IBS)
Irritable bowel syndrome (IBS)	Irritable bowel syndrome (IBS)
Irritable bowel syndrome (IBS)	–
Irritable bowel syndrome (IBS)	–
Irritable bowel syndrome (IBS)	–
Irritable bowel syndrome, post-infectious (IBS)	Irritable bowel syndrome, post-infectous (IBS)
Leukocytoclastic vasculitis	Leukocytoclastic vasculitis
Pain amplification syndrome	Pain amplification syndrome
Pancreatic insufficiency	–
Splenic cyst	Splenic cyst
Urticaria factitia	–
Vitamin B12 deficiency anemia	–
Yellow nail syndrome	–

New diagnoses are in bold. If no diagnosis after second opinion is mentioned, this means that the diagnosis at time of referral was not confirmed and that a new diagnosis was not established.

Supplementary Table 6: additional diagnoses established during second opinions

Additional diagnosis*	Prevalence (Total: n = 55 patients, n = 78 diagnoses)
Iron deficiency	11
Vitamin B12 deficiency	10
Folic acid (vitamin B11) deficiency	8
Urinary tract infection	8
Vitamin D deficiency	7
Hypertension	5
Dyslipidemia	4
Adrenal incidentaloma	1
Angina pectoris	1
Baker's cyst	1
Breast cyst	1
Bursitis	1
Candidiasis	1
Chronic idiopathic urticaria	1
Erythema chronicum migrans	1
Focal nodular hyperplasia of the liver	1
Gallbladder polyp	1
Gastric fundic gland polyp	1
IgG2 subclass deficiency	1
IgM monoclonal gammopathy of unknown significance (MGUS)	1
Kidney cyst	1
Mannose-binding lectin deficiency	1
Obstructive sleep apnea syndrome (OSAS)	1
Panniculitis mesenterica	1
Parasitic gastroenteritis	1
Peptic duodenitis	1
Pulmonary embolism	1
Pulmonary emphysema	1
Sjögren syndrome	1
Specific antibody deficiency	1

Supplementary Table 6: continued.

Additional diagnosis*	Prevalence (Total: n = 55 patients, n = 78 diagnoses)
Syndrome of inappropriate antidiuretic hormone secretion (SIADH)	1
Tubular adenoma	1

Prevalence of additional diagnoses is presented as number. A patient may have multiple additional diagnoses.
* Diagnosis established during second opinion, concerning a condition which cannot cause the chief complaint.

Supplementary Table 7: intercollegial consultation during second opinions

	n	% of referred patients (n = 62)	% of total population (N =173)
Patients referred for consultation	62		36%
Total number of consultations	92		–
Specialties			
Gastroenterology	20	32%	12%
Rheumatology	11	18%	6%
Dermatology	10	16%	6%
Neurology	9	15%	5%
Cardiology	8	13%	5%
General surgery	6	10%	3%
Hematology	6	10%	3%
Anesthesiology	4	7%	2%
Gynaecology	4	7%	2%
Otorhinolaryngology	3	5%	2%
Pulmonology	3	5%	2%
Urology	3	5%	2%
Other*	5	8%	3%
Diagnosis during consultation			
Diagnosis chief complaint	6	10%	3%
Additional diagnosis	4	6%	2%

Patients and consultations are presented as number (% of patients in category), with % of total population (N = 173). A patient may be referred to multiple specialties.
* Other specialties included: ophthalmology (n = 2), cardiothoracic surgery (n = 1), orthopedic surgery (n = 1) and oncology (n = 1).

Chapter 6

Redundant laboratory testing on referral from general practice to the outpatient clinic: a post-hoc analysis

B.E.L. Vrijsen, M.J. ten Berg, W.W. van Solinge, J. Westerink

BJGP Open 2022; 6(1):BJGPO.2021.0134

Abstract

Background

Inappropriately repeated laboratory testing is a commonly occurring problem. However, this has not been studied extensively in the outpatient clinic after referral by general practitioners.

Aim

The aim of this study was to investigate how often laboratory tests ordered by the general practitioner were repeated on referral to the outpatient clinic, and how many of the normal test results remained normal on repetition.

Design and Setting

This is a post-hoc analysis of a study on laboratory testing strategies in patients newly referred to the outpatient clinic.

Method

All patients who had a referral letter including laboratory test results ordered by the general practitioner were included. These results were compared to the laboratory test results ordered in the outpatient clinic.

Results

Data were available for 295 patients, 191 of which had post-visit testing done. In this group, 56% of tests ordered by the general practitioner were repeated. Tests with abnormal results were repeated more frequently than tests with normal results (65% vs 53%; $p < .001$). A longer test interval was associated with slightly smaller odds of tests being repeated (OR 0.97 [0.95 – 0.99]; $p = .003$). Of the tests with normal test results that were repeated, 90% remained normal. This was independent of testing interval or testing strategy.

Conclusion

Laboratory tests ordered by the general practitioner are commonly repeated on referral to the outpatient clinic. The number of test results remaining normal on repetition suggests a high level of redundancy in laboratory test repetition.

Introduction

When patients are referred from general practice to the outpatient clinic, history taking and physical examination are typically repeated by the physician in the hospital uncontroversially. However, repeating imaging or laboratory testing is generally deemed to be wasteful in many cases.[1] Inappropriate duplicate testing is a frequently occurring example of overutilization, with an estimated 21% of overall laboratory tests being unwarranted.[2] Ordering unnecessary tests not only leads to increased health care costs, but can also compromise the diagnostic process by leading to more false positive results.[3] Furthermore, overtesting poses a burden on patients, by subjecting them to more phlebotomies and potential follow-up testing, and leading to anxiety about test results.[4,5]

One of the causes of inappropriate repeat testing is unawareness that the test has already been ordered.[6] This can occur when the care for patients is being transferred from one health care provider to another. For example, in transfers between emergency departments, duplication rates of 32 to 88% have been found.[7-11]

Repetition rates for patients being transferred from general practice to the hospital have been studied less extensively. One study, investigating repetition rates for patients who were admitted to the medical department of a hospital found that 63% of tests ordered by the general practitioners in the previous year were repeated on admission.[12]

For patients referred to the outpatient clinic, there are only data on repetition rates in patients referred for second opinions, where up to 90% of laboratory tests were repetitions of tests ordered by the original physician.[13,14] Yet no data are available on patients newly referred to the outpatient clinic by their general practitioner. Given the paucity of the data, we performed a post-hoc analysis of a study we performed on laboratory test strategies in the outpatient clinic.[15]

The aim of the current study was to determine how often the laboratory tests ordered by the general practitioner are repeated in the outpatient clinic after referral, depending on the test, the test result, the test interval and the laboratory test strategy at the outpatient clinic, and secondarily, how often tests with normal test results remain normal on repetition, depending on the test, the test interval and the laboratory test strategy at the outpatient clinic.

Methods

Setting and patient selection

This study is a post-hoc analysis of a study investigating the effect of pre-visit laboratory testing on the time to diagnosis in outpatients.[15] In this previous study, patients newly referred to the internal medicine outpatient clinic of the University Medical Center Utrecht (UMC Utrecht), a tertiary hospital in the Netherlands, were allocated to either pre-visit or post-visit laboratory testing.

In the pre-visit arm, patients had a standardized panel of laboratory tests done one hour prior to the first visit to the outpatient clinic, so the test results were available to the treating physician during the visit. In the post-visit arm, laboratory testing was done after the first visit at the discretion of the treating physician.

For the current study we included all subjects for whom a referral letter from their general practitioner was available that included laboratory test results. From these referral letters all results from laboratory tests including testing dates were collected by one of the authors (BV). Data on the laboratory testing done in the outpatient clinic were already available from the previous study.

Measures

All test results were classified as being either normal (i.e. within the reference range) or abnormal (i.e. outside of the reference range). For this, we used the reference intervals of the laboratory of the UMC Utrecht, because the referral letters generally did not include reference intervals or information on which laboratory had performed the tests, so the reference intervals could not be obtained.

Both the leukocyte differentiation and the urine screening, which comprise several different tests, were considered as one test. They were considered to be abnormal if one or more of the constituent tests were outside the reference range.

In several cases the reference range was adjusted, as for some tests a value below or above the reference range is generally clinically irrelevant. The lower limit of normal was removed for the activated partial thromboplastin time (aPTT), C-reactive protein (CRP), creatinine, erythrocyte sedimentation rate (ESR), hemoglobin A1C (HbA1c), international normalized ratio (INR), prothrombin time (PT), and urea. The upper limit of normal was removed for vitamins B12 and D.

For every test ordered by the general practitioner, the test interval was calculated as the time in weeks between the test ordered by the general practitioner and the date of the first visit to the outpatient clinic.

The laboratory strategy (either pre-visit testing or post-visit testing) was recorded "as treated", meaning that all patients were assigned to the laboratory strategy they

actually underwent in the original study, rather than the strategy to which they were originally allocated.

Outcomes

The primary outcome was the number of tests ordered by the general practitioner that were repeated in the outpatient clinic. This was studied in the patients who had undergone post-visit testing only, because including the group with pre-visit testing would lead to inflated repetition rates, given that the standardized panels used in the pre-visit testing group inevitably lead to repeat testing. The secondary outcome was the number of repeated tests remaining normal on repetition. This was studied in all patients.

Statistical analysis

The difference in repetition rates for tests with normal versus abnormal test results was statistically tested using the χ^2 test. Odds ratios for tests being repeated and for normal tests remaining normal on repetition depending on the time interval were calculated using a logistic mixed-effects model with the individual patients as random effect. The number of repeated tests and repeats of normal tests per patient were compared between laboratory testing strategies using the Wilcoxon rank-sum test. Repetition rates and rates of normal tests remaining normal were also calculated for the ten most commonly ordered tests. All analyses were performed in R version 4.0.3.[16]

Results

Baseline characteristics

Of the 594 eligible subjects in the original study, referral letters could be retrieved in 449 cases (76%), 295 of which (66%) included laboratory test results. All 295 subjects were included in this study. The included patients visited the internal medicine outpatient clinic between April 2016 and November 2017. The mean age was 52.7 years (95% confidence interval 50.5 – 54.8) and 66% were female. Pre-visit testing was done in 104 subjects and post-visit testing in 191. Baseline characteristics of these patients are reported in Table 1.

TABLE 1: BASELINE CHARACTERISTICS

	Pre-visit testing	Post-visit testing
Total (n)	104	191
Age (years)	53.4 (49.7 – 57.1)	52.3 (49.6 – 54.9)
Female (n, %)	66 (63%)	129 (68%)
Referral reason* (n)		
- abdominal complaints	12 (12%)	23 (12%)
- abnormal laboratory test	16 (15%)	28 (15%)
- anemia	12 (12%)	28 (15%)
- fatigue	37 (36%)	47 (25%)
- lymphadenopathy / suspected malignancy	1 (1%)	8 (4%)
- weight loss	15 (14%)	29 (15%)
- other	24 (23%)	43 (23%)

* Categories are not exclusive.

Rate of repeated tests

In the post-visit testing group, the median number of laboratory tests ordered by the general practitioner was 14 (interquartile range [IQR] 10 – 18) per patient, of which 8 (IQR 4 – 11) were repeated in the outpatient clinic. Overall, 1,440 out of 2,587 tests (56%) were repeated. Median time interval was 32 days (IQR 19 – 67). Test with abnormal results were repeated more frequently than those with normal results (337/516 (65%) vs. 1,103/2,071 (53%); $p < .001$). A longer test interval was associated with slightly smaller odds of tests being repeated (OR 0.97 [0.95 – 0.99]; $p = .004$).

The number of repeated tests per patient for the two different laboratory testing strategies is reported in Table 2. Pre-visit testing led to more repetitions (median 9.5 [IQR 6 – 13]) compared to post-visit testing (median 8 [IQR 4 – 11]; $p < .001$).

The ten tests most commonly ordered by the general practitioners are listed in Table 3. At the top of this list are hemoglobin, mean corpuscular volume and white blood cell count. With the exception of glucose, the tests most commonly ordered by the general practitioners are also among the most commonly repeated tests, with common hematology parameters and creatinine being repeated in over 85% of cases.

Rate of repeated tests remaining normal

Of all tests with normal results that were repeated, 1,678 (90%) remained normal on repetition. This was independent of the testing interval (OR 0.99 [0.97 – 1.01]; $p = .1$) or testing strategy (OR 1.06 [0.73 – 1.53]; $p = .76$).

The rates of normal test results remaining normal on repetition were even higher for the most commonly ordered tests: more than 95% of the repeated tests for

Table 2: number of repeats of primary care physician's tests depending on laboratory testing strategy in the outpatient clinic

	Pre-visit testing (n = 104)	Post-visit testing (n = 191)	
Tests	12.5 (8.8 – 17)	14 (10 – 18)	
Normal results	9.5 (6 – 14)	11 (8 – 14)	
Repeats overall	9.5 (6 – 13)	8 (4 – 11)	p < .001
Repeats of normal results	8 (3.8 – 10)	5 (3 – 9)	p < .001

Statistically tested using the Wilcoxon rank-sum test.

Table 3: most commonly ordered tests with repetition rates and rates of normal tests remaining normal on re-testing

Test	Tests by primary care physician n	Repeats n (% of total)	Normal results on first test being repeated n (% of normal results)	Normal repeats remaining normal n (% of normal repeats)
Hemoglobin	247	217 (88%)	141 (84%)	134 (95%)
MCV	237	204 (86%)	170 (85%)	163 (96%)
Leukocytes	227	198 (87%)	154 (85%)	136 (88%)
Creatinine	213	185 (87%)	152 (85%)	147 (97%)
ESR	198	151 (76%)	96 (77%)	95 (99%)
TSH	192	141 (73%)	129 (75%)	125 (97%)
Glucose	191	114 (60%)	73 (57%)	57 (78%)
Thrombocytes	188	162 (86%)	142 (85%)	136 (96%)
ALT	164	136 (83%)	111 (80%)	103 (93%)
Leukocyte differential	156	114 (73%)	106 (72%)	82 (77%)

ALT: alanine transaminase; ESR: erythrocyte sedimentation rate; MCV: mean corpuscular volume; TSH: thyroid stimulating hormone.

normal results of hemoglobin, mean corpuscular volume, creatinine, erythrocyte sedimentation rate, thyroid stimulating hormone and thrombocyte count remained normal.

The tests with the highest rates of normal test results becoming abnormal on repetition are shown in Table 4. The urine screening was the only test in which normal test results became abnormal on repetition in the majority of cases.

TABLE 4: TESTS WITH THE HIGHEST RATES OF NORMAL TEST RESULTS BECOMING ABNORMAL ON RE-TESTING

Test	Normal results being repeated n (% of normal results)	Normal repeats remaining normal n (% of normal repeats)
Urine screening	14 (82%)	4 (29%)
Creatine kinase	2 (22%)	1 (50%)
Iron	3 (14%)	2 (67%)
Cholesterol	4 (6%)	3 (75%)
Reticulocytes	8 (29%)	6 (75%)
Leukocyte differential	106 (72%)	82 (77%)
Potassium	90 (74%)	70 (78%)
Glucose	73 (57%)	57 (78%)
Sodium	83 (75%)	67 (81%)
Ferritin	11 (31%)	9 (82%)

DISCUSSION

Summary

This study shows that repeated ordering of tests is common on referral from general practice, with 56% of tests ordered by the general practitioner being repeated in the outpatient clinic.

Our study is the first study to specifically investigate the rate of repeating laboratory tests on referral from the general practitioner to the outpatient clinic.

We found that tests with normal results are repeated only slightly less frequently than tests with abnormal results. In the majority of cases, tests with normal results remained normal on repetition, which suggests a high level of redundancy in laboratory test repetition. This may be especially true for tests such as the erythrocyte sedimentation rate, thyroid stimulating hormone or creatinine, which are generally known to be stable over time and subsequently almost never became abnormal on repetition in our study.

Strengths and limitations

A strength of our study is that it includes an unselected group of patients with a wide variety of referral reasons, as the original study included all newly referred patients to the outpatient clinic, which increases the external validity of the study. Furthermore, this is the first study we know of investigating laboratory test redundancy on referral from general practice to the outpatient clinic.

Our study also has several limitations. Firstly, we had no information to determine why the treating physicians at the outpatient clinic decided to repeat some tests and consequently, we could not quantify the appropriateness of these repetitions using retrospective data. Secondly, this study was conducted in a single tertiary center, so the results may not be applicable to other settings. Finally, our hospital's reference ranges were also used for the tests ordered by the primary care physicians in other laboratories, because those laboratories' reference ranges were not available. As a result, several laboratory tests may have been misclassified as either normal or abnormal.

Comparison with existing literature

The repetition rate found in our study is significantly higher than in a previous Dutch study which found only 0.5% of laboratory tests to be duplicates when studying data from patients who had laboratory testing done both at the primary care physician and at the hospital.[17] However, in this study the number of duplicate tests was compared to the total number of tests performed at the hospital's laboratory, not just the tests performed in newly referred patients. Furthermore, they only considered tests to be duplicates if repeated within a seven day time period.

Our results can also be compared to data from studies investigating a similar question but in a different setting. For example, one study focused on repetitions of laboratory tests ordered by the primary care physician when patients on admission to the hospital.[12] In this study, 300 consecutive patients who had been admitted to the medical department of a general hospital were included. Their primary care physicians were asked to provide information on laboratory testing performed in the previous year. Data were available for 202 patients and showed that 63% of the tests ordered by the primary care physician in the previous year were repeated on admission. However, as these patients were admitted to the hospital, there was an a priori higher chance of the results from the repeated test being abnormal. This underlines the notion that repetition of a laboratory test does not necessarily mean redundancy as the patient's condition might change rapidly.

The definition of whether or not a repeated test is redundant is hampered by the absence of good data and guidelines as well as reasons related to the psychology of

both the patient and the doctor. In many cases therefore, tests may not be completely appropriate or inappropriate, but fall into a grey zone.[18]

Appropriateness of repeated laboratory test is thus often simplified. For example, one study simply defined repeat tests as appropriate if the results of the initial tests were available to the physician ordering the repeat test, on the basis that there must have been a good reason for repeating.[12]

Two other studies which have focused on repetition rates on transfer between hospitals have also investigated the appropriateness of repeat testing, using expert panels to assess each test separately. They respectively found 63% and 99.5% of repetitions to be inappropriate.[8,10] We could find no studies on whether or not the physician requesting the tests thought of these tests of redundant retesting or what the reasons for the retesting were.

There are several possible causes of inappropriate repeat testing. The physician in the outpatient clinic may not be aware of the existing data due to insufficient preparation. This might be due to time constraints, or maybe the referral letter fail to be read because they are believed to be of poor quality.[19] Another reason could be that the results from the first test are not available to the doctor or only in an unsuitable format, such as a hard copy on paper, or embedded in a referral letter which has been faxed or sent as a PDF. Furthermore, doctors might distrust the results included in the referral letter based on biased attitudes towards other laboratories. This might be compounded by the differences in reference ranges between laboratories, which makes it more difficult for clinicians to correctly interpret the test result. This can also account for doctors wanting a baseline measurement in their own electronic patient file for future reference. Besides these reasons there might also be psychological reasons involved, including the physician's risk aversion and the perceived expectations of the patient.[20,21]

Implications for research and/or practice

The high prevalence of repeat testing, a significant proportion of which is likely to be redundant, provides an important target for improving laboratory test utilization, which consequently can lead to cost savings and improved patient care, by reducing the burden of venipunctures and unnecessary follow-up testing.

Possible interventions may consist of making the test results ordered by the general practitioner more readily available to other health care professionals, for instance by the coupling of different laboratory information systems, so the electronic medical records can provide clinicians with an integrated view of all laboratory tests performed either in the hospital, or elsewhere.[22] Whether this will actually reduce test repetitions warrants further study.

References

1. van Walraven C, Raymond M. Population-based study of repeat laboratory testing. *Clin Chem*. 2003;49:1997-2005.
2. 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(11):e78962.
3. Tapper EB, Strnad L, Shmerling R. Reducing the Burden of False-Positive Tests: It Is Time to Optimize Diagnostic Testing for Undifferentiated Conditions. *Mayo Clin Proc*. 2019;94(7):1143-6.
4. Thavendiranathan P, Bagai A Fau - Ebidia A, Ebidia A Fau - Detsky AS, Detsky As Fau - Choudhry NK, Choudhry NK. Do blood tests cause anemia in hospitalized patients? The effect of diagnostic phlebotomy on hemoglobin and hematocrit levels. *J Gen Intern Med*. 2005;20(6):520-4.
5. Korenstein D, Chimonas S, Barrow B, Keyhani S, Troy A, Lipitz-Snyderman A. Development of a Conceptual Map of Negative Consequences for Patients of Overuse of Medical Tests and Treatments. *JAMA Intern Med*. 2018;178(10):1401-7.
6. Procop GW, Keating C, Stagno P, Kottke-Marchant K, Partin M, Tuttle R, et al. Reducing Duplicate Testing: A Comparison of Two Clinical Decision Support Tools. *Am J Clin Pathol*. 2015;143(5):623-6.
7. Rogg JG, Rubin JT, Hansen P, Liu SW. The frequency and cost of redundant laboratory testing for transferred ED patients. *Am J Emerg Med*. 2013;31(7):1121-3.
8. Stewart BA, Fernandes S, Rodriguez-Huertas E, Landzberg M. A preliminary look at duplicate testing associated with lack of electronic health record interoperability for transferred patients. *J Am Med Inform Assoc*. 2010;17(3):341-4.
9. Brady H, Piggott L, Dunne SS, O'Connell NH, Dunne CP. Clustered interventions to reduce inappropriate duplicate laboratory tests in an Irish tertiary hospital. *Clin Biochem*. 2018;52:26-32.
10. Bertrand J, Fehlmann C, Grosgurin O, Sarasin F, Kherad O. Inappropriateness of Repeated Laboratory and Radiological Tests for Transferred Emergency Department Patients. *J Clin Med*. 2019;8(9).
11. Thomas SH, Orf J, Peterson C, Wedel SK. Frequency and costs of laboratory and radiograph repetition in trauma patients undergoing interfacility transfer. *Am J Emerg Med*. 2000;18(2):156-8.
12. Bülow PM, Knudsen LM, Staehr P. [Duplication in a medical department of tests previously performed in a primary sector]. Ugeskr Laeger. 1992;154(37):2497-501.
13. Burger PM, Westerink J, Vrijsen BEL. Outcomes of second opinions in general internal medicine. *PLoS One*. 2020;15(7):e0236048.
14. Mustafa MK, Bijl M, Gans R. What is the value of patient-sought second opinions? *Eur J Intern Med*. 2002;13(7):445-7.
15. Vrijsen BEL, ten Berg MJ, Naaktgeboren CA, Vis JY, Dijstelbloem HM, Westerink J, et al. The Impact of a Standardized Pre-visit Laboratory Testing Panel in the Internal Medicine Outpatient Clinic: a Controlled "On-Off" Trial. *J Gen Intern Med*. 2021.
16. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
17. van Loon D, Lagemaat O, Kooijman-Buiting AMJ. Onderzoek naar het voorkomen van dubbele diagnostiek bij verwijzing van de 1ste-lijn naar de 2de-lijn. *Ned Tijdschr Klin Chem Labgeneesk*. 2014;39:70-98.

18. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390(10090):156-68.

19. Berendsen AJ, Kuiken A, Benneker WHGM, Meyboom-de Jong B, Voorn TB, Schuling J. How do general practitioners and specialists value their mutual communication? A survey. *BMC Health Serv Res*. 2009;9:143-.

20. Kravitz RL, Cope DW, Bhrany V, Leake B. Internal medicine patients' expectations for care during office visits. *J Gen Intern Med*. 1994;9(2):75-81.

21. van der Weijden T, van Bokhoven MA, Dinant G-J, van Hasselt CM, Grol RPTM. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract*. 2002;52(485):974-80.

22. Ganslandt TA-O, Neumaier M. Digital networks for laboratory data: potentials, barriers and current initiatives. *Clin Chem Lab Med*. 2019;57(3):336-42.


LOT
A2313 2023-11-30
Exp.

Chapter 7

Inappropriate laboratory testing in internal medicine inpatients: Prevalence, causes and interventions.

B.E.L. Vrijsen, C.A. Naaktgeboren, L.M. Vos, W.W. van Solinge, H.A.H. Kaasjager, M.J. ten Berg

Ann Med Surg (Lond) 2020; 51: 48-53

Abstract

Background

To reduce overutilization of laboratory testing many interventions have been tried, but selecting the most effective intervention for a given setting is challenging. To be sustainable, interventions need to align with health care providers' needs and daily practices. This study aimed to assess the extent of overutilization and the perspectives of health care providers, which may be used to guide the choice of intervention.

Methods

The extent of inappropriate laboratory testing in internal medicine inpatients was evaluated using a database. Surveys and focus groups were used to investigate health care providers' perceptions on its causes and solutions.

Results

On average, patients had 5.7 laboratory orders done during the first week of admission, whereas guidelines advise performing laboratory testing no more than twice per week. Repeat testing of normal test results occurred in up to 85% of patients. The frequency of laboratory testing was underestimated by survey responders, even though the majority of responders (78%) thought that laboratory tests are ordered too frequently. Residents were considered to be most responsible for laboratory test ordering.

The primary causes of overutilization discussed were personal factors, such as a lack of awareness and knowledge, as well as feelings of insecurity. Regarding possible solutions, residents generally recommended educational interventions, whereas specialists tended to favour technical solutions such as lockouts.

Conclusion

Inappropriate laboratory testing is common in internal medicine. The most important causes are a lack of awareness and knowledge, especially in residents. The intervention most favoured by residents is education, suggesting educational interventions may be most applicable.

Introduction

Laboratory testing affects up to 70% of downstream treatment decisions.[1] The overutilization of laboratory tests is common and some estimate that one out of every five tests performed is unnecessary.[2] Inappropriate laboratory test utilization increases the potential for diagnostic errors when these tests give false-positive or false-negative results.[3,4] Additionally, given the large volume of laboratory testing, overutilization leads to substantial costs.[5,6]

The relevance of the issue of overutilization of laboratory tests is increasingly being recognised, as evidenced by the development of guidelines and campaigns aimed at reducing inappropriate test utilization. For example, the international Choosing Wisely campaign has encouraged professional societies to issue guidelines recommending targeted, deliberate laboratory testing.[7] However, adherence to guidelines is often poor.[8]

A systematic review of the literature revealed that many different interventions to reduce inappropriate testing have been investigated, such as educational methods, changes in the ordering system, audit and feedback methods.[9] While all interventions have been shown to reduce unnecessary laboratory testing initially, evidence on long-term sustainability is lacking. Additionally, which interventions are most (cost)-effective is unclear due to the lack of head-to-head comparisons.

At our institution, several small ad hoc initiatives to reduce overutilization have been undertaken, but these initiatives have not yet lead to any sustainable reduction of laboratory test utilization. Effective implementation of innovations in health care requires a systematic approach including an analysis of the target audience and the context they work in.[10,11]

So, when making a considered choice of which intervention to implement, information on the local practices and attitudes of health care professionals regarding laboratory testing is required.

In order to effectively implement interventions to increase appropriate laboratory testing, we investigated the current practice of laboratory testing at our department of internal medicine, and what health care professionals think about the causes of the surmised inappropriate test ordering as well as their ideas for potential solutions.

Methods

This study comprises three parts. Firstly, we performed a database study to investigate the appropriateness of laboratory testing. Secondly, we did a survey and thirdly a series of focus group interviews, both to evaluate health care workers' attitudes and perceptions of the barriers and facilitators of appropriate laboratory testing.

This study was deemed to be exempt from review by the Medical Research Ethics Committee of the University Medical Center Utrecht.

Database study

SETTING AND PATIENT POPULATION

For the database study we used data from the internal medicine department of the University Medical Center Utrecht, a 1,042-bed academic teaching hospital with about 28,000 clinical and 15,000 day-care hospitalizations and 334,000 outpatient visits annually.

At our hospital, laboratory tests are generally ordered by residents, who primarily manage the care for admitted patients. They are supervised by specialists daily. There are no restrictions on the laboratory tests the physicians can order. Laboratory tests are ordered through the electronic medical record. All tests have to be ordered individually: no fixed panels, e.g. sets of tests that are always ordered together, are used.

The venipunctures are performed by either specially trained laboratory staff or the nursing staff on the ward.

DATA SOURCE

We collected data on all laboratory requests for patients who had been hospitalised at the general internal medicine ward between June 2011 and December 2016.

Data were obtained from the Utrecht Patient Oriented Database (UPOD), an infrastructure of relational databases comprising data on patient characteristics, hospital discharge diagnoses, medical procedures, medication orders and laboratory tests for all patients treated at the University Medical Center Utrecht since 2004. The UPOD data acquisition and management is performed in accordance with current regulations concerning privacy and ethics. The structure and content of UPOD have been described in more detail elsewhere.[12]

DEFINING AND QUANTIFYING INAPPROPRIATE LABORATORY TESTING

Several measures of inappropriate overuse of laboratory testing were determined through plenary discussions among the authors. These were based on

recommendations from the Choosing Wisely campaign, such as the Netherlands Association of Internal Medicine's recommendation to perform laboratory tests no more than twice a week in clinically admitted patients,[13] and the recommendation of the American Society for Clinical Pathology to perform lipase testing instead of amylase in suspected pancreatitis,[14] and recommended minimal testing intervals.[15,16] The evaluated measures are presented in Table 1. All analyses were done in R version 3.1.2.

Survey

The survey (Table 2) was developed by consensus through discussions in our team, comprising four topics: perceptions of the frequency of overall laboratory testing, perceptions of who are involved in or responsible for the decision to perform laboratory testing, thoughts on the benefits and harms of laboratory testing, and thoughts on interventions to reduce excessive laboratory testing.

Invitations to fill out the questionnaire online were sent by e-mail to all nurses, residents and specialists working in the general internal medicine department of the University Medical Center Utrecht.

Focus groups

For the focus group participants a purposive sample was recruited from residents and specialists from the internal medicine department. Potential participants were those who had worked on the ward within the past six months. They were approached face-to-face and none declined to participate. The focus groups were organised between January and May 2018 and were prepared and conducted by three of the investigators (BV, MtB, and CN; an internist working in the same department as the focus group participants, a clinical pathologist and a clinical epidemiologist respectively).

The focus group discussions were set up according to the framework developed by Stalmeijer et al.[17] To encourage an open discussion, residents and consultants were included in separate groups, consisting of five to seven people. Prior to the semi-structured focus groups, a set of questions and topics was prepared based on the results of the database study and the survey. A summary of the survey results was presented to the focus group participants at the start of each meeting.

The number of focus groups was determined by the principle of thematic saturation.

Meetings were scheduled to last for 45 to 60 minutes and were held in a staff meeting room after working hours. No other people were present.

Data collection consisted of audio recordings and one of the investigators' taking notes.

The transcriptions of the audio recordings were coded by three investigators (CN, MtB, and BV) independently, using the methods described by Ose.[18] A conventional content analysis was used to analyse the data, meaning that coding categories were derived directly from the text.[19] The resulting codes were combined into one coding system and categorised by one researcher (BV). The categorization was checked by the other two coding investigators (CN and MtB). The emerging themes are discussed and supported by quotations.

Results

Database study

The results of the database study can be found in Table 1. In the study period there were 3,938 admissions to our ward for 3,122 unique patients. A total of 29,993 lab orders including 261,859 individual clinical chemistry tests were ordered. The median length of hospitalization was 4.1 days (interquartile range 1.8 – 8.3). The mean number of laboratory test orders was 5.7 during the first week of admission, which is well above the Dutch Society of Internal Medicine's recommendation to order lab no more than twice per week. The repeat rate for lab results within normal ranges differs per test, but is generally high, with sodium, potassium, and bicarbonate having the highest rates, at over 80%.

Of the 923 admissions via the emergency room who had a C-reactive protein (CRP) test repeated during admission, 548 patients had (59%) repeat CRP tests performed within 24 hours, the recommended minimum testing interval.[16] In 87.8% of repeat CRPs, no effect on patient management, defined as a change in antibiotic therapy, was seen.

Combinations of laboratory tests were common: sodium and potassium were combined in 95% of cases, alanine transaminase (ALT) and aspartate transaminase (AST) in 97%, lipase and amylase in 74%. Procalcitonin, on the other hand, was ordered in combination with CRP in only 8% of CRP test orders.

Table 1: DATABASE STUDY RESULTS

Measure	Result	Explanation
Average number of laboratory orders per patient per week	Week 1: 5.7 Week 2: 3.2 Week 3: 3.2 Week 4: 3.9	Per patient the number of lab orders were counted per week of admittance. Lengths of admittance was rounded up to the next full week.
Repetition of normal test results	Sodium 82% (n = 2,833) Potassium 85% (n = 3,032) Bicarbonate 83% (n = 597) Creatinin 75% (n = 2,142) Leukocytes 71% (n = 2,463) GGT 23% (n = 1,513) ALP 31% (n = 2,530) ALT 33% (n = 2,919) AST 34% (n = 2,476) LD 35% (n = 2,326) CRP 45% (n = 1,257)	The percentage of tests that are repeated when the test result is within the reference range.
Time from admission to repetition of CRP (n = 923)	167 (18%) within 12 hours 381 (41%) 12 – 24 hours 211 (23%) 24 – 48 hours 164 (18%) more than 48 hours	For all patients in whom a CRP is tested in the Emergency Department and in whom a repeat CRP test was performed during hospitalization, the time in hours between the CRP testing in the Emergency Department and the first subsequent CRP testing during admission.
Percentage of repeated CRP measurements that led to changes in patient management (n = 509)	6.9% start antibiotics 5.1% stop antibiotics 0.2% switch antibiotics 87.8% no effect on antibiotic treatment	The fraction of repeat CRP tests that led to initiating, discontinuing or changing antibiotic therapy (defined as a new medication order or a stopping order within 4 hours of the CRP test).
Inappropriate fixed combinations of tests	Sodium + potassium 95% ALT + AST 97% lipase + amylase 85% CRP and procalcitonin 8% Creatinine + BUN 74%	E.g. the fraction of lab orders with a sodium test that also include a potassium test

A lab order is defined as a single blood collection and can contain one or several individual laboratory tests.

ALP: alkaline phosphatase; ALT: alanine transaminase; AST: aspartate transaminase; BUN: blood urea nitrogen; CRP: C-reactive protein; GGT: gamma glutamyltransferase; LD: lactate dehydrogenase

Survey

Response rates for the survey were 59% (13/22) for specialists, 14% (10/71) for residents, and 85% (17/20) for nurses. The survey results are presented in Table 2 and Figure 1. On average, respondents underestimated the weekly number of laboratory orders, with nurses' estimates being the most accurate. The majority of respondents (78%) believed the frequency of lab ordering is too high. Also, the majority of respondents (78%) considered the responsibility for ordering lab tests to lie primarily with the residents.

Regarding the benefits and harms of laboratory testing, 52% of respondents were concerned about the negative consequences of overutilization of laboratory testing for patients. 36% were concerned about the costs of laboratory testing, while only 5% claimed to have insight into the costs. The residents' and specialists' answers were very similar for all but two items. First, residents were more likely than specialists to agree with the statement that frequent laboratory testing helps them to monitor their patients' condition (70% vs. 18%). Second, residents were less likely than specialists to agree with the statement that laboratory testing is discussed during supervision (33% vs. 69%).

Focus groups

Three focus group discussions were held: the first two with residents (R) and the third with specialists (S), comprising 15 participants in total. The causes and solutions of overuse of laboratory testing discussed were divided into three broad categories by the researchers: personal factors, organizational factors and technical factors and further categorized in sub-topics (Table 3).

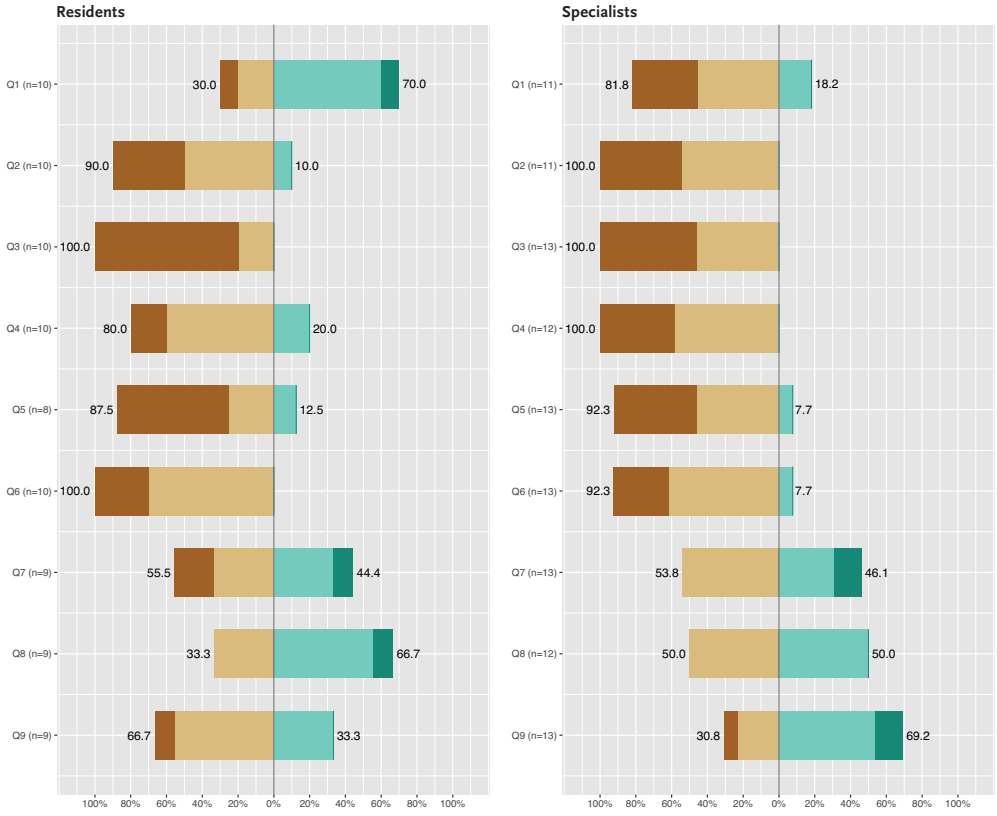
CAUSES: PERSONAL FACTORS

Several personal factors were said to affect inappropriate test ordering. Participants mentioned a lack of knowledge on laboratory testing in general. As one respondent put it: *"How much work [laboratory testing] is, how much it costs, how much normal results can fluctuate, things like that, I think we know very little about that. At least I don't."* (R8) Residents also noted that improving laboratory ordering is a learning process, in the sense that their lab ordering skills had improved over time during their residency. *"I do think about it more than before. Now, I try to consider whether I really need to order everything."* (R4) The specialists confirmed this, stating that a certain amount of inappropriate testing is acceptable in order to accommodate the residents' learning process. One specialist used a metaphor to make his point: *"If you let me run the ward, things will go faster and better. And I accept that we don't. Because we want to teach … My child's first time on a bike on the road is a hazard for scratches on other cars. But otherwise they never learn how to ride a bike."* (S1)

Table 2: survey results

	Resident	Specialist	Nurse	Overall	
How many times per week are laboratory tests ordered per patient?	2.9	3.8	4.8	3.9	n
On average, how many individual laboratory tests are in one order?	8.2	10.8	6.5	8.2	n
How many times per week are add-ons ordered after laboratory tests have been performed?	1.6	2.0	2.9	2.3	n
What do you think of the amount of laboratory testing that is being ordered?	0.80	0.85	0.71	0.79	% too many or far too many
To what extent do you personally decide which laboratory tests are being performed?	1.00	0.92	0.00	0.55	% most of the time or always
To what extent do you personally decide how often laboratory tests are being performed?	1.00	0.69	0.06	0.50	% most of the time or always
Who is mostly responsible for deciding which laboratory tests are being performed?	0.80	0.69	0.82	0.78	% choosing residents
Who is mostly responsible for the frequency of laboratory testing?	1.00	0.69	0.82	0.83	% choosing residents
To what extent do you agree or disagree with the following statements?					% agree or strongly agree
I can monitor patients better if laboratory tests are performed more often.	0.70	0.18	0.50	0.46	
Daily laboratory testing increases patient safety.	0.10	0.00	0.25	0.14	
On the day of discharge, patients should have laboratory tests performed.	0.00	0.00	0.14	0.05	
If less laboratory testing is performed, patient safety will be negatively impacted.	0.20	0.00	0.38	0.21	
If less laboratory testing is performed, patient satisfaction will be negatively impacted.	0.13	0.08	0.13	0.11	
I have insight into the costs of laboratory testing.	0.00	0.08	0.06	0.05	
I worry about the costs of laboratory testing.	0.44	0.46	0.21	0.36	
I worry about the negative consequences of laboratory testing for patients.	0.67	0.50	0.47	0.53	
Ordering laboratory tests is a standard topic of discussion during supervision.	0.33	0.69	0.43	0.52	

Figure 1: survey results



Questions

1. I can monitor patients better if laboratory tests are performed more often.
2. Daily laboratory testing increases patient safety.
3. On the day of discharge, patients should have laboratory testing performed.
4. If less laboratory testing is performed, patient safety will be negatively impacted.
5. If less laboratory testing is performed, patient satisfaction will be negatively impacted.
6. I have insight into the costs of laboratory testing.
7. I worry about the costs of laboratory testing.
8. I worry about the negative consequences of laboratory testing for patients.
9. Ordering laboratory tests is a standard topic of discussion during supervision.

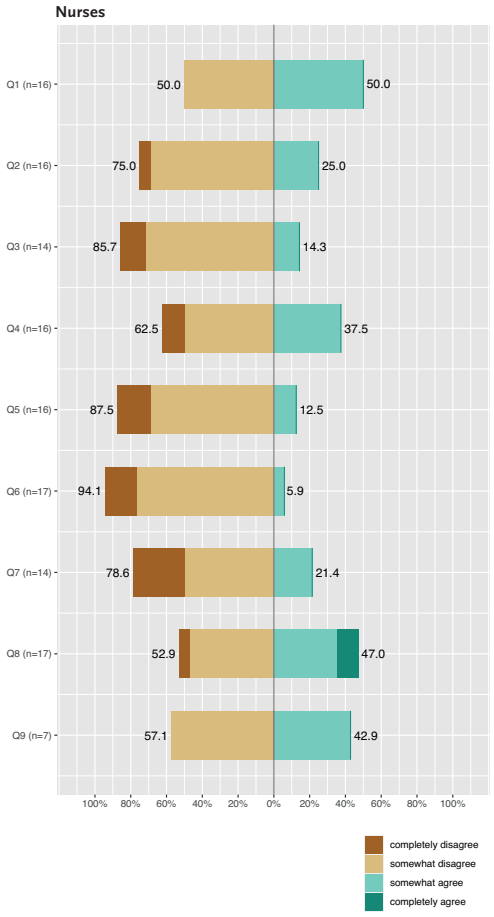

Nurses
Q1 (n=16)
Q2 (n=16)
Q3 (n=14)
Q4 (n=16)
Q5 (n=16)
Q6 (n=17)
Q7 (n=14)
Q8 (n=17)
Q9 (n=7)
50.0
50.0
75.0
25.0
85.7
14.3
62.5
37.5
87.5
12.5
94.1
5.9
78.6
21.4
52.9
47.0
57.1
42.9
100%
80%
60%
40%
20%
0%
20%
40%
60%
80%
100%
completely disagree
somewhat disagree
somewhat agree
completely agree

Table 3: summary of focus group results

	Causes	Solutions
Personal factors	- lack of knowledge regarding laboratory testing - overtesting being accepted in the context of the residents' learning curve - insecurity of the ordering physician - lack of awareness - lab testing being considered trivial	- creating awareness - reflecting on consequences - conferring with peers
Organizational factors	- specialists not providing feedback to residents - time constraints - lack of education	- more supervision by specialists - feedback on the amounts of testing - feedback and training by clinical pathologists - education
Technical factors	- ease of laboratory testing - not being able to cancel orders	- pop-ups - automated lock-outs

Participants also said often more tests were being ordered than strictly necessary due to the ordering physician's insecurity. As one supervisor said: *"Lab testing is often only done for the doctor's peace of mind."* (S2)

The personal factor that was being discussed the most was a lack of awareness about test ordering. *"I think that it is something you say very easily: oh, let's do some labs tomorrow. And you may not be aware that, when it's the afternoon, tomorrow is only twelve hours away."* (R9) Another admitted *"I don't give it much thought really. It's not completely unthinkingly, but to say that it is completely thought through, no."* (R6)

Many participants expressed that they consider overutilization to be a relatively unimportant issue. *"Nothing really can go wrong,"* said one participant (R9), and another considered it *"too trivial"* (R8).

At the same time, most participants did express concern about the consequences of inappropriate testing. Most participants were primarily concerned about the consequences for patients, such as *"It is a burden to have a venipuncture every day"* (S4) and *"We're making them anemic."* (S3), while the financial consequences were considered to be unimportant. As some participants said: *"The costs aren't that high."* (S4) and *"Adding a CRP test costs, I think, only a few euros."* (R6). Yet, these consequences do not appear to be an important factor in the actual decision-making on test ordering. As one participant stated: *"I think we don't have a lot of problems with it ourselves. It's more a problem financially and a burden on patients and nurses. But I don't think it affects me personally if too many lab tests are done."* (R8)

CAUSES: ORGANIZATIONAL FACTORS

Participants discussed multiple organizational factors that cause inappropriate test ordering. The most important factor, according to residents, was a lack of adequate supervision and feedback from their supervisors on their ordering behaviour and culture of not questioning which tests a supervisor suggests. For instance: *"Very little attention is paid to lab ordering. Also when I'm being supervised. That is my opinion at least."* (R7) or *"Well, often the supervisor just says to run some tests, and I just accept that without question."* (R3).

Supervisors agreed that they generally did not discuss test ordering with the residents. As one supervisor stated: *"The things I check on a detailed level are the things that make patients live longer or not. But not what lab tests to do tomorrow."* (S1) Time constraints were mentioned as an important cause of the supervisors not discussing lab ordering with the residents. As one supervisor said: *"Yes, when the supervisor comes running past, then the resident doesn't bring up lab tests, I think."* (S2) Not all residents expressed a desire to be supervised on laboratory testing. As one resident put it: *"Let me just figure some things out for myself."* (R3).

Both residents and supervisors said there was hardly any formal education on laboratory testing: *"You are totally dependent on your direct supervisor for what you learn about it."* (R3) Residents also indicated that there is variability between hospitals. One resident recalled: *"I recently ordered a lipase, but then the gastroenterologist called me and said: in this hospital, we always combine it with an amylase."* (R3)

CAUSES: TECHNICAL FACTORS

The technical factor that was mentioned most often by participants was the ease of laboratory testing, both with regards to the ordering as to the actual blood drawing, especially in patients with intravascular access readily available. *"Central venous access or an intra-arterial line does lower the barrier,"* one participant (R1) mentioned. The digital order form was perceived to be an important facilitator, because of the lack of any barrier for adding more tests. For instance residents stated that when *"Checking boxes on the lab form, I often go, let's do this one too, and that one..."* (R1) and that*"When you're ordering lab tests, it is easy to just order some more tests."* (R6)

Several of the residents also mentioned the inconvenient process of cancelling laboratory orders which involves calling the laboratory. "*Because you order a test, and then, later, you think, oh silly goose, let's cancel it, but then you have to call. And something else comes up and you forget to call*." (R2)

SOLUTIONS: PERSONAL FACTORS

The possible solutions that were discussed parallelled the causes. With respect to personal factors, most participants argued that creating awareness was essential. As one participant said: "*You have to think about every single laboratory test ordered.*" (S2). Another said: *"It is all about doing things consciously. And that consciousness has to be created."* (S3). Reflecting on the consequences of inappropriate testing was thought to be an effective way to increase awareness: *"I noticed a great difference when, for my research, I had to do venipunctures myself. Then you notice how much work it is, and what you're doing to a patient, especially if you have to do a second blood draw."* (R8)

Another way to increase awareness that was mentioned was to confer with peers. For instance: *"When you're unsure, you could ask your fellow resident. We do that sometimes, but not as often as we could."* (R2). Another resident suggested a weekly evaluation: *"On the ward, there are two or three residents. So, at the end of the week, you ask each other: how many lab tests did you order and looking back, was it all necessary?"* (R6)

SOLUTIONS: ORGANIZATIONAL FACTORS

With regards to organizational factors several possible solutions were discussed. In the focus groups with residents, the role of the supervisor was discussed at length. Many residents believed more supervision by the specialists could improve lab ordering behaviour. As one resident said: *"Becoming more aware of the problem only happens when someone points it out to you."* (R1) There were some concerns about the manner of feedback: *"It has to be practicable. I mean – I would become rather grouchy if it was hammered home every day during supervision."* (R3) and *"I don't want to be micro-managed."* (R9)

Residents generally felt that feedback on the volume and appropriateness of testing performed would stimulate them to be more critical, functioning as *"a wake-up call"* (R5). Both residents and specialists would appreciate feedback and training by clinical pathologists. For example:

"I believe that would be incredibly useful. Just more background information and more awareness." (R8)

"Yes, I think it would be very good if you [clinical pathologists] would be more critical about when new lab tests are needed and what to order." (S4)

Residents and specialists had different opinions on the effectiveness of education in reducing inappropriate testing. Residents were generally favourable. As one put it: *"I think education is extremely useful. Just more background knowledge and more awareness."* (R8). Most specialists were not as convinced that education would lead to changes in lab ordering behaviour: *"I have the feeling that, in that respect, education really has zero effect."* (S2)

SOLUTIONS: TECHNICAL FACTORS

Technical solutions that were suggested included pop-up messages and automated lockouts. Pop-up messages were thought by some to be potentially effective. As one specialist said: *"You might start to think: oh, maybe it's not necessary."* (S3). However, most participants were skeptical: *"I don't think that you will still give [pop-up messages] much thought if you see them every day."* (R4). Another concern was that pop-up messages would be *"rather annoying"* (R5).

Regarding lockouts, residents and specialists had different opinions. Most residents were skeptical, saying for instance that *"they only increase the work load."* (R8), whereas specialists were mostly in favour: *"With everything that's been tried, also in other hospitals, if you want this, it only works top down. And apparently, it has to be done with lockouts, because the other measures don't work."* (S4). Still, specialists did not think that lockouts alone would be a good strategy. *"The downside of only applying restrictions is that you lose the opportunity to actually teach the residents."* (S1).

Discussion

While the prevalence of laboratory test overutilization is known to be high,[2] this study also showed that clinicians often underestimate the actual extent of overutilization. We found health care providers to be ambivalent about the problem. On the one hand, most respondents indicated that they consider overutilization to be a relevant problem, both in terms of patient safety and financially. On the other hand, however, most focus group participants admitted to considering the problem relatively trivial in comparison to other aspects of medical care. Even when physicians profess to find laboratory test overutilization important, in daily practice they don't give laboratory test ordering much thought, citing time constraints.

The most important causes of overutilization, as identified through focus groups, were personal factors, such as a lack of awareness of overutilization and knowledge about appropriate testing, and feelings of insecurity. The causes of overutilization identified by our focus groups are similar to causes found in other studies, such as a lack of understanding of costs, diagnostic uncertainty, and fear of not having the lab results when requested by supervisors.[20-22] Fear of malpractice suits, which has also been found to be a driver of overtesting, could not be identified as such in this study.[23]

With regards to potential solutions, opinions differed on what would be the most effective interventions to reduce overtesting. Most residents said they would appreciate more education and direct feedback on appropriate laboratory test utilization, whereas most specialists were more in favour of technical solutions, such as lockouts.

Most survey respondents believed that the residents were most responsible for laboratory ordering, and therefore, it may be argued that residents should be targeted, and that the intervention most favoured by them, more education, may be the most applicable. As many focus group participants stated, clinical pathologists can play a vital role here.

On the other hand, residents frequently rotate between departments, so interventions that only target residents may not produce lasting effects. Therefore, specialists need to be involved and a multifaceted approach that addresses the needs of both residents and specialists may be warranted. It may be effective to combine educational measures with automated lockouts, which are relatively easy and inexpensive to implement, and have the bonus of providing feedback at the moment of test ordering.[9]

One of the strengths of this study is that we used multiple research modalities, including a database study, a survey and focus groups, to look at the topic from different angles. Also, we included nurses, residents and specialists in this study, which provides insight from most health care workers involved.

This study has several limitations. First of all, we could not give exact estimates of the amount of inappropriate lab ordering. Because of the large number of hospital admissions, it was not feasible to perform a chart review on all individual admissions to determine whether a test order was actually appropriate or not, but instead we looked at aggregates of the laboratory test results only.

Also, the question of what constitutes inappropriateness does not have a clear cut answer.[24] Not all of the measures we evaluated are covered by guidelines, and even when guidelines apply, they do not always provide unambiguous answers. For instance, in the full text of the Netherlands Association of Internal Medicine's recommendation to order laboratory tests no more than twice a week, the phrase "unless indicated" is added, which begs the question.[13]

Secondly, this is a single centre study conducted in a large academic hospital. The results may therefore not be generalisable to other settings, as overutilization of laboratory tests has been shown to be more common in teaching hospitals than in general hospitals.[25]

Thirdly, the response rate of residents in our survey was low. This may have affected the outcomes, as residents who are more concerned about the harms of overutilization could be more likely to fill out the questionnaire.

In conclusion, laboratory test overutilization is a common problem with many causes. The most important causes we found were a lack of awareness and knowledge.

Interventions to reduce overutilization that were most favoured in our study were education and automated lock-outs.

This study can be used as a template for others to identify local practices and causes of inappropriate laboratory test utilization, which can help identify which interventions are most likely to be successful.

References

1. Forsman RW. Why is the laboratory an afterthought for managed care organizations? *Clin Chem*. 1996;42(5):813-6.
2. 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(11):e78962.
3. Korenstein D, Chimonas S, Barrow B, Keyhani S, Troy A, Lipitz-Snyderman A. Development of a Conceptual Map of Negative Consequences for Patients of Overuse of Medical Tests and Treatments. *JAMA Intern Med*. 2018;178(10):1401-7.
4. Sarkar MK, Botz CM, Laposata M. An assessment of overutilization and underutilization of laboratory tests by expert physicians in the evaluation of patients for bleeding and thrombotic disorders in clinical context and in real time. *Diagnosis (Berl)*. 2017;4(1):21-6.
5. Strauss R, Cressman A, Cheung M, Weinerman A, Waldman S, Etchells E, et al. Major reductions in unnecessary aspartate aminotransferase and blood urea nitrogen tests with a quality improvement initiative. *BMJ Qual Saf*. 2019;28(10):809-16.
6. Vegting IL, van Beneden M, Kramer MH, Thijs A, Kostense PJ, Nanayakkara PW. How to save costs by reducing unnecessary testing: lean thinking in clinical practice. *Eur J Intern Med*. 2012;23(1):70-5.
7. Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf*. 2015;24(2):167-74.
8. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282(15):1458-65.
9. Bindraban RS, Ten Berg MJ, Naaktgeboren CA, Kramer MHH, Van Solinge WW, Nanayakkara PWB. Reducing Test Utilization in Hospital Settings: A Narrative Review. *Ann Lab Med*. 2018;38(5):402-12.
10. Jalbert R, Gob A, Chin-Yee I. Decreasing daily blood work in hospitals: What works and what doesn't. *Int J Lab Hematol*. 2019;41 Suppl 1:151-61.
11. Grol R, Wensing M, Eccles M, Davis D. Improving patient care: the implementation of change in health care: John Wiley & Sons; 2013.
12. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-9.
13. Nederlandse Internisten Vereniging. Verstandige keuzes bij interne geneeskunde [Available from: https://internisten.nl/werken-als-internist/beroepsinformatie-divers/publicaties/ verstandige-keuzes-bij-interne-geneeskunde].
14. Barbieri JS, Riggio JM, Jaffe R. Amylase testing for abdominal pain and suspected acute pancreatitis. *J Hosp Med*. 2016;11(5):366-8.
15. Bruns AH, Oosterheert JJ, Hak E, Hoepelman AI. Usefulness of consecutive C-reactive protein measurements in follow-up of severe community-acquired pneumonia. *Eur Respir J*. 2008;32(3):726-32.
16. The Royal College of Pathologists. National miminum retesting intervals in pathology. 2015 [Available from: https://www.rcpath.org/asset/253E8950-3721-4AA2-8DDD4 BD94F73040E/.
17. Stalmeijer RE, McNaughton N, Van Mook WN. Using focus groups in medical education research: AMEE Guide No. 91. *Medical teacher*. 2014;36(11):923-39.
18. Ose SO. Using Excel and Word to Structure Qualitative Data. *J Appl Soc Sci*. 2016;10(2):147-62.

19. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15(9):1277-88.

20. Sedrak MS, Patel MS, Ziemba JB, Murray D, Kim EJ, Dine CJ, et al. Residents' self-report on why they order perceived unnecessary inpatient laboratory tests. *J Hosp Med*. 2016;11(12):869-72.

21. Roman BR, Yang A, Masciale J, Korenstein D. Association of Attitudes Regarding Overuse of Inpatient Laboratory Testing With Health Care Provider Type. *JAMA Intern Med*. 2017;177(8):1205-7.

22. Hoffman JR, Kanzaria HK. Intolerance of error and culture of blame drive medical excess. *BMJ*. 2014;349:g5702.

23. Studdert DM, Mello MM, Sage WM, DesRoches CM, Peugh J, Zapert K, et al. Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *JAMA*. 2005;293(21):2609-17.

24. Hauser RG, Shirts BH. Do we now know what inappropriate laboratory utilization is? An expanded systematic review of laboratory clinical audits. *Am J Clin Pathol*. 2014;141(6):774-83.

25. Valencia V, Arora VM, Ranji SR, Meza C, Moriates C. A Comparison of Laboratory Testing in Teaching vs Nonteaching Hospitals for 2 Common Medical Conditions. *JAMA Intern Med*. 2018;178(1):39-47.

Chapter 8

General discussion

The general aim of this thesis was to investigate current issues with the quality of care at the crossroads of laboratory medicine and clinical medicine, and how to improve the value of laboratory testing in this setting.

The rationale for wanting to improve the value of laboratory testing lies in the increasing strains on health care, which threaten its accessibility, affordability and quality. Last year, the Dutch health ministry and several parties in the health care sector in the Netherlands have published the Integrated Healthcare Agreement (IZA) which aims to address these issues.[1] One of its goals is to improve the appropriateness of care. This can be done by reducing low value care, or by increasing the value of the care that is being provided. When it comes to the diagnostic process, value is generated by timely and accurate results that are being interpreted promptly and correctly, and that contribute to patient management, while resulting in a minimum of incidental findings, and imposing a similarly minimal burden on patients. In the clinical chemistry literature, this idea has been coined as the right test, using the right method, at the right time, to the right patient, with the right costs and for producing the right outcome.[2]

In this thesis, the focus has been on the following aspects that pertain to the value of laboratory testing: timeliness, correct interpretation and contribution to patient management. This general discussion is organized around these aspects, which will be addressed consecutively with an emphasis on the interplay between clinical and laboratory medicine. For each aspect, the results of the studies in this thesis will be discussed, including the implications for future research.

The studies in this thesis focus on internal medicine, as this is a specialty that relies heavily on laboratory testing, and we looked at three different health care settings: the emergency department, the ward and the outpatient clinic.

Timeliness

The timeliness of laboratory testing refers to the laboratory test results' being available when required. When considering the timeliness of laboratory testing, not only do the laboratory processes need to be taken into account, but rather the entire time period from the ordering of the laboratory tests to the results' becoming available to the clinician who ordered the tests and the clinician acting on them. The timeliness of laboratory testing is important as shorter turnaround times allows single-visit decision making in the setting of the outpatient clinic and can shorten the diagnostic process.

The effect of the timeliness of laboratory testing on patient outcomes is explored in chapters 2 and 4. In **chapter 2** the focus is on the timeliness of laboratory testing

in the emergency department. We found that a shorter laboratory turnaround time and a shorter time to testing (defined as the time between arrival in the emergency department and the ordering of laboratory testing) were associated with shorter emergency department length of stay.

Similar results have been found in two other studies.[3,4] The laboratory turnaround times in these studies were longer than in ours: for instance, in the study by Li et al. the median turnaround time was 58 minutes, but they only included the time from the sample's arriving at the laboratory. In our study, the median turnaround time from the sample's arriving at the laboratory was 32 minutes. This suggests that improving the timeliness of laboratory testing is beneficial even if turnaround times are already relatively short.

Turnaround times are not the most important drivers of extended emergency department length of stay, but they are an interesting target nonetheless, because they are modifiable, as opposed to other factors such as patients' acuity level.[5] We found that on average, turnaround times were 10.9 and 14.4 minutes shorter during the evenings and nights when compared to office hours.

The association between laboratory turnaround time and emergency department length of stay can in part be explained through confounding. The most obvious confounder is that patients who have a higher acuity and complexity on the one hand will have more elaborate laboratory testing done with more abnormal test results, and consequently longer turnaround times, and on the other hand will have longer emergency department length of stay, because they need more comprehensive treatments and consultations. In our study we tried to correct for these confounders by firstly only including patients for whom a fixed set of laboratory tests had been performed, and secondly by including acuity measures into the regression model.

Prospective studies or an interrupted time series design, which suffer less from confounding, may be better equipped to answer the question of causality.[6] Future studies should also focus on whether interventions to improve laboratory turnaround time actually improve clinical outcomes for patients other than just lead to shorter emergency department length of stay.

The consensus in laboratory medicine is that turnaround times for the emergency department should be less than one hour.[7,8] One way to reduce turnaround times, which was not addressed in this thesis but may still be of interest to this discussion, is through point-of-care testing. This entails the performance of laboratory tests at or in the direct vicinity of the patient's bedside.[9] Studies investigating the effect of point-of-care testing in the emergency department that have used measures such as admission rates and lengths of stay have shown conflicting results.[10-16] This may be

due to the fact that in most studies, point-of-care testing was only available for a limited number of tests and consequently, a large number of patients still required additional laboratory testing from the central laboratory. In recent years, point-of-care testing has become available for an increasing number of tests. Studies that have evaluated comprehensive point-of-care testing in the emergency department have shown reductions in emergency department length of stay.[17,18]

The most important barriers that impede the widespread use of point-of-care testing are the higher costs per test and concerns about the test accuracy, due to both the performance of the point-of-care devices themselves and their use by relatively untrained staff.[19,20] Future research should focus on addressing these implementation issues, for instance by identifying patients most likely to benefit from shorter turnaround times, and by determining what training could aid staff to better operate the point-of-care devices.

Besides through implementing point-of-care testing, reductions in turnaround times can also be achieved by innovations in the central laboratory. Many medical laboratories have increasingly automated their work processes, for instance through installing robot track systems for transport of samples or establishing auto-verification procedures to check out-of-range results. This has been shown to lead to shorter turnaround times.[21-23] Consequently, the increased speed of the laboratory processes has obviated the need to differentiate between regular and stat testing, as all turnaround times are short enough.[24]

In our own hospital, stat testing has already been abolished since 2014, as a consequence of the improved turnaround time of regular laboratory testing, with > 95% of routine laboratory test results' being reported within 1 hour. These shorter turnaround times enabled us to re-evaluate the diagnostic process in the outpatient clinic. If the turnaround time is short enough, patients can have blood tests done prior to their visit to the physician with an acceptable waiting time, rather than having to return to hear the laboratory test results on a separate occasion or come to the hospital for a phlebotomy in the days before their appointment. This could lead to a shorter time to diagnosis and fewer visits to the hospital, both of which would be beneficial from the patients' perspective. Whether this testing strategy would indeed lead to these improvements in patient care was studied in the POORT (Patient Outcomes Of Rapid Testing) study, which is covered in **chapter 4**.

The study population of the POORT study consisted of patients newly referred to the internal medicine outpatient clinic of our hospital, who were assigned to one of two groups. The first group had laboratory testing done directly prior to their visit to the outpatient clinic, so that the test results were available to the clinician during

the first visit. As the patients had not been seen by a physician yet, they all received the same comprehensive set of laboratory tests. The other group received standard care, which meant that they had laboratory tests done on request after the visit to the physician.

We found no differences in the primary outcome, which was the overall time to diagnosis, or in the median number of visits to the outpatient clinic, but in the pre-visit testing group, the chance of making a correct diagnosis during the first visit was increased, with a number needed to test of 15. This suggests that a pre-visit testing strategy is only effective in relatively simple cases that only rely on laboratory testing for their diagnosis, whereas more complicated cases require more extensive diagnostics such as imaging. Another possibility is that the standard pre-visit laboratory test panel was inadequate. This is supported by the fact that additional laboratory testing was ordered in 66% of patients in the pre-visit testing group. The most frequently added tests were ferritin, vitamin B12, and folic acid, suggesting diagnostic workups for anemia.

Something we have not addressed in our study is whether a reduced time to diagnosis would improve patients' outcomes. Although early recognition is associated with improved survival for time-sensitive conditions such as sepsis or myocardial infarction,[25,26] this remains an open question for the outpatient clinic. So, any future study on interventions that manage to reduce the time to diagnosis in the outpatient clinic should include evaluations of the effects on clinical outcomes.

Other potentially interesting patient outcomes are patient satisfaction, anxiety and resolution of symptoms. Even though we found no differences in patient satisfaction between the two groups, we did find that more than 90% of the patients preferred to get a diagnosis on the first day. This implies that pre-visit testing, which increased the chance of a diagnosis during the first visit, can add value for patients, even if overall the number of outpatient clinic visits did not differ.

Given the high rate of patients who received additional testing in the pre-visit testing group, a possible improvement could be to extend the standard test panel. However, this would bring about increased costs and more false positive test results, which might lead to additional downstream overtesting. A more promising strategy might be to individualize the laboratory panels based on the referral reason, or on the laboratory test results already ordered by the referring general practitioner. This process might be automated using artificial intelligence relying on big data and with access to all of patient's previous test results.[27] Future research is needed to see whether such automated test ordering will outperform clinicians and whether this actually leads to improved patient outcomes.

One thing to keep in mind when implementing pre-visit laboratory testing, is the risk of premature closure, which occurs when health care professionals make a diagnosis before sufficient information has been collected and plausible alternatives have been ruled out.[28] If the physician has access to the laboratory test results before they meet with the patient, the test results could push them in a certain direction and cause them to miss certain cues from the history.

Correct interpretation

Laboratory tests need to be interpreted correctly if they are to have value for patients. As described in the Introduction, clinicians frequently experience difficulty when interpreting test results. This can lead to treatment delay and diagnostic errors.[29,30] Up to 44% of diagnostic errors have been reported to be caused by errors in laboratory testing.[31]

Several potential solutions have been proposed. One of those is the so-called diagnostic management team: a team of both clinicians and pathologists, centered around a particular diagnostic problem.[32] Such teams are in line with the 2015 recommendation from the National Academy of Medicine in the United States that diagnostic processes be set up as interprofessional collaboratorive efforts between health care providers from different backgrounds.

The idea is that in a diagnostic management team, clinicians and clinical chemists can communicate and confer more easily, which leads to improved test selection and result interpretation. Of course, some of the benefits of cross-fertilization between clinicians and clinical chemists could also be reaped through other ways than a formal diagnostic management team, for instance through presence of clinical chemists during morning handovers or grand rounds.

In the studies that have evaluated diagnostic management teams, the focus has been exclusively on improved test selection.[33-35] While these studies have shown positive results, the effect of diagnostic management teams on result interpretation remains largely unknown. Also, these studies have not reported on the operating costs of such diagnostic management teams.

A perhaps more promising solution lies in digitalization, through artificial intelligence and decision support. The effects of decision support systems on diagnosis have not been studied extensively, and the studies performed so far have shown inconsistent results.[36,37] Even now, ChatGPT can answer 90% of the questions of the United States Medical Licencing Exam correctly.[38] However, in interpreting laboratory test results, ChatGPT was recently found to perform rather poorly.[39] In

any case, the field of artificial intelligence is evolving rapidly and artificial intelligence systems specifically designed for medical diagnosis are just around the corner.

Laboratory medicine has been at the forefront of digitalization in health care, which is unsurprising, given that it is a heavily data-driven specialty.[40,41] Artificial intelligence uses techniques such as data mining and machine learning to produce complex algorithms that can process data more efficiently.[42,43] Thus, we can make better use of data that are already available.

An example of how the diagnostic value of laboratory data can be improved through machine learning is given in **chapter 3**, in which we use data from the *CELL-DYN Sapphire* hematology analyzer which is used in the University Medical Center Utrecht (UMC Utrecht). Almost two decades ago a research group was started at the UMC Utrecht, focusing on investigating hematological parameters as biomarkers for better diagnosis, prediction and monitoring of diseases. To this end, a relational database system, the Utrecht Patient Oriented Database (UPOD) was established, in which all hematology data from hematology analyzers are linked to other clinical patient data.[44]

Like other modern hematology analyzers, the *CELL-DYN Sapphire* analyzer measures a variety of cell characteristics regarding their size and shape, besides just the cell counts. It provides approximately 80 parameters, which are normally not reported. We evaluated the diagnostic potential of these hematological parameters for the diagnosis of immune related adverse effects (irAE) caused by immune checkpoint inhibitors. This relatively new group of drugs is used in the treatment of various cancers, generally with high success rates, but at the cost of sometimes serious side effects. Differentiating these immune related adverse effects from other conditions, such as progression of cancer or intercurrent infections, can be challenging.

Using novel statistical methods and machine learning we could identify variables that are associated with irAE. We found that the eosinophil count, the red blood cell count measured with impedance, the coefficient of variance of the neutrophil depolarization and the red blood cell distribution width have added value beyond the judgment of the treating physician in determining whether irAE is present.

These hematological parameters have already been researched in prognostic research, for instance in predicting cardiovascular disease and overall mortality,[45-47] where they have shown modest, but consistent improvement in diagnostic accuracy on top of other clinical parameters. It is likely that there are more areas of diagnostic uncertainty where these parameters can be of value. Future research should focus on identifying these areas.

Rather than only using previous laboratory test results as input, which is frought with pitfalls such as data missing not at random, and a lack of clinical context,[48] future

applications of artificial intelligence systems should integrate laboratory data with other health care data. Decision support systems that incorporate laboratory results have already been developed, but their scope has been limited, and external validation has often been lacking.[49,50] Research should focus on the further development of decision support systems based on big data, to aid clinicians in the diagnosis and management of patients.

Decision support systems might also benefit from the analysis of trends in laboratory values within patients rather than comparisons between patients. Just like trends in vital parameters have been shown to improve the early detection of clinical deterioration,[51] trends in laboratory values might allow for the earlier detection of changes in patients' conditions.

An important issue is how such decision support systems will be used by health care providers. A lack of trust in decision support systems has been identified as a barrier against their implementation.[52] Furthermore, some clinicians feel that decision support systems reduce their professional autonomy and may introduce medico-legal issues.[53] If the diagnosis made by an AI system proves to be incorrect, who is responsible: the treating physician, the institution where they work or the software developer? Furthermore, the development of artificial intelligence using big data raises important questions about privacy. Most applications are controlled by commercial parties. Rigorous government oversight will be necessary to protect patients' rights.[54] Many more uncertainties remain about how the advances in AI will affect health care. In any case, health care professionals will need to be educated and involved to ensure a responsible development and implementation.[55,56]

Contribution to patient management

Laboratory testing has several roles in patient management: to make diagnoses, to screen for disease, to follow-up on chronic conditions and to monitor treatment. To what extent laboratory tests contribute to patient management depends on multiple factors, such as the timeliness and interpretation as mentioned above, but also whether the tests themselves are indicated. In this section the focus is on the diagnostic value of the tests themselves. Ordering tests that do not contribute to patient management is called overtesting. Over the past decades, cost awareness has been increasing, and this has led to many initiatives to reduce overtesting, such as the global Choosing Wisely campaign, and the national deimplementation programme "Doen of Laten" (To Do Or Not To Do) in the Netherlands.[57,58] It has been estimated that overtesting makes up 20% of all performed tests.[59] Overtesting

can be divided into two categories: inappropriate initial testing and inappropriate repeat testing. The latter refers to repeating a laboratory test at a higher frequency than warranted.[60]

In **chapter 6** the focus was on a particular example of overtesting: inappropriate repeat testing on referral from general practice to the outpatient clinic, which is covered only scantly in the literature. We performed a post-hoc analysis of the POORT study mentioned in chapter 4, comparing the laboratory tests ordered by the general practitioner (as indicated in the referral letter) and the laboratory tests ordered by the physician in the outpatient clinic. In the POORT-study, pre-visit testing was compared with post-visit testing. The pre-visit testing group used a standardized laboratory test panel, and therefore we mostly used data from the post-visit group only. We found that overall, 56% of tests were repeated, and that the majority of repetitions yielded no new information: of the 80% of tests with normal results when ordered by the general practitioner, 90% remained normal on repetition. This suggests that a large proportion of these test repetitions were redundant and thus did not contribute to patient management. This is in line with the findings from previous studies, which found 63 to 99.5% of repetitions to be inappropriate.[61,62]

In this retrospective study we did not have information on the reasons for the repetitions, so we could not quantify exactly how many of the repeated tests were inappropriate. This also reflects the more fundamental issue that there are few hard criteria for the appropriateness of laboratory testing. Several studies have used arbitrary cut-offs or expert panels whose interpretation may be hindered by hindsight bias and interobserver variability, suggested by a kappa of 0.57 on the appropriateness of laboratory test repetitions in one study.[62] In reality, the appropriateness of a laboratory test may not be a dichotomous measure, but rather a broad spectrum.[63] Nevertheless, it is evident that overutilization is highly common, notwithstanding the caveats regarding the precise amount.

Another example of laboratory test overutilization can be found in **chapter 5**, in which we report the outcomes of a retrospective study on second opinions in internal medicine. In this study, a new diagnosis was established in only 13% of patients. We found that many investigations that had been performed by the first physician were repeated during the second opinion. Laboratory tests in particular were often repeated: 89% of the laboratory testing ordered during the second opinion constituted repetitions. In only 4% of cases did the laboratory testing performed during the second opinion contribute to the diagnosis, suggesting a large rate of redundancy. This is in line with the results from a previous study, which found 86 – 90% of laboratory tests to be repeated during second opinions.[64]

The reason for this inappropriate repeat testing on referral may well lie in a failure to read the information in the referral letter, which may be thought to be of poor quality or available only in an unsuitable format.[65] Ideally, all laboratory tests should be integrally available to the clinician, regardless of where these tests were performed.[66] This requires integration of electronic health records between different health care providers. In the Dutch Integrated Healthcare Agreement (IZA), digital data exchange has been identified as a conditio sine qua non for successful cooperation between primary and secondary care.[1] By 2025, every person in the Netherlands should have digital access to their own health care data, to be easily shared with health care providers. When combining laboratory data from different laboratories, it is important to consider automated measures to reduce inappropriate testing, such as automated lockouts. In a setting where patients have laboratory tests performed in different laboratories, such measures need to operate across laboratory information systems to function efficiently.

Given the increasing strains on health care accessibility, as identified in the Integrated Healthcare Agreement (IZA), second opinions need to be performed more efficiently. Apart from improving the accessibility of laboratory test results and other patient data, possible solutions, which warrant further investigation, could be online second opinions or peer-to-peer consultations.[67]

A large number of measures aimed at reducing overutilization have been investigated.[68,69] There have been no head to head comparisons and data on long-term efficacy of these measures are mostly lacking. The success of a measure most likely relies on the local context and on whether it addresses the underlying causes.[70] Therefore, in **chapter 7**, we investigated the local practices of test overutilization, and its perceived causes in our local department of internal medicine, through a questionnaire and focus group interviews.

We found the prevalence of overtesting to be underestimated by physicians: residents' estimate of the weekly number of laboratory test orders was 49% lower than the actual incidence. Several different causes of overtesting were identified in the focus groups, which were in line with the results from other studies and included time constraints and personal factors as a lack of awareness and knowledge about appropriate testing and aversion to undertesting.[71-73] Residents were considered to be most responsible for laboratory test ordering. Arguably, measures to reduce overtesting should focus on residents.

In the focus groups, residents generally favoured educational measures to reduce overtesting. This can be done by supervisors, but also by clinical pathologists, for instance in diagnostic management teams as mentioned earlier.[74] Technical solutions,

such as automated lockouts, which were favoured by specialists in our survey, can also serve an educational purpose, if they provide feedback on why certain tests are locked out. Rather than educating physicians, a different approach to improve the efficiency of laboratory test ordering could be to use artificial intelligence to aid in the selection of the appropriate laboratory tests. Preliminary results have been promising, but the clinical application, including its acceptance by clinicians remains a topic for further study.[27,75] Arguably, the successful implementation of any intervention requires close collaboration between committed clinicians and pathologists.[76,77]

Conclusion

The value of laboratory diagnostics can be enhanced through shorter turnaround times and support in the selection and interpretation of tests, as demonstrated in this thesis. The need for improvements in the diagnostic process is made more acute by the perfect storm facing health care today, with an increased demand due to an ageing population and increased technical possibilities, and constraints on staffing.[78,79] This challenge can only be met by increased automation and digitalization.[80] Furthermore, integration of health care data across different platforms and collaboration between all health care providers involved are a requisite. The application of artificial intelligence in health care, while still in its infancy, is expanding rapidly and perhaps it can play a defining role in improving laboratory diagnostics. Applications that were suggested in this Discussion include selecting which tests to order, and interpreting the test results, using big data and previous test results for the patient in question. When evaluating the value of such decision support systems, the important question is to what extent they promote outcomes for patients. For therein lies the true value of diagnostics.

References

1. Ministerie van Volksgezondheid Welzijn en Sport. Integraal Zorgakkoord. 2022.
2. Lippi G, Bovo C, Ciaccio M. Inappropriateness in laboratory medicine: an elephant in the room? *Ann Transl Med*. 2017;5(4):82.
3. Kaushik N, Khangulov VS, O'Hara M, Arnaout R. Reduction in laboratory turnaround time decreases emergency room length of stay. *Open Access Emerg Med*. 2018;10:37-45.
4. Li L, Georgiou A, Vecellio E, Eigenstetter A, Toouli G, Wilson R, et al. The effect of laboratory testing on emergency department length of stay: a multihospital longitudinal study applying a cross-classified random-effect modeling approach. *Acad Emerg Med*. 2015;22(1):38-46.
5. Kreindler SA, Cui Y, Metge CJ, Raynard M. Patient characteristics associated with longer emergency department stay: a rapid review. *Emerg Med J*. 2016;33(3):194-9.
6. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther*. 2002;27(4):299-309.
7. Hawkins RC. Laboratory turnaround time. *Clin Biochem Rev*. 2007;28(4):179-94.
8. McKillop DJ, Auld P. National turnaround time survey: professional consensus standards for optimal performance and thresholds considered to compromise efficient and effective clinical management. *Ann Clin Biochem*. 2017;54(1):158-64.
9. Florkowski C, Don-Wauchope A, Gimenez N, Rodriguez-Capote K, Wils J, Zemlin A. Point-of-care testing (POCT) and evidence-based laboratory medicine (EBLM) – does it leverage any advantage in clinical decision making? *Crit Rev Clin Lab Sci*. 2017;54(7-8):471-94.
10. Murray RP, Leroux M, Sabga E, Palatnick W, Ludwig L. Effect of point of care testing on length of stay in an adult emergency department. *J Emerg Med*. 1999;17(5):811-4.
11. Lee-Lewandrowski E, Corboy D, Lewandrowski K, Sinclair J, McDermot S, Benzer TI. Implementation of a Point-of-Care Satellite Laboratory in the Emergency Department of an Academic Medical Center. *Arch Pathol Lab Med*. 2003;127(4):456-60.
12. Lewandrowski K, Flood J, Finn C, Tannous B, Farris AB, Benzerx TI, et al. Implementation of Point-of-Care Rapid Urine Testing for Drugs of Abuse in the Emergency Department of an Academic Medical Center: Impact on Test Utilization and ED Length of Stay. *Am J Clin Pathol*. 2008;129(5):796-801.
13. Goodacre S, Bradburn M, Cross E, Collinson P, Gray A, Hall AS. The RATPAC (Randomised Assessment of Treatment using Panel Assay of Cardiac Markers) trial: a randomised controlled trial of point-of-care cardiac markers in the emergency department. *Health Technol Assess*. 2011;15(23):1-102.
14. Kendall J, Reeves B, Clancy M. Point of care testing: randomised controlled trial of clinical outcome. *BMJ*. 1998;316(7137):1052-7.
15. Asha SE, Chan ACF, Walter E, Kelly PJ, Morton RL, Ajami A, et al. Impact from point-of-care devices on emergency department patient processing times compared with central laboratory testing of blood samples: a randomised controlled trial and cost-effectiveness analysis. *Emerg Med J*. 2014;31(9):714-9.
16. Lee EJ, Shin SD, Song KJ, Kim SC, Cho JS, Lee SC, et al. A point-of-care chemistry test for reduction of turnaround and clinical decision time. *Am J Emerg Med*. 2011;29(5):489-95.
17. Jang JY, Shin SD, Lee EJ, Park CB, Song KJ, Singer AJ. Use of a Comprehensive Metabolic Panel Point-of-Care Test to Reduce Length of Stay in the Emergency Department: A Randomized Controlled Trial. *Ann Emerg Med*. 2013;61(2):145-51.

18. Kankaanpää M, Raitakari M, Muukkonen L, Gustafsson S, Heitto M, Palomäki A, et al. Use of point-of-care testing and early assessment model reduces length of stay for ambulatory patients in an emergency department. *Scand J Trauma Resusc Emerg Med*. 2016;24(1):125.
19. Quinn AD, Dixon D, Meenan BJ. Barriers to hospital-based clinical adoption of point-of-care testing (POCT): A systematic narrative review. *Crit Rev Clin Lab Sci*. 2016;53(1):1-12.
20. Huddy JR, Ni MZ, Barlow J, Hanna GB. Qualitative analysis of stakeholder interviews to identify the barriers and facilitators to the adoption of point-of-care diagnostic tests in the UK. *BMJ Open*. 2021;11(4):e042944.
21. Jensen K, Haniff R, Kamarinos A, Rosenberg A, Santiago M, Laser J. Improving Turnaround Times through a Process Improvement Initiative Involving Barcoding, Floorplans, Dual Measuring Cells, Chemistry Analyzers, and Staff Shifts. *J Appl Lab Med*. 2019;4(3):311-22.
22. Ialongo C, Porzio O, Giambini I, Bernardini S. Total Automation for the Core Laboratory: Improving the Turnaround Time Helps to Reduce the Volume of Ordered STAT Tests. *J Lab Autom*. 2016;21(3):451-8.
23. Ellison TL, Alharbi M, Alkaf M, Elimam S, Alfaries M, Nounou RA, et al. Implementation of total laboratory automation at a tertiary care hospital in Saudi Arabia: effect on turnaround time and cost efficiency. *Ann Saudi Med*. 2018;38(5):352-7.
24. Singh G, Savage NM, Gunsolus B, Foss KA. Requiem for the STAT Test: Automation and Point of Care Testing. *Lab Med*. 2020;51(2):e27-e31.
25. Torsvik M, Gustad LT, Mehl A, Bangstad IL, Vinje LJ, Damås JK, et al. Early identification of sepsis in hospital inpatients by ward nurses increases 30-day survival. *Crit Care*. 2016;20(1):244.
26. Bates ER, Jacobs AK. Time to Treatment in Patients with STEMI. *N Engl J Med*. 2013;369(10):889-92.
27. Islam MM, Poly TN, Yang H-C, Li Y-C. Deep into Laboratory: An Artificial Intelligence Approach to Recommend Laboratory Tests. *Diagnostics [Internet]*. 2021; 11(6).
28. Krupat E, Wormwood J, Schwartzstein RM, Richards JB. Avoiding premature closure and reaching diagnostic accuracy: some key predictive factors. *Med Educ*. 2017;51(11):1127-37.
29. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med*. 2006;145(7):488-96.
30. Graber ML, Franklin N, Gordon R. Diagnostic Error in Internal Medicine. *Arch Intern Med*. 2005;165(13):1493-9.
31. 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(20):1881-7.
32. Graber ML, Rusz D, Jones ML, Farm-Franks D, Jones B, Gluck JC, et al. The new diagnostic team. *Diagnosis*. 2017;4(4):225-38.
33. Sarkar MK, Botz CM, Laposata M. An assessment of overutilization and underutilization of laboratory tests by expert physicians in the evaluation of patients for bleeding and thrombotic disorders in clinical context and in real time. *Diagnosis (Berl)*. 2017;4(1):21-6.
34. Rajendran R, Salazar JH, Seymour RL, Laposata M, Zahner CJ. Overutilization and underutilization of autoantibody tests in patients with suspected autoimmune disorders. *Diagnosis (Berl)*. 2021;8(4):497-503.
35. Seegmiller AC, Kim AS, Mosse CA, Levy MA, Thompson MA, Kressin MK, et al. Optimizing Personalized Bone Marrow Testing Using an Evidence-Based, Interdisciplinary Team Approach. *Am J Clin Pathol*. 2013;140(5):643-50.

36. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3(1):17.

37. Marcin T, Hautz SC, Singh H, Zwaan L, Schwappach D, Krummrey G, et al. Effects of a computerised diagnostic decision support tool on diagnostic quality in emergency departments: study protocol of the DDx-BRO multicentre cluster randomised cross-over trial. *BMJ Open*. 2023;13(3):e072649.

38. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med*. 2023;388(13):1233-9.

39. Cadamuro J, Cabitza F, Debeljak Z, Bruyne SD, Frans G, Perez SM, et al. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. An assessment by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI). *Clin Chem Lab Med*. 2023 Apr 24;61(7):1158-66.

40. Cadamuro J. Rise of the Machines: The Inevitable Evolution of Medicine and Medical Laboratories Intertwining with Artificial Intelligence-A Narrative Review. *Diagnostics (Basel)*. 2021;11(8).

41. Gruson D, Helleputte T, Rousseau P, Gruson D. Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation. *Clin Biochem*. 2019;69:1-7.

42. Rakha EA, Toss M, Shiino S, Gamble P, Jaroensri R, Mermel CH, et al. Current and future applications of artificial intelligence in pathology: a clinical perspective. *J Clin Pathol*. 2021;74(7):409-14.

43. Miller DD, Brown EW. Artificial Intelligence in Medical Practice: The Question to the Answer? *Am J Med*. 2018;131(2):129-33.

44. ten Berg MJ, Huisman A, van den Bemt PM, Schobben AF, Egberts AC, van Solinge WW. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-9.

45. Bindraban RS, Ten Berg MJ, Haitjema S, Hoefer IE, de Regt M, Kramer M, et al. Exploring the value of routinely measured hematology parameters for identification of elderly patients at high risk of death at the Emergency Department. *Acute Med*. 2018;17(4):188-202.

46. Gijsberts CM, den Ruijter HM, de Kleijn DPV, Huisman A, Ten Berg MJ, van Wijk RHA, et al. Hematological Parameters Improve Prediction of Mortality and Secondary Adverse Events in Coronary Angiography Patients: A Longitudinal Cohort Study. *Medicine (Baltimore)*. 2015;94(45):e1992.

47. Kofink D, Muller SA, Patel RS, Dorresteijn JAN, Berkelmans GFN, de Groot MCH, et al. Routinely measured hematological parameters and prediction of recurrent vascular events in patients with clinically manifest vascular disease. *PLoS One*. 2018;13(9):e0202682.

48. Overmars LM, Niemantsverdriet MSA, Groenhof TKJ, De Groot MCH, Hulsbergen-Veelken CAR, Van Solinge WW, et al. A Wolf in Sheep's Clothing: Reuse of Routinely Obtained Laboratory Data in Research. *J Med Internet Res*. 2022;24(11):e40516.

49. Mueller OS, Rentsch KM, Nickel CH, Bingisser R. Disposition Decision Support by Laboratory Based Outcome Prediction. *J Clin Med*. 2021;10(5).

50. Walter Costa MB, Wernsdorfer M, Kehrer A, Voigt M, Cundius C, Federbusch M, et al. The Clinical Decision Support System AMPEL for Laboratory Diagnostics: Implementation and Technical Evaluation. *JMIR Med Inform*. 2021;9(6):e20407.

51. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*. 2016;102:1-5.
52. Ankolekar A, van der Heijden B, Dekker A, Roumen C, de Ruysscher D, Reymen B, et al. Implications of Clinicians' Attitudes Towards Clinical Decision Support Systems. *Stud Health Technol Inform*. 2020;270:1319-20.
53. Catho G, Centemero NS, Catho H, Ranzani A, Balmelli C, Landelle C, et al. Factors determining the adherence to antimicrobial guidelines and the adoption of computerised decision support systems by physicians: A qualitative study in three European hospitals. *Int J Med Inform*. 2020;141:104233.
54. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics*. 2021;22(1):122.
55. Jongsma KR, van Solinge WW, Haitjema S. [Eight misconceptions about AI in healthcare]. *Ned Tijdschr Geneeskd*. 2023;167.
56. Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N Engl J Med*. 2023;388(13):1201-8.
57. Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf*. 2015;24(2):167-74.
58. To Do Or Not To Do [Available from: https://todoornottodo.nl/#sectie-2.
59. 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(11):e78962.
60. Lang T, Croal B. National minimum retesting intervals in pathology. *The Royal College of Pathologists*; 2021.
61. Stewart BA, Fernandes S, Rodriguez-Huertas E, Landzberg M. A preliminary look at duplicate testing associated with lack of electronic health record interoperability for transferred patients. *J Am Med Inform Assoc*. 2010;17(3):341-4.
62. Bertrand J, Fehlmann C, Grosgurin O, Sarasin F, Kherad O. Inappropriateness of Repeated Laboratory and Radiological Tests for Transferred Emergency Department Patients. *J Clin Med*. 2019;8(9).
63. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390(10090):156-68.
64. Mustafa M, Bijl M, Gans R. What is the value of patient-sought second opinions? *Eur J Intern Med*. 2002;13(7):445.
65. Berendsen AJ, Kuiken A, Benneker WHGM, Meyboom-de Jong B, Voorn TB, Schuling J. How do general practitioners and specialists value their mutual communication? A survey. *BMC Health Serv Res*. 2009;9:143-.
66. Ganslandt TA-O, Neumaier M. Digital networks for laboratory data: potentials, barriers and current initiatives. *Clin Chem Lab Med*. 2019;57(3):336-42.
67. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. *Health Services and Delivery Research*. 2018.
68. Müskens JLJM, Kool RB, van Dulmen SA, Westert GP. Overuse of diagnostic testing in healthcare: a systematic review. *BMJ Qual Saf*. 2022;31(1):54-63.
69. Bindraban RS, Ten Berg MJ, Naaktgeboren CA, Kramer MHH, Van Solinge WW, Nanayakkara PWB. Reducing Test Utilization in Hospital Settings: A Narrative Review. *Ann Lab Med*. 2018;38(5):402-12.

70. Jalbert R, Gob A, Chin-Yee I. Decreasing daily blood work in hospitals: What works and what doesn't. *Int J Lab Hematol*. 2019;41 Suppl 1:151-61.

71. Sedrak MS, Patel MS, Ziemba JB, Murray D, Kim EJ, Dine CJ, et al. Residents' self-report on why they order perceived unnecessary inpatient laboratory tests. *J Hosp Med*. 2016;11(12):869-72.

72. Roman BR, Yang A, Masciale J, Korenstein D. Association of Attitudes Regarding Overuse of Inpatient Laboratory Testing With Health Care Provider Type. *JAMA Intern Med*. 2017;177(8):1205-7.

73. Hoffman JR, Kanzaria HK. Intolerance of error and culture of blame drive medical excess. *BMJ*. 2014;349:g5702.

74. Laposata ME, Laposata M, Van Cott EM, Buchner DS, Kashalo MS, Dighe AS. Physician Survey of a Laboratory Medicine Interpretive Service and Evaluation of the Influence of Interpretations on Laboratory Test Ordering. *Arch Pathol Lab Med*. 2004;128(12):1424-7.

75. Cismondi F, Celi LA, Fialho AS, Vieira SM, Reti SR, Sousa JMC, et al. Reducing unnecessary lab testing in the ICU with artificial intelligence. *Int J Med Inform*. 2013;82(5):345-58.

76. Hueth KD, Prinzi AM, Timbrook TT. Diagnostic Stewardship as a Team Sport: Interdisciplinary Perspectives on Improved Implementation of Interventions and Effect Measurement. Antibiotics (Basel). 2022;11(2).

77. Bindraban RS, van Beneden M, Kramer MHH, van Solinge WW, van de Ven PM, Naaktgeboren CA, et al. Association of a Multifaceted Intervention With Ordering of Unnecessary Laboratory Tests Among Caregivers in Internal Medicine Departments. *JAMA Netw Open*. 2019;2(7):e197577-e.

78. de Meijer C, Wouterse B, Polder J, Koopmanschap M. The effect of population aging on health expenditure growth: a critical review. *Eur J Ageing*. 2013;10(4):353-61.

79. Marć M, Bartosiewicz A, Burzyńska J, Chmiel Z, Januszewicz P. A nursing shortage – a prospect of global and local policies. *Int Nurs Rev*. 2019;66(1):9-16.

80. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *Eur J Public Health*. 2019;29(Supplement_3):23-7.

Appendices

Nederlandse samenvatting

Laboratoriumonderzoek speelt een belangrijke rol in de geneeskunde, niet alleen bij het stellen van diagnoses, maar ook bij de follow-up van aandoeningen en het monitoren van behandelingen. Het wordt geschat dat 70% van de beslissingen in de geneeskunde tenminste deels berust op laboratoriumonderzoek.

Laboratoriumonderzoek gaat ook gepaard met kosten en een belasting voor de patiënt. Die belasting betreft niet alleen het ongemak van de bloedafname, maar ook mogelijk bloedarmoede door herhaalde bloedafnames en invasieve vervolgonderzoeken als gevolg van afwijkende uitslagen. De financiële kosten per test zijn in de regel laag: in Nederland liggen de prijzen van de meeste testen rond €1,00. Maar het is het enorme aantal testen dat de kosten hoog maakt. Zo worden in de Verenigde Staten meer dan 12 miljard testen per jaar verricht.

Door de stijgende zorgkosten wordt er ook kritischer gekeken naar laboratoriumonderzoek en wordt er geprobeerd om de doelmatigheid ervan te vergroten. Het is bekend dat ongeveer 20% van alle laboratoriumonderzoek overbodig is en niet bijdraagt aan de gezondheid van patiënten. Het terugdringen van overbodig laboratoriumonderzoek heeft daarom de laatste jaren steeds meer aandacht gekregen.

Maar er zijn meer zaken waar de doelmatigheid van laboratoriumonderzoek van afhangt. De resultaten van het laboratoriumonderzoek moeten snel beschikbaar zijn voor de behandelend arts, zodat er sneller op gehandeld kan worden. In meerdere situaties is aangetoond dat een snellere diagnose leidt tot betere gezondheidsuitkomsten. Daarnaast is het belangrijk dat testresultaten juist geïnterpreteerd worden. Dit blijkt in de praktijk soms best lastig. Zo wordt bijvoorbeeld 28% van de toxicologische screeningstesten verkeerd geïnterpreteerd door de aanvragend arts.

Dit proefschrift bevat meerdere onderzoeken naar de doelmatigheid van laboratoriumonderzoek, gekeken vanuit het perspectief van de interne geneeskunde. Dit is een medisch specialisme dat sterk leunt op laboratoriumonderzoek. De onderzoeken zijn gesitueerd op de spoedeisende hulp (SEH), de polikliniek interne geneeskunde en de verpleegafdeling interne geneeskunde van het UMC Utrecht.

Hoofdstuk 1 bevat de introductie van het proefschrift, waarin bovengenoemde zaken uitgebreider besproken worden.

Hoofdstuk 2 beschrijft de resultaten van een studie naar het effect van doorlooptijden van laboratoriumonderzoek op de ligduur van patiënten op de SEH. Lange ligduren op de SEH zijn onwenselijk, niet alleen omdat ze patiëntonvriendelijk zijn, maar ook omdat ze leiden tot slechtere gezondheidsuitkomsten. Een mogelijke oorzaak

van lange ligduren op de SEH is de doorlooptijd van het laboratoriumonderzoek. Met de doorlooptijd wordt de tijd bedoeld tussen het aanvragen van het laboratoriumonderzoek tot het moment dat de resultaten bekend worden.

In deze studie hebben we 23.718 SEH-bezoeken tussen 2017 en 2020 meegenomen van patiënten bij wie een standaard set aan laboratoriumonderzoek is verricht. Per SEH-bezoek werden onder meer de doorlooptijd van het laboratoriumonderzoek en de ligduur van patiënt op de SEH berekend. Het effect van de doorlooptijd op de ligduur werd bepaald door middel van een model waarbij werd gecorrigeerd voor andere mogelijke oorzaken van een langere ligduur, zoals de ernst van de klacht en het tijdstip van de dag.

We vonden dat de mediane ligduur van patiënten op de SEH 199 minuten bedroeg. De mediane doorlooptijd van het laboratoriumonderzoek was 51 minuten en deze bleek inderdaad geassocieerd met de ligduur op de SEH. Voor elke minuut dat de doorlooptijd van het laboratoriumonderzoek langer duurde, was de ligduur op de SEH 19 seconden langer.

Dat lijkt misschien maar een klein verschil, maar toch kan het wel uitmaken. De doorlooptijden die wij vonden zijn namelijk beduidend lager dan in andere studies. Verder zagen we in onze studie dat de doorlooptijd van het laboratorium 's avonds en 's nachts 11 respectievelijk 14 minuten korter was dan overdag, wat suggereert dat er overdag nog winst te behalen valt in de doorlooptijd van het laboratorium.

In **hoofdstuk 3** wordt een mogelijke diagnostische toepassing onderzocht van bepaalde hematologische laboratoriumwaarden. Als er hematologisch onderzoek wordt verricht, worden door de analyzer veel meer laboratoriumwaarden bepaald dan wat er is aangevraagd. Het gaat om ongeveer 80 waarden die normaliter niet worden gerapporteerd, maar wel worden opgeslagen in het informatiesysteem van het laboratorium. In eerder onderzoek is al aangetoond dat deze hematologische waarden de kans op overlijden kunnen voorspellen bij SEH-patiënten.

In deze studie is gekeken naar de diagnostische waarde van deze hematologische waarden bij patiënten die behandeld worden met checkpoint-remmers. Deze relatief nieuwe medicijnen zijn zeer effectief tegen meerdere vormen van kanker, maar kunnen gepaard gaan met ernstige auto-immuun-bijwerkingen. Deze bijwerkingen zijn soms lastig om te onderscheiden van andere aandoeningen en betere diagnostische mogelijkheden zijn daarom gewenst.

Voor dit onderzoek zijn de gegevens gebruikt van 409 SEH-bezoeken van patiënten die met checkpointremmers worden behandeld. Er zijn twee diagnostische modellen gemaakt: een basis-model waarin alleen de voorlopige diagnose van de arts op de

SEH, de leeftijd en het geslacht zijn meegenomen en een uitgebreid model. Dit uitgebreide model is vastgesteld door middel van machine learning, waarmee uit de onderdelen van het basis-model en de hematologische waarden vijf relevante parameters zijn gekozen. Dit zijn de voorlopige diagnose van de arts op de SEH en vier hematologische waarden: het aantal eosinofielen, het aantal rode bloedcellen gemeten d.m.v. impedantie, de variatiecoëfficiënt van de neutrofielendepolarisatie, en de spreiding van de grootte van de rode bloedcellen.

We vonden dat het uitgebreide model beter onderscheid maakte tussen patiënten met en zonder auto-immuuntoxiciteit dan het basis-model. De maat waarin dit verschil wordt uitgedrukt is de oppervlakte onder de "receiver operating characteristic" (ROC)-curve. De oppervlakte onder de ROC-curve van het basis-model was 0,67 en van het uitgebreide model 0,79. Dat betekent dat de vier hematologische waarden uit het uitgebreide model diagnostische waarde hebben bovenop de voorlopige diagnose van de arts. Deze resultaten moeten nog verder bestudeerd worden voordat ze in de praktijk kunnen worden toegepast, maar ze bieden wel een mogelijke inkijk in hoe auto-immuunbijwerkingen bij checkpointremmers ontstaan. Mogelijk zijn deze hematologische waarden ook te gebruiken in de diagnostiek van andere aandoeningen.

Hoofdstuk 4 gaat over de POORT-studie. Dit is een studie die is voortgekomen uit de verbetering van de doorlooptijden van het laboratorium, waardoor in meer dan 95% van de gevallen de uitslag van het laboratoriumonderzoek binnen één uur bekend is. Hierdoor kunnen patiënten die naar de polikliniek komen direct voorafgaand aan hun bezoek bloed laten prikken waarbij de resultaten bekend zijn tijdens het bezoek aan de arts.

We hebben onderzocht of deze werkwijze ertoe leidt dat artsen er minder lang over doen om de juiste diagnose te stellen. Hiervoor hebben we alle nieuwe patiënten op de polikliniek interne geneeskunde in twee groepen verdeeld: een groep bij wie van tevoren een standaard set laboratoriumonderzoek werd verricht zoals hierboven beschreven en een groep bij wie dat niet gebeurde. Bij hen besloot de arts tijdens het bezoek welk laboratoriumonderzoek nodig was en werd dat onderzoek dus pas na afloop verricht. Daardoor kon dus pas iets gedaan worden met de resultaten tijdens een volgend bezoek aan de polikliniek.

Er bleek geen verschil te zijn in de tijd tot het stellen van de diagnose: in beide groepen duurde dat gemiddeld 35 dagen. Ook het aantal bezoeken aan de polikliniek verschilde niet. Wel zagen we dat in de groep met van tevoren laboratoriumonderzoek bij 10% van de patiënten al tijdens het eerste bezoek een diagnose gesteld werd, terwijl dat in de groep met laboratoriumonderzoek achteraf maar bij 3% was.

Waardoor de verschillen tussen de groepen zo klein zijn, kan door deze studie niet bepaald worden. Mogelijk speelt laboratoriumonderzoek niet zo'n belangrijke rol bij het stellen van de diagnose. Een andere mogelijke verklaring is dat het laboratoriumonderzoek dat van tevoren werd verricht niet volledig was: zo vonden we dat in 66% van de gevallen waarin van tevoren laboratoriumonderzoek werd verricht, na afloop van het bezoek nog extra testen werden aangevraagd.

Hoofdstuk 5 beschijft een onderzoek naar de uitkomsten van second opinions. Een second opinion is een nieuwe beoordeling van een klacht of behandeling door een andere, onafhankelijke arts. Second opinions vinden meestal plaats omdat de eerste dokter geen diagnose kon stellen of als er geen effectieve behandeling wordt gegeven. Eerdere onderzoeken naar de opbrengst van second opinions laten sterk wisselende getallen zien. Binnen de interne geneeskunde zijn er sowieso maar weinig studies verricht. Daarom hebben wij in dit onderzoek gekeken naar de opbrengst en de kosten van second opinions in de interne geneeskunde.

We hadden gegevens van 173 patiënten. In 13% van de gevallen werd een nieuwe diagnose gesteld en in 56% van de gevallen werd een nieuwe behandeling gestart. Ongeacht behandeling werd in 28% van de patiënten verbetering van de klachten gezien. Bij de second opinions werd veel aanvullend onderzoek verricht. Zo onderging 89% van de patiënten laboratoriumonderzoek en 49% van de patiënten beeldvormend onderzoek, meestal röntgenfoto's. Een groot deel hiervan, tot wel 89%, betrof herhalingen van onderzoek dat al door de eerste arts was uitgevoerd. Slechts in een minderheid van de gevallen leidde dit herhaalonderzoek tot een nieuwe diagnose.

Hoewel second opinions dus van waarde zijn voor een aanzienlijk deel van de patiënten, suggereert het grote aantal niet-bijdragende herhalingen van aanvullend onderzoek dat de manier waarop second opinions worden uitgevoerd verbeterd kan worden.

Overbodige herhalingen van laboratoriumonderzoek zijn ook het onderwerp van **hoofdstuk 6**. Daarin is gekeken naar hoe vaak laboratoriumonderzoek dat door de huisarts is verricht later in het ziekenhuis herhaald wordt. We hebben hiervoor gebruik gemaakt van de verwijsbrieven van de patiënten uit de POORT-studie (hoofdstuk 4). In 66% van de verwijsbrieven in deze studie had de huisarts de verrichte onderzoeken gerapporteerd. Voor het aantal herhalingen is alleen gebruik gemaakt van de patiënten uit de studie-arm waarin het laboratoriumonderzoek achteraf plaatsvond. In de andere studie-arm werd namelijk een standaard set testen bepaald.

Het mediane aantal testen dat in de huisartsenbrief was vermeld was 14, waarvan er 8 werden herhaald op de polikliniek. Gemiddeld genomen betrof het 56% van alle laboratoriumonderzoeken van de huisarts. Testen met normale uitslagen werden iets minder vaak herhaald dan testen met afwijkende uitslagen. Van alle testen met normale uitslagen bleef meer dan 90% normaal bij herhaling. Voor de meest gebruikelijke testen was dat percentage nog hoger.

Dit alles suggereert dat een groot deel van de herhalingen van het laboratoriumonderzoek van de huisarts overbodig is. Dit gaat gepaard met overbodige kosten en ongemak voor de patiënt. Waarom de artsen op de polikliniek dit overbodige onderzoek aanvragen kan op basis van deze studie niet beantwoord worden. Het is in elk geval niet zo dat de artsen op de polikliniek geen beschikking hadden over de uitslag: ze stonden immers in de verwijsbrief.

Hoofdstuk 7 is opnieuw een studie naar overbodig laboratoriumonderzoek, maar dan op de verpleegafdeling. Dit onderzoek bestaat uit verschillende onderdelen. Ten eerste hebben we op basis van database-onderzoek geïnventariseerd hoeveel overbodig laboratoriumonderzoek er plaatsvindt bij opgenomen patiënten, en ten tweede hebben we onderzocht hoe artsen en verpleegkundigen hiertegen aankijken. Voor dat laatste hebben we een vragenlijst uitgezet en focusgroepgesprekken gevoerd.

Aan de hand van landelijke en internationale richtlijnen hebben we een aantal voorbeelden van overbodig laboratoriumonderzoek opgesteld. Zo is er bijvoorbeeld een landelijke richtlijn die stelt dat bij opgenomen patiënten in principe niet vaker dan twee keer per week laboratoriumonderzoek moet worden aangevraagd. Wij vonden dat het in werkelijkheid gemiddeld 5,7 keer per week gebeurde.

Zowel artsen als verpleegkundigen onderschatten de hoeveelheid laboratoriumonderzoek, en nog steeds vond 78% van degenen die de vragenlijst hadden ingevuld dat er te veel laboratoriumonderzoek werd verricht. De consensus was dat de arts-assistenten het meest verantwoordelijk zijn voor het aanvragen van laboratoriumonderzoek op de afdeling.

Uit de focusgroepgesprekken kwam naar voren dat artsen een ambivalente houding hebben ten aanzien van overmatig laboratoriumonderzoek: aan de ene kant vinden ze het een relevant probleem, maar aan de andere kant geven ze het maar weinig aandacht, onder andere door tijdsdruk en omdat er andere, belangrijkere problemen zijn. De belangrijkste oorzaken voor overmatig laboratoriumonderzoek zijn onwetendheid en onzekerheid. Over wat de beste aanpak van het probleem zou zijn, verschilden arts-assistenten en specialisten van mening: arts-assistenten

wilden graag meer onderwijs over het onderwerp, terwijl specialisten daar weinig van verwachtten en meer zagen in ICT-oplossingen.

De toenemende druk op de gezondheidszorg, als een gevolg van een vergrijzende bevolking, toegenomen technische mogelijkheden en krapte aan personeel, maakt dat de noodzaak tot het verbeteren van de doelmatigheid van laboratoriumonderzoek steeds groter wordt. In dit proefschrift zijn meerdere aspecten aan bod gekomen die noodzakelijk zijn voor een doelmatige inzet van laboratoriumonderzoek: kortere doorlooptijden en ondersteuning bij de selectie en de interpretatie van testen. Deze zaken worden uitgebreid besproken in de algemene discussie van dit proefschrift in **hoofdstuk 8**.

Verbetering van de doelmatigheid van laboratoriumonderzoek kan alleen door middel van automatisering en digitalisering. Ook is het nodig dat gezondheidsinformatie, die nu vaak nog versnipperd is over verschillende systemen, beter geïntegreerd wordt. We zien al dat automatisering van het klinisch-chemisch labotarorium tot duidelijke verbeteringen in de doorlooptijd heeft geleid. Toekomstige toepassingen kunnen mogelijk bestaan uit beslisondersteuning door kunstmatige intelligentie, voor het selecteren van de juiste testen om te bepalen, of voor de interpretatie van de resultaten. Deze toepassingen staan nu nog in de kinderschoenen, maar het gebruik ervan neemt snel toe. Voordat deze toepassingen in de praktijk kunnen worden gebruikt, zijn er nog veel vragen die beantwoord moeten worden.

Dankwoord

Dit proefschrift zou er niet zijn geweest zonder de hulp van heel veel mensen. Een paar daarvan wil ik in bijzonder bedanken.

Wouter, je hebt me in mijn promotietraject alle ruimte – en tijd – gegeven, maar er steeds voor gezorgd dat de structuur overeind bleef. Je hebt het overzicht bewaard en waar dat hier en daar nodig was focus aangebracht. Dank voor je begeleiding, het geregel achter de schermen en je altijd enthousiaste reacties in onze meetings.

Maarten, het was een waar genoegen om jou als mijn copromotor te hebben. Hoewel het eigenlijk meer voelde als een partner in crime. Je hebt een uitzonderlijke dossierkennis, waar ik dankbaar gebruik van heb gemaakt. Ik heb genoten van onze ongedwongen overleggen, waarin de wetenschap altijd moest vechten om aandacht omdat een van ons weer een zijpaadje was ingeslagen. Laten we gewoon blijven afspreken om bij een kopje koffie te praten over van alles en nog wat. Bedankt voor je vriendelijkheid, je openheid en de gezelligheid.

Jan, bedankt dat je in deze rijdende trein bent gestapt. Je bent een van de erudietste mensen die ik ken. Ik mis je verhandelingen over het Pools-Litouwse gemenebest nu je naar Zwolle bent gegaan. Voor je humor en je vlijmscherpe analyses, waarbij je het perspectief van de internist altijd in het oog hield, ben ik je bijzonder erkentelijk.

Christiana, bedankt voor je begeleiding aan het begin van mijn promotie. Ik heb dankbaar gebruik gemaakt van je methodologische kennis en je ervaring in R. Helaas ziet mijn code er nog altijd niet zo netjes uit als die van jou.

Cornelia, Saskia en Mark, jullie wil ik van harte bedanken voor het beschikbaar stellen van de data uit UPOD, het kritisch meedenken en specifiek Saskia omdat je me er steeds aan herinnert dat niet elke patiënt een witte man is.

Michael, dank voor het oppakken van onze gezamenlijke studie toen ik er niet helemaal meer uit kwam. Je hebt het artikel echt naar een hoger plan getrokken.

Karijn, bedankt voor je enthousiaste inzet en het geduldig beantwoorden van al onze vragen over checkpointremmers.

Pascal en Thérèse, jullie ijver en inzet tijdens jullie wetenschapsstages waren aanstekelijk en we mogen trots zijn op de resultaten.

Laura, bedankt dat je samen met mij tijdens de Internistendagen in Maastricht een sessie hebt willen houden over doelmatig laboratoriumonderzoek, die het startpunt was van het artikel dat nu als hoofdstuk 7 in dit proefschrift is opgenomen.

De POORT-studie is een project geweest van heel veel verschillende mensen. Annemieke Tinholt, Dirk Tjwa, Jan Willem Uffen, Linda Everlo, Maaike Weijmans, Maria de Winter, Marieke de Regt, Mathilde Nijkeuter, Matthijs van Oevelen, Rebecca Knobbe, Roos Boerman, Saskia Briedé, Shanna de Groot, Tom Wouters

en Willemijn Huiszoon, door jullie, deels vrijwillige, inzet in het expert panel verliep de data-analyse uiteindelijk een stuk sneller dan gedacht. Joris de Groot en Hans Reitsma, bedankt voor jullie waardevolle advies bij de studieopzet. Ad van Mullem, Jantien Woudstra-de Liefde, Marjon Wesseling en Ronald Swinkels, door jullie verliep de logistiek aan de laboratoriumkant vlekkeloos. En Bianca Silvius, Maartje Woelders en Marianne Schurink, bedankt voor de hulp bij de dataverzameling.

Jolande, jij stond aan de wieg van de POORT-studie. Je hebt het oorspronkelijke idee bedacht en bent voortvarend aan de slag gegaan om dat verder uit te werken, waarbij je out-of-the-box oplossingen niet schuwde. Bedankt dat je jouw project aan mij hebt durven toevertrouwen en dat je als eerste het zaadje hebt geplant voor mijn promotie.

De leden van de beoordelingscommissie, Joost Frenkel, Prabath Nanayakkara, Jelle Ruurda, Frank Visseren en Dorien Zwart, hartelijk dank voor het lezen en beoordelen van mijn proefschrift.

Mathilde en Douwe, mijn teamgenoten van het eerste uur, het was een hele achtbaan. Ik wil jullie bedanken voor onze hechte band, waarin we alles tegen elkaar kunnen zeggen. Ik waardeer jullie steun en onze goede gesprekken, onder andere in Tienhoven en op het Eemmeer.

Mijn collega internisten, ik wil jullie van harte bedanken voor de altijd fijne samenwerking en de gemoedelijke sfeer. Mijn kamergenoot Marjolein wil ik bij naam noemen: bedankt voor je altijd geduldige luisterende oor en je enthousiasme voor ansichtkaarten.

In mijn opleiding tot internist heb ik veel leraren en voorbeelden gehad. In het bijzonder wil ik Wiek Rensma, mijn opleider in Tilburg, bedanken. Ik heb veel geleerd van je kijk op het vak en je ogenschijnlijke onbezorgdheid.

Marieke, ik waardeer onze vriendschap enorm. Je weet me altijd een spiegel voor te houden en als we afspreken is het altijd bijzonder, of het nu op een rooftop in New York is (zonder zonsverduistering), in een wijnbar in Napels of gewoon om de hoek. Bedankt dat je mijn paranimf wilt zijn.

Mijn vrienden van de Piemelparty, Bram, Dook, Jesse en Marinus, jullie zijn de beste vriendengroep die iemand zich kan wensen. Ik kan met alles bij jullie terecht. En als we samen zijn is het altijd feest. Bedankt voor de gezelligheid, de warmte en de geweldige verhalen. Ik pis. En Jesse, bedankt dat je mijn paranimf wilt zijn.

Promoveren in combinatie met een fulltime baan en een master zorgt ervoor dat er nogal wat werk in de avonden en weekends is verzet. Matijn en Joris, ik wil jullie bedanken voor de ruimte die jullie me daarin gegeven hebben.

Hoewel sommige andere mensen dit ook over hun eigen ouders zeggen, zijn mijn ouders écht de allerliefste ouders die er zijn. Bedankt dat jullie vertrouwen in mij en jullie onvoorwaardelijke steun.

Ook de rest van mijn familie, Roel, Sandra, Krijn, Dominique, Janneke, Max, Jaap, Anna, Stach, Mare, Dieter, Juultje, Freeke en Gijs, bedankt voor alle gezelligheid.

Nova en Mika, ik ben zo ongelofelijk trots op jullie en hoe snel jullie aan het uitgroeien zijn tot volwassen personen met een sterke eigen mening en een positieve kijk op de wereld. Bedankt dat jullie mijn leven waardevoller maken.

En tenslotte Manu, jouw gedrevenheid en doorzettingsvermogen zijn een voorbeeld voor me. Je steun en aanmoedigen hebben ervoor gezorgd dat ik deze promotie heb kunnen afronden, en dan heb ik het nog niet eens over de wedijver over wie als eerste klaar zou zijn. Ik ben zo gelukkig met jou. Ik hou van je.

Curriculum vitae

Bram Vrijsen is op 5 december 1978 geboren in Tilburg. Voor zijn zesde verjaardag kreeg hij van zijn ouders pianolessen (en een piano). Vijf jaar later is hij met de pianolessen gestopt vanwege het uitblijven van succes.

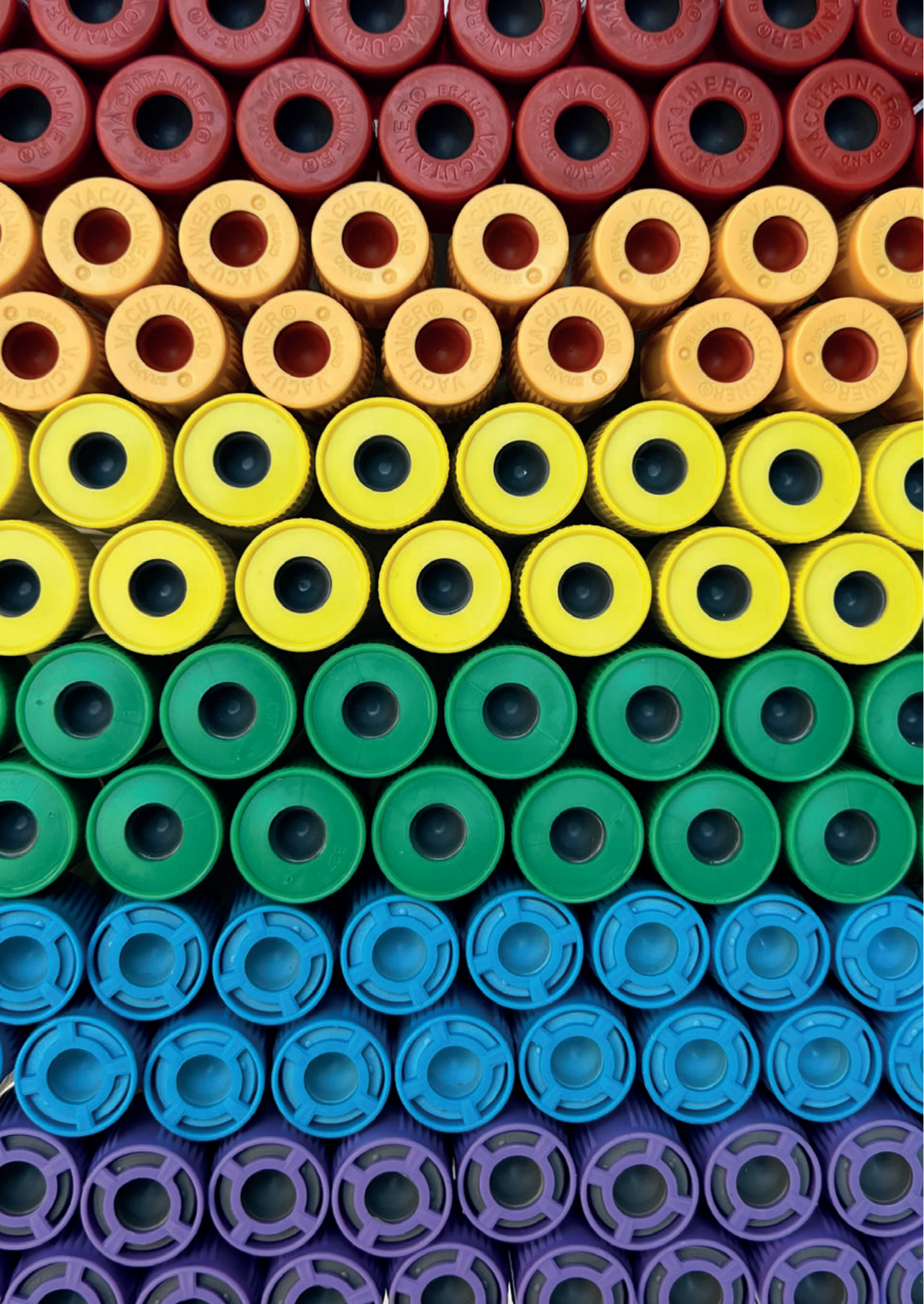
In de jaren daarna behaalde hij zijn gymnasiumdiploma aan het Mill-Hillcollege in Goirle en rondde hij zijn studie geneeskunde aan de Universiteit Leiden af. Na omzwervingen in Rotterdam en Zwijndrecht volgde hij de opleiding interne geneeskunde in het St. Elisabethziekenhuis in Tilburg en het UMC Utrecht, waarbij hij een meervoudig profiel heeft gekozen. Vervolgens heeft hij even als internist in het St. Antoniusziekenhuis gewerkt, waarna hij gevraagd is om terug te keren naar het UMC Utrecht. Daar werkt hij nu nog steeds.

Tijdens zijn loopbaan in het UMC Utrecht heeft hij in 2015 de opleiding acute geneeskunde afgerond, werkte hij van 2020 tot 2022 als medisch afdelingshoofd van de spoedeisende hulp en is hij in 2020 medisch achterwacht geworden bij het Nationaal Vergiftigingen Informatie Centrum.

Van 2016 tot 2019 was hij weekendpleegvader van Joris en vanaf 2019 is hij pleegvader van Nova en Mika. Hij woont met zijn vriend Manu samen in Utrecht.

Naast dit alles is hij in 2017 begonnen met een promotietraject. Kort na het starten hiervan besloot hij, geïnspireerd door de piano in de hal van het ziekenhuis, om een piano te kopen en zijn oude lesboeken weer ter hand te nemen. Zijn studiegenoten van de master epidemiologie die hij tijdens zijn promotie volgde hield hij op de hoogte van zijn voortgang, of het gebrek daaraan.

Zijn promotietraject heeft uiteindelijk geresulteerd in dit proefschrift. Hij kan nog steeds geen piano spelen.


VACUTAINER